Title
X-ray Absorption Spectroscopy of Biologically Relevant Systems

Author
Uejio, Janel Sunayo

Publication Date
2010

Peer reviewed|Thesis/dissertation

X-ray Absorption Spectroscopy of Biologically Relevant Systems

by

Janel Sunayo Uejio

A thesis submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

Of the

University of California, Berkeley

Committee in Charge:

Professor Richard J. Saykally, Chair
Professor David Chandler
Professor Robert Glaeser

Spring 2010

X-ray Absorption Spectroscopy of Biologically Relevant Systems

by

Janel Sunayo Uejio

Abstract

X-ray Absorption Spectroscopy of Biologically Relevant Systems

by

Janel Sunayo Uejio

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Richard J. Saykally, Chair

The relationship of biologically relevant molecules to their aqueous environment remains an active area of current research. In this dissertation, I present research biologically relevant molecules in an aqueous environment using X-ray spectroscopy and accompanying calculations.

Chapter 1 is a more thorough introduction to the topic of near edge X-ray absorption fine structure (NEXAFS) spectroscopy with stress on the experimental set up used, and computational methods. Additionally, in Chapter 1 there is a more detailed outline on the topics in subsequent chapters.

Chapters 2 and 3 detail an examination of the model system of carboxylate (CH_3COO^-) with a variety of metals. These studies were to examine the preferential ion pairing of different cations by NEXAFS. In chapter 2, I investigate the interactions of carboxylate with the monovalent cations lithium, sodium, and potassium; chapter 3, is an examination of the strength of interactions of carboxylate with the following divalent cations: magnesium, calcium, strontium, copper, manganese, and zinc.

During the course of my graduate work, the need for improved computational results became apparent. The development of the current technique employed within my research group is detailed within Chapter 4 using prototypical gas phase molecules as example systems.

In chapter 5, I compare both experimental and computational results of a monopeptide, alanine, compared to one of its synthetic analogues, a monopeptoid, sarcosine. These results compare the spectral importance of hydration versus conformation, in addition to conclusions concerning the spectral impact of microhydration versus bulk hydration.

I present preliminary results for other biological molecules in Chapter 6. The molecules studied are the amino acids arginine and valine, and the RNA nucleobases uracil and cytosine, and their respective nucleosides, uridine and cytidine. The chapter includes both preliminary experimental and theoretical spectra.

In Chapter 7, I look to the future with a brief discussion of improvements and advice to future coworkers.

Acknowledgements

Acknowledgements are difficult to write. This is my attempt to address the points of who to thank and how to thank them, and the details to include. Thanks goes to my adviser, Richard Saykally, for having me join the group and giving me the time to develop into the headstrong scientist I am today. My gratitude to the Saykally group, especially the microjet side, Alice, Andrew, Craig, and Walter; in particular, Craig for always being the one willing to argue and saving me the trouble. My gratitude to David Prendergast of the Molecular Foundry, due to his patience and guidance and pressure to do one more calculation my work is magnitudes more insightful. Thanks to Kathy Durkin, of the Molecular Graphics Laboratory, for always being willing to troubleshoot and install just one more program. Particular thanks goes to my friends, who have been there for the roller coaster of graduate school.

And finally, and most importantly thanks goes to my family.

This work was supported by the auspices of the Chemical Sciences Division, Office of Basic Energy Sciences, U.S. Department of Energy and at the Molecular Foundry, Lawrence Berkeley National Laboratory, supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Thanks to the Advanced Light Source (ALS) Fellowship fund and the ALS itself, which is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under contract no. DE-AC02-05CH11231 at Lawrence Berkeley National Laboratory. Computational resources required for this work were provided by NERSC.

Table of Contents

Chapter 1.
Introduction and Overview

Water has been important for a very long time,[1] and as essentially all biological systems are aqueous, understanding the nature of an aqueous environment is vital. Hydration impacts the conformer populations for biomolecules,[2] the stability of different charge states,[3] and reactivity.[4] The work described in this thesis attempts to characterize several biologically relevant aqueous systems using X-ray absorption spectroscopy with a chapter detailing the development of a new theoretical approach that has greatly enhanced the quality of the chemical information that can be extracted.

1.1 Experimental Aspects

The primary experimental technique employed was near edge X-ray absorption fine structure (NEXAFS) spectroscopy. This is a core-level technique, an incident X-ray excites a core level electron on an atom to anti-bonding molecular orbitals or low-lying Rydberg states.[5] In studying biological molecules, most studies yielded information on the nitrogen K-edge (promotion from the 1s orbital refers to the K-edge), oxygen K-edge, and carbon K-edge. The lowest unoccupied molecular orbitals are centered primarily on the molecule itself, and provide information concerning structural aspects such as conformation.[6] High-energy anti-bonding molecular orbitals, and Rydberg states have large spatial extents, and accordingly, environmental changes impact the energy and strength of transitions, such as for hydrogen bonding networks.[7]

The experiments were performed at the Advanced Light Source, beamline 8.0.1. This beamline provided an intense flux, on the order of 10^{15} photons/sec, and resolution of E/ΔE 8000. Also this beamline permitted us to operate without windows; windows are usually prepared out of silica nitride and reduce the flux on the edges of interest.

NEXAFS is primarily a vacuum technique, as the energies required for studying low Z elements such as carbon, nitrogen, and oxygen are absorbed in air. To study liquids in a vacuum setting, a liquid microjet is used.[8] The microjet nozzle is a piece of fused silica capillary with an inner diameter of 30 μm. The liquid sample typically flows at 50 m/s, this number varying with the backing pressure from the Teledyne Isco syringe pump. As the syringe pump is at room temperature, according to multiple techniques and modeling the temperature of the liquid jet is approximately 15-20°C.[7-11] Using a liquid microjet permits not only vacuum studies on a liquid sample, but as the sample constantly refreshes, it avoids the problem of sample damage caused by the incident X-ray beam.[12]

The chamber itself (Figure 1) is primarily constructed of stainless steel. The jet is placed through an XYZ manipulator on the right of the schematic and flows from right to left. In the main chamber, there is a liquid nitrogen (abbreviated LN_2) trap and a large throughput turbopump (Leybold 1000 L/s) to maintain base pressures of 10^{-4} Torr or lower. The jet is subsequently skimmed and condensed on a cryogenic trap connected to a molecular drag rough pump. Before entering the main chamber, the X-rays pass through the differential pump section, which comprises two Kimball cubes, three small 70 L/s Varian turbopumps, and two pinhole apertures to limit backflow pressures from the main chambers. The X-rays would travel out of the page in the diagram provided. The incident X-ray beam intersects the liquid microjet, and excites the electrons from atoms of interest. From the excitation, both primary and secondary electrons (such as those from Auger effects, and ionization) are produced; these electrons are collected by a positively biased copper electrode. As the energy is scanned the signal is

collected, creating a spectrum of intensity versus energy. Signal from the incident beam on a gold mesh positioned farther up the beamline is collected and used for normalization to account for any beamline fluctuations. For a more thorough description of the construction, please consult Wilson et al.'s work in the *Review of Scientific Instruments*.[11]

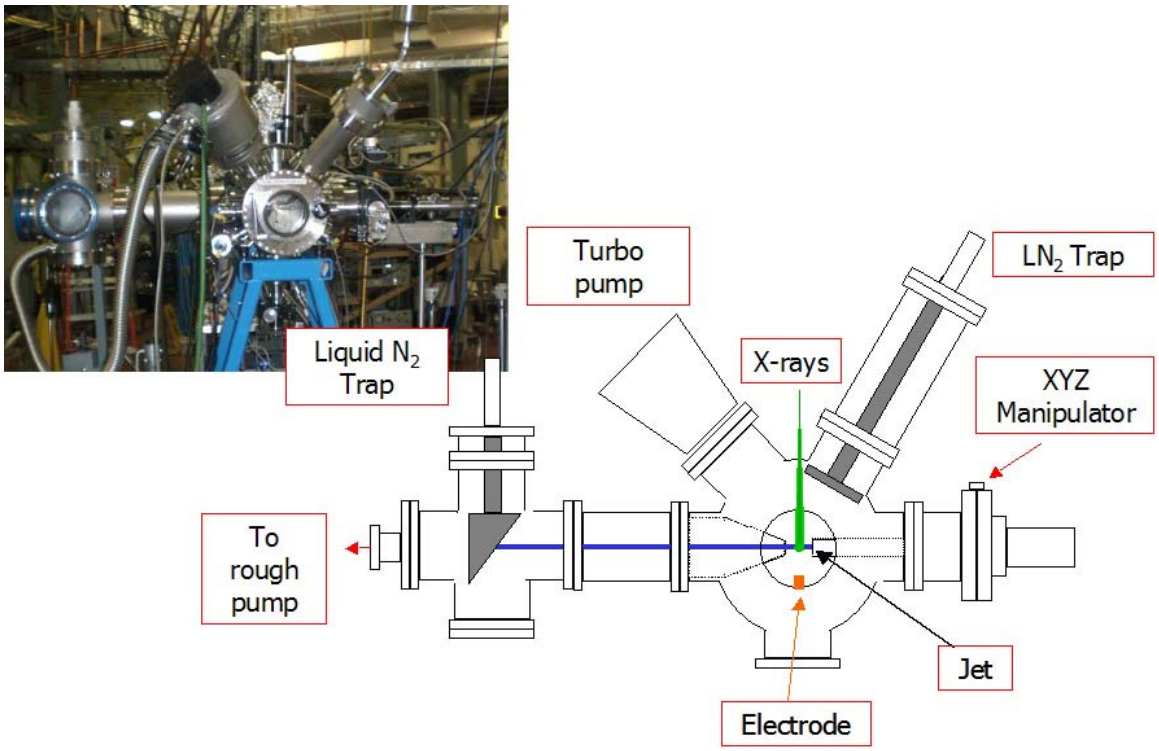


Figure 1. The upper left shows a picture of the experimental design. Below and to the right is a simple schematic of the chamber. The jet would flow from the right to the left, and the X-ray beam would travel out of the page. The differential pump section is not shown.

1.2 Theoretical Background

NEXAFS spectra are convoluted by the inherent broadening of the technique, the hydrated environment, and conformational changes.[5, 13] To decipher the experimental spectra and extract chemical insights, comparisons with theoretical calculations are made. During the course of this thesis work, two methods were used, the first being StoBe deMon, and the second being the excited core hole (XCH) method. Notably, both use the formalism of density functional theory (DFT) using the dipole operator with Fermi's Golden rule:

$$P_{if} = \frac{2\pi}{\hbar} |< f | \overline{V} | i >|^2 \ \rho_f(E) \qquad (1)$$

Equation 1 is a simple form of Fermi's Golden Rule wherein the probability is expressed in terms of constants, the magnitude of overlap between the states (within the bra-ket notation), and the density of final states (ρ_f).

StoBe deMon

StoBe was developed and is maintained by Lars Pettersson and Klaus Hermann, from Stockholm and Berlin, respectively.[14] The method is a self-consistent field (SCF) DFT method, and is relatively straight forward to use and can be run on an average personal computer at this time, requiring limited computational resources from the user. The code is based on the use of Slater's transition potential, also known as the transition state method,[5, 15] and has two basic approximations. The first approximation is that only the transition state needs to be calculated (not the initial state nor any of the final states) and the interactions of the excited electron and the static electrons are ignored.[15] The other approximation is the treatment of the X-ray induced core-hole. Using the transition state method, transitions fall into two categories: above or below the ionization potential. For transitions to final states below the ionization potential, the electron is deemed long-lived and relatively localized to the target molecule, allowing for some degree of relaxation. Whereas, for transitions to final states above the ionization potential they are considered "sudden," with relatively little time for relaxation, or changes to the energetics of the system; thus, the binding energy of which can be described by Koopman's theorem.[16] To accommodate the range of relaxation times, the usage of non-integral occupation values for electrons are used, such that a half occupied core-hole (HCH) approximates all transitions and is the second large approximation used in StoBe.[5]

Additionally, StoBe is a cluster code, and produces very distinct surface states at the edge of clusters. The existence of surface bound states is exacerbated by the use of a Gaussian basis set; the excited atom is modeled using the IGLO basis set.[17] An aspect of StoBe which has garnered criticism, [7, 18, 19] is that a linearly increasing broadening scheme is used. It is the users' arbitrary choice as to width of the broadening and how much it should increase over a spectrum, allowing for a large degree of fitting. This is shown schematically in Figure 2

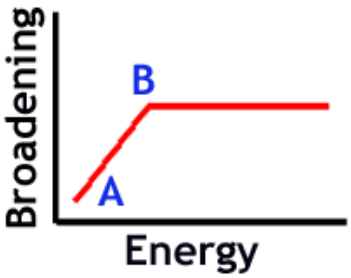


Figure 2. A diagram of the broadening used in StoBe. The x-, and y- coordinates of A and B are chosen arbitrarily by the user.

For a more detailed discussion of StoBe and its usage, please consult Benjamin Messer's thesis,[20] and the StoBe manual.[14]

Excited State Core Hole (XCH) Method

The excited state core hole technique was developed primarily by Dr. David Prendergast, et al..[21, 22] The method employs DFT, using an SCF calculation to model the electronic structure of the lowest-energy core-level excited state with a full core hole (as opposed to the HCH used in StoBe) with explicit inclusion of the excited electron. Norm-conserving pseudopotentials are used to model the atoms in the system,[21] though some work has now been tested using ultra-soft pseudopotentials, as a means for reducing computational cost.[23] The calculations are performed

using a plane wave basis set with periodic boundary conditions; by employing these conditions with Brillouin zone sampling, accurate calculations of the density and the spatial extent of continuum states can be reached.[22] The electronic structure calculation is performed using PWSCF,[24] which uses the open-source quantum Espresso package. The transition amplitudes are then calculated within the dipole and single particle approximation. Since all of the systems of study in this thesis are isotropic, the polarization directions are averaged together. Due to more accurate sampling of continuum states, a homogenous broadening is used for the computational spectra produced by XCH.

Unfortunately, the calculations require large amounts of CPU time; in order to have convergence for a single set of coordinates with over 100 atoms, it can take up to 700 computational hours on Franklin, a super computing cluster maintained by the National Energy Research Scientific Computing Center (NERSC).

In this section, there has been no discussion of input coordinates for either computational method. The importance of reasonable input coordinates and sampling techniques with a comparison of the two methods are discussed later in Chapter 4, and previously published work.[18, 25]

1.3 Thesis Overview

A brief overview of the subsequent chapters is provided here, thus giving context to each chapter, as well as highlighting its role experimentally and theoretically.

Chapter 2 is based on previously published work,[26] and describes a new approach for characterizing selective binding between oppositely charged ionic functional groups. NEXAFS spectra measured for aqueous formate and acetate solutions containing lithium, sodium, and potassium cations reveal stronger binding of the lighter metals, supporting recent results from molecular dynamics simulations and other experiments.[4] These results are in accord with the "Law of Matching Water Affinities," proposed by K.D. Collins relating relative hydration strengths of ions to their respective tendencies to form contact ion pairs.[27, 28]

Looking beyond the "Law of Matching Water Affinities," a preliminary examination of divalent cations with carboxylate is detailed in Chapter 3. A variety of divalent metals were examined, some with biological relevance, along with others to complete periodic trends, including: calcium, strontium, magnesium, copper, manganese and zinc. The metals did not follow predicted trends based on hydration energy or size arguments; current theoretical approaches are hampered by a lack of accurate potentials for all the metals cations under study, thus precluding a current complete theoretical analysis.

The development of the computational method, XCH, briefly described earlier in this chapter is detailed more thoroughly in Chapter 4, which is based on previously published work in *Chemical Physics Letters*.[18] A computational approach, XCH, is described in terms of previous work on water[18, 21, 22] and is validated through the prediction and interpretation of several gas phase organic molecules, sampling a range of chemical bonding and structural motifs. The chapter details the importance of appropriate sampling by proving that the use of lowest energy structures within the rigid molecule approximation is inadequate, and agreement with experiment is significantly improved upon approximating vibrational motions with molecular dynamics sampling. Additionally, assignment of spectral features, and general agreement with experiment is shown to improve with the new methodology.

Chapter 5 is a discussion of the application of XCH to the aqueous interactions of peptides and their synthetic analogues called peptoids (poly-N-substituted glycines) and is based

on a recently published paper in the *Journal of Physical Chemistry B.*[13] Peptoids have recently emerged as a promising biomimetic material, particularly due to their robust secondary structure and resistance to denaturation. By employing XCH, and comparing it to experimental NEXAFS spectra of aqueous sarcosine, the simplest peptoid, and alanine, its peptide isomer, insights were gathered about the relative impact of hydration on these systems. Importantly, a predicted solvent dependence for alanine, as compared to sarcosine is examined in the light of experimental results and a variety of theoretical calculations.

Chapter 6 is a discussion of preliminary data for small biological molecules. This work was performed earlier than that discussed in Chapters 3-5, thus it employs the older computational method of StoBe deMon[14] for calculations. Two amino acids are examined, arginine and valine, under different charge conditions imposed by changing the pH. Two nucleobases, uracil and cytosine, are examined with their nucleoside counterparts (nucleobase and sugar) uridine and cytidine, respectively. Experimental results and preliminary theoretical calculations are discussed in terms of their strengths and weaknesses.

In Chapter 7, I include suggestions for future work, these include: minor upgrades to current experimental equipment, possible future systems of study, and lastly, I leave a bit of advice for any interested current or feature colleagues.

References:

(1) Genesis. *Bible*, 1:6.

(2) Shin, S. B. Y.; Kirshenbaum, K., *Org. Lett.* **2007,** 9, (24), 5003-5006.

(3) Ahn, D. S.; Park, S. W.; Jeon, I. S.; Lee, M. K.; Kim, N. H.; Han, Y. H.; Lee, S., *J. Phys. Chem. B* **2003,** 107, (50), 14109-14118.

(4) Vrbka, L.; Vondrasek, J.; Jagoda-Cwiklik, B.; Vacha, R.; Jungwirth, P., *Proc. Natl. Acad. Sci. U. S. A.* **2006,** 103, (42), 15440-15444.

(5) Stöhr, J., *NEXAFS Spectroscopy*. Springer: New York, 1996; 20-31.

(6) Messer, B. M.; Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B* **2005,** 109, (11), 5375-5382.

(7) Smith, J. D.; Cappa, C. D.; Messer, B. M.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B* **2006,** 110, (40), 20038-20045.

(8) Faubel, M.; Schlemmer, S.; Toennies, J. P., *Zeitschrift Fur Physik D-Atoms Molecules and Clusters* **1988,** 10, (2-3), 269-277.

(9) Smith, J. D.; Cappa, C. D.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J., *J. Am. Chem. Soc.* **2006,** 128, (39), 12892-12898.

(10) Wilson, K. R.; Rude, B. S.; Catalano, T.; Schaller, R. D.; Tobin, J. G.; Co, D. T.; Saykally, R. J., *J. Phys. Chem. B* **2001,** 105, (17), 3346-3349.

(11) Wilson, K. R.; Rude, B. S.; Smith, J.; Cappa, C.; Co, D. T.; Schaller, R. D.; Larsson, M.; Catalano, T.; Saykally, R. J., *Rev. Scientific Instrum.* **2004,** 75, (3), 725-736.

(12) MacNaughton, J. B.; Moewes, A.; Lee, J. S.; Wettig, S. D.; Kraatz, H. B.; Ouyang, L. Z.; Ching, W. Y.; Kurmaev, E. Z., *J. Phys. Chem. B* **2006,** 110, (32), 15742-15748.

(13) Uejio, J. S.; Schwartz, C. P.; Duffin, A. M.; England, A. E.; Prendergast, D.; Saykally, R. J., *J. Phys. Chem. B* **2010,** 114, 4702-4709.

(14) Hermann, K.; Pettersson, L.; Casida, M.; Daul, C.; Goursot, A.; Koester, A.; Proynov, E.; St-Amant, A.; Salahub, D. *StoBe-deMon version 3.0*, Version 3.0; StoBe Software: 2007.

(15) Slater, J. C., *Adv. Quant. Chem.* **1972,** 6, 1.

[16] Atkins, P.; de Paulo, J., *Physical Chemistry, 7th ed.* W.H. Freeman and Company: New York, 2002; 569.
[17] Kutzelnigg, W.; Fleischer, U.; Schindler, M., The IGLO-Method: ab-initio calculation and interpretation of NMR chemical shifts and magnetic susceptibilities. In *NMR- Basic Principles and Progress*, ed.; Diehl, P., Springer Verlag: New York, 1991; 23, 165-262.
[18] Uejio, J. S.; Schwartz, C. P.; Saykally, R. J.; Prendergast, D., *Chem. Phys. Lett.* **2008,** 467, (1-3), 195-199.
[19] Smith, J. D.; Cappa, C. D.; Messer, B. M.; Cohen, R. C.; Saykally, R. J., *Science* **2005,** 308, (5723).
[20] Messer, B. M. *X-ray Absorption Spectroscopy of Aqueous Amino Acids.* **Doctoral**, University of California Berkeley, Berkeley, 2005.
[21] Prendergast, D.; Galli, G., *Phys. Rev. Lett.* **2006,** 96, (21), 21522-21526.
[22] Prendergast, D.; Grossman, J. C.; Galli, G., *J. Chem. Phys.* **2005,** 123, (1), 14501-14505.
[23] Schwartz, C. P., *Manuscript in Preparation* **2010**.
[24] Baroni, S.; Corso, A. D.; de Gironcoli, S.; Gianozzi, P. PWSCF. www.pwscf.org
[25] Schwartz, C. P.; Uejio, J. S.; Saykally, R. J.; Prendergast, D., *J. Chem. Phys.* **2009,** 130, (18), 184109.
[26] Uejio, J. S.; Schwartz, C. P.; Duffin, A. M.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (19), 6809-6812.
[27] Collins, K. D., *Methods* **2004,** 34, (3), 300-311.
[28] Collins, K. D., *Biophysical Journal* **1997,** 72, (1), 65-76.

Chapter 2.
Characterization of Selective Binding of Alkali Cations with Carboxylate by X-Ray Absorption Spectroscopy of Liquid Microjets

2.1 Introduction

The discovery of the selective interactions between simple ions and proteins dates back over a century to Hofmeister's studies with chicken egg protein; proteins could be selectively "salted in" or "salted out" by addition of various salts to the solution.[1] This "Hofmeister effect" has subsequently been observed with more salts and many more proteins, with relative magnitudes following the "Hofmeister series" ordering, as do various related phenomena.[2, 3] Nevertheless, it remains to understand the origin of "Hofmeister effects," despite enormous effort.[4, 5]

To rationalize biological ion specificity, such as the physiologically important preferential binding of sodium versus potassium with proteins, the "Law of Matching Water Affinities" was proposed by Collins.[6, 7] Based on charge densities and electrostatic arguments, this law holds that ions with similar hydration free energies form the most stable (insoluble) contact ion pairs. In the case of proteins, the carboxylate group is considered to have a hydration energy much closer to that of sodium than to that of potassium, which is manifested in the sodium binding free energy being larger by 2.22 kcal/mol, as determined by Vrbka et al.[8] via simulations and conductivity measurements. The simulations indicated that the interaction is localized on the carboxylate groups of the protein. Conductivity measurements were performed on protein solutions for experimental support, revealing a larger relative decrease upon addition of sodium chloride compared to potassium chloride, indicative of sodium being more efficiently removed from solution than potassium. The rationalization of the preferential interaction was again that there was a closer match of hydration energy, which was reflected in more stable contact ion pairing between the protein carboxylate groups and sodium, versus potassium ions.

In this chapter, I examine the selective interactions of alkali cations with the carboxylate groups of the model carboxylate anions formate and acetate, using novel liquid microjet technology to introduce the aqueous solutions into a soft X-ray spectroscopy experiments and finding strong support for the above conclusions reached by Vrbka, et al..[8] The carboxylate K-shell spectra comprise transitions from the atomic 1s core levels to π*carbonyl antibonding orbitals, the latter shifting to higher energies in response to stronger interaction with the cation. This experiment thus provides a direct probe of such cation-anion interactions in aqueous systems, which are difficult to access by other methods, e.g. mass spectroscopy. Similar experiments, but with quasi-static liquid cells, were recently reported for NaCl solutions by Aziz et al.,[9] wherein X-ray "spectral fingerprints" of contact ion pairs were found. However, their approach is not likely to work with complicated solutes, as in the present study, due to the usage of a windowed coupling; the Si_3N_4 used would absorb in the biologically relevant energy ranges and drastically reduce photon flux.

2.2 Materials and Methods

Samples:

All samples were purchased from Sigma-Aldrich in a crystalline powder forms at a stated purity of 98% or higher and used without further purification. Solutions were prepared with 18 MΩ (Millipore) water to compose a solution of 2.0 M.

Experimental Details:

The NEXAFS spectra were obtained on the relevant edges at Beamline 8.0.1 of the Advanced Light Source (ALS), at Lawrence Berkeley National Lab. Carbon K-edge and oxygen K-edge spectra were collected both from the liquid jet as well as from the gas phase background. Detailed carbon spectra were obtained in the range of 285-289 eV, and were subsequently calibrated to HOPG peaks previously recorded by beamline scientists. A gold mesh was used to collect I_o farther up the beamline, which was used to account for intensity fluctuations. The spectra were subsequently area normalized.

The experimental apparatus comprises: a main chamber where the X-ray beam intersects with the microjet stream, a section for trapping the microjet with a skimmer to reduce pressures, and a differential pump section to allow for a windowless coupling to the beamline. The measurements in this chapter were obtained using a chopped fused silica microjet of 30 μm attached to an HPLC pump with an average flow rate of .9 mL/min. The signal was collected by total electron yield (TEY) with a biased copper electrode of 1.8 kV; the usage of TEY has been shown to be a bulk technique due to the larger escape depths of electrons.[10] A more detailed description of the experimental apparatus has been previously published by Wilson et al..[11]

The formate carbon K-edge spectra were smoothed due to a low signal to noise. The data was treated with a five point second order Savitzky- Golay smoothing scheme;[12] it was empirical verified that this did not broaden rather reduced the noise in the signal.

Computational Details:

StoBe DeMon,[13] a density functional theory based commercial package, for computing core-level spectra, was used to aid interpretation of the experimental spectra. The functionals developed by Perdew, Berke and Erzenhoff [14] were used in addition to the revised exchange functional of Hammer, Hansen and Norkov.[15] The atom of interest was modeled using the IGLO basis set,[16] the remaining non-hydrogen atoms were replaced with model core potentials provided in the program.[13] The design of the calculation gives the energetic isolation of the 1s orbital of interest, which is then partially excited by replacing its electron with a half core hole. The use of a half core hole is a result of the sudden approximation, which has been shown when used in conjunction with the transition potential method to approximate the relaxation of the molecule under the influence of the core hole.[10, 17, 18] A stick spectrum is produced and then uniformly broadened with Gaussians of width .5 eV. The spectra are not globally shifted, as would be the normal procedure to compare with experimental spectra, so absolute transition values between theory and experiment cannot be compared; the magnitude of the shifts between theory and experiment are instead compared. StoBe has been shown to reproduce relative shifts well; errors of note have been found in the relative intensities rather than the relative energies.[19]

Kohn-Sham corrections were calculated for the acetate gas phase system by calculating both the ground state and the excited state.[20] These were found not to affect the relative shift between the potassium and sodium acetate systems; accordingly, these effects are not included in the figures given in this chapter.

The structures noted as gas phase systems were taken after geometry optimization; for the solvated systems, molecular dynamics were performed to provide snapshots for subsequent calculations. An initial geometry of a single ion pair was built using a commercial software package, MacroModel from Schrodinger Inc., and was subsequently placed in an 18.62 Å cube with over 200 water molecules and employing periodic boundary conditions. A molecular dynamics calculation was performed at 298K for 20 nanoseconds. From the molecular dynamics

trajectories, snapshots were taken over 70 picoseconds apart as to assure that the snapshots were uncorrelated. These snapshots were subsequently "ice cream scooped" to include the system of interest and the 20 closest waters, due to the need for a manageable sized cluster for the core level calculations.

StoBe is a cluster code and thus does not have the ability to use periodic boundary conditions. Thus, in a solvated system it is important to note these calculations are a guideline and a tool to help interpretations as opposed to a comprehensive predictive tool.

2.3 Results

Several carboxylate salts were examined in order to study the trend of preferential interactions between alkali metal cations and carboxylate ions using the technique of near edge X-ray absorption fine structure (NEXAFS) spectroscopy, introducing the liquid solutions into the high vacuum X-ray experiment with a liquid microjet system. This novel approach, in conjunction with theory, has recently been used to reveal subtle conformational and geometric interactions in aqueous systems, comprising both biological molecules and ionic solutions.[21, 22] Important for the current study, the rapid sample flow (ca. 50 m/s) minimizes X-ray photo damage, which has been found to be severe .

Simple carboxylate salts were employed due to their straightforward structure, high solubility, and easily distinguished 1s $\rightarrow \pi^*_{C=O}$ feature. Formic acid and acetic acid are the simplest organic acids, minimizing possible alkane chain effects. Additionally, both formate and acetate anions have shown preferential interactions with sodium versus potassium in previous experimental and theoretical studies.[8] The current experimental study was expanded to include the lithium ion as an additional test of the "Law of Matching Water Affinities," although technical difficulties were encountered with the associated supporting calculations.

An overlay of the NEXAFS spectra of the three different aqueous acetate solutions is shown in Figure 1. This figure is an expansion of the carbon K-edge spectra displaying the peak that is the result of the 1s electron being promoted to the antibonding π orbital.[25] There are small features on the high-energy edge of each of the peaks, most likely due to associated vibrational states of the carbonyl group. It was not possible to compare the σ^* region, above 290 eV, of the respective spectra due to potassium having transitions which obscure the underlying features at that energy range. Between each of the scans on the overlay, there is a spacing of about .05 eV on the leading edge of the feature; the largest blue shift is found in the comparison of lithium to the other two cations, noting that lithium has an ionic radius of 90 pm, compared with those of sodium and potassium, 116 and 152 pm, respectively.[26]

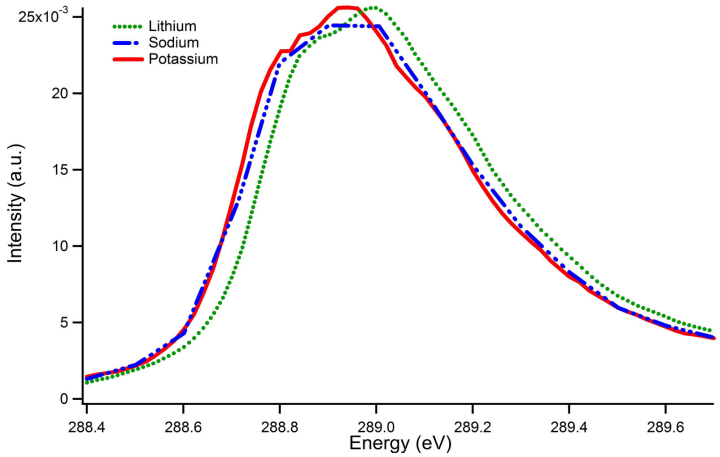


Figure 1. An overlay of the experimental π* feature in the K-shell spectrum of the acetate anion for 2 M solutions of potassium acetate, sodium acetate, and lithium acetate revealing distinct shifts between the cations, indicative of preferential interactions.

The carbon K-edges of formate complexes exhibit a marked shift between lithium and the other two cations, as evident in Figure 2; lithium exhibits a blue shift of approximately 0.2 eV. However, in these scans, there is no discernable shift between the potassium and sodium complexes; hence, preferential interactions are difficult to assess in this case. The difference in sizes between the formate and the acetate anions affects their hydration and electrostatic properties in terms of the "Law of Matching Water Affinities," thus determining the strength of the cation- anion interaction and the resulting shifts in their K-shell spectra.

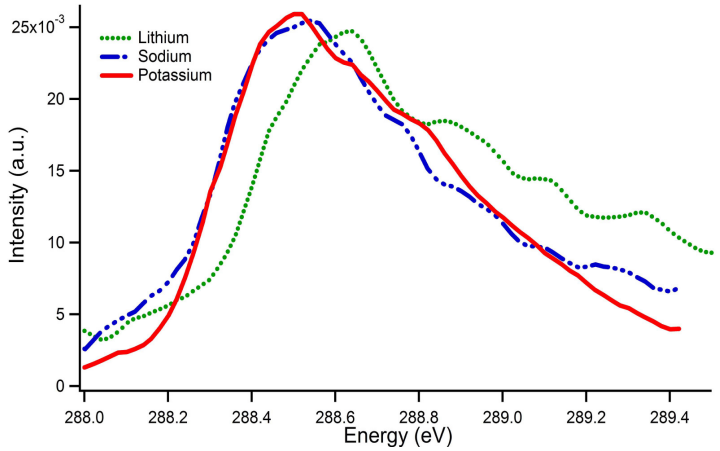


Figure 2. An overlay of the experimental carbon K-edge spectra of lithium formate, sodium formate, and potassium formate. The sodium and potassium spectra are nearly indistinguishable, whereas lithium is distinctly blue shifted. Clearly, the observed shifts between sodium and potassium formate are smaller than those found for acetate.

Using previously calculated radial distribution functions[8] as a guideline for approximate equilibrium distances (2.5 A° and 2.9 A°, respectively), structures were optimized for the acetate and formate complexes of both sodium and potassium. Unfortunately, the StoBe deMon software, [13] the program used for the NEXAFS calculations, does not include an effective core

potential basis set for lithium. Calculations performed without this potential exhibited Mulliken populations that did not accurately reflect the charges of the system of interest, and thus could not produce comparable energy values, although they did reproduce the general shape of the spectra. This behavior was discovered for sodium, when comparing results with and without an effective core potential; the Mulliken population did not correspond with the charges of the system and it produced differences in calculated spectra. This is important in the current study, considering the small shifts, whereas such a treatment is quite suitable for systems where these small shifts are inconsequential. The usage of Dolg pseudopotentials was also attempted and found did not accurately reflect the experimental spectra due to errors in the magnitude of the interaction of the carboxylate with lithium; this is logical due to notoriously difficult nature of lithium, and the errors that are found in its total valence correlation energy.[27] Figure 3 shows an expanded view of the $\pi^*_{C=O}$ feature from the calculation of the carbon K-edge of the isolated acetate ion pair with both potassium and sodium. The size of the relative spectral shift, (< 0.1 eV) is an expected result due to the different inter-ion distances of complexes with the respective cations. It should be noted that a calculation was done to verify that even when choosing equal distances, sodium transitions were blue shifted with respect to those of potassium.

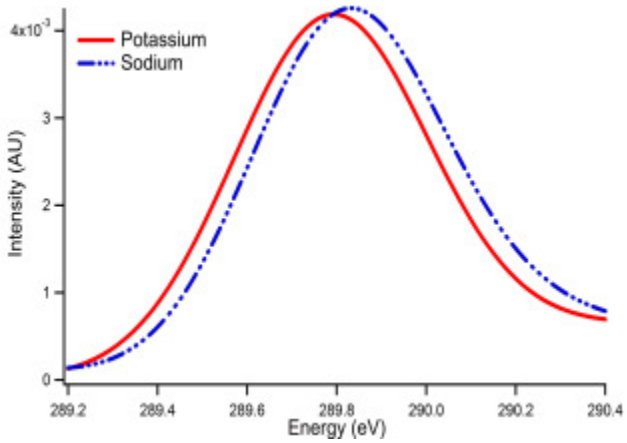


Figure 3. Theoretical calculations of the isolated ion pairs of acetate with alkali cations, produced with StoBe.[13] The magnitude of the spectral shift is ca. 0.07 eV between sodium and potassium. Lithium was not calculated due to a lack of core potential basis sets.

As mentioned above, the magnitude of the experimental K-shell shifts differed between formate and acetate; this is confirmed by our theoretical calculations, and reflects the corresponding ion binding energies. Such differences in the strengths of the interactions for formate and acetate were also found in other studies.[8] Shown in Figure 4 are the StoBe results for isolated formate ion pairs with sodium and potassium. The two spectra are very similar, but display a reversal of the shifts found for acetate. The magnitude of this small theoretical shift is beyond the precision of the current experiment (ca. 0.005 eV), but both the experimental and theoretical formate spectra clearly display smaller shifts than do the corresponding acetate spectra.

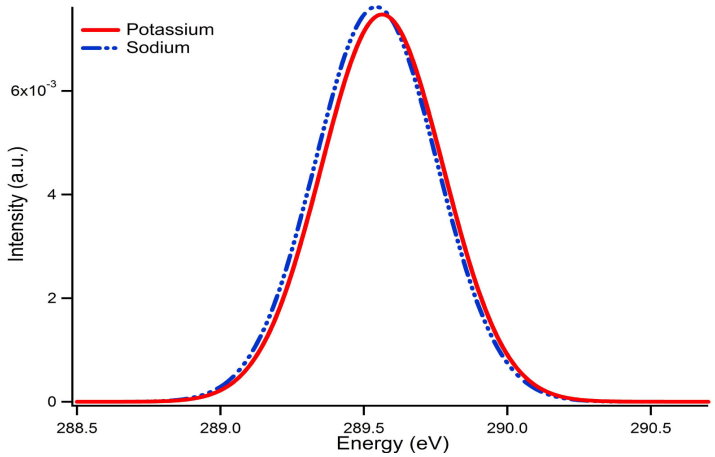


Figure 4. Computed spectrum for isolated ion pairs of formate with sodium and potassium. The two are very similar, but show a slight reversal of the trend previously noted. Lithium was not calculated due to a lack of core potential basis sets.

Figure 5 shows an enlarged section of theoretical spectra of hydrated acetate ion pairs, computed using equilibrated molecular dynamics snapshots. At a higher computational cost, the same result of a global shift between the sodium and potassium spectra was found. There is a blurring of some of the higher energy σ* features (not shown in the figure). The distances between the cation and the acetate (sodium closer than potassium) vary among the different MD snapshots and corresponding calculations, which does affect the magnitude of the spectral shift from snapshot to snapshot. In previous results, there was a ca. 2 kcal/mol difference in the association energy between the sodium and potassium acetate complexes;[8] this would correspond to a difference of about 0.08 eV, which is similar to the spectral shifts observed in the present study.

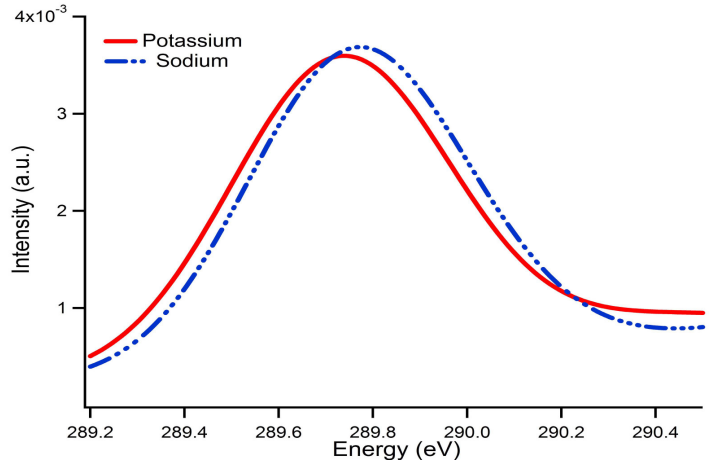


Figure 5. Calculated K-edge spectrum for hydrated ion pairs of acetate with sodium and potassium. The coordinates were taken from molecular dynamics snapshots. The energy range is slightly altered from previous calculations. The relative shift between the cations is larger than 0.05 eV.

The form of aqueous acetate and formate ions, viz. protonated or deprotonated, is governed by the acid/base equilibrium. Noting that pK_b for acetate is 9.243 and for formate

10.255, a pH of approximately 9.5 is obtained for the ca. 2 molar solutions used in this work, implying that the deprotonated forms dominate, as they would at biological pH . All of the solutions studied here were held at the same concentration. Therefore, I can discount variations in the composition of the solutions as influencing the above results.

2.4 Discussion and Conclusions:

The relative interaction strengths of the monovalent cations of sodium, potassium, and lithium with acetate and formate anions in aqueous solutions were examined using NEXAFS of liquid microjets. An unambiguous difference between the effects of lithium and those of the two other cations is evident on the carbon K-edge for both acetate and formate; a preferential interaction of sodium versus potassium is evidenced by a small but reproducible relative blue-shift in the carbonyl feature, reflecting a shift in the π^* orbital to higher energies. In the formate system, the corresponding sodium/potassium spectral shifts were barely discernable. Calculated K-shell spectra support the experimental findings, with a slight overestimation of the shifts, but unambiguously matching the observed trends. At the pH used in these experiments, the deprotonated forms of the anions are predominant, as they would be under biological conditions. Future experiments will include extending the study to other alkali ions, examining relevant divalent cations (Mg^{2+}, Ca^{2+}), as well as concentration and pH dependent studies. In any case, the current results establish a new way of probing selective interactions of ions with biological molecules in aqueous environments under relevant conditions, and provides support for the "Law of Matching Water Affinities," which invokes ion pairing to explain Hofmeister effects on proteins.

References:

(1) Hofmeister, F., *Arch Exp Pathol Pharmakol* **1888,** 24, 247-260.
(2) Inouye, K.; Kuzuya, K.; Tonomura, B., *J. Biochem.* **1998,** 123, (5), 847-852.
(3) Baldwin, R. L., *Biophys. J.* **1996,** 71, (4), 2056-2063.
(4) Kunz, W.; Lo Nostro, P.; Ninham, B. W., *Curr. Opin. Colloid Interface Sci.* **2004,** 9, (1-2), VII-VII.
(5) Pegram, L. M.; Record, M. T., *J. Phys. Chem. B.* **2007,** 111, (19), 5411-5417.
(6) Collins, K. D., *Biophys. J.* **1997,** 72, (1), 65-76.
(7) Collins, K. D., *Methods* **2004,** 34, (3), 300-311.
(8) Vrbka, L.; Vondrasek, J.; Jagoda-Cwiklik, B.; Vacha, R.; Jungwirth, P., *Proc. Natl. Acad. Sci. USA* **2006,** 103, (42), 15440-15444.
(9) Aziz, E. F.; Zimina, A.; Freiwald, M.; Eisebitt, S.; Eberhardt, W., *J. Chem. Phys.* **2006,** 124, (11).
(10) Stöhr, J., NEXAFS Spectroscopy. In *NEXAFS Spectroscopy*, Springer: New York, 1996; 20-31.
(11) Wilson, K. R.; Rude, B. S.; Smith, J.; Cappa, C.; Co, D. T.; Schaller, R. D.; Larsson, M.; Catalano, T.; Saykally, R. J., *Rev. Sci. Instrum.* **2004,** 75, (3), 725-736.
(12) Savitzky, A.; Golay, M. J. E., *Anal. Chem.* **1964,** 36, (8), 1627-&.
(13) Hermann, K.; Pettersson, L.; Casida, M.; Daul, C.; Goursot, A.; Koester, A.; Proynov, E.; St-Amant, A.; Salahub, D. *StoBe-deMon version 3.0*, Version 3.0; StoBe Software: 2005.
(14) Perdew, J. P., *Phys. Rev. B* **1986,** 33, (12), 8822-8824.
(15) Hammer, B.; Hansen, L. B.; Norskov, J. K., *Phys. Rev. B* **1999,** 59, (11), 7413-7421.

[16] Kutzelnigg, W.; Fleischer, U.; Schindler, M., The IGLO-Method: ab-initio calculation and interpretation of NMR chemical shifts and magnetic susceptibilities. In *NMR- Basic Principles and Progress*, ed.; Diehl, P., Springer Verlag: New York, 1991; Vol. 23, 165-262.
[17] Kolczewski, C.; Hermann, K., *Surf. Sci.* **2004,** 552, (1-3), 98-110.
[18] Triguero, L.; Pettersson, L.; Agren, H., *Phys. Rev. B* **1998,** 58, (12), 8097-8110.
[19] Prendergast, D.; Galli, G., *Phys. Rev. Lett.* **2006,** 96, (21).
[20] Kolczewski, C.; Puttner, R.; Plashkevych, O.; Agren, H.; Staemmler, V.; Martins, M.; Snell, G.; Schlachter, A. S.; Sant'Anna, M.; Kaindl, G.; Pettersson, L. G. M., *J. Chem. Phys.* **2001,** 115, (14), 6426-6437.
[21] Messer, B. M.; Cappa, C. D.; Smith, J. D.; Drisdell, W. S.; Schwartz, C. P.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B.* **2005,** 109, (46), 21640-21646.
[22] Messer, B. M.; Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B.* **2005,** 109, (11), 5375-5382.
[23] Cappa, C. D.; Smith, J. D.; Messer, B. M.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B.* **2006,** 110, (3), 1166-1171.
[24] Cappa, C. D.; Smith, J. D.; Messer, B. M.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B.* **2006,** 110, (11), 5301-5309.
[25] Urquhart, S. G.; Ade, H., *J. Phys. Chem. B.* **2002,** 106, (34), 8531-8538.
[26] Miessler, G. L., Tarr, Donald A., *Inorganic Chemistry*. Pearson Education: Upper Saddle River, N.J., 2004; 706.
[27] Dolg, M., *Chem. Phys. Lett.* **1996,** 250, (1), 75-79.

Chapter 3.
An X-ray Absorption Spectroscopy Investigation of the Interaction of Divalent Cations with Carboxylates

3.1 Introduction

Over a century ago, Franz Hofmeister performed studies with chicken egg protein; he found that the solubility of the proteins could be increased or decreased by the addition of various salts to the solution.[1] The "Hofmeister effect" has been observed with a variety of salts and proteins, following the pattern of the "Hofmeister series;" this phenomenon has been related to other biological effects.[2, 3] Nevertheless, the Hofmeister series remains incompletely understood despite enormous effort.[4, 5]

Recently, the "Law of Matching Water Affinities" was proposed by Collins;[6, 7] the law is based on charge densities and electrostatic arguments, and states that ions with similar hydration free energies form the most stable contact ion pairs. Previous work has examined the carboxylate group and its interaction with sodium and potassium ions;[8, 9] the hydration energy of the carboxylate is closer to that of sodium than to that of potassium, thus a stronger interaction of hydrated carboxylate with hydrated sodium is predicted. This was confirmed through near edge X-ray absorption fine structure (NEXAFS) spectroscopy and theoretical calculations; a preferential trend was found in interaction strengths, viz. $Li^+ > Na^+ > K^+$ with acetate, a model for protein backbones.[9]

In this chapter, preliminary results on the preferential interactions of various divalent cations with the carboxylate groups of the model system acetate are presented. Multivalent cations are the next challenging step to Collins' law, as there have been recent papers discussing its failure to accurately predict trends when dealing with large molecular ions, or the polarizability of multivalent ions.[10] The cations under examination are both alkaline earth metals: magnesium, calcium, and strontium, and transition metals: copper, manganese, and zinc; the divalent cations examined allowed for an experimental study of periodic trends as well as the inclusion of biologically relevant metals. A description of the data and the computational impediments follows.

3.2 Materials and Methods

Samples:

All samples were purchased from Sigma-Aldrich in a crystalline powder forms at a stated purity of 98% or higher and used without further purification. Solutions were prepared with 18 MΩ (Millipore) water to compose a solution of 1.0 M of the salt (following the form of $M(CH_3COO)_2$), or 2.0 M of the acetate and 1.0 M of the metal ion.

Experimental Details:

The NEXAFS spectra were obtained on the relevant edges at Beamline 8.0.1 of the Advanced Light Source (ALS), at Lawrence Berkeley National Lab. Carbon K-edge and oxygen K-edge spectra were collected both from the liquid jet and from the gas phase background. The carbon spectra were calibrated to HOPG peaks previously recorded by beamline scientists. Water dominates the oxygen K-edge making accurate analysis difficult, and therefore not included here. A gold mesh was used to collect I_o farther up the beamline, which was used to normalize for intensity fluctuations. The spectra were subsequently area normalized.

The experimental apparatus primarily comprises: a main chamber where the X-ray beam intersects with the microjet stream, additions for skimming and liquid nitrogen traps to reduce the pressures produced by the jet, and a differential pumping section to produce low enough pressures for windowless coupling to the beamline. The spectra were obtained using a chopped fused silica microjet of 30 μm backed a syringe pump. The signal was collected as total electron yield (TEY) with a biased copper electrode of 1.7 kV. A more detailed description of the experimental apparatus has been previously published by Wilson et al..[11]

The carbon K-edge spectra were smoothed due to a low signal to noise. The spectra shown in figures is an average of multiple scans and subsequent treatment with a five point second order Savitzky- Golay smoothing scheme;[12] it was empirically verified that this did not broaden the spectra.

Computational

The planned computational approach was to follow that of previous published work and is detailed in Chapters 4 and 5;[13, 14] this involves the sampling through molecular dynamics techniques[14] and subsequent density functional calculations with PWSCF.[15] As NEXAFS is very sensitive to changes in the immediate environment of the selected atom, reasonably correct MD sampling is crucial for accurate theoretical calculations. In the past, AMBER[16] had been used, but this was found to give a poor representation of the polarizable metals and their interactions with acetate, in particular for manganese and strontium. Below is a table of the first peak from the radial distribution functions (RDF) of the model systems of $MnCl_2$ and $SrCl_2$ with the experimental information from X-ray diffraction data.[17] As is shown in the data, classical molecular dynamics did not adequately describe the first peak, which represents the distances between the metal (M) and the anion. In comparison, the Vienna *ab initio* simulation package (VASP)[18, 19] yielded improved agreement with the first peak of the RDF. Unfortunately, at the time of writing, producing 100 uncorrelated molecular dynamics snapshots is difficult using *ab initio* techniques. An approach which could be used in future situations for uncorrelated snapshots for larger systems using *ab initio* techniques would be to take uncorrelated starting structures and running 100 short molecular dynamics trajectories, as opposed to one larger one.

	Strontium ($SrCl_2$)	**Manganese ($MnCl_2$)**
Experimental RDF 1st peak[17]	2.64	2.21
AMBER (% error)	2.85 (8.0%)	2.42 (9.5%)
VASP (% error)	2.59 (1.9%)	2.24 (1.4%)

Table 1. A comparison between the classical molecular dynamics, AMBER, and the *ab initio* molecular dynamics, VASP by examining agreement of the first peak in an experimental radial distribution function.

3.3 Results and Discussion

In the course of studies of monovalent cations, a correlation was verified between the hydration energy and the strength of the interaction.[8, 9] In the examination of the divalent cations, the hydration energies were consulted, and are recorded in Table 2. The trends indicate strontium as having the strongest affinity for the acetate, according to Collins' predictions, as the hydration energies are the most similar in magnitude. The alkaline metals will increase in preference down

the column, whereas the ordering for the transition metals in preference should be manganese, zinc, and then copper.

Group	Hydration Energy (ΔH)
Magnesium (Mg^{2+})	-1931 kJ/mol
Calcium (Ca^{2+})	-1584 kJ/mol
Strontium (Sr^{2+})	-1452 kJ/mol
Manganese (Mn^{2+})	-1856 kJ/mol
Copper (Cu^{2+})	-2105 kJ/mol
Zinc (Zn^{2+})	-2052 kJ/mol
Acetate (CH_3COO^-)	-423 kJ/mol

Table 2. A comparison of hydration energies for the given metals. The experimental values are from Marcus' book *Ion Solvation*.[20]

In examining the $1s \rightarrow \pi_{CO}^*$ feature from the acetate for the alkaline earth metals, there appears to be no distinct trend for the strength of the interaction, as seen in Figure 1. Previous work suggested that the magnitude of the blue shift of the transition would correlate with the strength of the interaction[9] yet no discernable shifts are evident. Size arguments have also been made in the past,[21] such that she smallest ions would have the strongest interaction, viz. the order of preference would be $Mg^{2+}>Ca^{2+}>Sr^{2+}$, but neither does this trend fit the experimental data.

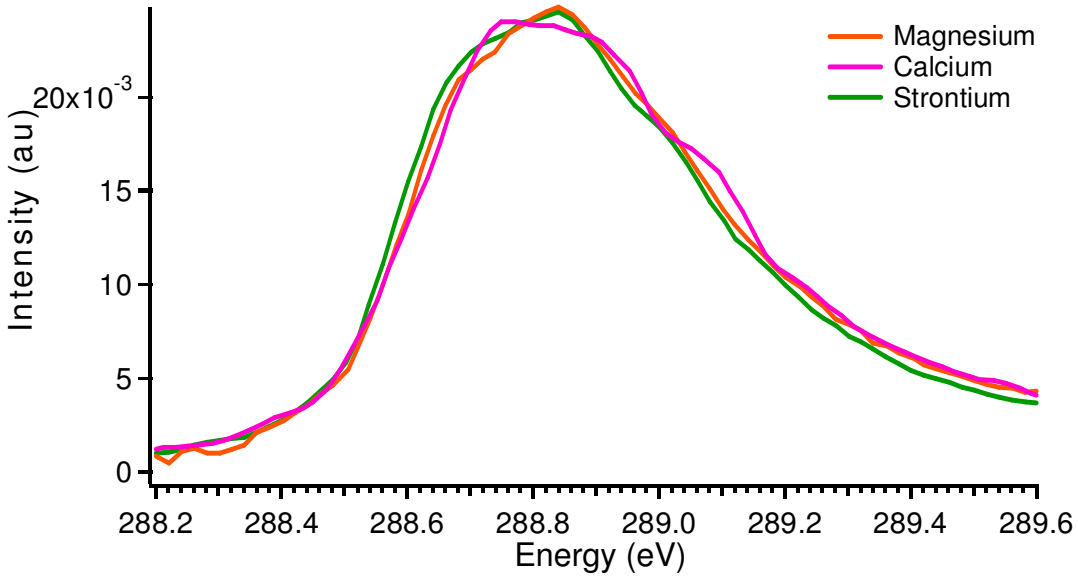


Figure 1. The carbon K-edge π* feature of the acetate of the alkaline metals, magnesium (orange), calcium (pink), and strontium (green).

In examining the transition metals, the expected trend in interaction strength, and thus blue shift, as stated before, would be manganese, copper, and then zinc. Figure 2 is a close-up on the carbon K-edge of the π* feature of the carboxylate in acetate solutions. The trend does not follow this prediction; instead the order $Zn^{2+}< Mn^{2+}<Cu^{2+}$. A more distinct trend is evident for the transition metals than for the alkaline earth metals. The order of the preference of the

interaction may be different due to the number of waters that can complex with each of the metals, the possibility of one or two acetates interacting directly with the metal, and the polarizability of the water and the metals themselves.

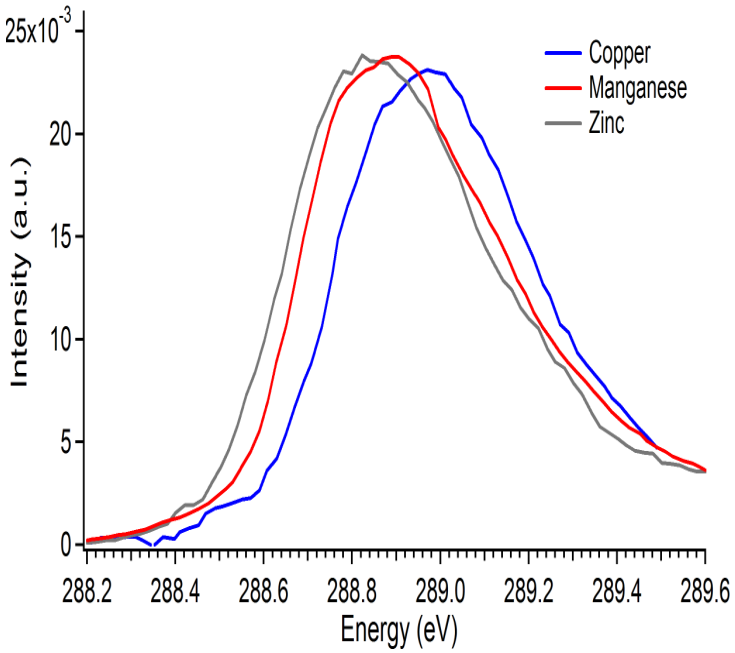


Figure 2. The carbon K-edge π* feature of the acetates of the transition metals, copper (blue), manganese, (red) and zinc (gray).

In order to understand the exceptions to the predicted trends, it is important to have information about the most common arrangements of the metals with water and acetate. This requires accurate molecular dynamics sampling, which is beyond the scope of this thesis work. . A further experimental study would be to examine concentration effects, in that an increase in the acetate concentration may impact the number of anions directly associating with the cations, shifting the average experimental spectrum.

References:

(1) Hofmeister, F., *Arch Exp Pathol Pharmakol* **1888,** 24, 247-260.
(2) Inouye, K.; Kuzuya, K.; Tonomura, B., *Journal of Biochemistry* **1998,** 123, (5), 847-852.
(3) Baldwin, R. L., *Biophysical Journal* **1996,** 71, (4), 2056-2063.
(4) Kunz, W.; Lo Nostro, P.; Ninham, B. W., *Current Opinion in Colloid & Interface Science* **2004,** 9, (1-2), VII-VII.
(5) Pegram, L. M.; Record, M. T., *J. Phys. Chem. B* **2007,** 111, (19), 5411-5417.
(6) Collins, K. D., *Biophysical Journal* **1997,** 72, (1), 65-76.
(7) Collins, K. D., *Methods* **2004,** 34, (3), 300-311.
(8) Vrbka, L.; Vondrasek, J.; Jagoda-Cwiklik, B.; Vacha, R.; Jungwirth, P., *Proc. Natl. Acad. Sci. U. S. A.* **2006,** 103, (42), 15440-15444.
(9) Uejio, J. S.; Schwartz, C. P.; Duffin, A. M.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (19), 6809-6812.
(10) Fedorov, M. V.; Goodman, J. M.; Schumm, S., *J. Am. Chem. Soc.* **2009,** 131, (31), 10854-+.

[11] Wilson, K. R.; Rude, B. S.; Smith, J.; Cappa, C.; Co, D. T.; Schaller, R. D.; Larsson, M.; Catalano, T.; Saykally, R. J., *Rev. of Scientific Instrum.* **2004,** 75, (3), 725-736.
[12] Savitzky, A.; Golay, M. J. E., *Anal. Chem.* **1964,** 36, (8), 1627-&.
[13] Uejio, J. S.; Schwartz, C. P.; Saykally, R. J.; Prendergast, D., *Chem. Phys. Lett.* **2008,** 467, (1-3), 195-199.
[14] Schwartz, C. P.; Uejio, J. S.; Saykally, R. J.; Prendergast, D., *J. Chem. Phys.* **2009,** 130, (18), 184109.
[15] Baroni, S.; Corso, A. D.; de Gironcoli, S.; Gianozzi, P. PWSCF. www.pwscf.org
[16] D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R.; Luo, K. M. M., D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S.; Hayik, A. R., G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H.; Gohlke, L. Y., C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C.; Schafmeister, W. S. R., and P.A. Kollman *Amber 9*, 2006.
[17] Magini, M., *X-ray Diffraction of Ions in Aqueous Solutions: Hydration and Complex Formation*. CRC Press: Boca Raton, FL, 1988.
[18] Kresse, G.; Furthmuller, J., *Phys. Rev. B* **1996,** 54, (16), 11169-11186.
[19] Kresse, G.; Hafner, J., *Phys. Rev. B* **1993,** 48, (17), 13115-13118.
[20] Marcus, Y., *Ion Solvation*. John Wiley & Sons Ltd.: New York, 1985.
[21] Fennell, C. J.; Bizjak, A.; Vlachy, V.; Dill, K. A., *J. Phys. Chem. B* **2009,** 113, (19), 6782-6791.

Chapter 4.
Effects of vibrational motion on core-level spectra of prototype organic molecules

4.1 Introduction and Methods

When applied to molecular systems, core level spectroscopies are powerful probes of both occupied and unoccupied electronic states, uniquely revealing intimate details of both intra- and inter-molecular interactions.[1] Methods involving X-ray absorption (XAS, NEXAFS, XANES) or X-ray photo-electron spectroscopy (XPS) are increasingly being applied to complex molecular systems, including nucleotides, peptides and large organic molecules.[2] However, a major limitation of this technology is the fact that extraction of molecular information from these experiments often depends explicitly on comparisons with theoretical calculations, which are extremely challenging to perform at experimental accuracy. This chapter describes the extension of a recently developed method for predicting core-level spectra of condensed phases[3] to isolated organic molecules -- pyrrole, *s*-triazine, pyrrolidine and glycine – which demonstrates qualitative improvements over existing methods[4-6] in comparison with experiment and provides new insights into the origins of particular spectral features in terms of coupling of electronic and vibrational degrees of freedom.

The challenges for simulating gas phase core-level spectra are maintaining accuracy in the following areas: (1) description of the core-hole excited state; (2) representation of both bound excitonic states below the ionization potential (IP) and resonance states in the continuum above the IP; and (3) inclusion of vibrational effects, either due to experiments being performed near room temperature, or from intrinsic zero-point motion.

Density functional theory (DFT)[7, 8] has proved accurate in reproducing the excitation energies associated with core-level spectra via total energy differences (so-called ΔSCF or ΔKS).[9] Accordingly, the lowest energy core-level excited is modeled state self-consistently using a full core-hole and excited electron (XCH).[3] This is particularly important for molecular systems, where screening of the core-hole excitation is greatly enhanced by the presence of the excited electron, which can be strongly bound to the core-hole in the lowest energy excited state. In contrast, for non-molecular condensed phases, such as covalent and ionic crystalline solids, the inherent dielectric screening of the valence charge density often dominates, and so, explicit inclusion of the excited electron may prove insignificant in such cases.[5] The calculation uses the PBE form of the generalized gradient approximation to the exchange-correlation potential.[10] Transition amplitudes are estimated in the single-particle and dipole approximations and excitations to states above this first excited state are approximated using the unoccupied Kohn-Sham eigenstates computed from the XCH self-consistent potential. This is in contrast to the closely related full core hole (FCH) approximation,[4, 5] which ignores the excited electron, or replaces it with a uniform background charge density, and the half core hole (HCH) approach[6] related to Slater's transition-state potential (TP).[11] The HCH (or TP) approach has been applied extensively to molecular and cluster models of materials using linear combinations of atomic orbitals (LCAO) to describe the electronic structure. These have included applications to isolated molecules,[9] molecules on surfaces,[12] and condensed phase molecular liquids.[13] In our XCH implementation,[3] norm-conserving pseudopotentials are used. Core-hole matrix elements with valence electrons are calculated by reconstructing the core region of the pseudostates within an atomic frozen core approximation. Other approaches often treat the core-excited atom at the all-electron level, while using effective core potentials for the surrounding unexcited atoms.

There are other approaches to modeling core-level excitations: The static exchange approach (STEX) of Ågren[14] describes final state of the core-level excitation by freezing the orbitals of the molecular ion and calculates their exchange interaction with the excited electron; The multiple scattering approach of Ankudinov and Rehr[15] has been used extensively to examine the X-ray absorption of compounds at the edges of (typically) heavier elements using real-space cluster models;[16] accurate solutions to the Bethe-Salpeter equation have been used effectively by Shirley[17] in the context of core-level spectra of crystalline solids.

The XCH approach is applied within periodic boundary conditions using a plane-wave basis, enabling uniform convergence in accuracy for representing both localized and delocalized states. In contrast, LCAO approaches have difficulty in describing delocalized states, particularly those appearing above the IP in core-level spectra of isolated molecules. This has been mitigated in the past using Stieltjes imaging techniques[18] or some *ad hoc* numerical broadening. However, in our approach, the use of plane-waves to represent electronic states of isolated molecules engenders a significant computational cost. Spurious interactions between periodic images must be reduced by increasing the size of the supercell, which, in turn, increases the size of the basis and the density of electronic states to be determined at energies above the ionization potential. Our compromise is to use supercells large enough to represent the excited states below the IP, close to the absorption onset. Then full advantage of the periodic boundary conditions is taken to approximate the high energy continuous electronic density of states by numerically converging an integration over the first Brillouin zone (BZ). This will have no impact on bound states localized fully within the supercell. However, for states which span the supercell, an accurate determination of the electronic density of states can be achieved by BZ sampling.[19] Such delocalized states should be close analogues of free electron states scattered from the molecule. The weakness of this approach is in describing bound states below the IP having a spatial extent larger than the supercell, but increasing the size of our supercells mitigates this effect.

For each of the molecules studied in this work supercell volumes of $(20)^3$ Å^3 and a plane-wave kinetic energy cut-off of 85 Ry were used. In all cases, approximately 100 Kohn-Sham eigenstates are used in constructing transition matrix elements. This is only sufficient to extend our spectra approximately 3 eV above the estimated IP. To reduce the significant computational cost of a numerically converged BZ sampling, a recently implemented interpolation scheme (based on an approach by Shirley[20]) is exploited that requires only the electronic states at the zone center as input. Furthermore, this scheme also increases the number of electronic states beyond the ~ 3 eV limit. The accuracy of these states is not guaranteed (see fixed-nuclei spectra in Figure 1), but finite-temperature sampling improves the agreement with experiment. The zone-center electronic structure is calculated using the PWSCF code.[21]

Typically, core-level spectra of isolated molecules are simulated within the fixed-nuclei approximation, particularly for molecules in their vibrational ground state under the experimental conditions. A candidate (lowest energy) structure is chosen and the electronic structure is calculated while modeling the atomic nuclei as fixed point charges, located at the mean of the nuclear distribution or, more commonly, at an energy minimum derived from an formalism which models the electrons as quantum particles and the nuclei as classical point charges. Usually, the core-excitation transition amplitudes are estimated without the impact of nuclear dynamics on the electronic subsystem. Often finite temperature effects are approximated by increased numerical broadening of calculated spectral peaks. More detailed approaches calculate Franck-Condon factors based on a vibrational eigenmode analysis in the ground and excited

states. These factors are used to modulate a single electronic transition and help to reproduce asymmetric lineshapes associated with accompanying transitions to excited vibrational modes.[9, 22, 23] However, the Franck-Condon approximation ignores the impact of nuclear motion on the electronic transition amplitude. To first order, this impact is referred to as the Herzberg-Teller effect.[24]

In this chapter, I attempt to include the impact of nuclear motion on core-level excited electronic states and transition amplitudes. The nuclear degrees of freedom in these molecules are modeled using classical molecular dynamics (MD) performed at 300 K using a Langevin thermostat. I used AMBER 9 with the generalized AMBER force field and Antechamber.[25, 26] The resulting distribution of nuclear coordinates is sampled at regular intervals, spaced at least 10 ps apart to reduce correlation between nuclear snapshots, for at least 100 snapshots. For the small molecules studied here, the molecular dynamics calculations represent an insignificant computational overhead with respect to the 100 or more plane-wave electronic structure calculations required to simulate the core-level spectra. For larger systems, or for first principles molecular dynamics sampling, such long trajectories may not be computationally tractable. In this case, Monte Carlo sampling may prove more efficient in sampling accessible molecular configurations, but it was not used in this work. I recognize that for each of these molecules, some (or all) of the vibrational eigenmodes (estimated using DFT calculations in good agreement with experiment), are in their quantum ground states at room temperature and will exhibit systematically different spatial distributions from the classical model. The use of a Langevin thermostat leads to distributions of nuclear coordinates which resemble quantum distributions (they are not peaked at the classical turning points), although their mean-square displacements are typically underestimated. Nevertheless, using the computationally inexpensive uncorrelated sampling approach, significant improvements in experimental agreement are possible over using only the vibrationless mean nuclear positions of the fixed-nuclei approximation. For molecules occupying their vibrational ground states at experimental temperatures, this improved agreement is an indicator that features missing from the fixed-nuclei spectra, appear in experiment due to zero-point vibronic coupling effects, which might be well-reproduced by the Herzberg-Teller approximation. Evidence for Herzberg-Teller effects is seen in the spectra of *s*-triazine and glycine and future work will address these in more detail.

The approach of sampling molecular dynamics trajectories has been applied already in simulating core-level spectra of molecular clusters and liquids, particularly for photoelectron spectroscopy,[27, 28] X-ray absorption spectroscopy,[3] and X-ray emission spectroscopy,[13] where distinct changes have been observed based on configurational changes. For molecules with multiple low-energy conformations or significant anharmonicity, the molecular dynamics approach has clear advantages over vibrational eigenmode analysis. Calculation of eigenmodes is only possible about a minimum in the potential energy surface and for more complex systems there can be many such minima. This is seen for the case of glycine, shown below. All calculated spectra are numerically broadened using Gaussians of 0.2 eV full width at half maximum (FWHM). Previous simulations have used larger and nonuniform numerical broadening in order to simultaneously approximate electronic and vibronic coupling. For example, in Ref. 9 a 0.3 eV FWHM broadening was used below the IP, and this was linearly increased to 4.5 eV for the next 30 eV and then held constant. (In the same work a smaller broadening of 0.15 eV is used when discussing vibronic effects.) In contrast, a relatively small and uniform broadening was used with the aim of simulating and distinguishing electronic and vibrational effects explicitly; thereby arriving at a predictive computational approach.

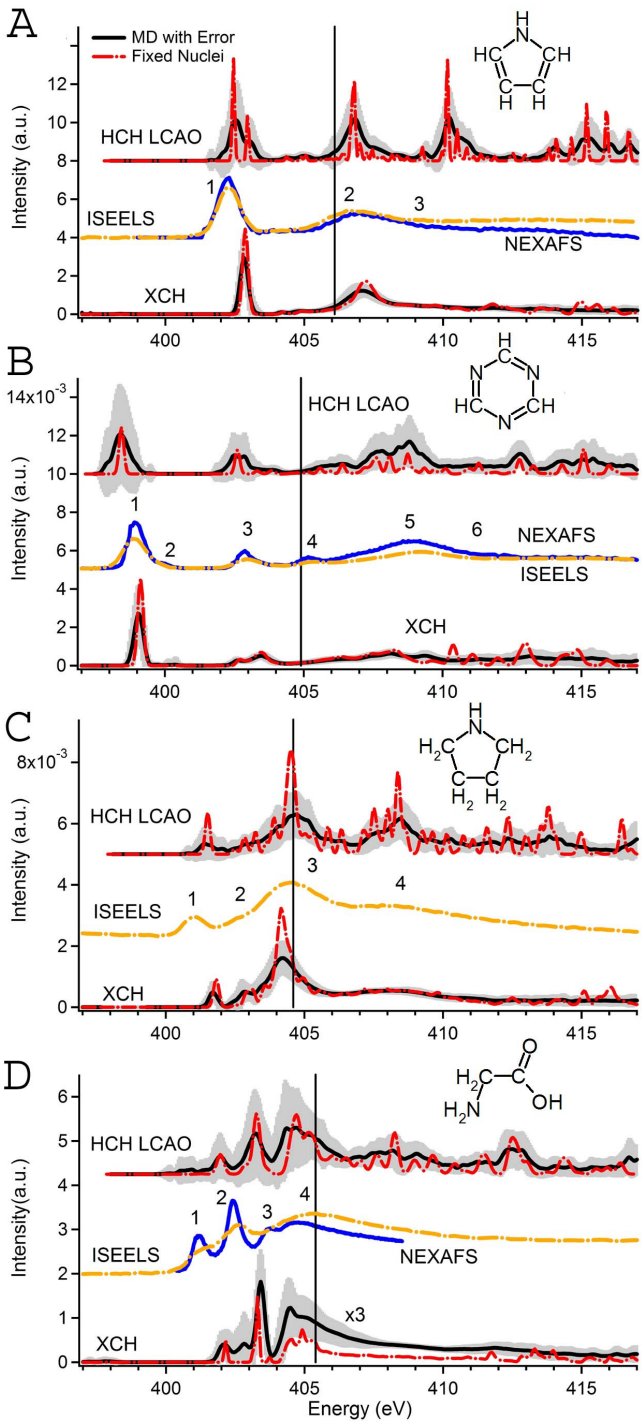


Figure 1. Comparison of gas phase experimental and calculated core-level spectra of (A) pyrrole, (B) *s*-triazine, (C) pyrrolidine, and (D) glycine. NEXAFS (solid blue) and ISEELS (dash-dot orange) data are from previous experiments (see text) except for pyrrole NEXAFS. Calculated spectra using HCH LCAO (top) and XCH (bottom) are compared with (solid black) and without vibrations (dash-dot red), with standard deviations indicated by gray shading. Vertical lines indicate measured or calculated IP positions.

Figure 1 provides comparison between measured core-level spectra from NEXAFS and inner-shell electron energy loss (ISEELS) and calculated spectra using the HCH LCAO and

XCH approaches both with and without sampling of nuclear degrees of freedom. All calculated spectra are aligned by their IP estimates. An observable systematic contraction of XCH spectra along the energy axis is seen, consistent with previous work.[3, 29] The HCH LCAO calculations were performed with a commercially available package, StoBe deMon.[30] The excited nitrogen of interest was modeled using the IGLO-III basis set,[31] the hydrogens were modeled with a diffuse basis set and the remaining non-hydrogen atoms were modeled with double zeta valence plus polarization basis sets included in the StoBe deMon package. The same PBE functional was used in both XCH and HCH calculations. I found that using larger basis sets and functionals led to only minor spectral changes in HCH calculations.

4.2 Results and Discussion

Pyrrole

Found in biological systems (Fig. 1(a)), this molecule has an aromatic rigid ring structure and occupies its vibrational ground state at the experimental temperature. The NEXAFS experimental data were collected using total electron yield at beamline 8.0.1 at the Advanced Light Source, using prior methods[32] and the ISEELS was previously taken.[33] Peaks 1 and 2 are reproduced well by XCH without vibrations, but with an incorrect peak height ratio. HCH LCAO calculations produce a spurious peak after the first main peak and the second peak is too intense compared to experiment. Previous experiment has identified a shape resonance,[34] labeled 3 here, which has some oscillator strength above the decay that occurs well above the ionization potential; some features are apparent there, but nothing definitive emerges from either calculation.

Averaging 100 MD snapshots, XCH is able to produce an improved intensity ratio between features 1 and 2 and a smoother continuum region above 410 eV. MD sampling broadens out the features in the HCH LCAO calculation, but incorrect peak height ratios and an overly structured continuum region remain. The latter is most likely due to basis set limitations.

S-triazine

A very rigid prototypical aromatic molecule; much like pyrrole, this molecule is in its vibrational ground state at the experimental temperature. The fixed-nuclei XCH spectrum (Fig. 1(b)) does not compare well with experiment.[35, 36] Peak 1 is reproduced by XCH but with overestimated oscillator strength. The small shoulder labeled 2 corresponds to the LUMO(+1) and is attributed to vibronic coupling. HCH LCAO and XCH are unable to reproduce this peak without inclusion of vibrations, but they capture peak 3. Both methods are plagued by spurious features in the continuum region in the absence of vibrations.

It was expected that using MD snapshots would produce only small changes due to *s*-triazine being in the vibrational ground state at the experimental temperatures. However, inclusion of MD sampling results in a clear improvement with extra broadening induced by small displacements of the nuclei. The XCH approach more accurately captures peaks 1, 3, 5 and 6; the continuum is smooth and in better agreement with experiment. HCH LCAO is similarly broadened, with large sampling error bars around peak 1; the continuum region remains overly structured.

Peak 2 is evident as a large sampling error bar from MD sampling and is visible for individual snapshots for both HCH LCAO and XCH calculations. One such snapshot is examined in detail in Figure 2. Here, two electronic transitions which are forbidden in the absence of vibrations, are turned on for displaced nuclei. Breaking the in-plane mirror symmetry modifies the electronic structure at the excited nitrogen atom. For the LUMO(+1), this disrupts a

nodal plane through the excited nitrogen, while for the LUMO(+2) a ring-like state in the molecular plane shifts adding to the p-character at the excited nitrogen.

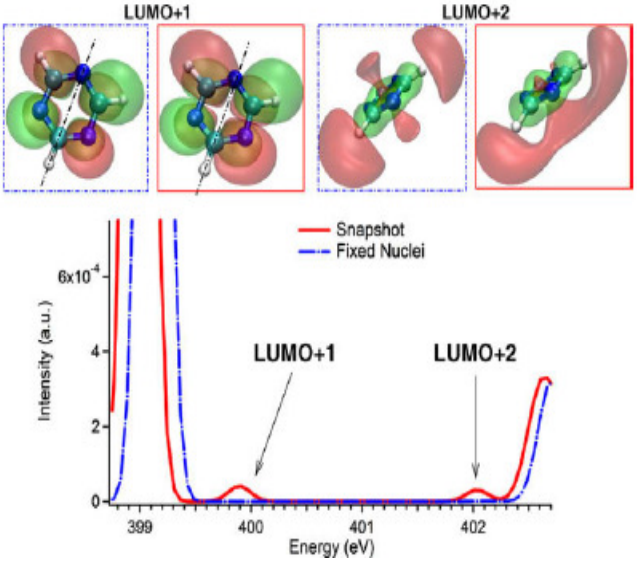


Figure 2. Bottom: Initially dark electronic transitions near the N K-edge onset of *s*-triazine (dash-dot blue), which become allowed (red) upon inclusion of deviations from mean nuclear positions of the vibrational ground state, corresponding to the first 2 states above the lowest unoccupied molecular orbital (LUMO). Top: Isosurfaces of the LUMO(+1) (left) and LUMO(+2) (right) with (red) or without (dash-dot blue) nuclear displacements; green (red) indicates positive (negative) values.

Note that none of our calculations reproduce peak 4. It is believed, by comparison with a similar feature in molecular nitrogen,[1, 35] that this is multi-electron in origin. Our calculations do not include excitations of more than one electron. However, I have performed test calculations which include an additional electron excitation from near the top of the valence band (shake-up), indicating some transitions in the right energy range, such as HOMO(-2) or HOMO(-3) to LUMO.

Pyrrolidine

Structurally similar to pyrrole but non-aromatic, pyrrolidine has an active low frequency vibrational mode centered on its nitrogen atom. The XCH fixed-nuclei approximation predicts three distinct large peaks, corresponding to the three experimental features (1,3, and 4 respectively);[33] it also produces the small shoulder to the main feature centered at 404 eV, labeled 2. Using HCH LCAO without vibrations produces unrealistically sharp features over the entire spectrum, particularly around peaks 2, and 3.

Introducing vibrations via MD sampling greatly improves experimental agreement. This is expected due to the population of excited vibrational modes at experimental temperatures localized near the core-hole excitation. The XCH method accurately reproduces the continuum region in contrast to the overstructured continuum resulting from HCH LCAO with MD. The broadening of peak 3 in both spectra shows that the peak width is due not to core-hole lifetime effects but rather to the variety of structures that exist at room temperature. The shoulder at peak 2 is reproduced using XCH.

Glycine

In Figure 1(d), the spectrum of glycine, the simplest amino acid, comprises two well resolved peaks, 1 and 2, a smaller less defined peak 3, and a broad peak 4.[37, 38] In the absence of vibrations, the XCH method reproduces all 4 peaks, as does the traditional HCH-LCAO. However, neither method yields the correct general shape of the experimental spectra. Note that the MD averaged XCH spectrum has been rescaled by a factor of three.

A recent study on core-level spectra of glycine, examined only its four lowest energy conformers, and was expected to represent all individual conformers with populations >2% of the total.[38] Due to their atomistic sensitivity, core-level spectra can be strongly influenced by sparsely populated conformations.[39] By comparison with a Boltzmann-averaged spectrum of the four dominant conformers (not shown) it was apparent, in this case, that merely analyzing the four lowest energy conformers might not be adequate to accurately reproduce core-level spectra.

MD sampling has the advantage of not merely sampling the lowest energy state conformers, but also phase space the molecule would have to occupy as it changes conformations. Including MD sampling in HCH LCAO calculations produces a spurious low energy feature with low intensity, a second peak that is much too intense, and the fourth feature is not resolved. However, the overall spectral appearance is much closer to that of experiment, particularly in the low energy range. At higher energy, the same nonphysical peaks present for the fixed-nuclei structure are also present with MD. Using the IP for alignment is beneficial, providing objective alignment when the first peak is not experimentally observed.

Using a 0.2 eV FWHM broadening scheme there appears to be an extra peak just above the onset in the XCH MD spectrum, however when the peak widths are increased to 0.4 eV FWHM (Fig. 3), this extra peak merges with the onset, and the measurement is reproduced. This indicates insufficient or incorrect sampling of the full configuration space of glycine using our classical MD approach. Perhaps more sampling might broaden this new peak correctly. This example indicates how tempting it can be to broaden theoretical spectra in order to "fit" to experiment, at the sacrifice of predictability and correct interpretation. At higher energies, the XCH approach performs reasonably for glycine, where there is a reasonable decay, devoid of any unphysically sharp features.

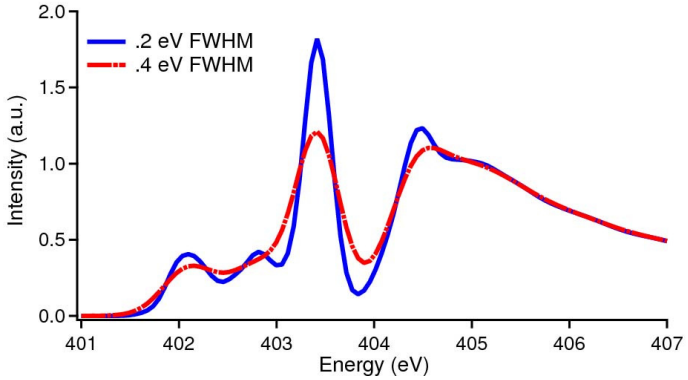


Figure 3. The effect of numerical broadening on the calculated XCH spectrum of glycine at the nitrogen K edge. Broadening by convolution with a Gaussian of 0.2 eV (0.4 eV) FWHM is show in blue (red). The larger broadening obscures the second peak at 402.8 eV.

Using a finer 0.2 eV FWHM numerical resolution, it is clear from Figure 1(d) that the "new" spectral feature between the measured peaks 1 and 2 is not present in the absence of vibrations, for the lowest energy structure of glycine. A survey of those molecular configurations

within our molecular dynamics ensemble which contain this extra peak indicates that it results from a non-zero dihedral angle along the molecular backbone from the nitrogen atom to the carbonyl oxygen (NCC=O), which breaks the mirror symmetry of glycine. Characterization of the first four core-level electronic transitions is provided in Figure 4. Using the XCH approach, it is found the first allowed transition of the lowest energy structure has σ_{NH}^* character, in agreement with previous work[37] using the STEX approach.[14] However, the agreement stops here. The next two allowed transitions, computed using XCH, have σ_{NC}^* character and $2b_2$-like symmetry respectively. (I imagine the NH_2 group is like the water molecule, which also possesses a true $2b_2$ unoccupied state.) Each of these three transitions persists upon variation of the NCC=O dihedral angle, albeit with some small energy-reordering and shifting in position and oscillator strength which can be understood by analysis of expansion or contraction of the excited electron density in the presence of these structural perturbations.

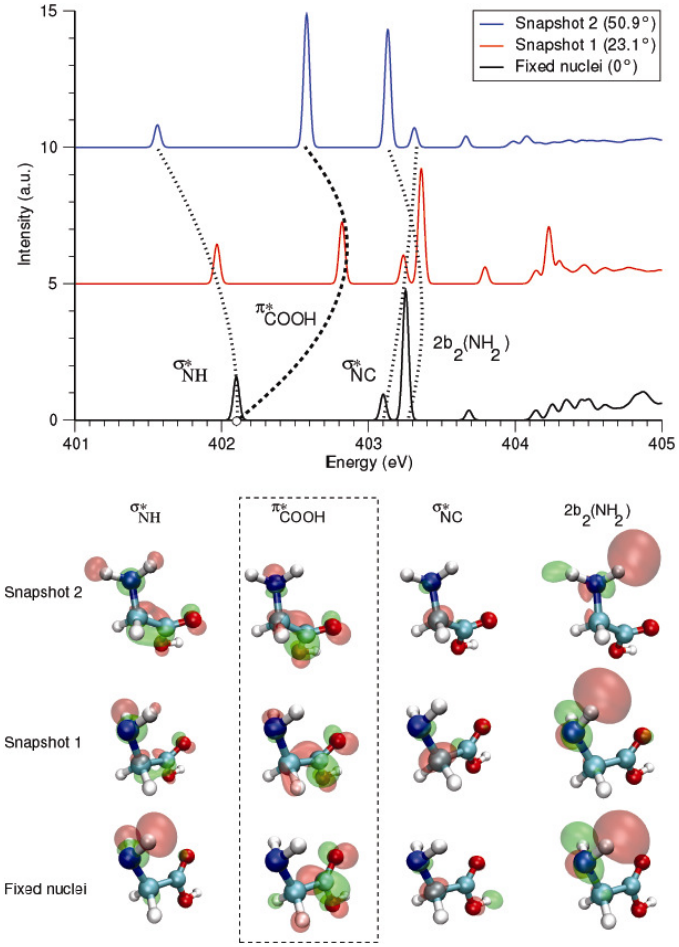


Figure 4.Top: A comparison of spectra of glycine conformations with increasing NCC=O dihedral angle: the lowest energy (fixed-nuclei) conformation with a 0° dihedral angle (black), and two snapshots sampled from a 300 K molecular dynamics trajectory with dihedral angles of 23.1° (red) and 50.9°(blue). Spectral features with similar symmetry indicated are joined by dotted lines. A dashed line indicates the forbidden transition which becomes allowed for nonzero dihedral angles. Bottom: Molecular orbital isosurfaces corresponding to each excited electronic state of specified symmetry (see text) for each of the molecular conformations mentioned above. The forbidden transition of the fixed-nuclei structure is highlighted along

with the allowed states of similar symmetry from the perturbed structures. Green (red) indicates positive (negative) values of the electronic wave function.

However, there also exists an optically forbidden transition in the lowest energy structure, only 14 meV higher in energy than the first excited state. From the electronic wave function isosurfaces, this forbidden state is of π_{COOH}^{*} character and is forbidden due to its lack of significant overlap with the core-excited nitrogen. Upon increasing the dihedral angle, this state leaks onto the core-excited nitrogen leading to an allowed transition with large oscillator strength, lying in energy in the gap between the first and second allowed transitions of the lowest energy structure. The variation in energy position and oscillator strength of this transition indicates how it leads to a broad peak intermediate between features 1 and 2 of the experiment, which in our calculations is no longer resolvable at 0.4 eV FWHM. The connection between the strength and position of this transition and the angle of the carboxyl group relative to the amine is likely what leads to the broad, featureless peak observed for glycine both in crystalline solid and solvated phases.

In conclusion, this chapter has demonstrated an accurate computational approach for prediction and interpretation of core-level molecular spectra, which we have applied to isolated organic molecules. Our plane-wave calculations permit faithful representations of both bound and unbound electronic excited states, thereby producing accurate spectral lineshapes both above and below the ionization potential (in contrast to those approaches which use localized basis sets). By sampling vibrational degrees of freedom using molecular dynamics, significant improvements with respect to experiment are observed. It is now possible to isolate the vibrational contribution to the broadening of spectral features. In particular cases, where molecules are in their vibrational ground state, improvements due to molecular dynamics sampling indicate that zero-point effects have significant impacts on core-level spectra, enabling new electronic transitions which are forbidden when nuclear quantum effects are ignored. Future work in which the nuclear degrees of freedom are sampled correctly with respect to their quantum ground state distribution will further illuminate this issue.

References:

(1) Stöhr, J., *NEXAFS Spectroscopy*. Springer: New York, 1996; 20-31.
(2) Stewart-Ornstein, J.; Hitchcock, A. P.; Cruz, D. H.; Henklein, P.; Overhage, J.; Hilpert, K.; Hale, J. D.; Hancock, R. E. W., *J. Phys. Chem. B* **2007,** 111, (26), 7691-7699.
(3) Prendergast, D.; Galli, G., *Phys. Rev. Lett.* **2006,** 96, (21), 21522-21526.
(4) Cavalleri, M.; Ogasawara, H.; Pettersson, L. G. M.; Nilsson, A., *Chem. Phys. Lett.* **2002,** 364, (3-4), 363-370.
(5) Taillefumier, M.; Cabaret, D.; Flank, A. M.; Mauri, F., *Phys. Rev. B* **2002,** 66, (19).
(6) Triguero, L.; Pettersson, L.; Agren, H., *Phys. Rev. B* **1998,** 58, (12), 8097-8110.
(7) Hohenberg, P.; Kohn, W., *Phys. Rev. B* **1964,** 136, (3B), B864.
(8) Kohn, W.; Sham, L. J., *Phys. Rev.* **1965,** 140, (4A), 1133.
(9) Kolczewski, C.; Puttner, R.; Plashkevych, O.; Agren, H.; Staemmler, V.; Martins, M.; Snell, G.; Schlachter, A. S.; Sant'Anna, M.; Kaindl, G.; Pettersson, L. G. M., *J. Chem. Phys.* **2001,** 115, (14), 6426-6437.
(10) Perdew, J. P.; Burke, K.; Ernzerhof, M., *Phys. Rev. Lett.* **1996,** 77, (18), 3865-3868.
(11) Slater, J. C., *Adv. Quant. Chem.* **1972,** 6, 1.

[12] Ostrom, H.; Ogasawara, H.; Naslund, L. A.; Andersson, K.; Pettersson, L. G. M.; Nilsson, A., *J. Chem. Phys.* **2007,** 127, (14).
[13] Tokushima, T.; Harada, Y.; Takahashi, O.; Senba, Y.; Ohashi, H.; Pettersson, L. G. M.; Nilsson, A.; Shin, S., *Chem. Phys. Lett.* **2008,** 460, (4-6), 387-400.
[14] Agren, H.; Carravetta, V.; Vahtras, O.; Pettersson, L. G. M., *Chem. Phys. Lett.* **1994,** 222, (1-2), 75-81.
[15] Ankudinov, A. L.; Rehr, J. J., *Phys. Rev. B* **2000,** 62, (4), 2437-2445.
[16] Ankudinov, A. L.; Ravel, B.; Rehr, J. J.; Conradson, S. D., *Phys. Rev. B* **1998,** 58, (12), 7565-7576.
[17] Shirley, E. L., *Phys. Rev. Lett.* **1998,** 80, (4), 794-797.
[18] Langhoff, P. W., *Electron-Molecule and Photon-Molecule Collisions*. Plenum: New York, 1979.
[19] Prendergast, D.; Grossman, J. C.; Galli, G., *J. Chem. Phys.* **2005,** 123, (1), 14501-14505.
[20] Shirley, E. L., *Phys. Rev. B* **1996,** 54, (23), 16464-16469.
[21] Baroni, S.; Corso, A. D.; de Gironcoli, S.; Gianozzi, P. PWSCF. www.pwscf.org
[22] Minkov, I.; Gel'mukhanov, F.; Agren, H.; Friedlein, R.; Suess, C.; Salaneck, W. R., *J. Phys. Chem. A* **2005,** 109, (7), 1330-1336.
[23] Minkov, I.; Gel'mukhanov, F.; Friedlein, R.; Osikowicz, W.; Suess, C.; Ohrwall, G.; Sorensen, S. L.; Braun, S.; Murdey, R.; Salaneck, W. R.; Agren, H., *J. Chem. Phys.* **2004,** 121, (12), 5733-5739.
[24] Herzberg, G.; Teller, E., *Z. Phys. Chem. B-Chem. Elem. Aufbau. Mater.* **1933,** 21, (5/6), 410-446.
[25] Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., *J. Comput. Chem.* **2005,** 26, (16), 1668-1688.
[26] Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A., *J. Mol. Graph. Model.* **2006,** 25, (2), 247-260.
[27] Abu-samha, M.; Borve, K. J.; Saethre, L. J.; Ohrwall, G.; Bergersen, H.; Rander, T.; Bjorneholm, O.; Tchaplyguine, M., *Phys. Chem. Chem. Phys.* **2006,** 8, (21), 2473-2482.
[28] Felicissimo, V. C.; Minkov, I.; Guimaraes, F. F.; Gel'mukhanov, F.; Cesar, A.; Argen, H., *Chemical Physics* **2005,** 312, (1-3), 311-318.
[29] Lin, J. F.; Fukui, H.; Prendergast, D.; Okuchi, T.; Cai, Y. Q.; Hiraoka, N.; Yoo, C. S.; Trave, A.; Eng, P.; Hu, M. Y.; Chow, P., *Phys. Rev. B* **2007,** 75, (1).
[30] Hermann, K.; Pettersson, L.; Casida, M.; Daul, C.; Goursot, A.; Koester, A.; Proynov, E.; St-Amant, A.; Salahub, D. *StoBe-deMon version 3.0*, Version 3.0; StoBe Software: 2007.
[31] Kutzelnigg, W.; Fleischer, U.; Schindler, M., The IGLO-Method: ab-initio calculation and interpretation of NMR chemical shifts and magnetic susceptibilities. In *NMR- Basic Principles and Progress*, ed.; Diehl, P., Springer Verlag: New York, 1991; **23**, pp 165-262.
[32] Wilson, K. R.; Rude, B. S.; Catalano, T.; Schaller, R. D.; Tobin, J. G.; Co, D. T.; Saykally, R. J., *J. Phys. Chem. B* **2001,** 105, (17), 3346-3349.
[33] Newbury, D. C.; Ishii, I.; Hitchcock, A. P., *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **1986,** 64, (6), 1145-1155.
[34] Duflot, D.; Hannay, C.; Flament, J. P.; Hubin-Franskin, M. J., *J. Chem. Phys.* **1998,** 109, (13), 5308-5318.
[35] Vall-llosera, G.; Gao, B.; Kivimaeki, A.; Coreno, M.; Ruiz, J. A.; de Simone, M.; Agren, H.; Rachlew, E., *J. Chem. Phys.* **2008,** 128, (4).

[36] Apen, E.; Hitchcock, A. P.; Gland, J. L., *J. Phys. Chem.* **1993,** 97, (26), 6859-6866.
[37] Plekan, O.; Feyer, V.; Richter, R.; Coreno, M.; de Simone, M.; Prince, K. C.; Carravetta, V., *J. Electron Spectrosc. Relat. Phenom.* **2007,** 155, (1-3), 47-53.
[38] Gordon, M. L.; Cooper, G.; Morin, C.; Araki, T.; Turci, C. C.; Kaznatcheev, K.; Hitchcock, A. P., *J. Phys. Chem. A* **2003,** 107, (32), 6144-6159.
[39] Godfrey, P. D.; Brown, R. D., *J. Am. Chem. Soc.* **1995,** 117, (7), 2019-2023.

Chapter 5.
Monopeptide vs. Monopeptoid: Insights on Structure and Hydration of Aqueous Alanine and Sarcosine via X-ray Absorption Spectroscopy

5.1 Introduction

Peptoids have recently emerged as an important subject of research.[1, 2] They can be readily synthesized, [3] and are resistant to enzymatic, chemical and thermal denaturation.[4, 5] Peptoids are currently being evaluated for numerous applications, including cell penetration for drug delivery,[6] protein-protein interaction inhibitors,[7] lung surfactants,[8] and other biomimetics.[1, 2, 9] Salient characteristics of peptoids include their robust higher-order structure, folding dynamics, increased resistance to denaturation relative to proteins, and the means by which the structure and interactions can be intelligently designed by the primary sequence of the chain.[2, 3, 10-12] Previous methods of studying peptides and peptoids have included circular dichroism of oligomers,[5, 11, 13] infrared spectroscopy,[12] NMR,[11] mass spectrometry,[14, 15] and computational modeling;[16, 17] these methods primarily focused on design and characterization of the secondary structure.

I sought to understand the structural aspects of aqueous peptoids on a fundamental level by comparing sarcosine with its complementary peptide, alanine, using near-edge X-ray absorption fine structure (NEXAFS) spectroscopy, which is sensitive to details of charge interactions, presence of counter ions, conformation, and hydration.[18-24] Comprehensive and predictive design requires an understanding of the environment and its impact on the molecule of interest, and in the case of biological applications, this necessitates addressing the aqueous systems. Previous NEXAFS studies of peptides have primarily been limited by technical requirements to experiments on the gas, solid phase, thin films, or small static samples.[23-31] To enable direct measurements of aqueous systems, liquid microjet technology was employed, where a thin stream of liquid is injected into vacuum, and is intersected by a synchrotron X-ray beam.[19, 20, 32, 33] With results from this system, I present the first direct comparison of the properties of aqueous alanine and sarcosine using NEXAFS.

To understand a solute embedded in a solvent, several factors must be initially considered, including but not limited to solute dynamics, intramolecular interactions of the solute (interactions between the carboxylate and charged nitrogen group), solute-solute, and solute-solvent interactions. As these effects are hard to separate, calculations are necessary for interpretation and predictive capability. Our first principles density functional theory (DFT) approach has been shown to be successful for addressing both isolated organic molecules and for water.[34-37] Many X-ray calculations done in the past on hydrated systems have used clusters to simulate bulk hydration;[38-42] the question arises as to the number of waters necessary to accurately represent solvation. Further, previous practice has placed waters around a moiety of interest, instead of the entire molecule but whether this is an appropriate practice has not been addressed.[18-20] Additionally, finite clusters of water possess surface electronic states due to dangling hydrogen bonds; these states are not present for calculations with periodic boundary conditions. For these reasons, a periodic simulation cell is used to model bulk hydration for X-ray calculations.

In this chapter, I present the first NEXAFS spectra of aqueous alanine and sarcosine, along with accompanying DFT calculations. I attempt to untangle the spectral impact of molecular conformation versus the strength of hydration, and contrast these effects between alanine and sarcosine. I also find that for NEXAFS simulations care must be exercised in how

one approximates the system under consideration— e.g. the spectra of clusters do not represent the electronic structure of the condensed phase and vice versa.

5.2 Methods

Samples:

Samples of alanine and sarcosine were purchased from Sigma Aldrich at stated purity of 99% or higher and were used without further purification; 1 M solutions were prepared using 18 MΩ (Millipore) water.

Experimental:

The NEXAFS spectra were obtained at Beamline 8.0.1 of the Advanced Light Source (ALS), at Lawrence Berkeley National Lab. Nitrogen K-edge spectra were collected both from the liquid jet as well as from the gas phase background. A baseline intensity, I_o, was collected farther up the beamline using a gold mesh, which was used to account for intensity fluctuations. The spectra were subsequently area normalized for ease of comparison. Spectra were collected for the carbon, nitrogen and oxygen K-edge for both sarcosine and alanine, but found the largest and most interesting differences on the nitrogen K-edge. Water dominates the oxygen K-edge spectra; the carbon K-edge spectra only minimally differ between sarcosine and alanine, being relatively insensitive to the change in the location of the methyl group.

The experimental apparatus comprises: a main chamber wherein the X-ray beam intersects with the microjet stream, a skimmer, a section for cryogenically trapping the microjet reduces pressures, and a differential pumping section allows for windowless coupling to the beamline. The measurements in this chapter were obtained using a chopped fused silica capillary of 30 μm attached to an HPLC pump with an average flow rate of 1 mL/min. The temperature of the sample was approximately 300 K. The signal was collected by total electron yield (TEY) with a copper electrode biased at +1.8 kV; the use of TEY has been shown to provide bulk characterization due to the large escape depths of electrons.[43] A more detailed description of the experimental apparatus has been previously published by Wilson *et al.*[33]

Theoretical:

As has previously been described,[34-37] accurately simulating NEXAFS of molecular systems requires numerical spectral convergence from adequate conformational space sampling. Classical molecular dynamics snapshots, in addition to fully relaxed structures, were input to first principles density functional theory (DFT) calculations for theoretical K-shell spectra.[44, 45] AMBER 9 with generalized AMBER force field and Antechamber were used for the construction of the solute, box, and the MD.[46, 47] The MD simulations were calculated in an approximately $(12\ \text{Å})^3$ orthorhombic simulation cell under periodic boundary conditions, where the solute and solvent are replicated. I note that the presence of spurious interactions between periodic images of the same molecule and its surrounding solvent does introduce some false correlation in the structural and electronic contributions to the X-ray absorption spectrum; however, by analogy with previous work on water,[36] this effect should be spectrally insignificant for cell volumes of this size, although this has not been explicitly investigated. The system was first relaxed, then equilibrated at constant standard pressure followed by equilibration at constant temperature before recording a trajectory at 300K using a Langevin thermostat. The simulation cell used contained 78 TIP3P waters;[48] this corresponds to a concentration of .7 M, slightly less concentrated than experiment. Higher concentrations necessitate either smaller simulation cells with fewer waters, but incomplete solvation shells or much larger simulation cells including more waters plus an additional solute molecule, making the first principles calculations too

expensive. Published studies establish that polarizable force fields for water are critical for simulations to achieve correct structure and dynamics for an ionic solute.[49, 50] The effect of a first principles molecular dynamics on the spectra of ionic solutes is the subject of future work. Periodic boundary conditions also eliminate surface electronic states; surface electronic states occur in finite cluster X-ray calculations due to dangling hydrogen bonds.

X-ray absorption spectra for each MD snapshot were calculated using DFT, where the lowest energy core-level excited state was self-consistently calculated with a full core hole and explicitly including the excited electron (XCH).[35-37] Transition amplitudes are approximated within the single-particle and dipole approximations. I employed the Perdew-Burke-Ernzerhof form of the generalized gradient approximation to the exchange-correlation potential.[51] These simulations employed the same periodic boundary conditions as the MD simulations-- explicitly including all 78 waters and using a plane wave basis set to represent the electronic structure. Norm-conserving pseudopotentials were used in the course of the calculation, necessitating a plane-wave kinetic energy cut-off of 85 Rydberg. Under periodic boundary conditions, the electron wavevector k is well defined and expectation values require an integration over the first Brillouin zone in k-space. Upon numerical convergence, this approach has been found to accurately approximate the spectra of disordered condensed phases.[36, 37] A grid of 5 x 5 x 5 k-points were used to converge the Brillouin zone integration for the X-ray absorption calculations. For more details about the theoretical approach and its development, see previous work.[34-37] Self-consistent charge densities and states were computed using PWSCF.[52] Calculations were performed on the Franklin supercomputer at NERSC, requiring over 700 CPU hours per hydrated snapshot.

Each calculated spectrum was homogeneously broadened by Gaussian convolution with a constant full width at half maximum of .2 eV. By numerical experience, this broadening is considered small with respect to inhomogeneous broadening introduced by nuclear motion for molecular systems, where non-degenerate molecular orbital energies are typically well-separated at the onset of absorption but can vary within a 1 eV range with respect to molecular conformation.[35] These numerically broadened spectra were then averaged over all available snapshots. In this work, I found that using 100 snapshots is more than sufficient for spectral convergence; that is, after 60 snapshots, the addition of more snapshots did not alter the average spectrum. Unlike all-electron calculations, DFT pseudopotential calculations lack an absolute energy reference. Consequently, I employ a relative energy scale, wherein the spectra are aligned to one another using total energy differences.[34] Individual features within a given spectrum should have the correct alignment, but with an underestimation of their energy spacing due to typical DFT underestimation of excited state energies.[36] Finally, for comparison with experiment, the computed spectra are rigidly shifted to the experimental onset.

5.3 Discussion

Figure 1 shows the structures for both sarcosine and alanine, illustrating their chemical differences when in water. In Figure 2, the experimental nitrogen K-edge spectra for solvated alanine and sarcosine is compared with the calculated spectra. The theoretical spectra are computed using 100 snapshots from an MD trajectory where the solute is free to move and is fully hydrated. Upon inspection, the main difference between the two spectra is the slight blue-shift apparent in the leading edge (the lower energy features) of the sarcosine spectrum with respect to alanine, indicating a change in the electronegativity or degree of bonding with neighboring atoms. This is analogous to what is seen for a carbonyl group; when changing its

neighboring moiety to a more electronegative functional group, a blue shift occurs on the oxygen K-edge.[53] The methyl group substitution on the nitrogen in sarcosine altered the bonding around the nitrogen, shifting transitions to a lower energy. The computed sarcosine spectra reproduce the general shape as well as the slight blue-shift found in the experimental spectra. Spectral features for both simulated alanine and sarcosine appear closer together in energy than in the experiment due to the known DFT bandwidth underestimation. I deliberately chose a narrow Gaussian broadening to expose distinct spectral features underlying the entire spectrum. The good agreement of experimental trends with the calculated spectra, Figure 2, is a validation of the computational method used and of the predictions made in the following sections.

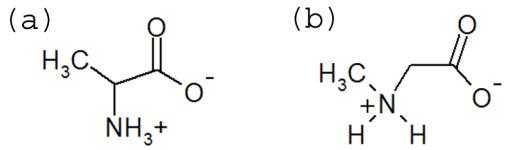


Figure 1. Structures of (a) alanine and (b) sarcosine in their zwitterionic forms; the negative charge is resonant between the two oxygens.

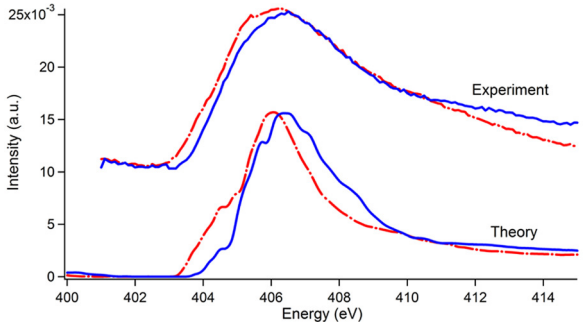


Figure 2. Measured nitrogen K-shell NEXAFS spectra of alanine and sarcosine and the computed spectra aligned to experiment are shown above. For both experiment and theory, sarcosine (blue) is shifted slightly to higher energy than alanine (red and dashed); the slight shift is reproduced well in the computed spectra, and assigned primarily to methyl group substitution on the nitrogen in sarcosine.

Spectral Effects of Conformation and Hydration

The MD trajectory sampled for the spectra shown in Figure 2 comprised 78 explicit waters and an unconstrained solute, this is referred to as "hydrated-free," and includes both conformation and hydration effects. To examine and enumerate the contributions of different factors to the spectrum (e.g. hydration, conformation), MD simulations were run to create different sets of coordinates under various conditions. Previous work has shown that the inclusion of internal molecular motion is very important for computational X-ray absorption studies;[35] this was confirmed by first performing calculations of the completely relaxed gas phase structures of alanine and sarcosine for both zwitterion and neutral forms, and finding these to be a poor representation of experiment (spectra not shown). To isolate the impact of solute conformation on the spectrum, the unconstrained and fully solvated MD trajectory (hydrated-free) was stripped of waters, leaving a trajectory with only the solute present in different conformations. These snapshots include all the motions of the solute that occur in the presence of

water; thus, it can be examined whether the changes in the spectrum result from the conformations induced by interactions with the water, or from the presence of the surrounding solvent itself. These structures were used for the DFT calculations, are shown in Figure 3, and are referred to as "dehydrated." Notably, the dehydrated calculations are not physically meaningful, as both alanine and sarcosine would be neutral, not zwitterionic in the gas phase. Nor does the dehydrated reproduce a gas phase MD trajectory based on the zwitterion, which would lack conformational changes induced by solvent interaction. Finally, in order to isolate the impact of hydration without the inclusion of conformation effects, another MD trajectory was generated where the solute structure was constrained, while the surrounding waters were allowed to move freely; these calculations are labeled as "hydrated-fixed." The chosen solute conformation of the hydrated fixed trajectory was taken from the original hydrated-free MD simulation and chosen for its typical conformation with regard to intermolecular angles (such as the dihedral angle), distances, and being a reasonably low energy structure. I recognize that both the hydrated-fixed and the dehydrated simulations are unphysical situations, but they serve as models to isolate the effects of conformation and hydration on the calculated spectra.

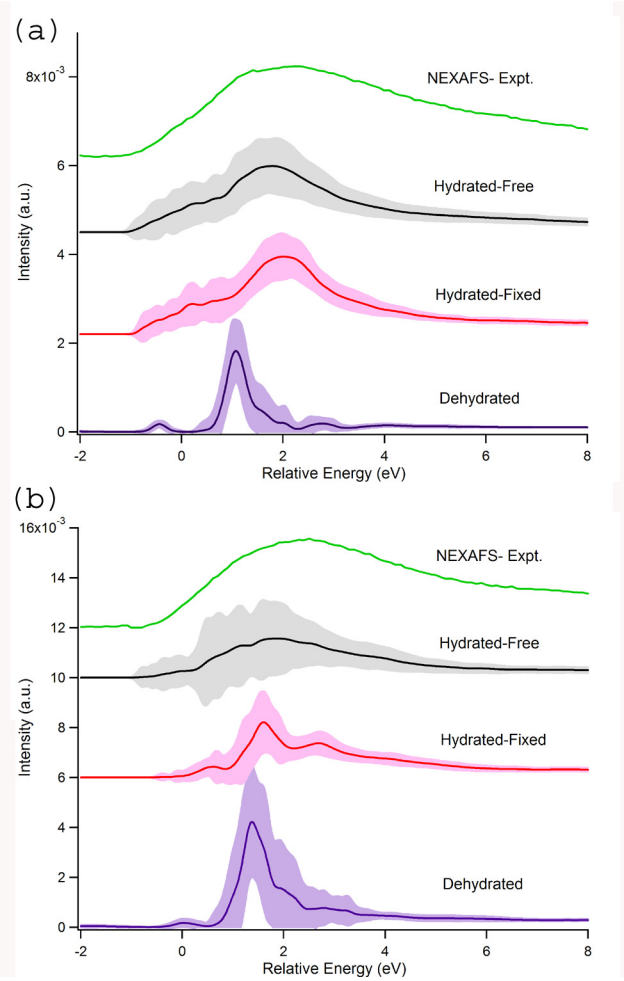


Figure 3. Computed (a) alanine (b) sarcosine nitrogen K-shell spectra with hydrated-free, hydrated- fixed, and dehydrated coordinates (see text) compared to experiment. The error (shown as shaded area) indicate one standard deviation based on MD sampling. The experimental spectra were shifted to align with the onset of the hydrated-free spectra.

Comparing spectra generated by the three MD sampling techniques and experimental spectrum for alanine, Figure 3(a), it is clear that holding the hydrated molecule rigid (hydrated-fixed) or removing solvation (dehydrated) produce inadequate representations of the measured spectrum. The most accurate calculated spectrum is, unsurprisingly, the unconstrained and fully hydrated MD-based spectrum (hydrated-free); this approach reproduces the general shape and the overall shift between sarcosine and alanine observed in experiment (Figure 2); verifying that this is the best technique for hydrated NEXAFS. However, in the case of alanine, the hydrated-fixed coordinates approximate both the experimental spectrum and the hydrated-free spectrum better than the corresponding dehydrated spectrum. This suggests that hydration has a larger impact on the electronic structure of alanine than conformational changes. The strong effect of hydration implies that given a different solvent, the electronic structure would again be altered; i.e., a solvent dependence. Alanine's solvent dependence has been found in other studies; in vacuum, a polyalanine molecule assumed an α-helix, and was only marginally stable once it was hydrated.[15] Additionally, a blocked alanine was found to have solvent dependence in FT-IR experiments.[54] Notably, more broadening occurs in the hydrated-free spectrum due to the additional structural changes (versus the hydrated-fixed) and the inclusion of hydrogen bonding structures (versus the dehydrated spectrum).

Figure 3(b) shows a similar comparison for sarcosine, between experiment and simulation for the hydrated unconstrained solute (hydrated-free), hydrated with constrained solute (hydrated-fixed), and dehydrated solute scenarios. The best approximation for the experimental spectrum is, similar to alanine, the unhindered fully solvated case. Yet, for sarcosine there is less agreement between the hydrated-free and hydrated-fixed scenarios implying conformation is a larger spectral contributor than it is for alanine. As there are numerous conformations, it is impossible to say whether this is true for each; rather, this typical (in terms of structure and energy) structure, is an indicator of lesser solvent contribution to the electronic structure. This will be examined further later in the text with regard to hydrogen bonding. Intramolecular interaction is possible for both sarcosine and alanine in the gas phase for both the neutral and zwitterion.[15, 55] Based on visual inspection of the snapshots intramolecular interaction is reduced upon hydration; still, the degree of interaction is greater for sarcosine than alanine based on the proximity brought about by the methyl substitution. Given increased intramolecular interaction any change in angle would result in changes in the strength of these interactions and the electronic structure, and would be seen spectrally.

To further examine the relative impact of conformation on spectra, individual configurations with various dihedral angles were chosen and their spectra calculated; Figures 4(a) and (b) present the results for a variety of angles for both sarcosine and alanine, respectively; these measurements are based on the trajectories from the dehydrated calculations. These snapshots are from the dehydrated trajectory to isolate the impact of conformation alone and simplify interpretation. The angles indicated are the sum of dihedral angles illustrated in the insets of Figure 4; dihedral angles were used because they provide a better representation of possible intramolecular interaction and overall molecular strain. Previous computational work has shown that for gas phase glycine the nitrogen K-edge is sensitive to such conformational changes, therefore some angle dependence was predicted.[22, 34, 35] The distribution of dihedral angles for sarcosine (standard deviation approximately 90°) is much wider than for alanine (standard deviation approximately 50°). The free rotation of the NH_3^+ of alanine gives somewhat degenerate spectral features; whereas, the methyl on the nitrogen of sarcosine results in lower symmetry around the nitrogen and the sampling of more unique positions such that conformation

has a greater spectral role. Examining Figure 4, it is evident that sarcosine exhibits larger spectral variation depending on the dihedral angle as opposed to alanine; this implies that sarcosine is more conformation dependent than alanine.

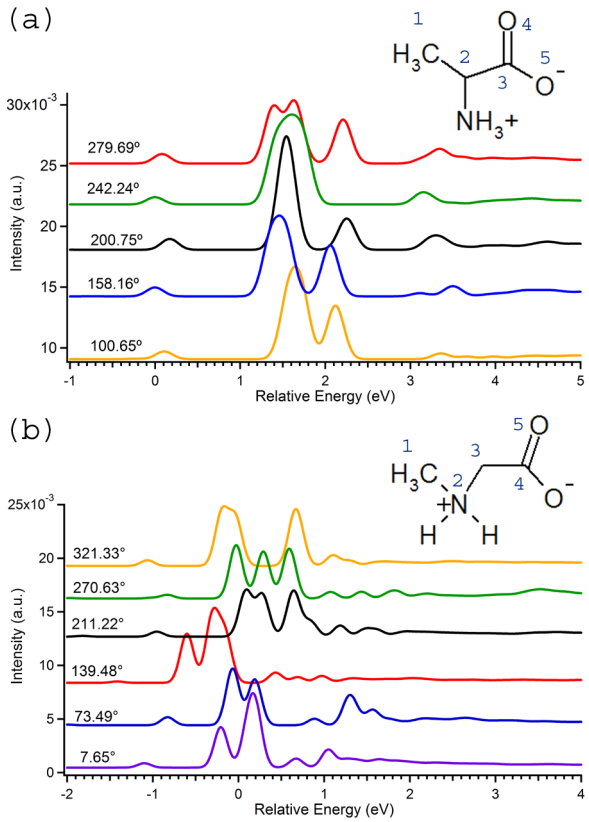


Figure 4. Calculated N K-edge spectra of (a) alanine (b) sarcosine for a range of dihedral angles; the number shown on the left is the sum of the two dihedral angles. These spectra do not include any water molecule, but are from the dehydrated trajectory, and serve to isolate the impact of changing the dihedral angles. The indicated values are the sum of ∠1,2,3,4 and ∠1,2,3,5 for alanine, and ∠1,2,3,4 combined with ∠2,3,4,5 for sarcosine.

As a further test of the effect of solvent on sarcosine and alanine different configurations were compared from the hydrated-fixed calculations to examine the impact of the hydrogen bonding without conformational changes (Figure 5). For sarcosine and alanine, two snapshots each are chosen with high and low number of hydrogen bonds. A hydrogen bond was defined as having a distance less than 2.5 Å, an acceptor angle of 120° or greater, and a donor angle of 90° or more; a schematic is shown in Figure 5(c). For alanine, Figure 5(a), there are large spectral differences between the two snapshots, despite having the same solute coordinates. For alanine, neither of the snapshots resemble the ensemble-averaged spectrum, which requires an average of many snapshots. Sarcosine's two hydrated snapshots, Figure 5(b), show little change between the two spectra other than minor shifts and a slight change in the intensity of the shoulder to the blue of the main peak. The two spectra from individual snapshots seem to be narrower versions of the average sarcosine hydrated-fixed spectrum. The 2 snapshots for alanine and sarcosine are representative, and indicative of the larger spectral impact of hydration on alanine than sarcosine, and the importance of the strength or type of interaction. This suggests if sarcosine were placed

in another polar solvent, the states would be less impacted than for alanine. This has been seen in previous studies of the effect of other polar solvents on the secondary structure of peptoid oligomers using circular dichrosim.[5]

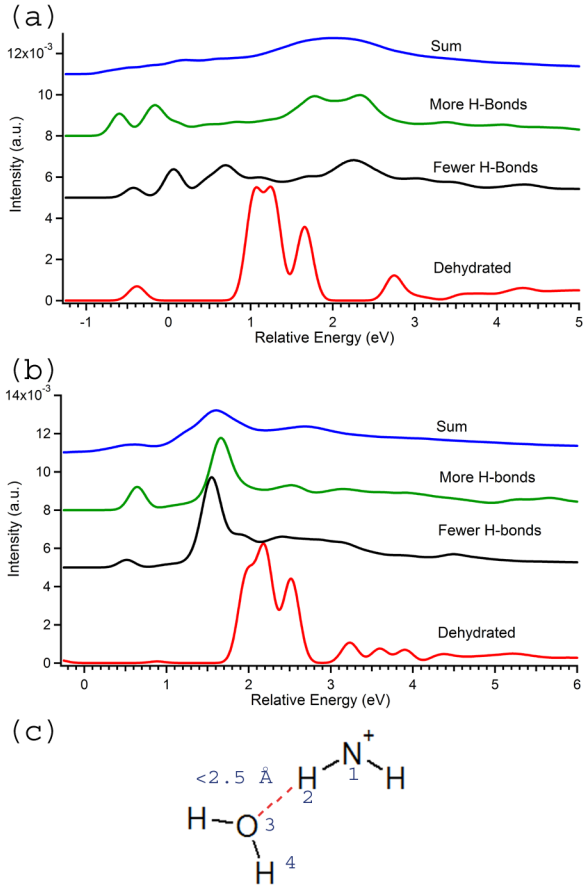


Figure 5. Using snapshots from the MD trajectory with a constrained solute, nitrogen K-edge spectra were calculated for cases where there is a higher or lower number of hydrogen bonds for (a) alanine, and (b) sarcosine. The sum is the hydrated-fixed average also shown in Figure 3. A hydrogen bond, shown in (c) was defined as having a distance between acceptor and donor of less than 2.5 Å, a donor angle greater than 120° (∠1,2,3), and an acceptor angle greater than 90° (∠2,3,4).

Hydration has a large spectral impact, particularly for alanine. Certain excited states are more sensitive to hydration; the more sensitive states appear to be those that have more density on the hydrogens of the nitrogen. For sarcosine, most of the low-lying excited states are molecular in nature (dispersed over the entire molecule) with relatively more density along the backbone of the molecule as opposed to alanine where low-lying excited are more localized on the terminal NH_3^+ group. This can be seen in Figure 6, which contrasts the LUMOs of alanine and sarcosine with and without water. These LUMOs are representative of the excited states as our XCH calculations explicitly place the excited electron in the LUMO in addition to the core hole. The LUMO of alanine without water, Figure 6(a), is extremely localized on the amine group; whereas, sarcosine's LUMO without water, 6(b), has localized density on the nitrogen but also density along the methyl group and the carboxylate group. Both dehydrated LUMOs alter

upon hydration; a large change occurs in the hydrogens upon hydrogen bonding-- there is no density on the involved hydrogen. The hydrated LUMO of alanine, figure 6(c), shows a large migration to the backbone of the molecule, with increased density near the carboxylate group. Alanine's LUMOs exhibit a density difference upon hydration particularly along the backbone; this can be attributed to the localization of the dehydrated LUMO on the dangling nitrogen group. For sarcosine, the change to the LUMO upon hydration is less than alanine as the dehydrated LUMO has character that is more molecular than alanine. Molecular states change less upon hydration, as the hydrogen bonding tends to affect specific terminal moieties in the molecule as opposed to the molecule in its entirety. Yet, a more molecular state will be impacted more by dihedral angle variations than a more localized state, such as what was found for sarcosine and alanine. Overall, the low-lying unoccupied states of sarcosine are more molecular than alanine; thus, giving sarcosine less sensitivity to the hydrogen bonding network in its entirety, but increased sensitivity to conformation.

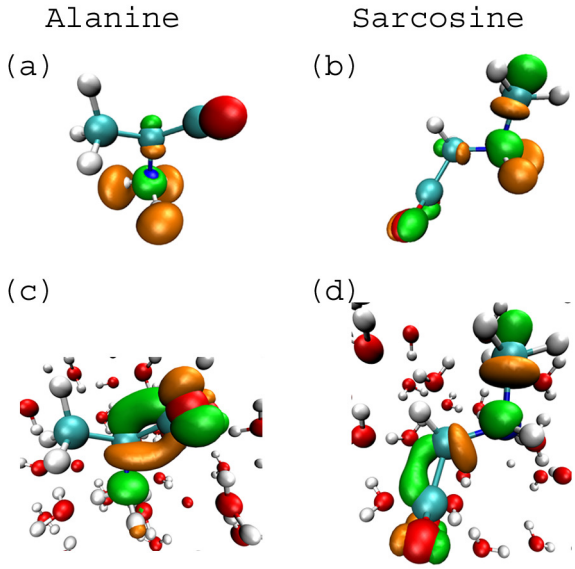


Figure 6. LUMOs from a single snapshot for: (a) alanine without water, (b) sarcosine without water (c) alanine with water and (d) sarcosine with water. For the lower-lying unoccupied orbitals, sarcosine shows more molecular character than do the states in alanine, and are therefore less affected by hydration. These states have isosurface values of 30% of the total integrated value.

Partial versus Bulk Hydration

Hydration can be a vague term—partial or bulk hydration? Are the waters around the atom of interest the only ones important, and how many solvation shells are important? I examine the question of using microhydration (a limited but localized number of water molecules) in simulations to approximate the bulk. Similarly, can bulk calculations or experiments be used to approximate real microhydration situations such as confined water? These concerns need to be addressed to approximate experimental conditions for NEXAFS, maintain a tractable calculation, and to move towards predictive computational work.

Clusters of various sizes were constructed for alanine and sarcosine using the nitrogen atom as the center and selecting waters by minimum distance to a specified center, the spectra of

these clusters were calculated using DFT and are shown in Figure 7; alanine is discussed below and sarcosine is discussed later in the text.

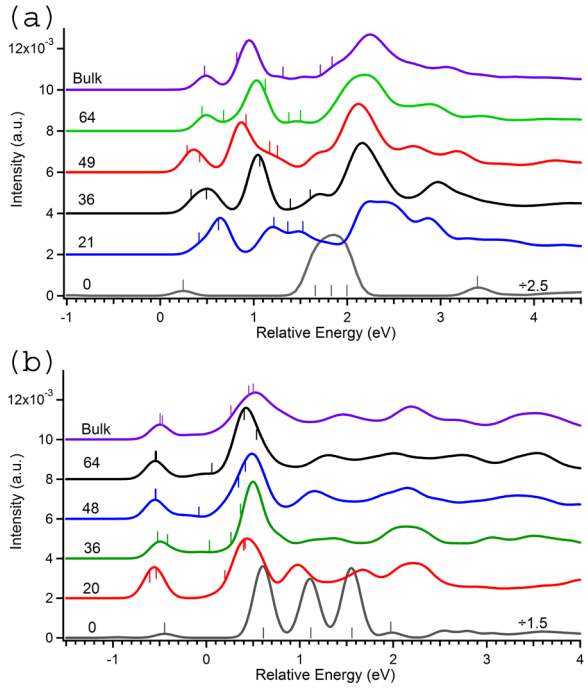


Figure 7. Computed N K-edge spectra for various aqueous cluster sizes solvating (a) alanine and (b) sarcosine, with the focal point of the cluster being the nitrogen position. The number of waters for each cluster are labeled, with 78 filling the periodic box (bulk). Spectra were aligned to a reference snapshot as described in the test. The vertical dashes identify the first several states.

Inspecting the spectra of the clusters for alanine, Figure 7(a), there is a marked difference between the spectra with 21 waters versus the other cluster sizes. In that configuration, there are only enough waters molecules to hydrate the amine group; thus, the carboxylate group is not hydrated until 36 waters are present. Thus, attempting to examine bulk hydration simulations with incomplete solvation may lead to incorrect interpretation.[56]

To examine the stability of our calculations and the effects of specific interactions on the nitrogen K-edge, smaller clusters of alanine were constructed, and their spectra calculated. The water clusters are centered on the NH_3^+ or around the carboxylate group, in Figure 8(a) and (b), respectively. The waters included were those closest to the functional group in question. As expected, the spectra of the NH_3^+ centered cluster exhibit definite spectral changes upon hydration. There is an increase in the number of states, changes in the intensity and energies of transitions, and in the density of states with addition of each water molecule in the NH_3^+ centered small cluster. The character (spatial extent and relative density on a particular atom) of the states change with the increased interaction, this, in turn, affects the oscillator strength and thus the spectrum. This can be seen in Figure 6 in the case of the LUMO; even small alterations to these final states can strongly affect their overlap with the initial electronic core state and thus alter the transition strength. The addition of waters to the COO^- centered water cluster impacts the energy spacing of the states, removing degeneracy and lowering the energy of excited states

of this charged species, as opposed to the creation of new spectral features. In the dehydrated spectrum, the largest feature (at approximately 1.8 eV) comprises three smaller transitions. After hydrating the carboxylate, this approximate degeneracy is broken and the 3 smaller features become distinct. Hydration of the carboxylate group leads to a decrease in perturbations of the electronic structure, particularly in the case of a small zwitterion; this is notable by a slight lowering in the energy of the states. Whether this effect is important for polypeptides, where charge groups may be more spatially separated is the subject of future research. In cases of microsolvation (solvation by a limited number of waters) or clusters, this effect would be important if other nearby charged groups were unsolvated, such as in proteins or other polymers exhibiting preferential hydration. I recognize that while illustrative the small cluster studies presented here are unphysical, since the zwitterion is unstable *in vacuo* or with limited waters; research has suggested that at least 7-8 waters are required to stabilize the alanine zwitterion.[16, 17]

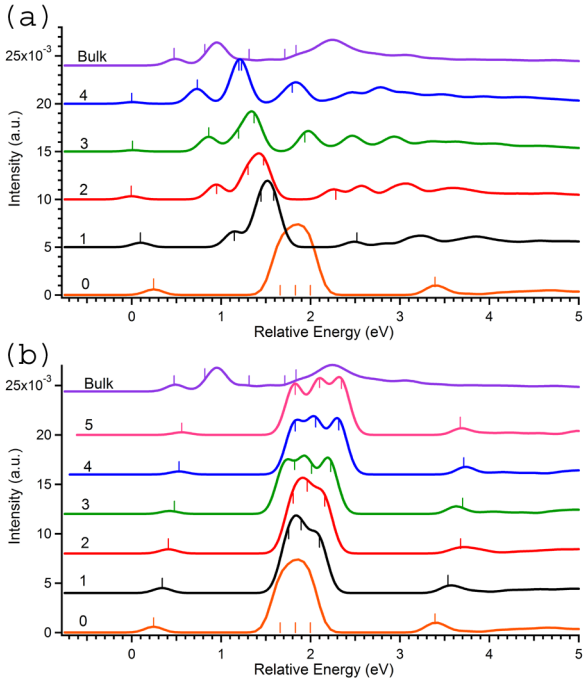


Figure 8. Nitrogen K-edge spectra of (a) NH_3^+ and (b) COO^- centered clusters of water molecules around alanine, examining the effects of microhydration. The nitrogen-centered clusters show differences in the appearance of new spectral features and large-scale shifts of features, indicated by the vertical dashes. In the case of the carboxylate-centered clusters, no new large peaks appear and there is more splitting of the main feature at approximately 2 eV. Bulk refers to 78 waters, the fully hydrated case, where the periodic box is full.

Hydrated sarcosine was similarly examined using various water cluster sizes; their spectra are shown in Figure 7(b). These clusters show similarities to those of alanine, Figure 7(a), with regard to the spacing of spectral features. The spectral change produced when the cluster size is increased to 36 waters is less evident for sarcosine than for alanine; this is likely due to the nearest waters around sarcosine being closer to the carboxylate group than for alanine due to the difference in molecular structures. In order to isolate the spectral impact of water on sarcosine further, smaller clusters were constructed (Figure 9) for both nitrogen and carboxylate

centered clusters. A reasonable approximation to the bulk hydrated spectrum is reached by having 4 or 5 waters in the nitrogen centered cluster for sarcosine, implying that the impact of hydration on the nitrogen K-edge spectrum of sarcosine converges within the first solvation shell, as opposed to the case of alanine. The shifting of states' energies and the appearance of new transitions occurs with the addition of each water for the small nitrogen-centered clusters' spectra, similar to what was found for alanine; this is particularly visible for the LUMO and LUMO+1. There are intensity changes of the features but the number of states within the first 1.5 eV of features remains constant after the spectrum with 3 waters for the nitrogen-based small clusters, corroborating a higher dependence on short range interactions than was found for alanine. In the case of hydration around the carboxylate, there appears to be a contraction of the existing states, as well as small energy shifts, rather than the appearance of new states, suggesting that carboxylate-based hydration only modifies the energy levels of the states. For both sarcosine and alanine, it is important to account for the fact that there are two charged groups nearby and this will primarily affect the states and not their spectral intensities. This implies that groups that have only one such charged group would not yield good representations of those with two, and vice versa.

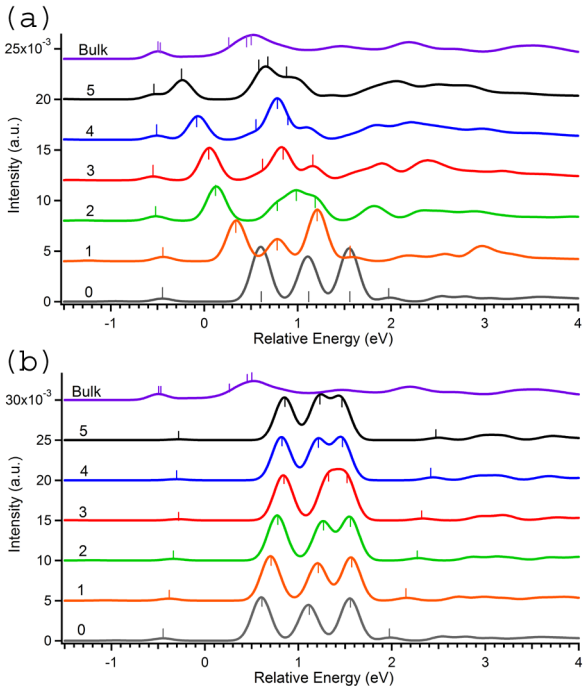


Figure 9. Nitrogen K-edge spectra of small clusters around sarcosine with (a) focused around the nitrogen and (b) centered around the carboxylate. The first few unoccupied states are labeled on each trace by small vertical lines. As found for alanine, the nitrogen-based clusters exhibit the appearance of new spectral features with increased numbers of water molecules, whereas the carboxylate based clusters exhibit changes in the energy of the states, rather than the appearance of new transitions. 78 waters (bulk) is the fully solvated case, where the periodic box is full.

5.4 Conclusions

This chapter presents the NEXAFS measurements of aqueous sarcosine and alanine, and accompanying first principles DFT calculations and analysis; calculations are in good agreement with experiment, further validating our recently developed theoretical (XCH) approach. Based on our theoretical analysis, it is shown that the effect of hydration of alanine has a larger spectral effect than does conformation; the reverse is true for sarcosine. This was further supported by examining spectra of alanine and sarcosine conformers over a range of dihedral angles, and with different hydrogen bonding environments. Part of alanine's higher sensitivity to hydrogen bonding is likely related to the structural difference of having a terminal nitrogen versus an embedded nitrogen, as in the case of sarcosine. It is important to note that some of the findings revealed in this study will change upon the formation of a polypeptide or polypeptoid due to the altered bonding on the nitrogens.[29] Presently, I can only predict what the spectral implications of secondary structure would be. In the case of alanine, the restrictions of motion effected by a longer peptide chain would likely affect the excited states, and thus the spectrum, less than if the solvent was changed or mixed, as has been found previously for alanine in different environments.[15, 54] In general, side chains that have dangling hydrogen bonds would probably be more affected by hydration than by conformation, and this would be reflected in computed and likely experimental NEXAFS spectra. This is supported by many of the excited states of sarcosine, a molecule with a non-terminal nitrogen, resembling delocalized molecular orbitals more than moiety-specific orbitals, and therefore being less explicitly dependent on the solvent and more dependent on conformation. For polypeptoids, the existence of a longer chain may constrain the conformational movements due to the establishment of secondary structure in the absence of amino-group hydrogen-bond donors along the backbone, thus leading to more rigid higher-order structure than found in polypeptide analogues. I predict that this rigidity should exhibit a change in spectral signature given the increased population of certain dihedral angles. In addition, this work also demonstrates the sensitivity of nitrogen K-edge NEXAFS to details of the solvation environment, including degree of hydration, and solvent identity. Thus, NEXAFS spectroscopy, in combination with accurate excited state theory, appears to be a useful tool for the examination of the secondary structure of polypeptoids, polypeptides, and quite probably DNA and RNA.

References:

(1) Bautista, A. D.; Craig, C. J.; Harker, E. A.; Schepartz, A., *Curr. Opin. Chem. Biol.* **2007,** 11, (6), 685-692.

(2) Yoo, B.; Kirshenbaum, K., *Curr. Opin. Chem. Biol.* **2008,** 12, (6), 714-721.

(3) Wu, C. W.; Kirshenbaum, K.; Sanborn, T. J.; Patch, J. A.; Huang, K.; Dill, K. A.; Zuckermann, R. N.; Barron, A. E., *J. Am. Chem. Soc.* **2003,** 125, (44), 13525-13530.

(4) Miller, S. M.; Simon, R. J.; Ng, S.; Zuckermann, R. N.; Kerr, J. M.; Moos, W. H., *Bioorg. Med. Chem. Lett.* **1994,** 4, (22), 2657-2662.

(5) Sanborn, T. J.; Wu, C. W.; Zuckerman, R. N.; Barron, A. E., *Biopolymers* **2002,** 63, (1), 12-20.

(6) Wender, P. A.; Mitchell, D. J.; Pattabiraman, K.; Pelkey, E. T.; Steinman, L.; Rothbard, J. B., *Proc. Natl. Acad. Sci. U. S. A.* **2000,** 97, (24), 13003-13008.

(7) Hara, T.; Durell, S. R.; Myers, M. C.; Appella, D. H., *J. Am. Chem. Soc.* **2006,** 128, (6), 1995-2004.

(8) Seurynck, S. L.; Patch, J. A.; Barron, A. E., *Chem. Biol.* **2005,** 12, (1), 77-88.

[9] Patch, J. A.; Barron, A. E., *J. Am. Chem. Soc.* **2003,** 125, (40), 12092-12093.
[10] Lee, B. C.; Chu, T. K.; Dill, K. A.; Zuckermann, R. N., *J. Am. Chem. Soc.* **2008,** 130, (27), 8847-8855.
[11] Gorske, B. C.; Blackwell, H. E., *J. Am. Chem. Soc.* **2006,** 128, (44), 14378-14387.
[12] Gomez-Zavaglia, A.; Fausto, R., *Vib. Spectrosc.* **2003,** 33, (1-2), 105-126.
[13] Shin, S. B. Y.; Kirshenbaum, K., *Org. Lett.* **2007,** 9, (24), 5003-5006.
[14] Lemoff, A. S.; Williams, E. R., *J. Am. Soc. Mass Spectrom.* **2004,** 15, (7), 1014-1024.
[15] Kohtani, M.; Jones, T. C.; Schneider, J. E.; Jarrold, M. F., *J. Am. Chem. Soc.* **2004,** 126, (24), 7420-7421.
[16] Aikens, C. M.; Gordon, M. S., *J. Am. Chem. Soc.* **2006,** 128, (39), 12835-12850.
[17] Ahn, D. S.; Park, S. W.; Jeon, I. S.; Lee, M. K.; Kim, N. H.; Han, Y. H.; Lee, S., *J. Phys. Chem B.* **2003,** 107, (50), 14109-14118.
[18] Uejio, J. S.; Schwartz, C. P.; Duffin, A. M.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (19), 6809-6812.
[19] Messer, B. M.; Cappa, C. D.; Smith, J. D.; Drisdell, W. S.; Schwartz, C. P.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem B.* **2005,** 109, (46), 21640-21646.
[20] Messer, B. M.; Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem B.* **2005,** 109, (11), 5375-5382.
[21] Ide, M.; Maeda, Y.; Kitano, H., *J. Phys. Chem B.* **1997,** 101, (35), 7022-7026.
[22] Gordon, M. L.; Cooper, G.; Morin, C.; Araki, T.; Turci, C. C.; Kaznatcheev, K.; Hitchcock, A. P., *J. Phys. Chem. A* **2003,** 107, (32), 6144-6159.
[23] Aziz, E. F.; Ottosson, N.; Bonhommeau, S.; Bergmann, N.; Eberhardt, W.; Chergui, M., *Phys. Rev. Lett.* **2009,** 102, (6), 068103.
[24] Kaznacheyev, K.; Osanna, A.; Jacobsen, C.; Plashkevych, O.; Vahtras, O.; Agren, H., *J. Phys. Chem. A* **2002,** 106, (13), 3153-3168.
[25] Boese, J.; Osanna, A.; Jacobsen, C.; Kirz, J., *J. Electron Spectrosc. Relat. Phenom.* **1997,** 85, (1-2), 9-15.
[26] Carravetta, V.; Plashkevych, O.; Agren, H., *J. Chem. Phys.* **1998,** 109, (4), 1456-1464.
[27] Feyer, V.; Plekan, O.; Richter, R.; Coreno, M.; Prince, K. C.; Carravetta, V., *J. Phys. Chem. A* **2008,** 112, (34), 7806-7815.
[28] Nyberg, M.; Hasselstrom, J.; Karis, O.; Wassdahl, N.; Weinelt, M.; Nilsson, A.; Pettersson, L. G. M., *J. Chem. Phys.* **2000,** 112, (12), 5420-5427.
[29] Otero, E.; Urquhart, S. G., *J. Phys. Chem. A* **2006,** 110, (44), 12121-12128.
[30] Plekan, O.; Feyer, V.; Richter, R.; Coreno, M.; de Simone, M.; Prince, K. C.; Carravetta, V., *J. Electron Spectrosc. Relat. Phenom.* **2007,** 155, (1-3), 47-53.
[31] Zubavichus, Y.; Shaporenko, A.; Grunze, M.; Zharnikov, M., *J. Phys. Chem B.* **2008,** 112, (15), 4478-4480.
[32] Wilson, K. R.; Rude, B. S.; Catalano, T.; Schaller, R. D.; Tobin, J. G.; Co, D. T.; Saykally, R. J., *J. Phys. Chem B.* **2001,** 105, (17), 3346-3349.
[33] Wilson, K. R.; Rude, B. S.; Smith, J.; Cappa, C.; Co, D. T.; Schaller, R. D.; Larsson, M.; Catalano, T.; Saykally, R. J., *Rev. Sci. Instrum.* **2004,** 75, (3), 725-736.
[34] Schwartz, C. P.; Uejio, J. S.; Saykally, R. J.; Prendergast, D., *J. Chem. Phys.* **2009,** 130, (18), 184109.
[35] Uejio, J. S.; Schwartz, C. P.; Saykally, R. J.; Prendergast, D., *Chem. Phys. Lett.* **2008,** 467, (1-3), 195-199.
[36] Prendergast, D.; Galli, G., *Phys. Rev. Lett.* **2006,** 96, (21), 21522-21526.

[37] Prendergast, D.; Grossman, J. C.; Galli, G., *J. Chem. Phys.* **2005,** 123, (1), 14501-14505.
[38] Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Messer, B. M.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem B.* **2005,** 109, (15), 7046-7052.
[39] Cappa, C. D.; Smith, J. D.; Messer, B. M.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem B.* **2006,** 110, (11), 5301-5309.
[40] Cappa, C. D.; Smith, J. D.; Messer, B. M.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem B.* **2006,** 110, (3), 1166-1171.
[41] Cavalleri, M.; Odelius, M.; Nilsson, A.; Pettersson, L. G. M., *J. Chem. Phys.* **2004,** 121, (20), 10065-10075.
[42] Wernet, P.; Nordlund, D.; Bergmann, U.; Cavalleri, M.; Odelius, M.; Ogasawara, H.; Naslund, L. A.; Hirsch, T. K.; Ojamae, L.; Glatzel, P.; Pettersson, L. G. M.; Nilsson, A., *Science* **2004,** 304, (5673), 995-999.
[43] Stöhr, J., *NEXAFS Spectroscopy*. Springer: New York, 1996; 20-31.
[44] Hohenberg, P.; Kohn, W., *Phys. Rev. B* **1964,** 136, (3B), B864.
[45] Kohn, W.; Sham, L. J., *Phys. Rev.* **1965,** 140, (4A), 1133-&.
[46] Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A., *J. Mol. Graphics. Modell.* **2006,** 25, (2), 247-260.
[47] Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., *J. Comput. Chem.* **2005,** 26, (16), 1668-1688.
[48] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J. Chem. Phys.* **1983,** 79, (2), 926-935.
[49] Warshel, A.; Kato, M.; Pisliakov, A. V., *J. Chem. Theory Comput.* **2007,** 3, (6), 2034-2045.
[50] Johnson, M. E.; Head-Gordon, T.; Louis, A. A., *J. Chem. Phys.* **2007,** 126, (14).
[51] Perdew, J. P.; Burke, K.; Ernzerhof, M., *Phys. Rev. Lett.* **1996,** 77, (18), 3865-3868.
[52] Baroni, S.; Corso, A. D.; de Gironcoli, S.; Gianozzi, P. PWSCF. www.pwscf.org
[53] Urquhart, S. G.; Ade, H., *J. Phys. Chem B.* **2002,** 106, (34), 8531-8538.
[54] Lee, M.-E.; Lee, S. Y.; Joo, S.-W.; Cho, K.-H., *J. Phys. Chem B.* **2009,** 113, (19), 6894-6897.
[55] Cocinero, E. J.; Villanueva, P.; Lesarri, A.; Sanz, M. E.; Blanco, S.; Mata, S.; Lopez, J. C.; Alonso, J. L., *Chem. Phys. Lett.* **2007,** 435, (4-6), 336-341.
[56] *Notably, we are using plane waves in these calculations, as opposed to the localized atomic orbitals used in previous work, which means we can better describe delocalized final states and avoid unintentional localization based on basis set choice.*

Chapter 6.
Preliminary Results of Small Biological Molecules

Early in my graduate career, I examined a number of small biomolecules using NEXAFS to characterize their hydration. These included amino acids studied under a variety of pH conditions, and the examination of hydrated nucleobases and nucleosides (the nucleobase and a sugar). This chapter details experimental results, preliminary conclusions and suggestions for future work to yield a more complete understanding of these interesting fundamental systems.

6.1 Amino Acids Introduction

As previously discussed, almost all biological systems are aqueous; thus, understanding the interactions of water molecules with prototypical biomolecules is crucial.[1, 2] Numerous studies have looked at the interactions of amino acids with water by various methods including Raman and infrared spectroscopy and mass spectrometry.[3-5] Previous work on amino acids using NEXAFS has focused mainly on studies of powder form, gas phase, or spun thin film preparations.[6-8] These samples present problems concerning sample preparation, and sample radiation damage.[9] By employing liquid jet technology and combining it with NEXAFS, information about the hydration of species has been extracted while avoiding sample damage.[1, 2, 10-12] By previous work in the Saykally group, it was shown that glycine in its anionic form assumes an "acceptor only" position; the molecule only accepting hydrogen bonds from solvating water molecules, as opposed to donating them.[2] Several other amino acids were subsequently investigated, in an effort to ascertain the generality of this behavior. It was argued due to the steric hindrance of larger R groups that not all amino acids would exhibit acceptor only behavior.[1]

In this chapter, those results are expanded upon by looking at two different amino acids: arginine, and valine, for which the structures are shown in Figure 1. These amino acids contain two very different R groups. Each amino acid was examined at varying pHs, in order to compare their hydration characteristics in different charge states.

(a) (b)

Figure 1. Structural representations of given amino acids (a) arginine (b) valine. The sites that change with pH are circled, and include amine groups and carboxylic acids.

6. 2 Methods for Amino Acids

Experimental

The methods employed were similar to those described in previous chapters. Solutions were made with arginine and valine from Sigma Aldrich at stated purity of 98% or higher and without further purification. Millipore water was used to make solutions of 0.5 M arginine, and 0.3 M valine, the pH was then adjusted accordingly with sodium hydroxide, or hydrochloric acid. The NEXAFS spectra were obtained on the relevant edges, carbon, nitrogen, and oxygen, at Beamline 8.0.1 of the Advanced Light Source (ALS), at Lawrence Berkeley National

Laboratory. The carbon spectra were calibrated to highly ordered pyrolitic graphite recorded previously by beamline scientists. A gold mesh collected incident intensity farther up the beamline for normalization to account for variations in photon flux. The spectra were subsequently area normalized. The experimental apparatus has been detailed in earlier chapters, and was previously published by Wilson et al.[13]

Computational

Impact,[14] a program produced by Macromodel that is commercially available, was used to solvate and then perform molecular dynamics (MD) on the hydrated amino acids. The simulations used a solvated cube of 18.62 Å with periodic boundary conditions and 205 water molecules. The twenty nearest neighbors from the center of mass in the MD snapshots were determined, and then 15 selected snapshots were used as input coordinates for StoBe deMon;[15] a similar method was used for the carboxylate study.[16] A composite of the StoBe spectra from different snapshots is the theoretical comparison given in subsequent figures. A description of StoBe and its inherent assumptions was discussed in Chapters 1 and 2; at the time of the work described herein, the more advanced methods detailed in Chapters 1 and 4 had yet to be developed.

6.3 Results and Discussion of Amino Acids Work

Amino acids can exist in multiple charge states; for example, arginine has four distinct charge states, +2, +1, neutral, -1. Through pH control, the predominant state in solution can be chosen; a systematic comparison was made of arginine and valine in their different charge states, the sites of protonation and deprotonation are shown in Figure 1. Analysis of the nitrogen K-edge and the carbon K-edge follow; oxygen K-edge spectra were also collected, but were dominated by the signal from water.

Changing the charge state of the amino acids had been previously seen to have little to no effect on the carbon K-edge spectrum.[1, 2] This is seen in Figure 2, where a comparison of the four different charge states of arginine is shown. The peak at approximately 288.5 eV results from the $1s \rightarrow \pi^*_{CN}$ transition; the subsequent peak is assigned to the carbonyl transition centered at 290 eV. The change from an overall positive charge of two to an overall negative charge seems to have no noticeable trend on the carbon K-edge. Previously, a slight red shift has been noted for the $\pi_{C=O}^*$ feature[1, 2] but with arginine this was not seen, likely due to the $\pi^*_{C=N}$ feature overshadowing any small blue shifts and not being as influenced by pH changes.

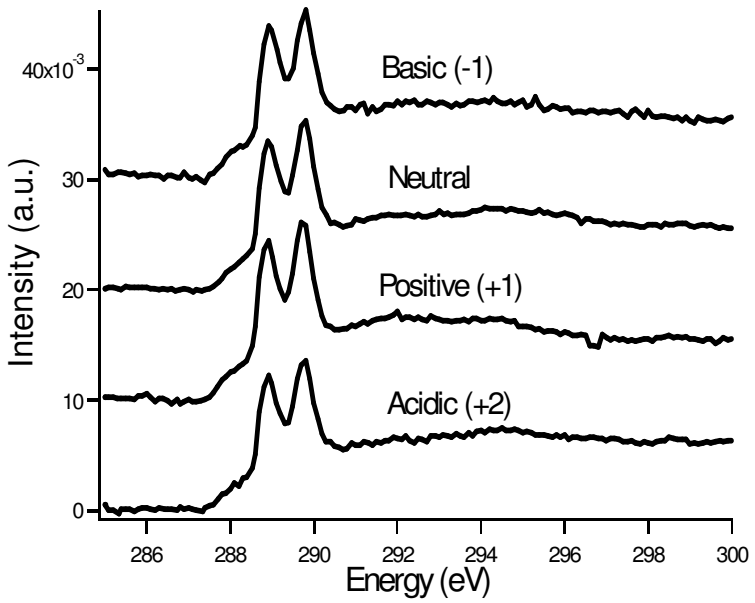


Figure 2. Carbon K-edge comparison of arginine in four different charge states (fully protonated, after the loss of the carboxylic hydrogen, lost of the first nitrogen hydrogen, and the loss of the second amine hydrogen).

A theoretical spectrum was calculated using StoBe[15] and compared to experiment, and is shown in Figure 3. The simulated spectrum reasonably approximates the experimental data for the π* resonances. Unfortunately, the energies after the π* features, after 290 eV, appear to increasingly disagree with the experimental spectrum; this appears to be an artifact of StoBe's use of a localized model. This effect has been seen before in other theoretical calculations of small biological molecules.[17] Additionally, there is a question of the overall fit due to recent literature which states that a spectrum calculated with StoBe will be affected by a change in basis set;[18, 19] this would add another fitting parameter to the theoretical spectrum.

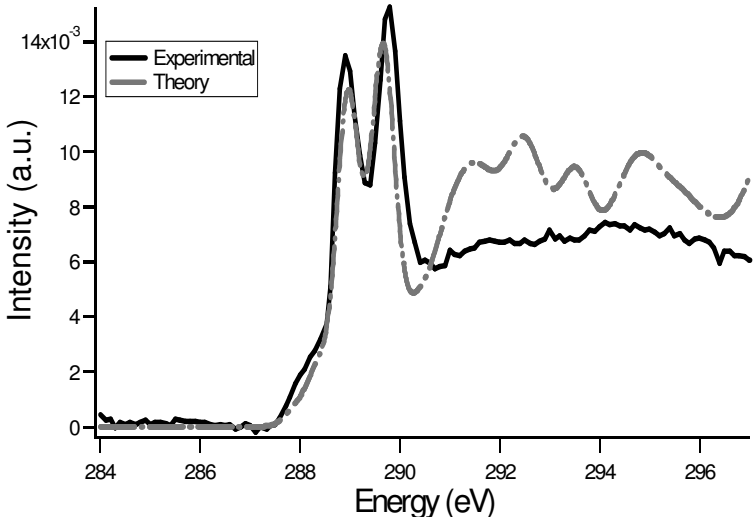


Figure 3. Carbon K-edge comparison of experiment versus theoretical calculations. The features due to the π bonds present in arginine are well matched. Theory and experiment become increasingly divergent in the σ* bonding region (above 290 eV).

The nitrogen K-edge spectra of arginine display a general invariance with charge state (Figure 4). The 1s→ π* at approximately 401.5 eV is particularly insensitive to changes in the charge state. This corresponds to the first feature on the carbon K-edge spectrum, at around 288.5 eV. In previous work, changes were seen on the nitrogen K-edge of glycine due to pH changes, a shift from deprotonation and small features on the leading edge labeled "acceptor only," but in the σ* region of the spectrum (above 403 eV) for arginine no shift is apparent. This is likely due to the four nitrogens with σ bonding making it unlikely to see a small shift of the other features. There is also a possibility of contamination due to atmospheric nitrogen leaks impacting the lower energy feature, but this is unlikely due to the relatively high concentrations of the amino acids.

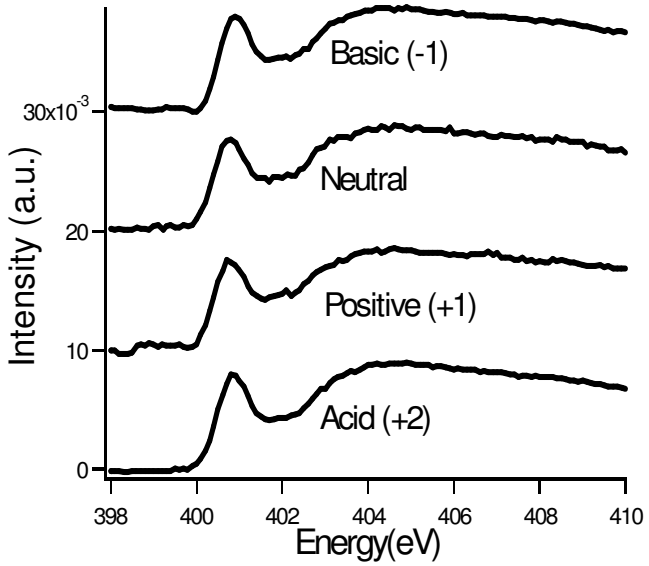


Figure 4 Nitrogen K-edge spectra of arginine with increasing positive charge from the top. The peak at 401 eV is the result of the 1s→π*$_{CN}$ transition. The following large feature is due to the nitrogen σ bonds and cannot be individually assigned at this time.

Valine exhibits few changes on the nitrogen K-edge, see Figure 5(a) below. The lack of "acceptor only" behavior was expected due to valine having similar steric hindrance to lysine and diglycine, which has been previously cited in lacking acceptor only behavior.[1] Yet, there lacks even a shift of the main feature, transitions to σ* states, that was seen previously.[1, 2] This is surprising in that changing the electronic environment generally impacts the energy of the transitions.[20] There also exists the possibility of contamination due to a leak in the jet feed-through of the system during the data collection; this would affect valine's spectrum more than arginine's due to its lower experimental concentration and can be seen in the increased noise of the spectrum. On the carbon K-edge of valine, Figure 5(b), there seems to be little change with charge state, as expected and seen previously.[1, 2] The π* feature (ca. 289 eV) remains unaffected by changes in the pH and the σ* region (above 290 eV) remains undecipherable at this time.

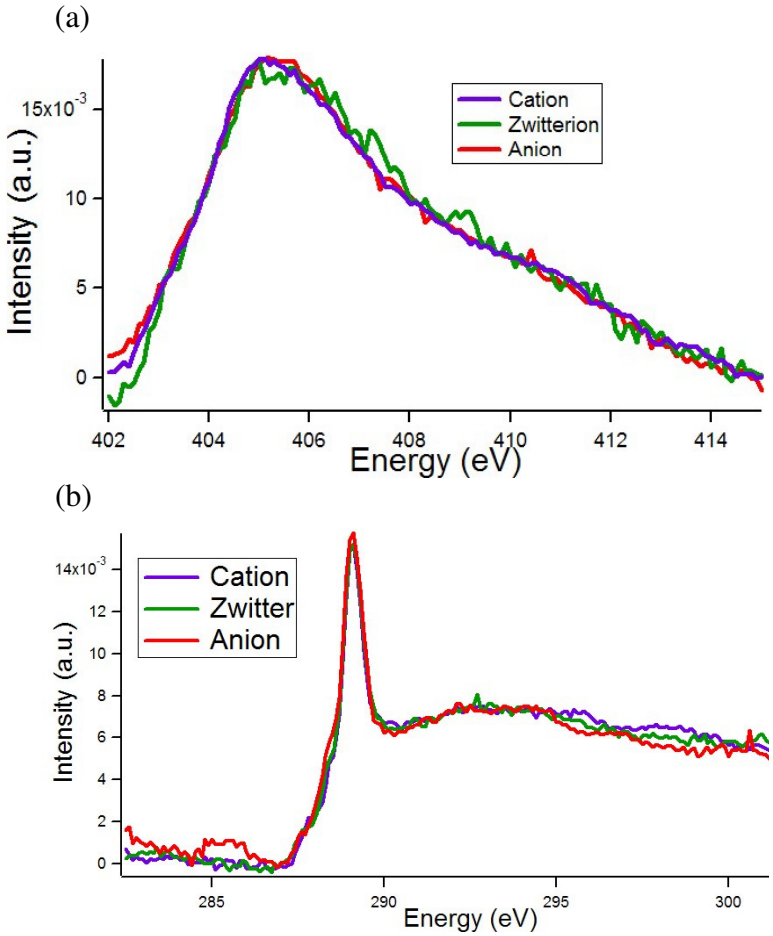


Figure 5. (a) Nitrogen K-edge spectra (b) carbon K-edge spectra of valine, under the various pH conditions; the cation corresponds to an acidic solution of pH 2, the zwitterion due to a pH 7 solution, the anion a basic solution of approximately pH 10.

As shown, the spectra of arginine and valine on the nitrogen K-edge show no significant changes with pH. This is surprising, as before there were indications of the "acceptor only" structures of glycine.[2] The lack of "acceptor only" features on the pre-edge was also seen in diglycine and lysine,[1] and was attributed to steric hindrance effects. It seems that this conclusion needs to be reevaluated with a larger library of amino acids and in light of current theoretical developments.[17, 21]

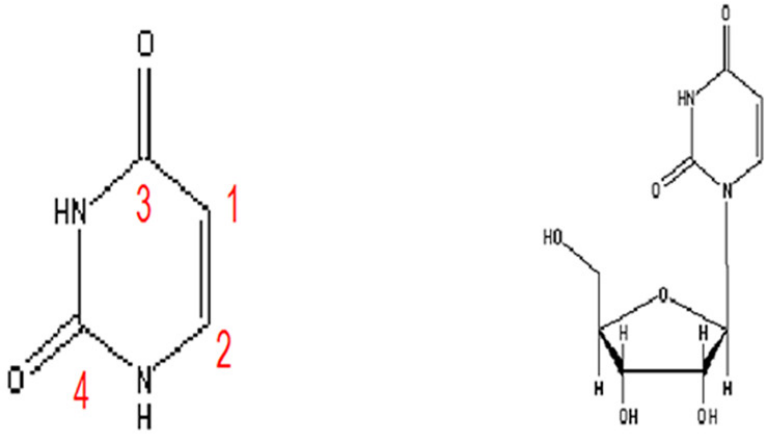
6.4 Nucleobases and Nucleosides

During the course of study of amino acids, work began on other small biomolecules. Examining the bases of DNA and RNA to understand the details of hydration and changes in hydrogen bonding was a logical next project. Uracil and cytosine were the only two bases soluble enough to yield experimental data. The only previous NEXAFS work addressing nucleic acids used powder form, and thin films,[9, 22] thus examining the impact of hydration on DNA and RNA was an important next step. A comparison was made of the nucleobase to its respective nucleoside (the nucleobase and ribose in these studies), e.g. uracil and uridine, and cytosine and cytidine.

The methods, both experimental and theoretical, employed are similar to that for the amino acids detailed previously in this chapter and similar to that of the carboxylate system which was detailed further in Chapter 2. The chemicals were purchased from Sigma Aldrich at a stated purity of 97% or higher; they were used without further purification with Millipore water to prepare 0.3 M solutions.

6.5 Uracil and Uridine

Uracil is a naturally occurring nucleobase found in RNA. The structures of both uracil, and its nucleoside, uridine, are given in Figure 6. The data were obtained only for the carbon K-edge due to a leak in the bellows of the jet feed-through disrupting data collection on the nitrogen K-edge.



(a) (b)

Figure 6. Structures of (a) uracil and (b) uridine. For uracil, the peaks are labeled for the four major peaks in the spectra, and can be translated to uridine.

The results for the carbon K-edge, Figure 7 (a), are in reasonable agreement with previously published work on uracil.[9, 22] The assignment of the π* peaks for both uracil and uridine are relatively straightforward, similar to NMR, the more electronegative the substituents attached the carbon are, the lower the transition energy.[20] The labels in Figure 6 on uracil, correspond to each peak; the lowest energy peak is labeled 1, and so on, and are directly transferable to uridine. Experimental and theoretical agreement for both uracil and uridine is reasonable within the localized π* region, but agreement is poorer within the overly structured σ* region; this is a known fault of the StoBe method, with its localized basis sets producing an overly structured σ* and continuum regions (above 290 eV).[17]

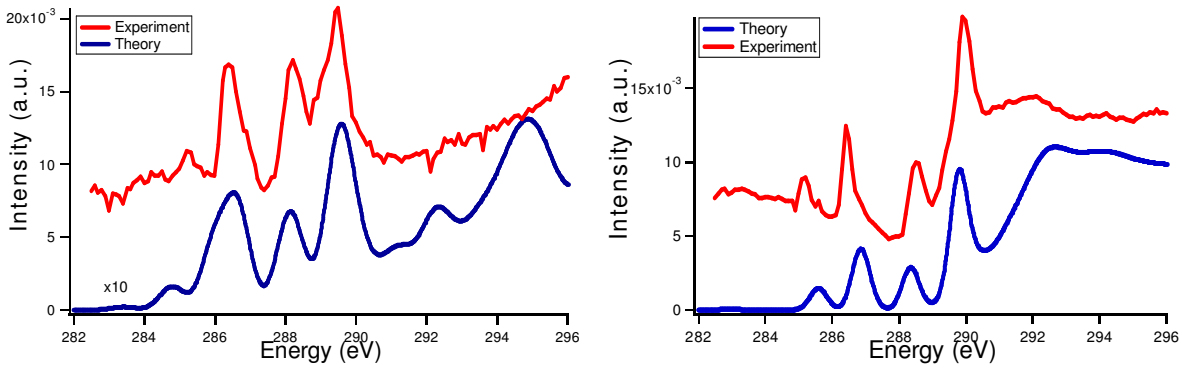


Figure 7. Comparison on the carbon K-edge spectra of (a) uracil and (b) uridine experimental(red) data versus theoretical(blue) calculations.

In comparing uracil to uridine, a few changes are evident, as seen in Figure 8. There are changes in the relative intensities of certain peaks. This can seen in the increase in signal in the σ* region, above 290 eV, and is due to an increase in the number of carbon σ bonds from the addition of the sugar. Additionally, the peak centered around 298.5 eV, the third peak from the left, is lessened upon the addition of ribose. Intuitively this seems strange, as this is due to the farthest carbon from the ribose. However, the addition of the ribose alters the tautomer populations,[23, 24] and thus it is possible that the comparison is not as straightforward as originally considered. Using pH controlled solutions to adjust tautomer populations would be a logical, but difficult next step in the studies of uracil and uridine. Additionally, improved theoretical studies would be useful in establishing primary configurations and the strength and impact of hydrogen bonding on the spectra, particularly on the nitrogen K-edge. And, of course, it is important to obtain a good experimental spectrum of the N-edges.

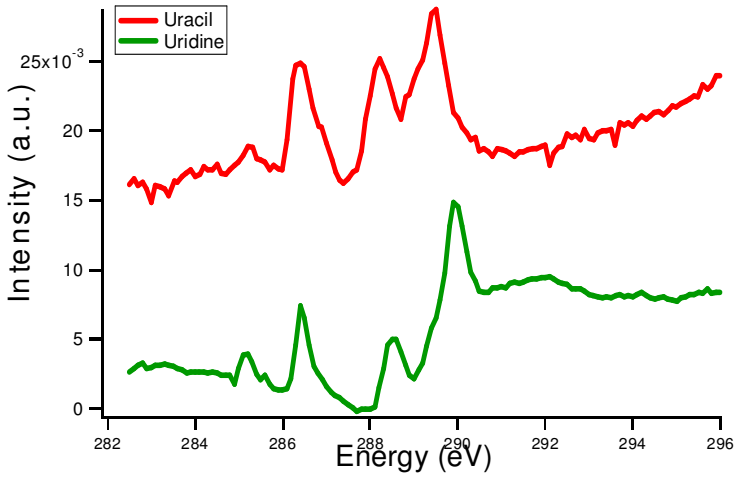


Figure 8. Comparison of uracil and uridine on the carbon K-edge, with uracil in red above, and uridine below in green.

6.6 Cytosine and Cytidine

Cytosine is a nucleobase found in both RNA and DNA, where cytidine is cytosine with ribose attached, structures are shown in Figure 9. As with the study of uracil compared to uridine, no measurements were possible on the nitrogen K-edge due to a leak to atmosphere.

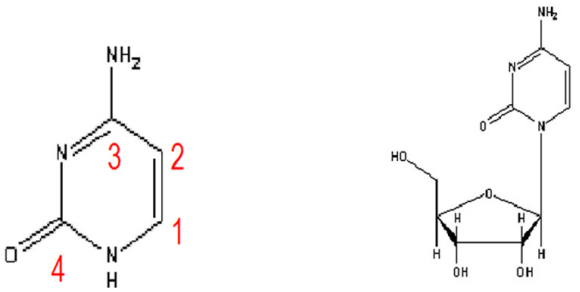


Figure 9. Structures of (a) cytosine and (b) cytidine. The carbons are labeled 1-4 to correspond with peak assignment. The labels on cytosine are directly transferable to cytidine.

The experimental results compared to theoretical results are shown below in Figure 10. Cytosine shows three π* features, and cytidine presents four π* features, agreeing with theoretical calculations. The three features from cytosine are from carbons 1 and 2 forming the first feature, and then carbons 3 and 4 forming individual features. The relative degeneracy of the first feature is broken for cytidine, where each carbon relates more directly to a π* feature. A further description of the states involved would provide useful information. Notably for both cytosine and cytidine, as for uracil and uridine, the region above 290 eV is overly structured in the theoretical calculations due to the localized basis set treatment.[17] Additionally, the relative heights of cytidine in the experimental spectrum appear to be in disagreement with the preliminary theoretical results shown. This shows the need for the new computational methods for such extended systems, particularly when a large degree of motion is involved.[21]

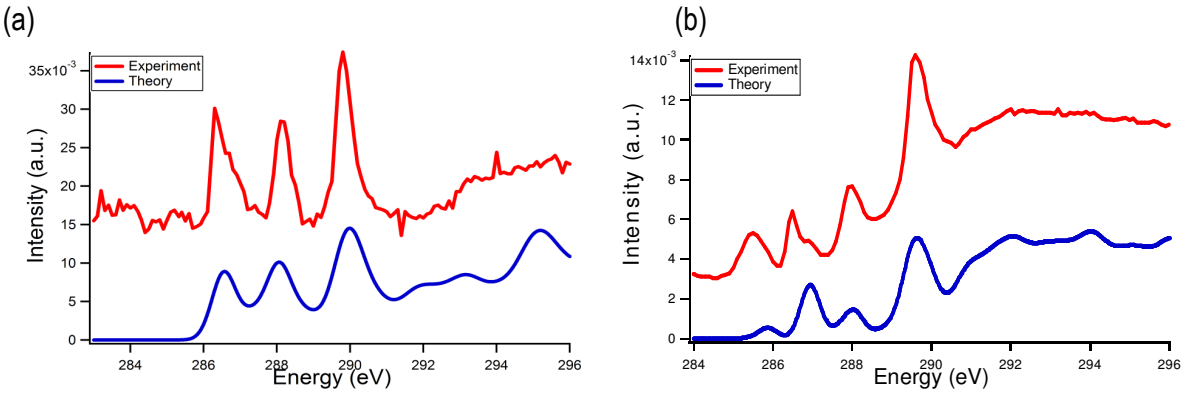


Figure 10. Comparison of experimental(red) (a) cytosine (b) cytidine data on the carbon K-edge to theoretical calculations(blue)

In comparing the cytosine to cytidine spectrum, Figure 11, there are two main differences to note. The first difference is the increase in signal above approximately 291 eV, corresponding to an increase in the number of carbons present by the addition of ribose. The second difference, as mentioned earlier, is the separation of the first feature into two distinct π* features. There appears to be a quasi-degeneracy in for the carbons labeled 1 and 2 which is broken upon the addition of the sugar. A more detailed look at the states involved, and how these change upon the addition of the sugar would be useful and could now be completed with current theoretical

techniques. Further sampling of the experimental data on all edges is needed to achieve a better comparison and understanding the impact of hydration.

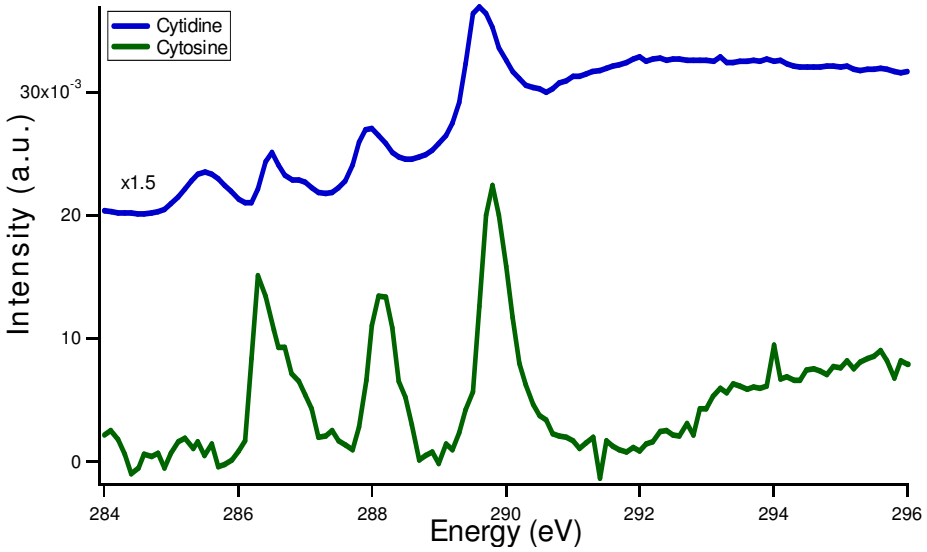


Figure 11. Comparison of experimental cytosine and cytidine on the carbon K-edge. Cytidine is in blue above, and cytosine is below in green.

References:

(1) Messer, B. M.; Cappa, C. D.; Smith, J. D.; Drisdell, W. S.; Schwartz, C. P.; Cohen, R. C.; Saykally, R. J., *J. Physc Chem. B* **2005,** 109, (46), 21640-21646.

(2) Messer, B. M.; Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B* **2005,** 109, (11), 5375-5382.

(3) Kolczewski, C.; Hermann, K., *Surface Science* **2004,** 552, (1-3), 98-110.

(4) Kamariotis, A.; Boyarkin, O. V.; Mercier, S. R.; Beck, R. D.; Bush, M. F.; Williams, E. R.; Rizzo, T. R., *J. Am. Chem. Soc.* **2006,** 128, (3), 905-916.

(5) Ide, M.; Maeda, Y.; Kitano, H., *J. Phys. Chem. B* **1997,** 101, (35), 7022-7026.

(6) Zubavichus, Y.; Fuchs, O.; Weinhardt, L.; Heske, C.; Umbach, E.; Denlinger, J. D.; Grunze, M., *Radiation Research* **2004,** 161, (3), 346-358.

(7) Kaznacheyev, K.; Osanna, A.; Jacobsen, C.; Plashkevych, O.; Vahtras, O.; Agren, H., *J. Phys. Chem. A* **2002,** 106, (13), 3153-3168.

(8) Gordon, M. L.; Cooper, G.; Morin, C.; Araki, T.; Turci, C. C.; Kaznatcheev, K.; Hitchcock, A. P., *J. Phys. Chem. A* **2003,** 107, (32), 6144-6159.

(9) MacNaughton, J. B.; Moewes, A.; Lee, J. S.; Wettig, S. D.; Kraatz, H. B.; Ouyang, L. Z.; Ching, W. Y.; Kurmaev, E. Z., *J. Phys. Chem. B* **2006,** 110, (32), 15742-15748.

(10) Cappa, C. D.; Smith, J. D.; Messer, B. M.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B* **2006,** 110, (11), 5301-5309.

(11) Cappa, C. D.; Smith, J. D.; Messer, B. M.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B* **2006,** 110, (3), 1166-1171.

(12) Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Messer, B. M.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B* **2005,** 109, (15), 7046-7052.

(13) Wilson, K. R.; Rude, B. S.; Smith, J.; Cappa, C.; Co, D. T.; Schaller, R. D.; Larsson, M.; Catalano, T.; Saykally, R. J., *Rev. Scientific Instrum.* **2004,** 75, (3), 725-736.

[14] Saunders, J. *Impact*, MacroModel: 2005.
[15] Hermann, K.; Pettersson, L.; Casida, M.; Daul, C.; Goursot, A.; Koester, A.; Proynov, E.; St-Amant, A.; Salahub, D. *StoBe-deMon version 3.0*, Version 3.0; StoBe Software: 2005.
[16] Uejio, J. S.; Schwartz, C. P.; Duffin, A. M.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (19), 6809-6812.
[17] Uejio, J. S.; Schwartz, C. P.; Saykally, R. J.; Prendergast, D., *Chem. Phys. Lett.* **2008,** 467, (1-3), 195-199.
[18] Smith, J. D.; Cappa, C. D.; Messer, B. M.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B* **2006,** 110, (40), 20038-20045.
[19] Leetmaa, M.; Ljungberg, M.; Ogasawara, H.; Odelius, M.; Naslund, L. A.; Nilsson, A.; Pettersson, L. G. M., *J. Chem. Phys.* **2006,** 125, (24).
[20] Urquhart, S. G.; Ade, H., *J. Phys. Chem. B* **2002,** 106, (34), 8531-8538.
[21] Schwartz, C. P.; Uejio, J. S.; Saykally, R. J.; Prendergast, D., *J. Chem. Phys.* **2009,** 130, (18), 184109.
[22] MacNaughton, J.; Moewes, A.; Kurmaev, E. Z., *J. Phys. Chem. B* **2005,** 109, (16), 7749-7757.
[23] Samijlenko, S. P.; Yurenko, Y. P.; Stepanyugin, A. V.; Hovorun, D. M., *J. Phys. Chem. B* 114, (3), 1454-1461.
[24] Feyer, V.; Plekan, O.; Richter, R.; Coreno, M.; Vall-Ilosera, G.; Prince, K. C.; Trofimov, A. B.; Zaytseva, I. L.; Moskovskaya, T. E.; Gromov, E. V.; Schirmer, J., *J. Phys. Chem. A* **2009,** 113, (19), 5736-5742.

Chapter 7.

Future Directions

NEXAFS spectroscopy of liquid microjets remains a promising technique, but new subjects for study and technical improvements always need to be considered to stay at the forefront of a field. This chapter is a discussion of possible technical improvements, systems to study, and limitations to consider.

7.1 Endstation Improvements

Currently, a member of the Saykally group is working on expanding the endstation capabilities to examine free radical systems. This should open new chemical systems and avenues of study, such as radical carbonate, and radical hydroxide. Still, some basic chamber modifications would be useful for easier maintenance. For a basic outline of the current endstation configuration please consult Chapter 1 and Wilson et al..[1]

During the course of this study, the liquid microjet feedthrough leaked upon multiple occasions, which limited our ability to study systems on the nitrogen K-edge. This equipment, produced by VG Scienta, comprising a z-bellows, and an xy manipulator, was damaged multiple times by drops of liquid falling and irreparably damaging the bellows. This will continue to be a problem as long as we run solutes, particularly at high concentrations. An alternative was proposed by Jonathan Spear,[2] an associate beamline scientist, suggested altering the current bellows arrangement to a long 2 ¾" tube with a magnetic mechanism to position in the z-direction. The current xy manipulator would remain with the Teflon insert to protect it, but a significantly cheaper stainless steel tube would replace the vulnerable z-bellows.

Currently, a main source of frustration when preparing to run experiments is attaching the six-way cross, a piece of equipment that holds the connection to the icicle breaker, a connection to the rough pump, and a liquid nitrogen trap. This piece is heavy and unwieldy, making it difficult to attach properly. The piece needs to be removed to move the chamber in and out of the Advanced Light Source and to clean adequately; with some redesign, the chamber would not need to be dismantled and could be easier to clean. At this time, a redesign of the chamber in regards to a more vertical orientation would make for both easier storage and cleaning.

7.2 Possible New Chemical Systems for Study

As a continuation of previous work,[3-6] it would be helpful to examine certain aqueous biological samples. Further peptide, and peptoid comparison would be useful, particularly in examining the impact of different R groups, and also longer chains, such as di- or tripeptide, and peptoid linkages.[4] Creating a library of soluble amino acids under different pH conditions, analogous to what was completed by Kaznacheyev et al. for solid samples, would be very useful for designing further study of these systems.[7] Similarly, examining the DNA and RNA bases, taking into account tautomer populations and using improved theoretical methods could lend further insight.

Due to improved theoretical methods,[8, 9] new systems for study are accessible and previously studied samples could be explored in greater detail. For example, the earlier mixing studies of methanol and ethanol could now be calculated within the framework of periodic systems, yielding a much better representation of the liquid samples.[10] Another project could be the continued study of the hydration of divalent cations, using a proposed approach for improved molecular dynamics sampling discussed in Chapter 3. To start a new direction, work by Dr.

David Prendergast of the Molecular Foundry has progressed to the point of enabling L-edge systems to be examined in detail, thus opening additional elements to study.

7.3 Advice to Future Coworkers

There are many directions to follow with liquid NEXAFS, but a scientist should consider a few points before proceeding with experiments. The first concern for measurements is the concentration available for experimentation; the concentration that will work with a liquid jet is always less than the literature maximum due to the backing pressure of the syringe pump. The current detection limit, ca. 0.15 M for a given element, should be considered before attempting to run a sample, though this limit varies if attempting to examine an edge with a lower photon flux. The next point to consider is the ability to reliably model the system of interest. Given a large system, or the types of atoms, check the availability of computing resources and potentials available to accurately model the experiment. Finally, consider the limitations of the beamline, for example the inherent resolution of the instruments available and the contamination by absorption by particular elements on the beamline optics.

(1) Wilson, K. R.; Rude, B. S.; Smith, J.; Cappa, C.; Co, D. T.; Schaller, R. D.; Larsson, M.; Catalano, T.; Saykally, R. J., *Rev. Scientific Instrum.* **2004,** 75, (3), 725-736.

(2) Spears, J. *Personal Communication*. Berkeley, CA, 2009.

(3) Uejio, J. S.; Schwartz, C. P.; Duffin, A. M.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J., *Proc. Natl. Acad. Sci. U. S. A.* **2008,** 105, (19), 6809-6812.

(4) Uejio, J. S.; Schwartz, C. P.; Duffin, A. M.; England, A. E.; Prendergast, D.; Saykally, R. J., *J. Phys. Chem. B* **2010,** 114, 4702-4709.

(5) Messer, B. M.; Cappa, C. D.; Smith, J. D.; Drisdell, W. S.; Schwartz, C. P.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B* **2005,** 109, (46), 21640-21646.

(6) Messer, B. M.; Cappa, C. D.; Smith, J. D.; Wilson, K. R.; Gilles, M. K.; Cohen, R. C.; Saykally, R. J., *J. Phys. Chem. B* **2005,** 109, (11), 5375-5382.

(7) Kaznacheyev, K.; Osanna, A.; Jacobsen, C.; Plashkevych, O.; Vahtras, O.; Agren, H., *J. Phys. Chem. A* **2002,** 106, (13), 3153-3168.

(8) Schwartz, C. P.; Uejio, J. S.; Saykally, R. J.; Prendergast, D., *J. Chem. Phys.* **2009,** 130, (18), 184109.

(9) Uejio, J. S.; Schwartz, C. P.; Saykally, R. J.; Prendergast, D., *Chem. Phys. Lett.* **2008,** 467, (1-3), 195-199.

(10) Wilson, K. R.; Rude, B. S.; Catalano, T.; Schaller, R. D.; Tobin, J. G.; Co, D. T.; Saykally, R. J., *J. Phys. Chem. B* **2001,** 105, (17), 3346-3349.