Measuring and Modeling NMR and Emission Spectra to Gain New Insight into Challenging Organic Compounds

2017

Jacob Powell
University of Central Florida

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

Part of the Chemistry Commons

STARS Citation

https://stars.library.ucf.edu/etd/5708

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
MEASURING AND MODELING NMR AND EMISSION SPECTRA TO GAIN NEW INSIGHT INTO CHALLENGING ORGANIC COMPOUNDS

by

JACOB R. POWELL
B.S. University of Florida, 2009
M.S. University of Central Florida, 2013

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Chemistry
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2017

Major Professor: James K. Harper
ABSTRACT

The advancement of theoretical methods in recent years has allowed the calculation of highly accurate spectroscopic parameters. Comparing these values to the corresponding experimental data can allow molecular structures to be elucidated. This dissertation details the use of experimental and theoretical data from nuclear magnetic resonance (NMR) and fluorescence spectroscopy to determine structure. Herein the NMR focus is on measuring & modeling chemical shift anisotropy and one-bond carbon-carbon J-coupling constants ($^{1}\text{J}_{\text{CC}}$). The fluorescence analysis models vibrationally resolved fluorescence spectra.

Chemical shift anisotropy techniques were used to study two conflicting crystal structures of the $n$-alkyl fatty acid, lauric acid. These two crystal structures differ only in their COOH conformation. Lattice-including density functional theory (DFT) refinements of each crystal structure failed to match experimental data leading to the proposal of a third crystal structure with a hydrogen disordered COOH moiety. This disorder strengthens the hydrogen bond providing a new rationalization to the long observed non-monotonic melting behavior of fatty acids having even and odd numbers of carbons.

The INADEQUATE is a NMR experiment that directly establishes the skeleton of organic compounds by measuring the $^{1}\text{J}_{\text{CC}}$ throughout a molecule. The low occurrence of $^{13}\text{C}-^{13}\text{C}$ pairs (1 in 10,000) and breaks in connectivity due to the presence of heteroatoms causes challenges to INADEQUATE analysis. Here, the insensitivity problem is overcome using analysis software that automatically processes data and identifies signals, even when they are comparable in magnitude to noise. When combined with DFT $^{1}\text{J}_{\text{CC}}$ predictions, configuration and confirmations of the natural products 5-methylmellein and hydroheptelidic acid are elucidated.
Vibrationally resolved fluorescence spectra of high molecular weight PAHs can be accurately calculated through time-dependent density functional theory (TD-DFT) methods. Here, the theoretical spectral profiles of certain PAHs are shown to match experimental high-resolution fluorescence spectra acquired at cryogenic temperatures. However, in all cases, theoretical spectra were systematically offset from experimental spectra. To decrease these uncertainties spectra were empirically corrected and an automated scheme employed to match theoretical spectra with all possible experimental spectra. In all cases the theoretical spectra were correctly matched to the experimental spectra.
To my mother, Mary
ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude and appreciation to my research advisor Dr. James K. Harper for his guidance and support throughout this work. Without him my knowledge of NMR would not be where it is today. I would also like to express my gratitude to all my committee members, Dr. Andres Campiglia, Dr. Melanie Beazley, Dr. David Richardson, and Dr. Richard Blair for their time, consideration, and assistance throughout my candidacy and dissertation.

I thank Dr. Robbie Iuliucci for the opportunities that allowed me to further develop my skills in NMR. Recognition goes to UCF’s Stokes Advance Research Computing Center (URL http://webstokes.ist.ucf.edu) for the computational resources that allowed me perform my calculations in a timely manner.

I would like to thank Dr. Campiglia’s research group for providing the state of the art Shpol’skii spectra of the polycyclic aromatic hydrocarbons used in this dissertation. I would also like to thank Dr. Moore’s research group for conducting the DNA and PCR analysis for the fungi described in the dissertation.

Special thanks goes to my fellow group members Yuemin Wang, Keyton Kalakewich, Geraldine Versfeld, Julia Nedimyer, Alex Elliot, and Luther Wang for their assistance in various research projects.

My deepest gratitude goes to all my family and friends for their endless support during my time as a graduate student here at UCF.
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................................................................. x

LIST OF TABLES ..................................................................................................................... xvi

LIST OF ABBREVIATIONS ....................................................................................................... xvii

CHAPTER 1: INTRODUCTION ................................................................................................. 1

1.1 Overview of Dissertation ................................................................................................. 1

1.2 Nuclear magnetic resonance (NMR) measurement of chemical shift anisotropy (CSA) .... 2

1.3 Measuring & modeling one bond carbon-carbon J-coupling constants ......................... 7

1.4 Vibrationally resolved fluorescence spectra ..................................................................... 9

1.5 Chapters that appear as publications .............................................................................. 11

1.6 References ....................................................................................................................... 12

CHAPTER 2: SOLID-STATE NMR AND DFT PREDICTIONS OF DIFFERENCES IN COOH HYDROGEN BONDING IN ODD AND EVEN NUMBERED N-ALKYL FATTY ACIDS ... 18

2.1 Abstract ......................................................................................................................... 18

2.2 Introduction ..................................................................................................................... 19

2.3 Experimental .................................................................................................................. 22

2.3.1 Materials ................................................................................................................... 22

2.3.2 Instrumention ............................................................................................................ 22

2.3.3 Computational details ............................................................................................... 25

2.4 Results and discussion .................................................................................................. 27

2.4.1 Refining lauric acid crystal structure ......................................................................... 27

2.4.2 Evaluating disordered hydrogen .............................................................................. 31
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 Conclusion</td>
<td>41</td>
</tr>
<tr>
<td>2.6 References</td>
<td>43</td>
</tr>
<tr>
<td>CHAPTER 3: EVALUATING STRUCTURE WITH $^{1}J_{cc}$ SCALAR COUPLING</td>
<td>49</td>
</tr>
<tr>
<td>3.1 Abstract</td>
<td>49</td>
</tr>
<tr>
<td>3.2 Introduction</td>
<td>49</td>
</tr>
<tr>
<td>3.3 Experimental</td>
<td>51</td>
</tr>
<tr>
<td>3.3.1 Materials</td>
<td>51</td>
</tr>
<tr>
<td>3.3.2 Taxonomic assignment of fungi</td>
<td>53</td>
</tr>
<tr>
<td>3.3.3 Instrumentation</td>
<td>54</td>
</tr>
<tr>
<td>3.3.4 Computational details</td>
<td>55</td>
</tr>
<tr>
<td>3.4 Results and discussion</td>
<td>56</td>
</tr>
<tr>
<td>3.4.1 Establishing the accuracy of theoretical methods.</td>
<td>56</td>
</tr>
<tr>
<td>3.4.2 Selecting structure with $^{1}J_{cc}$ in the presence of strong hydrogen bonding.</td>
<td>59</td>
</tr>
<tr>
<td>3.4.3 Characterization of hydroheptelicid acid with $^{1}J_{cc}$ data.</td>
<td>63</td>
</tr>
<tr>
<td>3.5 Conclusion</td>
<td>67</td>
</tr>
<tr>
<td>3.6 References</td>
<td>69</td>
</tr>
<tr>
<td>CHAPTER 4: PREDICTING ACCURATE FLUORESCENT SPECTRA FOR HIGH MOLECULAR WEIGHT POLYCYCLIC AROMATIC HYDROCARBONS USING DENSITY FUNCTIONAL THEORY</td>
<td>76</td>
</tr>
<tr>
<td>4.1 Abstract</td>
<td>76</td>
</tr>
<tr>
<td>4.2 Introduction</td>
<td>77</td>
</tr>
<tr>
<td>4.3 Experimental</td>
<td>82</td>
</tr>
<tr>
<td>4.3.1 Materials</td>
<td>82</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Instrumentation</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Computational details</td>
</tr>
<tr>
<td>4.4</td>
<td>Results and discussion</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Experimental fluorescence spectra at room temperature and 4.2 K.</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Predicting fluorescence spectra with PBE0 and CAM-B3LYP.</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Improving comparisons between predicted and experimental spectra.</td>
</tr>
<tr>
<td>4.4.4</td>
<td>The influence of solvent of calculated spectra</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusion</td>
</tr>
<tr>
<td>4.6</td>
<td>References</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>SUPPORTING INFORMATION FOR CHAPTER 2</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>SUPPORTING INFORMATION FOR CHAPTER 4</td>
</tr>
<tr>
<td>APPENDIX C</td>
<td>COPYRIGHT PERMISSION</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1–1. Illustration of CSA principal values as a function of orientation for the C=O moiety.

The $\delta_{11}$ component has the least electron density leading to the largest shift (a). $\delta_{22}$ has a higher electron density than $\delta_{11}$ causing an intermediate shift (b). $\delta_{33}$ has the largest electron density leading to the lowest shift (c). 4

Figure 1–2. Pulse sequence of FIREMAT. CP represents a cross polarization pulse transferring magnetization from $^1$H to $^{13}$C. The square rectangles on the $^{13}$C channel represent 180° pulses and the initial pulse on the $^1$H channel is a 90° pulse. T is the rotor period, $t_1$ and $t_2$ are, respectively, the evolution and acquisition times. 6

Figure 1–3. Pulse sequence of INADEQUATE. The narrow and wide square rectangles represent respectively 90° and 180° pulse on the $^{13}$C. The $\tau$ is a delay set to one fourth of the $^1$J$_{CC}$ coupling, $t_1$ is the evolution time delay, and $t_2$ is the acquisition time. 8

Figure 2–1. (a) Lauric acid showing the numbering used herein. Differences in the trans and cis orientations involve a 180° rotation about the O=C-Cα-Cβ dihedral angle as illustrated by structures (b) and (c), respectively. 20

Figure 2–2. The $^{13}$C FIREMAT spectrum obtained for lauric acid, phase C. A total of 11 lines were resolved in the isotropic spectrum (blue) and high signal-to-noise powder patterns obtained for each line to provide tensors. Black patterns indicate $sp^3$ sites while red denotes the COOH. 28

Figure 2–3. The root-mean-squared difference (rmsd) between calculated and experimental NMR $^{13}$C tensors before (blue) and after (yellow) refinement. The fit at $sp^3$ sites is significantly better than the agreement at the $sp^2$ (COOH) position. The expected error at
$^{13}$COOH moieties in well-defined structures is approximately 7.0 ppm (indicated by the 1σ dashed line) based on independent lattice-including computations of the $^{13}$COOH tensors in geometry refined structures of formic and acetic acid. These data suggest that neither the x-ray single crystal nor the powder diffraction structures of lauric acid are correct at the COOH moiety.

Figure 2–4. The agreement between experimental and computed $^{13}$C principal values for the COOH in lauric acid versus O–H separation distances. Both the cis and trans structures with a hydrogen located at a distance similar to that found in neutron diffraction (r = 1.0 Å) are rejected as feasible models at > 98% confidence. Two high probability locations for the hydrogen are identified at O–H distances of r = 1.16 ± 0.02 Å and 1.46 ± 0.04 Å. This outcome indicates a hydrogen that is dynamically disordered over two sites on the NMR timescale.

Figure 2–5. A comparison of calculated isotropic shifts (1/3*$[\delta_{11}+\delta_{22}+\delta_{33}]$) and spans ($\delta_{11}-\delta_{33}$) for the COOH hydrogen at different O–H bond lengths. These data both support the conclusion that proton disorder in the COOH creates stronger hydrogen bonds. All data points were computed at the B3LYP/D95** level of theory.

Figure 2–6. Differences between C–OH and C=O bond lengths in C7 to C14 n-alkyl monocarboxylic acids. All bond lengths are taken from x-ray single crystal data.

Figure 2–7. The agreement between experimental and computed 13C principal values for the COOH in undecanoic acid versus O–H separation distances. The cis structure with a localized hydrogen at an O–H separation of 1.70 ± 0.07 Å gives best agreement with
experimental 13C tensor data. This minimum corresponds to an O–H bond length of 1.02 Å.

Structures having a disordered hydrogen are rejected at > 95% statistical confidence.

Figure 3–1. Structures of (A) α-pinene, (B) β-pinene, and (C) carene showing the numbering employed herein.

Figure 3–2. A comparison between calculated and experimental $^1J_{CC}$ couplings for α-pinene, β-pinene and carene (N = 27) using the functionals B3LYP, B3PW91, and PBE0. The EPR-III basis set was employed in all computations.

Figure 3–3. The 8 model structures of 5-methylmellein considered herein. Newman projections viewed along the C3–C4 bond are included for each figure to illustrate the conformations of the C9 methyl group that were evaluated.

Figure 3–4. Calculated versus experimental $^1J_{CC}$ couplings of the 8 model structures of 5-methylmellein (left). The plot on the right includes only the best-fitting model to more clearly illustrate agreement between experimental and theoretical data. Model structures considered differences in OH hydrogen bonding, ring heteroatom identity and lactone ring conformation.

Figure 3–5. The 4 basic model structures considered for hydroheptelidic acid. For each structure, multiple conformational variations were also evaluated involving the C2–C3 and C4–C5 bonds.

Figure 3–6. Predicted conformations at the C2–C4 (top plot) and C4–C5 bonds (bottom) of hydroheptelidic acid.

Figure 3–7. Structures of hydroheptelidic acid found to be compatible with $^1J_{CC}$ data. Structures designated as major contributors constitute nearly 80% of the composition in solution.
Figure 4–1. Structures studied herein include benzo[a]pyrene (top left), dibenzo[a,e]pyrene (top right), dibenzo[a,i]pyrene (middle left), dibenzo[a,l]pyrene (middle right) and Benzo[a]pyrene–7,8,9,10-tetrol (bottom).

Figure 4–2. Fluorescence spectra for the PAHs studied here. In each plot, the top spectrum (blue) is obtained in n-octane at room temperature while the bottom spectrum (black) represents the same sample after cryogenic freezing in liquid helium. All Shpol’skii spectra were recorded under site-selective excitation ($\lambda_{\text{exc}}$) at the following wavelengths: $\lambda_{\text{exc}} = 298.0$ nm (benzo[a]pyrene), $\lambda_{\text{exc}} = 306.4$ nm (dibenzo[a,e]pyrene), $\lambda_{\text{exc}} = 312.0$ nm (dibenzo[a,i]pyrene) and $\lambda_{\text{exc}} = 318.0$ nm (dibenzo[a,l]pyrene). No Shpol’skii spectrum was acquired for benzo[a]pyrene tetrol, thus only the solution spectrum in methanol is displayed.

Figure 4–3. A comparison of experimental fluorescence spectra (black) in an n-octane matrix and calculated spectra (blue). The experimental spectrum for benzo[a]pyrene tetrol was acquired at room temperature in methanol. Theoretical spectra are calculated at the PBE0/cc-pVDZ level of theory and include n-octane or methanol as a solvent. The wavelength of the $S_{1,0} \rightarrow S_{0,0}$ transition is marked in both experimental and computed spectra.

Figure 4–4. A comparison of experimental emission spectra (black) and calculated spectra (blue). Theoretical spectra are obtained at the CAM-B3LYP/cc-pVDZ level of theory and include n-octane or methanol as a solvent. The wavelength of the $S_{1,0} \rightarrow S_{0,0}$ transition is indicated in both experimental and computed spectra.
Figure 4–5. A plot of experimental emission wavelengths for all major peaks in the 5 PAH samples studied versus the corresponding peaks in theoretical spectra. Computed spectra and are strongly correlated with experimental data with respective $R^2$ values of 0.925 and 0.932, respectively, for PBE0 and CAM-B3LYP.

Figure 4–6. A comparison of calculated and experimental fluorescence spectra. All computed spectra have been adjusted to correct for systematic over or underestimations in wavelengths and overestimations in the range of wavelengths over which the transitions occur. Computed spectra from the PBE0 and CAM-B3LYP methods are represented, respectively, by blue and orange plots.

Figure 4–7. Residuals derived from a comparison of the theoretical spectrum of dibenzo[\textit{a,e}]pyrene with experimental data from the 5 PAHs studies herein. Only the experimental spectrum of dibenzo[\textit{a,e}]pyrene matches this theoretical data, demonstrating that accurate vibronic spectra can be computed using DFT methods.

Figure 4–8. A comparison of the DFT predicted spectra for dibenzo[\textit{a,e}]pyrene before (left) and after (right) adjusting the theoretical spectra to minimize the residuals. The initial theoretical spectra (left) have previously been partially adjusted utilizing the empirical correction described in the text.

Figure 4–9. A plot showing the influence of solvent on DFT predicted spectra of PAHs. Spectra in the left and right columns show, respectively, data computed with the PBE0 and CAM-B3LYP methods. Green, orange and black spectra denote, respectively, spectra calculated in the gas phase, spectra computed with solvent included and experimental data. Most experimental spectra were acquired at 4.2 K in \textit{n}-octance, but the benzo[\textit{a}]pyrene tetrok
spectrum was acquired at room temperature in methanol. The computed spectra include either $n$-octane or methanol to match experimental conditions.
LIST OF TABLES

Table 2–1. Experimental $^{13}$C chemical shift tensor principal values for lauric acid. .................. 27

Table 2–2 Typical chemical shift tensor principal values$^a$ at $^{13}$COOH and $^{13}$COO$^-$ moieties
measured in 74 amino acids compared to tensors from lauric acid, phase C. ....................... 36

Table 2–3. Experimental 13C chemical shift tensor principal values (ppm) for undecanoic acid,
phase C’. ..................................................................................................................................... 39

Table 3–1. Taxonomic assignment to the genus level for fungi studied herein. ....................... 54

Table 3–2. Selected analysis parameters for all 2D INADEQUATE analysis performed in this
study........................................................................................................................................... 55

Table 3–3. The $^{13}$C Chemical shift and experimental and theoretical $^1$J$\text{CC}$ couplings for the model
terpenes. .................................................................................................................................... 57

Table 3–4. Uncertainty in theoretical $^1$J$\text{CC}$ values for three functionals. ......................... 58

Table 3–5. Carbon-13 chemical shifts (ppm) and $^1$J$\text{CC}$ couplings (Hz) for 5-methylmellein. ..... 61

Table 3–6. Agreement between experiment and calculated $^1$J$\text{CC}$ couplings for the 8 model
structures considered for 5-methylmellein............................................................................. 62

Table 3–7. Carbon-13 chemical shifts (ppm) and $^1$J$\text{CC}$ couplings (Hz) for hydroheptelic acid. 64

Table 3–8. Agreement between experiment and calculated $^1$J$\text{CC}$ couplings for the 4 model
structures of hydroheptelic acid. ............................................................................................... 65

Table 4–1. Linear correlations relating computed and experimental emission spectra............ 92

Table 4–2. Experimental spectra matching calculated DFT spectra........................................ 96
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^1$H</td>
<td>Hydrogen isotope 1</td>
</tr>
<tr>
<td>$^1$J&lt;sub&gt;CC&lt;/sub&gt;</td>
<td>One-bond carbon-carbon J-couplings</td>
</tr>
<tr>
<td>$^1$J&lt;sub&gt;CH&lt;/sub&gt;</td>
<td>One-bond carbon-hydrogen J-couplings</td>
</tr>
<tr>
<td>$^2$J&lt;sub&gt;CC&lt;/sub&gt;</td>
<td>Two-bond carbon-carbon J-couplings</td>
</tr>
<tr>
<td>$^2$J&lt;sub&gt;CH&lt;/sub&gt;</td>
<td>Two-bond carbon-hydrogen J-couplings</td>
</tr>
<tr>
<td>$^3$J&lt;sub&gt;CH&lt;/sub&gt;</td>
<td>Three-bond carbon-carbon J-couplings</td>
</tr>
<tr>
<td>$^3$J&lt;sub&gt;HH&lt;/sub&gt;</td>
<td>Three-bond hydrogen-hydrogen J-couplings</td>
</tr>
<tr>
<td>$^{13}$C</td>
<td>Carbon isotope 13</td>
</tr>
<tr>
<td>ADEQUATE</td>
<td>Adequate sensitivity double-quantum spectroscopy</td>
</tr>
<tr>
<td>ARCC</td>
<td>Advance Research Computing Center</td>
</tr>
<tr>
<td>CASTEP</td>
<td>Cambridge Serial Total Energy Package</td>
</tr>
<tr>
<td>CSA</td>
<td>Chemical shift anisotropy</td>
</tr>
<tr>
<td>DFT</td>
<td>Density functional theory</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>EPA</td>
<td>Environmental Protection Agency</td>
</tr>
<tr>
<td>FIREMAT</td>
<td>Five-pi replicated magic angle turning</td>
</tr>
<tr>
<td>GIAO</td>
<td>Gauge including atomic orbital</td>
</tr>
<tr>
<td>GIPAW</td>
<td>Gauge including projector augmented wave</td>
</tr>
<tr>
<td>HETCOR</td>
<td>Heteronuclear correlation</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>HPLC</td>
<td>High performance liquid chromatography</td>
</tr>
<tr>
<td>INADEQUATE</td>
<td>Incredible Natural Abundance Double Quantum Transfer Experiment</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>LETRSS</td>
<td>Laser-excited time-resolved Shpol’skii spectroscopy</td>
</tr>
<tr>
<td>MeOH</td>
<td>Methanol</td>
</tr>
<tr>
<td>m.w</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>P2DSS</td>
<td>Pseudo 2D spinning side-band suppression</td>
</tr>
<tr>
<td>PAH</td>
<td>Polycyclic aromatic hydrocarbons</td>
</tr>
<tr>
<td>PCM</td>
<td>Polarizable continuum model</td>
</tr>
<tr>
<td>PHORMAT</td>
<td>Phase-corrected magic angle turning</td>
</tr>
<tr>
<td>ppb</td>
<td>Parts per billion</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-mean-squared difference</td>
</tr>
<tr>
<td>RT</td>
<td>Room Temperature</td>
</tr>
<tr>
<td>RTF</td>
<td>Room Temperature fluorescence</td>
</tr>
<tr>
<td>SCF</td>
<td>Self-consistent field</td>
</tr>
<tr>
<td>SSNMR</td>
<td>Solid-state Nuclear magnetic resonance</td>
</tr>
</tbody>
</table>
SOPPA  Second-order polarization propagator approach
SPINAL  Small phase incremental alteration
\( t_1 \)  Evolution time
\( t_2 \)  Acquisition time
TDDFT  Time-dependent density functional theory
TIGER  Technique for importing greater evolution resolution
TPPM  Two pulse phase-modulation
CHAPTER 1: INTRODUCTION

1.1 Overview of Dissertation

The theoretical prediction of experimental results has long been a focus in chemistry. Early methods were often limited in their ability to provide accurate results and in the kinds of observations that could be predicted.\(^1\) In recent years significant progress has been made and, at present, very accurate predictions can be calculated for a large range of experimental observables such as identifying minimum energy structures,\(^2\) bond dissociation enthalpies,\(^3\) and IR and Raman spectra.\(^4\) Several types of calculations are now sufficiently accurate that theoretical modeling results can be employed to interpret unusual experimental observations and even predict new experimental outcomes. The work described in this dissertation emphasizes a combination of experimental and theoretical results to interpret and predict spectroscopic measurements with the primary emphasis on NMR and fluorescence data. In the following chapters, calculated NMR and fluorescence spectroscopic methods are employed to resolve discrepancies of crystal structures, enhance structural characterizations of several natural compounds, and differentiate between spectra of structurally similar environmental toxins. The remainder of this overview chapter provides a technical background needed to understand the material that follows. Chapter 2 focuses on hydrogen bonding differences in \(n\)-alkyl fatty acids with odd and even numbers of carbons. Discrepancies between alternative crystal structures of lauric acid are resolved herein by modeling various hydrogen bond lengths in lauric acid and comparing the respective calculated NMR tensor values to experimental data. The theoretical model developed is quite general and can be applied to other \(n\)-alkyl fatty acids to determine whether the hydrogen bond is disordered or not. Chapter 3 describes how comparisons of
experimental and calculated NMR one-bond carbon-carbon J couplings ($J_{CC}$) can be used to solve structural problems in organic products. This work relies on generating a wide variety of models involving changes in configuration, conformation, and heteroatom identity then comparing the calculated J coupling values to experimental data. Poor agreement with experimental values allow improbable models to be rejected with a known statistical confidence. Chapter 4 focuses on the calculation of accurate vibrationally resolved fluorescence spectra of unusually high molecular weight polycyclic aromatic hydrocarbons (PAHs). The match between a given theoretical spectrum and all known experimental spectra is quantified automatically using a program developed for this purpose. This Python code allows a library of calculated spectra of different PAHs to be compared to all possible experimental spectra, thus narrowing down potential matches.

1.2 Nuclear magnetic resonance (NMR) measurement of chemical shift anisotropy (CSA)

One advantage of solid-state NMR over solution NMR is that it provides information about the chemical shift anisotropy (CSA). The CSA arises from the dependence of chemical shifts on the orientation of molecule in the magnetic field and is mathematically described by a second-rank tensor represented by a $3 \times 3$ matrix, (equation 1–1). In powders only the diagonal terms, $\delta_{xx}$, $\delta_{yy}$, and $\delta_{zz}$, can be experimentally measured and as the powders contain molecules in every possible orientation, directional information relative to the magnetic field is lost. Tensors in powders are therefore listed in the so-called “Mehring notation” $\delta_{11} \geq \delta_{22} \geq \delta_{33}$. These terms are known as principal components or principal values.
Figure 1–1 illustrates the principal components for a carboxyl carbon atom. The electron cloud surrounding the C=O moiety creates electronic shielding represented by an ellipsoid in which the axis with the least electron density has the largest shift, $\delta_{11}$, and the axis with the most electron density has the smallest shift, $\delta_{33}$. The three principal components can be averaged together through by spinning a sample at an angle of 54.7˚ relative to the magnetic field (i.e. the “magic angle”) to form an isotropic shift, $(\delta_{\text{iso}} = 1/3(\delta_{11} + \delta_{22} + \delta_{33}))$. The isotropic shift is the parameter commonly measured in solution NMR. The three principal component shift tensors are able to provide more structural information for the molecule than just the isotropic shift. This is evident with a carbonyl carbon’s hydrogen bonding where the isotropic shift can vary between a range of ±6 ppm depending on the hydrogen bond length.$^9,10$ In contrast, when using the three principal component shift tensors variation as large as ± 18 ppm can be observed, allowing COOH and COO$^-$ moieties to be distinguished.$^{11}$ This ability to improve upon isotropic measurements is generally observed and gives CSA a large advantage in elucidating molecular and crystal structures.
$\delta_{11}$ component has the least electron density leading to the largest shift (a). $\delta_{22}$ has a higher electron density than $\delta_{11}$ causing an intermediate shift (b). $\delta_{33}$ has the largest electron density leading to the lowest shift (c).

$^{13}$C principal shift tensors can be experimentally measured by slow spinning magic angle turning experiments. The most accurate slow spinning experiments are phase corrected magic angle turning (PHORMAT), and Five-pi replicated magic angle turning (FIREMAT). These 2D experiments display the isotropic shift peaks along one dimension and powder patterns containing principal value information for each of the isotropic shift peaks along the second dimension. For the purposes of this dissertation only FIREMAT will be discussed. The pulse sequence of FIREMAT is illustrated in Figure 1–2 where after enhancing $^{13}$C magnetization via
cross polarization, five 180° pulses are applied on the $^{13}$C channel uniformly spaced through one rotor period. The pulses at one sixth, one half, and five sixths of the rotor period can vary in their positions. The $^1$H channel is decoupled throughout the pulse sequence. FIREMAT implements two other methodologies including a variation of the pseudo 2D spinning side-band suppression (P2DSS) data rearrangement scheme$^{15}$ and the technique for importing greater evolution resolution (TIGER) data processing technique$^{16}$ to improve the quality of data over previous methods. The P2DSS works by copying and rearranging the data collected in the directly observed dimension (i.e. the acquisition dimension ($t_2$)) for every rotor period. This is feasible because the $t_2$ signal repeats every rotor period. This allows for an extended indirect (or “evolution”) dimension with far less data collection. TIGER is a linear least-squares method for extracting slices from a 2D data set and replaces Fourier transformation. TIGER reduces the amount of acquired data needed, greatly reducing overall analysis time for FIREMAT. The FIREMAT has been shown to be a very accurate experiment with a RSMD of 0.7 ppm for the $^{13}$C chemical shift tensor when compare to benchmark values collected from highly accurate single crystal NMR methods.$^{17}$
Figure 1–2. Pulse sequence of FIREMAT. CP represents a cross polarization pulse transferring magnetization from $^1$H to $^{13}$C. The square rectangles on the $^{13}$C channel represent 180° pulses and the initial pulse on the $^1$H channel is a 90° pulse. T is the rotor period, $t_1$ and $t_2$ are, respectively, the evolution and acquisition times.

In the work described here, theoretical calculation of CSA is included to facilitate interpretation of experimental data. The theoretical computation of CSA, calculates the chemical shielding rather than shift, (Equation 1–2).\(^{18}\)

$$\sigma_{\alpha\beta} = \frac{\partial^2 E}{\partial \mu_\alpha \partial B_\beta}$$  \hspace{1cm} (1–2)

where $\sigma_{\alpha\beta}$ is the shielding tensor, $\mu_\alpha$ is the magnetic moment of the nucleolus in direction $\alpha$ and $B_\beta$ is the external magnetic field in direction $\beta$. The major theoretical methods for the calculation of shielding tensors include gauge including atomic orbital (GIAO)\(^{19}\) and the gauge including projector augmented wave (GIPAW).\(^{20}\) For shielding calculations of powders, the GIPAW method includes lattice effects and is incorporated in the software CASTEP.\(^{21}\) Prior work shows that highly accurate molecular and crystal structure refinements can be made by comparing CSA values from calculated theoretical models to experimental tensor data.\(^{5,17,22}\)
1.3 Measuring & modeling one bond carbon-carbon J-coupling constants

The interaction of NMR active nuclear spins between chemical bonded sites results in a phenomenon known as J-coupling. J-coupling is sensitive to bonded partners within 1-3 bonds. The strength of the coupling is mediated by the interactions of the electrons in the bond with the magnetic field. J-coupling is also known as indirect spin-spin coupling and occurs through bonding elections rather than through space. This dissertation focuses on one bond J-couplings between two carbon atoms (i.e. $^1J_{CC}$). The magnitude of $^1J_{CC}$ coupling ranges from below 10 Hz to over 250 Hz, allowing $^1J_{CC}$ couplings to be used to elucidate molecular structure. Examples of the factors influencing $^1J_{CC}$ include stereochemical configuration, torsion angle conformation, the identity of atoms bonded to the two carbons of interest and the percent s and π character of the bond between two carbons.

Herein, $^1J_{CC}$ coupling constants are experimentally measured by the 2D experiment known as Incredible Natural Abundance Double Quantum Transfer Experiment (INADEQUATE) developed by Bax. This is an experiment displaying in the directly observed dimension a peak split into a doublet by the $^1J_{CC}$ values but positioned at the 1D $^{13}$C resonance. Signal in the indirect dimension appears at the sum of two chemical shifts of the bonded $^{13}$C pairs. The strength of INADEQUATE is that it is designed for natural abundance compounds and signals from isolated $^{13}$C peaks are suppressed with only bonded $^{13}$C pairs being detected. The pulse sequence for INADEQUATE is shown in Figure 1–3. The sequence consists of a total of four pulses and three delays on the $^{13}$C channel. The first pulse is a 90° pulse on the $^{13}$C channel followed by a delay $\tau$ which is one fourth of the $^1J_{CC}$ coupling. Next is an 180° pulse on the $^{13}$C channel followed by another $\tau$ of the same length. A second 90° pulse is applied on the
$^{13}$C channel and to generate a two dimensional spectrum a variable evolution delay, $t_1$, is used before the final 90° pulse on the $^{13}$C channel. The proton channel is decoupled throughout the pulse sequence. In this dissertation the $^1J_{CC}$ coupling constants are extracted from the spectrum using the software NMR Analyst developed by Dr. Reinhard Dunkel$^{38}$ to detect low S/N signals.

![Diagram](image)

Figure 1–3. Pulse sequence of INADEQUATE. The narrow and wide square rectangles represent respectively 90° and 180° pulse on the $^{13}$C. The $\tau$ is a delay set to one fourth of the $^1J_{CC}$ coupling, $t_1$ is the evolution time delay, and $t_2$ is the acquisition time.

Theoretical $^1J_{CC}$ coupling constants are used for comparison with experimental data and are calculated in the program Gaussian$^{39-42}$ The J-coupling values are calculated through Equations (1–3) and (1–4):$^{39}$

\[
\begin{align*}
J_{KL} &= \hbar \frac{\gamma_K \gamma_L}{2\pi} \frac{d^2E}{dM_K dM_L} \\
M_K &= \gamma_K \hbar I_k
\end{align*}
\]

(1–3)

(1–4)

where $J_{KL}$ is the tensor of the J-coupling, $\gamma_K$ is the gyromagnetic ratio of the nucleolus, $M_K$ is the nuclear magnetic momentum and $I_K$ is the nuclear spin. Like the chemical shift, J-couplings are
second rank tensors, but only the average value is measured in INADEQUATE. The magnitude of the $^{1}J_{CC}$ coupling constant reflects four electronic mechanisms: $^{40,43}$ Diamagnetic spin-orbit, paramagnetic spin-orbital, Fermi contact, and spin-dipole. The Fermi contact is largest contributor of the four mechanisms completely overshadowing the other three mechanisms in lighter elements (i.e. organics). $^{39-41}$ The calculation of J-couplings thus requires methods that focus primarily on the Fermi contact term. As the Fermi contact reflects electronic properties at the nucleus of the atom, basis sets used require highly constrained s-functions for accurate calculations. Herein, the basis set EPR-III is demonstrated to be a suitable basis set. $^{44,45}$ Prior studies have shown the effectiveness of combining theoretical and experimental data to refine molecular structures. $^{46-48}$

1.4 Vibrationally resolved fluorescence spectra

The resolution of fluorescence spectra can be improved to the point that they become vibrationally resolved through a process known as Shpol’skii spectroscopy. $^{49,50}$ This high resolution spectroscopy method works by diluting a guest molecule in a solvent host matrix that has been frozen at 77 K or below and reduces typically broad spectral bands to show the vibronic transitions. $^{49}$ High quality Shpol’skii spectroscopy requires that three conditions be met. $^{51}$ The first is that the solvent host matrix must be frozen into an ordered polycrystalline matrix. This insures that only one peak per vibronic transition appears. A disorder random glass would create multiple peaks per site. The second condition requires that the guest-host interactions to be minimal to prevent band shifts due to solvent effects. This is achieved by using n-alkanes as solvents. The final condition is that the guest molecule’s dimensions match those of the host
alkane. For example, if the guest molecule is the PAH naphthalene, then the best host would be n-pentane as both molecules have a length along their greatest dimension of 5 carbons. When the dimensions of both the guest and host match, significantly improved resolution are observed in the spectra. Shpol’skii spectroscopy can be used in the identification and differentiation of PAHs.\(^{49, 51-59}\)

Vibrationally resolved fluorescence spectra can be theoretically calculated through a process developed by Barone et al.\(^{60-62}\) Gaussian is used for the calculations of all theoretical spectra described herein.\(^{63-66}\) The transitions from the excited state to the ground state’s vibrational levels are known as vibronic transitions. Previously, calculations of vibrationally resolved fluorescence spectra has been restricted to small molecules due to the extremely large number of vibrational states that must be evaluated. However, the majority of these transitions do not actually contribute to the spectrum and recently a set of rules has been established to determine the non-negligible transitions. Restricting calculation to only these levels now allows calculations on large structures to be completed.\(^{67, 68}\) Barone’s computational methods have been used to predict all fluorescence spectra described in this dissertation.\(^{69-75}\) While theoretical spectra with accurate relative vibronic transitions do not initially match experimental spectra due to a systematic offset, a computer program is described that empirically adjusts the calculated spectra to match experimental values. This computational correction is done without user bias or input and ultimately produces reasonably accurate spectra in all cases.
1.5 Chapters that appear as publications

The following chapters appear as they were published in peer reviewed journals. Chapter 2 was published in Physical Chemistry Chemical Physics.\textsuperscript{76} Chapter 4 was published in Journal of Molecular Spectroscopy.\textsuperscript{77}
1.6 References


CHAPTER 2: SOLID-STATE NMR AND DFT PREDICTIONS OF DIFFERENCES IN COOH HYDROGEN BONDING IN ODD AND EVEN NUMBERED N-ALKYL FATTY ACIDS


2.1 Abstract

For nearly 140 years n-alkyl monocarboxylic acids have been known to exhibit unusual non-monotonic melting between odd and even numbered acids. This behavior has been rationalized in terms of packing density at the hydrocarbon tails, with COOH hydrogen bonding considered to be invariant among different acids. A recent ambiguity involving the COOH conformation between two crystal structures of lauric acid suggests that COOH structure and hydrogen bonding may play a role in these differences. Here, the two conflicting lauric acid crystal structures are further refined using lattice-including DFT refinement methods. Solid-state NMR (SSNMR) $^{13}$C chemical shift tensor data are employed to monitor refinement quality by comparing experimental and computed tensors. This comparison provides a more sensitive measure of structure than X-ray data due to SSNMR’s ability to accurately locate hydrogens. Neither diffraction structure agrees with SSNMR data and an alternative is proposed involving a hydrogen disordered COOH moiety. The disordered hydrogen dynamically samples two most probable positions on the NMR timescale with O–H bond lengths of 1.16 and 1.46 Å. This disordered structure is consistent with SSNMR, IR and X-ray C–O and C=O bond lengths. The hydrogen disorder appears to be restricted to even numbered acids based on undecanoic acid’s
13COOH tensor data and C–O and C=O bond lengths for other n-alkyl acids. This disorder in even numbered acids results in stronger hydrogen bonds than are found in odd acids and invites a reevaluation of the melting behavior of n-alkyl acids that includes these differences in hydrogen bonding.

2.2 Introduction

Lauric acid is a twelve carbon n-alkyl monocarboxylic acid (Figure 2–1a) occurring naturally in certain plant oils1–4 and in the milk of some mammals.5 Lauric acid’s long-studied influence on human health and diet together with its well-known antimicrobial activity6, 7 has made it the focus of a vast number of studies. While much of this work centers on lauric acid’s properties and behavior in solution, there is also interest in lauric acid in the solid state where it exhibits polymorphism, with 5 known phases8–14 Phase C,15 which occurs immediately before melting, is of particular interest, in part, because lauric acid has a high enthalpy of fusion and may therefore be useful as a thermal energy storage material16, 17 Of perhaps greater general interest is the fact that solid n-alkyl monocarboxylic acids present an intellectual puzzle that has intrigued scientists for over a century. As the number of carbons in these acids increase, an unusual non-monotonic increase in melting point is observed with the even numbered acids melting at higher temperatures.18 Densities, sublimation enthalpies and solubilities also exhibit this behavior with even numbered acids showing higher values.12 Crystallography has been employed to elucidate this unusual behavior and this work has resulted in the publication of reasonably accurate crystal structures for most of these acids. Curiously, lauric acid is an exception where two crystal structures have recently been reported for phase C and these
structures disagree on the orientation of the COOH moiety. The single crystal x-ray structure\textsuperscript{13} specifies a \textit{trans} orientation having a O=C-\textit{C}\textalpha-C\textbeta dihedral angle of 180\degree while the x-ray powder structure\textsuperscript{12} indicates a \textit{cis} orientation with an angle of 0\degree (Figures 2–1b and 2–1c).

![Diagram of Lauric Acid with Numbering]

Figure 2–1. (a) Lauric acid showing the numbering used herein. Differences in the \textit{trans} and \textit{cis} orientations involve a 180\degree rotation about the O=C-\textit{C}\textalpha-C\textbeta dihedral angle as illustrated by structures (b) and (c), respectively.

Recent studies have demonstrated that SSNMR provides an extremely sensitive way to monitor crystal structure refinements.\textsuperscript{19-29} These SSNMR investigations accurately locate
hydrogens in many cases and therefore provide a critical piece of structural information often lacking in x-ray diffraction. These refinement studies typically employ DFT methods that account for lattice effects (e.g. CASTEP\textsuperscript{30}). Improvement is indicated by better agreement between experimental and calculated SSNMR parameters and a decrease in forces on the atoms of the lattice. In a few cases, where a high quality reference structure is independently known\textsuperscript{23}-\textsuperscript{25, 29} (e.g. single crystal neutron diffraction of the same phase), it has been demonstrated that these refinements consistently create coordinates that are in better agreement with the reference structure than were with the original unrefined values. This SSNMR work is of considerable interest because it offers an experimentally verified non-diffraction method for further improving low-resolution structures in a manner that provides accurate coordinates for all atomic sites \textit{including} hydrogens.

One of the challenges to current DFT refinement studies is the observation that, while large changes are observed in the computed SSNMR parameters, only small movements occur in most atom positions. Indeed, in many of the structures studied, the new atomic positions lie within the error reported for the original diffraction coordinates. Questions thus arise about whether such refinements can actually lead to new insights. Recent work has demonstrated that, at least in a few cases, new insights can be obtained. For example, refinement of the structure of cellulose I\textsubscript{α}\textsuperscript{31} clarified the hydrogen bonding arrangement at 4 positions and eliminated an alternative hydrogen-bonding scheme thought to be feasible based on diffraction data.\textsuperscript{32} A second study demonstrated that refinement could establish coordinates of atoms that were missed in a diffraction study of a pentapeptide.\textsuperscript{33} This result is of particular interest because high-resolution data is difficult to obtain for many proteins and nucleic acids. Here, DFT and other
refinement methods are employed to reconcile the differences found in lauric acid’s crystal structures. In this case, refinement of a single hydrogen in lauric acid’s crystal structure significantly improves agreement with SSNMR shift tensor data and provides new understanding into differences in hydrogen bonding in \( n \)-alkyl monocarboxylic acids.

2.3 Experimental

2.3.1 Materials

Lauric acid was purchased from Aldrich and prepared in the C form by melting the solid and then quenching it in liquid nitrogen. The resulting solid was gently ground to prevent a phase change and this powder was used for all solid-state NMR analyses.

2.3.2 Instrumentation

An x-ray powder diffraction was conducted to ensure that the lauric acid powder obtained from the melt was phase C (\( P2_1/c \), monoclinic). Analysis was performed using a Rigaku Miniflex 600 diffractometer, with \( \theta-2\theta \) Bragg-Brentano geometry, and a 600 W (40 kV, 15 mA) Cu X-ray tube source using Ka (\( \lambda = 1.5418 \text{ Å} \)) radiation. Samples were measured from \( 2\theta = 4 \) to \( 40^\circ \) using a step size of 0.02\(^\circ \) and a scan rate of 1.5 s per step. Samples were prepared by placing the powder on a glass sample holder. The powder pattern obtained closely matched the pattern derived from prior x-ray diffraction studies. Phases other than phase C exhibit large differences in their powder patterns, thus a visual comparison of the powder pattern was sufficient to clearly distinguish phase C from alternative polymorphs.
A five-pi replicated magic angle turning (FIREMAT)\textsuperscript{34} analysis was performed on the C phase of lauric acid in order to obtain $^{13}$C tensor principal values. FIREMAT data was acquired on a Chemagnetics CMX400 spectrometer using 7.5 mm PENCIL probe and operating at a frequency of 100.61916 MHz. Acquisition parameters included evolution and acquisition spectral widths of 26.9 and 65.4 kHz, respectively, a recycle time of 20 s, pulse widths of 4.5 and 9.4 $\mu$s for the $^1$H 90° and $^{13}$C 180° pulses, respectively, a 3 ms cross-polarization time, spinning speed of 527 Hz, decoupler frequency of 400.18800 MHz and a TPPM\textsuperscript{35} phase angle of 33.6°. The spectrum was externally referenced to the methyl resonance of hexamethyl benzene at 17.35 ppm. A total of 51 evolution dimension increments were acquired of 192 transients each. The digital resolution of the acquisition dimension was 16.0 Hz per point. The evolution dimension was significantly extended using a data rearrangement process described elsewhere\textsuperscript{34} to ultimately provide a digital resolution of 16.0 Hz per point. The FIREMAT data were TIGER\textsuperscript{36} processed.

The $^{13}$C chemical shift tensor for the COOH moiety in undecanoic acid phase C'\textsuperscript{15} was measured on an Agilent DD2 500 MHz narrow-bore spectrometer operating at 125.68166 MHz with an Agilent 1.6 mm T3 probe. Acquisition parameters for this 1D spectrum included a spectral width of 50.0 kHz, a 3.0 ms cross-polarization time, a $^1$H proton pulse width of 1.0 $\mu$s, a digital resolution was 48.8 Hz per point, and SPINAL decoupling at a $^1$H frequency of 499.77670. The spectrum was externally referenced to the methyl resonance of hexamethyl benzene at 17.35 ppm. Analysis was conducted at 23 °C using a non-spinning sample and a 25 s recycle time because phase C’ has stability over only a small temperature range of 17.2 °C to 28.5 °C\textsuperscript{37} and the risk of melting the sample or inducing a phase change by spinning or pulsing
too quickly was considered high. The $^{13}$COOH powder pattern was completely resolved from all other signals allowing principal values to be determined by visually inspecting the spectrum. The error in principal values is estimated to be $\pm 2$ ppm based on the width of the singularities representing the principal values in the powder pattern.

An x-ray powder diffraction analysis was conducted on undecanoic acid to verify that the powder utilized was phase C’ ($P2_1/c$, monoclinic). All analysis conditions are identical to those described above for lauric acid. The observed powder pattern closely matched a pattern derived from the prior single-crystal diffraction study.

All $^1$H/$^{13}$C heteronuclear correlation (HETCOR) spectra were acquired on an Agilent DD2 500 MHz narrow-bore spectrometer operating at 125.68166 MHz together with an Agilent 1.6 mm T3 probe. Acquisition parameters include acquisition and evolution dimension spectral widths of 29.8 kHz and 14.0 kHz, respectively, 128 evolution increments of 64 transients each, a recycle time of 16 s, and a spinning speed of 12.0 kHz. SPINAL decoupling was used for all spectra using a frequency of 499.77528 MHz and a 165° pulse of 1.84 $\mu$s. Each evolution step employed a $^1$H frequency offset of $\pm 175.0$ kHz and 360° pulse widths of 3.3 $\mu$s. A series of HETCOR spectra were acquired using Lee-Goldburg cross-polarization and contact times of 100 $\mu$s, 200 $\mu$s, 300 $\mu$s and 400 $\mu$s. All other acquisition parameters are as listed above. The $^{13}$C dimensions of all spectra were externally referenced to the methyl peak in hexamethyl benzene at 17.35 ppm. The $^1$H dimensions of all spectra were externally referenced to a non-spinning sample of liquid DMSO in a sealed capillary with a resonance at 2.49 ppm. All $^1$H shifts were scaled by 0.577 as required for optimal Lee-Goldburg decoupling.
The isotropic $^1$H chemical shift of the COOH moiety in lauric acid phase C was measured by acquiring a 1D spectrum on an Agilent DD2 500 MHz narrow-bore spectrometer operating at 499.77528 MHz. An Agilent 1.6 mm T3 probe was employed and data was acquired at a spinning speed of 25.0 kHz. Other acquisition parameters included a 1.0 µs 90° pulse width, a spectral width of 147.1 kHz and a recycle time of 20.0 s. The spectrum was externally referenced to a non-spinning sample of liquid dimethyl sulfoxide sealed in a glass capillary with a $^1$H resonance at 2.49 ppm.

An elemental analysis of lauric acid phase C gave % C = 71.88, % O = 16.13, and H = 12.10 % versus theoretical values of % C = 71.95, % O = 15.97, % H = 12.08. All analyses were performed at Atlantic Microlab, Inc. and are reported to have an uncertainty of ± 0.3 %.

2.3.3 Computational details

All geometry refinement and NMR tensor calculations were performed at the University of Central Florida’s Advanced Research Computing Center (ARCC). Both of lauric acid’s crystal structures were initially refined using the planewave DFT code CASTEP together with the PBE functional and ultrasoft pseudopotentials. The “ultra-fine” level was selected using a planewave basis set cut-off energy of 610 eV. The threshold convergence for SCF tolerance was $5 \times 10^{-7}$ eV/atom and a k-point spacing of 0.071/Å was employed. The minimizing approach of Broyden, Fletcher, Goldfarb and Shanno\textsuperscript{41} was used for geometry optimizations. Optimizations were considered converged when a change in energy threshold of $5 \times 10^{-6}$ eV/atom was reached, a maximum Cartesian force of 0.01 eV/Å on all atoms was achieved, and the maximum displacement of $5 \times 10^{-4}$ Å for each atom was observed. The unit cell dimensions from the single
crystal\textsuperscript{13} and powder diffraction\textsuperscript{12} studies agree within 1\% with respective values of $a = 27.563$ Å, $b = 4.9627$ Å, $c = 9.5266$ Å, $\beta = 98.006^\circ$ and $a = 27.54$ Å, $b = 4.953$ Å, $c = 9.604$ Å, $\beta = 97.28^\circ$. The DFT methods employed are known to overestimate these parameters,\textsuperscript{19} thus refinement of both structures did not include adjustment of these values. After geometry refinement, lattice-including NMR calculations were performed (PBE/ultrafine) using the gauge including projector augmented wave (GIPAW)\textsuperscript{42} method.

Model structures were prepared involving a pair of butyric acid molecules to evaluate COOH hydrogen positions. Butyric acids were manually placed in a $R_2^2(8)$hydrogen bonding arrangement. Non-hydrogen atoms were placed at positions determined from lauric acid’s single crystal diffraction structure. All hydrogens in C–H bonds were located from the CASTEP geometry refinement as described above. The COOH hydrogens were placed at a series of positions ranging from 1.0 to 1.7 Å in steps ranging in size from 0.03–0.1 Å and $^{13}$C NMR shift tensors computed at the B3LYP/D95** level of theory for each structure without further geometry optimization. The resulting computed shieldings for the COOH carbons were converted to shifts using previously established\textsuperscript{43} slope and intercept values of 1.01 and 194.93 which have been found to give accurate shifts for many $sp^2$ carbons.

The influence of disorder on hydrogen bond strength was evaluated by calculating the COOH $^1$H chemical shielding tensor at a series of O–H separation distances ranging from 1.0–1.7 Å in steps of 0.1 Å. All calculations were performed at the B3LYP/D95** level of theory using the butyric acid dimer model described above. The $^1$H shielding values for tetramethyl silane were also computed (B3LYP/D95**) after geometry optimization (B3LYP/D95*) and employed as a reference to convert shieldings to shifts.
2.4 Results and discussion

2.4.1 Refining lauric acid crystal structure

A wide variety of SSNMR parameters have now been utilized to monitor the refinement of crystal structure. Among these, the chemical shift tensor principal values have been reported to be one of the more sensitive parameters, in part because three shifts are measured for each nuclear position. Accordingly, $^{13}$C shift tensors were acquired for lauric acid using the FIREMAT tensor measurement experiment. A total of 11 isotropic resonances were resolved and corresponding tensors obtained for all lines (Table 2–1). An illustration of the FIREMAT spectrum obtained is given in Figure 2–2. All resonances were assigned to molecular positions by a process described in Appendix A.

Table 2–1. Experimental $^{13}$C chemical shift tensor principal values for lauric acid.

<table>
<thead>
<tr>
<th>Position</th>
<th>$\delta_{11}$ (ppm)</th>
<th>$\delta_{22}$ (ppm)</th>
<th>$\delta_{33}$ (ppm)</th>
<th>$\delta_{iso}$ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1$^a$</td>
<td>238.8</td>
<td>199.4</td>
<td>107.6</td>
<td>181.9</td>
</tr>
<tr>
<td>2$^a$</td>
<td>55.9</td>
<td>42.7</td>
<td>8.0</td>
<td>35.5</td>
</tr>
<tr>
<td>3$^a$</td>
<td>37.0</td>
<td>28.2</td>
<td>9.8</td>
<td>25.0</td>
</tr>
<tr>
<td>4$^a$</td>
<td>49.0</td>
<td>35.6</td>
<td>11.7</td>
<td>32.1</td>
</tr>
<tr>
<td>5</td>
<td>52.4</td>
<td>36.8</td>
<td>13.1</td>
<td>34.1</td>
</tr>
<tr>
<td>6$^b$</td>
<td>49.1</td>
<td>35.6</td>
<td>13.1</td>
<td>32.6</td>
</tr>
<tr>
<td>7$^b$</td>
<td>50.8</td>
<td>36.6</td>
<td>13.0</td>
<td>33.5</td>
</tr>
<tr>
<td>8$^b$</td>
<td>50.8</td>
<td>36.6</td>
<td>13.0</td>
<td>33.5</td>
</tr>
<tr>
<td>9$^b$</td>
<td>50.2</td>
<td>36.2</td>
<td>13.2</td>
<td>33.2</td>
</tr>
<tr>
<td>10</td>
<td>52.8</td>
<td>39.2</td>
<td>12.9</td>
<td>35.0</td>
</tr>
<tr>
<td>11$^a$</td>
<td>35.7</td>
<td>28.0</td>
<td>12.7</td>
<td>25.5</td>
</tr>
<tr>
<td>12$^a$</td>
<td>24.0</td>
<td>20.0</td>
<td>2.0</td>
<td>15.3</td>
</tr>
</tbody>
</table>

$^a$Experimentally determined from $^1$H/$^{13}$C correlation experiments. $^b$Assigned are interchangeable.
Figure 2–2. The $^{13}\text{C}$ FIREMAT spectrum obtained for lauric acid, phase C. A total of 11 lines were resolved in the isotropic spectrum (blue) and high signal-to-noise powder patterns obtained for each line to provide tensors. Black patterns indicate $sp^3$ sites while red denotes the COOH.

When multiple crystal structures have been reported for the same phase, lattice-including DFT refinements usually reconcile structural differences and cause convergence to a single structure.\textsuperscript{24, 25} It was therefore hypothesized that the differences in lauric acids COOH structure would be resolved by refinement. Both of lauric acid’s crystal structures were therefore refined and $^{13}\text{C}$ tensors computed using lattice-including DFT methods (see 2.3 Experimental). This process significantly improved agreement between computed and experimental tensors at CH$_2$ and CH$_3$ positions (i.e. C2–C12). In fact, the agreement at all $sp^3$ sites after refinement is comparable to the expected experimental error,\textsuperscript{47} suggesting that the structure at C2–C12 is correct. In contrast, the COOH conformation in both refined structures remained unchanged and each $^{13}\text{COOH}$ gave poor agreement with experimental data. Figure 2–3 illustrates agreement between experimental and calculated NMR tensors (root-mean-squared difference, rmsd) before
and after refinement. The poor fit at the COOH may indicate that both the cis or trans structures are incorrect or may simply reflect a greater difficulty in calculating accurate shift tensors at COOH moieties. To distinguish between these two possibilities, $^{13}$C shift tensors were computed for the COOH sites in formic and acetic acid. These model compounds have well-established crystal structures$^{48,49}$ with COOH hydrogens purported to be localized. Of equal importance is the fact that the $^{13}$C shift tensor data is known for both molecules.$^{50,51}$ The geometry of both crystal structures was refined and $^{13}$C tensors calculated using the lattice-including methods employed for lauric acid. An rmsd of 7.0 ppm was observed at the $^{13}$COOH in the model compounds. The corresponding rmsd in the cis structure of lauric acid was 18.7 ppm, eliminating it as a feasible structure. The trans structure with an rmsd of 10.5 ppm is also improbable with an rmsd falling outside of 86.6% of a Gaussian distribution (i.e. 1.5 standard deviations). The poor fit at both the cis and trans structures suggests that other structural models should be considered for the COOH. However, it is notable that the trans structure cannot be eliminated at higher confidence levels (e.g. > 99%), thus any model considered must include the possibility of selecting this structure.
Figure 2–3. The root-mean-squared difference (rmsd) between calculated and experimental NMR $^{13}$C tensors before (blue) and after (yellow) refinement. The fit at $sp^3$ sites is significantly better than the agreement at the $sp^2$ (COOH) position. The expected error at $^{13}$COOH moieties in well-defined structures is approximately 7.0 ppm (indicated by the 1σ dashed line) based on independent lattice-including computations of the $^{13}$COOH tensors in geometry refined structures of formic and acetic acid. These data suggest that neither the x-ray single crystal nor the powder diffraction structures of lauric acid are correct at the COOH moiety.

One possible source of error in the COOH moiety involves misplacement of the acidic hydrogen. Hydrogen disorder is suggested from the single crystal x-ray structure of the C10 and C12 $n$-alkyl acids in form C where C–O and C=O bond lengths are found to be similar.\textsuperscript{13} In contrast, the C11 and C13 acids exhibit C–O and C=O bond lengths more typical of a single and double bond and a localized COOH hydrogen. Disorder in the even numbered acids may involve either the occurrence of both cis and trans conformations in different parts of the lattice (static disorder) or the presence of COOH hydrogens that can transfer between partners in the dimer
over time in a dynamic process. The previous x-ray analyses of lauric acid phase C\textsuperscript{12,13} were unable to unambiguously assign the hydrogen position. Here, static disorder is rejected as a possibility because only a single COOH resonance is observed experimentally (see Figure 2–2). Two signals would likely be observed in a \textsuperscript{13}C isotropic (1D) spectrum of a sample with static disorder.

An infrared spectroscopy study on lauric acid phase C previously identified both the \textit{cis} and \textit{trans} conformations in the solid and measured the ratio as a function of temperature.\textsuperscript{52,53} However, this study failed to distinguish static from dynamic disorder and designated the structures only as \textit{cis} or \textit{trans}, providing only rough estimates of hydrogen position. Solid-state NMR has the potential to provide a more accurate position for the disordered hydrogen and to gauge differences in hydrogen bond strength between the disordered and ordered acids.

2.4.2 Evaluating disordered hydrogen

The dynamic disorder in lauric acid was characterized using a theoretical approach that considers a variety of positions for the COOH proton. A model was created by preparing a pair of butyric acid molecules hydrogen bonded in the head-to-head R\textsubscript{2}(8) arrangement\textsuperscript{54} found in the crystal structure of lauric acid and many other \textit{n}-alkyl acids. Additivity rules predict that substituents beyond the \textit{γ}–position in carboxylic acids will have a very small influence on \textsuperscript{13}COOH shift,\textsuperscript{55} thus, truncating lauric acid to butyric acid is unlikely to introduce significant errors. Moreover, prior SSNMR work has demonstrated that \textsuperscript{13}C tensors computed from such COOH dimers are comparable in accuracy to those described herein obtained from lattice-including calculations (i.e. GIPAW) on model structures.\textsuperscript{56} Here, the dimeric model was
preferred over GIPAW for two reasons. First, the butyric acid dimer provides a general model applicable to all n-alkyl acids involved in $R^2_2(8)$ hydrogen bonding. Second, this model includes only 12 non-hydrogen atoms and is therefore computationally inexpensive, allowing a large variety of candidate structures to be evaluated. All non-hydrogen atom positions in the dimeric model were defined using bond distances and angles from lauric acid’s single crystal structure. The C–H hydrogens were placed in positions obtained from the CASTEP refinement. A series of candidate structures were then generated by lengthening each O–H bond in the dimer from an initial separation of $r = 1.0$ Å in steps varying in size from 0.03 to 0.1 Å to a final separation of 1.7 Å. The initial structure corresponds to the trans conformation with an O–H bond length matching the neutron diffraction value in similar compounds.$^{57}$ The final 1.7 Å separation mimics the cis orientation. At each of the hydrogen position, $^{13}$C tensors were computed at the B3PW91/D95** level of theory. This approach has been employed previously by others and found to give accurate results.$^{58,59}$ The agreement with experimental data is shown in Figure 2–4 and indicates that hydrogen’s environment is characterized by two sites separated by an energy barrier that allows the hydrogen to sample each position on the NMR timescale. The best agreement is achieved at O–H distances of $r = 1.16 \pm 0.02$ and $r = 1.46 \pm 0.04$ Å. To ensure that the dimeric model was valid, a similar analysis was performed with lattice-including methods using lauric acid’s crystal structure. These results are included in appendix A and demonstrate that nearly identical O–H distances are selected as the most probable structures. While O–H bond lengths longer than 1.0 Å (i.e. the prototypical value found by neutron diffraction)$^{60}$ are less commonly encountered, hydrogen disorder is carboxyl dimers is well known from experimental
diffraction data.\textsuperscript{61-65} Our analysis indicates that lauric acid phase C belong to this class of hydrogen disorder solids.

The proposal for disordered hydrogens implies an energy barrier between the two positions. The height of this energy barrier was evaluated by computing the energy at each point using CASTEP (PBE/ultrafine) and was found to be 36.7 kJ mol\textsuperscript{-1}. A list of energy at each point and Boltzmann populations is included in appendix A. This energy barrier is significantly larger than RT at room temperature, suggesting that proton transfer in lauric acid occurs through a tunneling process rather than classical hopping. These calculations predict energy minima at O–H separations of $r = 1.04$ Å and $1.60$ Å corresponding to O–H bond lengths of 1.04 Å and 1.12 Å for the trans and cis structures, respectively. It is notable that these lengths which are slightly longer than the value of 1.0 Å expected for localized hydrogens.\textsuperscript{60} The predicted bond lengths generally agree with the NMR/DFT values but are consistently shorter by an average of 0.13 Å.
Figure 2–4. The agreement between experimental and computed $^{13}$C principal values for the COOH in lauric acid versus O–H separation distances. Both the cis and trans structures with a hydrogen located at a distance similar to that found in neutron diffraction ($r = 1.0$ Å) are rejected as feasible models at > 98% confidence. Two high probability locations for the hydrogen are identified at O–H distances of $r = 1.16 \pm 0.02$ Å and $1.46 \pm 0.04$ Å. This outcome indicates a hydrogen that is dynamically disordered over two sites on the NMR timescale.

Prior work has demonstrated that benzoic acid also forms hydrogen bonds in an $R_{2}^{2}(8)$ arrangement and that COOH proton transfer occurs through coherent tunneling.$^{66, 67}$ Moreover, benzoic acid’s $^{13}$COOH principal values have been measured$^{68}$ and are similar to those reported here for lauric acid. Perhaps the most relevant metric for comparing $^{13}$COOH tensors is $\delta_{11}-\delta_{22}$ because this value has been shown to be strongly correlated with O–H bond length.$^{58}$ Such a comparison yielded $\delta_{11}-\delta_{22}$ values of 39 and 36 ppm for lauric acid and benzoic acid, respectively. This strong similarity between $^{13}$COOH tensors indicates a similar electronic environment at the COOH moieties of these acids and suggests that tunneling makes a key contribution to hydrogen disorder in lauric acid, phase C.

These tunneling results taken together with the observation of a single $^{13}$COOH isotropic resonance and the fit to the dimer model (Figure 2–4) suggests a dynamic hydrogen transfer that
occurs through tunneling between the two COOH sites in lauric acid phase C. Thus the measured $^{13}$COOH shift tensor likely represents a weighted average of the $r = 1.16 \text{ Å}$ and $1.46 \text{ Å}$ structures. The rmsd of such an averaged structure versus experiment is 5.6 ppm, a value statistically indistinguishable from the two minima in Figure 2–4.

Another kind of evidence also supports the contention that lauric acid’s hydrogen is dynamically disordered. Prior work by Gu et al. demonstrated that COO$^-$ and COOH groups in amino acids could be distinguished by their $^{13}$C principal values. A summary of typical tensor values for COOH and COO$^-$ moieties measured in 74 amino acids and peptides is given in Table 2–2 together with principal values measured at C1 in lauric acid. These data show a close match between the typical COO$^-$ tensors and C1 of lauric acid. However, elemental analysis of phase C measuring %C, %H and %O unambiguously demonstrates that lauric acid contains the COOH functional group (see 2.3 Experimental). These data indicate that the electronic environment at lauric’s COOH is very similar to that found in COO$^-$ groups. Proton disorder appears to lessen the difference between C–O and C=O bond lengths. A correlation between the difference in these bond lengths (i.e. C–O minus C=O) and O–H bond length has previously been predicted by Facelli et al. This early work also predicted a decrease in the magnitude of $\delta_{11} - \delta_{22}$ as O–H bond length increased, an outcome that is observed in lauric acid. Thus the prediction of COOH hydrogen disorder explains the similarity between lauric acid’s $^{13}$C COOH data and the COO$^-$ data in amino acids reported previously.
Table 2–2 Typical chemical shift tensor principal values\(^a\) at \(^{13}\)COOH and \(^{13}\)COO\(^-\) moieties measured in 74 amino acids compared to tensors from lauric acid, phase C.

<table>
<thead>
<tr>
<th></th>
<th>(\delta_{11})</th>
<th>(\delta_{22})</th>
<th>(\delta_{33})</th>
<th>(\delta_{\text{iso}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>COOH</td>
<td>257 ± 6</td>
<td>158 ± 18</td>
<td>109 ± 8</td>
<td>175 ± 6</td>
</tr>
<tr>
<td>COO(^-)</td>
<td>241 ± 4</td>
<td>183 ± 18</td>
<td>108 ± 4</td>
<td>177 ± 6</td>
</tr>
<tr>
<td>C1 (lauric acid, phase C)</td>
<td>238.8</td>
<td>199.4</td>
<td>107.6</td>
<td>181.9</td>
</tr>
</tbody>
</table>

\(^a\)Average values are obtained from 31 COOH and 43 COO\(^-\) moieties.

It has been known since 1877\(^70\) that \(n\)-alkyl carboxylic acids display alternation in their melting points between odd and even numbered acids. Considerable effort has focused on explaining this difference.\(^71\)–\(^77\) One unvarying feature of all these explanations is the assumption that the hydrogen bonding at the COOH position is of equal strength in both even and odd numbered acids. One way to evaluate hydrogen bond strength is to measure the \(^1\)H shift in the COOH moiety. It is well known that \(^1\)H shifts in O–H···O moieties in the range of 12–21 ppm indicate strong hydrogen bonding with higher frequency shifts indicating stronger hydrogen bonding.\(^78\) A \(^1\)H spectrum of lauric acid exhibited a COO\(^1\)H resonance at 13.6 ppm, consistent with moderately strong hydrogen bonding. Unfortunately, all attempts to measure a similar \(^1\)H spectrum of undecanoic acid phase C’ were unsuccessful because the high spinning speeds needed for analysis (i.e. 25–30 kHz) appeared to melt the sample based upon the unusually narrow \(^1\)H lines observed upon spinning. This outcome cannot easily be avoided by decreasing temperature because phase C’ changes phase at 17.2° C. To overcome this limitation, a second analysis was included involving computing \(^1\)H shifts for the various disordered model structures.

The computed shifts (described above) for the butyric acids model structures include \(^1\)H isotropic and tensor principal values for all disordered structures. A comparison of calculated isotropic shifts is given in Figure 2–5 and indicates that hydrogen disordered structures have
higher frequency shifts and thus stronger hydrogen bonds. The strongest hydrogen bonding occurs for a hydrogen positioned at an equal distance from each oxygen atom (i.e. \( r = 1.3 \) Å). This prediction is consistent with prior studies of very strong O–H···O hydrogen bonds that find hydrogens in such bonds equally positioned between oxygen atoms in bonds that exhibit significant covalent character.\(^7\) Further support for the conclusion that disorder creates stronger hydrogen bonds comes from the computed \(^1\)H shift tensor data. It is known that the span (i.e. \( \delta_{11} - \delta_{33} \)) of \(^1\)H shift tensors in X–\(^1\)H···Y bonds are sensitive to hydrogen bond strength with larger spans indicating stronger hydrogen bonding.\(^8\) For the dimeric model structure considered, the span increases as the O–H bond length increases with a maximum value observed at \( r = 1.3 \) Å (Figure 2–5). Thus the data on span also indicate that stronger hydrogen bonding results from proton disorder.

Figure 2–5. A comparison of calculated isotropic shifts \((1/3[\delta_{11}+\delta_{22}+\delta_{33}])\) and spans \((\delta_{11}-\delta_{33})\) for the COOH hydrogen at different O–H bond lengths. These data both support the conclusion that proton disorder in the COOH creates stronger hydrogen bonds. All data points were computed at the B3LYP/D95** level of theory.
Overall, these analyses of $^1$H shifts suggest that lauric acid phase C has stronger hydrogen bonding than undecanoic acid phase C’. However, predictions regarding the magnitude of this difference cannot be made from the present data. We note that a more quantitative approach is often performed in order to compute hydrogen bond strength involving comparing the energy for the dimer in the optimized hydrogen bonding position against isolated monomers. In the current study such calculations are infeasible because the O–H bond lengths in the disordered structures do not represent stationary points.

The above discussion on differences in COOH hydrogen disorder in odd and even numbered acids is based primarily on bond length differences at C–O and C=O in diffraction structures. These differences (i.e. C–O minus C=O) are said to be observed in odd and even numbered acids from C6 to C15\textsuperscript{12} with the odd acids having larger values on average (Figure 2–6). However, these differences are consistently smaller than the ideal value of 0.10 Å found in localized COOH moieties\textsuperscript{57} and in some cases (e.g. C9/C10) are negligible. It is therefore desirable to look at other evidence to more fully establish that even and odd numbered acids differ in hydrogen disorder and hydrogen bond strength. Accordingly, the $^{13}$COOH tensor principal values were measured for the phase of undecanionic acid that occurs immediately before melting (i.e. phase C’\textsuperscript{15}) using an approach described in Experimental (Table 2–3). The agreement between undecanionic acid’s experimental COOH tensors and computed values from the dimeric model structure described previously is illustrated in Figure 2–7. These data establish that undecanionic acid’s COOH hydrogen is localized with an O–H bond length of 1.02 ± 0.07 Å and a Cβ–Cα–C=O moiety in a cis orientation. This prediction agrees with the single crystal x-ray diffraction structure where a cis conformation is observed. Our prior discussion of hydrogen
bonding in the dimer model and $^1$H chemical shifts indicates that localized $R^2(8)$ hydrogen bonds are weaker than delocalized bonds. Thus significant differences in hydrogen disorder and hydrogen bond strength are observed between lauric and undecanoic acid. Admittedly, the nature of this comparison is limited and a more extensive comparison involving other $n$-alkyl acids is desirable.

Table 2–3. Experimental 13C chemical shift tensor principal values (ppm) for undecanoic acid, phase C'.

<table>
<thead>
<tr>
<th>Position</th>
<th>$\delta_{11}$</th>
<th>$\delta_{22}$</th>
<th>$\delta_{33}$</th>
<th>$\delta_{iso}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>COOH</td>
<td>250</td>
<td>160</td>
<td>111</td>
<td>173.8</td>
</tr>
</tbody>
</table>

Figure 2–6. Differences between C–OH and C=O bond lengths in C7 to C14 $n$-alkyl monocarboxylic acids. All bond lengths are taken from x-ray single crystal data.\(^{13}\)
Figure 2–7. The agreement between experimental and computed $^{13}$C principal values for the COOH in undecanoic acid versus O–H separation distances. The cis structure with a localized hydrogen at an O–H separation of 1.70 ± 0.07 Å gives best agreement with experimental $^{13}$C tensor data. This minimum corresponds to an O–H bond length of 1.02 Å. Structures having a disordered hydrogen are rejected at > 95% statistical confidence.

It is interesting to speculate on why COOH hydrogen disorder is observed in even but not in odd numbered acids. Several studies have demonstrated that the key difference between even and odd $n$-alkyl acids lies in the packing density.$^{12, 13, 76}$ Even acids are able to achieve favorable packing at the methyl-methyl interface that occurs between bilayers. In contrast, odd acids show a void volume between methyl groups that is between 17 and 27% larger than that found in even acids.$^{13}$ At other points in the lattice, the packing efficiency for odd and even acids is similar. Prior work by Bond$^{13}$ has established that even acids are able to maintain this favorable methyl-methyl packing by translating the dimeric hydrogen bonded pairs along the $n$-alkyl direction relative to an adjacent dimer pair to relieve repulsive O···O contacts that would occur if optimal packing were enforced. This translation also allows neighboring $n$-alkyl chains to move closer to maximize the dispersion forces. Perhaps more importantly, this decrease in O···O repulsion
appears to be a crucial factor in allowing hydrogen disorder to develop. Prior theoretical work suggests that the key factor in developing low-barrier hydrogen bonds capable of proton disorder is the formation of a double-well hydrogen bond having degenerate or near-degenerate energy minima.\textsuperscript{82} Matched pKa’s for the conjugate bases of the two moieties involved in the hydrogen bond can help achieve this degeneracy, but this match appears to not be strictly required.\textsuperscript{82} In even numbered acids this disorder creates more dense solids with higher melting points due to both increased dispersion interactions in the \textit{n}-alkyl chains and stronger hydrogen bonding. In odd numbered acids, this translation process is prohibited by the dimer geometry and optimization of the unfavorable O\cdots O intermolecular interactions and energetically favorable \textit{n}-alkyl chain associations is not possible. This important difference results in an inability to create the degenerate double-well minima needed for proton disorder. This structural constraint in odd numbered acids creates localized O–H bonds that form weaker hydrogen bonds and less favorable dispersion interaction in the \textit{n}-alkyl chains. Taken together, these changes create lower melting solids for odd numbered acids.

\textbf{2.5 Conclusion}

The work described herein demonstrates that the accurate SSNMR characterization of COOH hydrogen position and the detection of dynamic disorder in lauric acid resolves an ambiguity in prior crystal structures. The refinement process includes a combination of techniques with CASTEP lattice-including DFT first relaxing the majority of atomic positions followed by manual adjustment of a single hydrogen to further alter structure near C1 where the agreement between experiment and theory remains poor. This study also invites a reevaluation of
the almost universally held assumption that all R$_2^2$(8) hydrogen-bonding interactions in n-alkyl acids are the same. The SSNMR and DFT modeling evidence supports the presence of a stronger R$_2^2$(8) hydrogen bond in lauric acid than is found in undecanoic acid. A variety of other data (e.g. x-ray bond lengths and melting points) suggest that similar differences in COOH structure and hydrogen bond strength occur in other odd and even numbered acids. This analysis thus provides a new and significant piece of information that should further improve understanding of the behavior of n-alkyl acids. It is noteworthy that several other classes of compounds (e.g. $\alpha,\omega$-alkanediols$^{83}$ $\alpha,\omega$-alkanediamines$^{83}$ $\alpha,\omega$-alkanedithiols$^{84}$ and $\alpha,\omega$-alkanedicarboxylic acids$^{85}$) have also been found to exhibit alternating density and melting point behavior. Solid-state NMR studies aimed at improving hydrogen positions and more fully elucidating the role of hydrogen disorder may also prove beneficial in understanding these materials.
2.6 References


15. For even-numbered acids, the high-temperature phase is referred to as the C form. For odd-numbered acids the high-temperature polymorph is called the C’ or C’’ phase.


CHAPTER 3: EVALUATING STRUCTURE WITH $^{1}\text{J}_{\text{CC}}$ SCALAR COUPLING

3.1 Abstract

Solution NMR studies that direct establish the skeleton of organic compounds are feasible using experiments such as INADEQUATE that measure the one-bond carbon-carbon J-coupling (i.e. $^{1}\text{J}_{\text{CC}}$). Previously, such studies have seldom been pursued due to a lack of sensitivity arising from the infrequent occurrence of $^{13}\text{C}-^{13}\text{C}$ pairs at natural abundance and the presence of heteroatoms that break connectivity in the skeleton. Here we demonstrate that sensitivity concerns can now be partially overcome by the use of software that automatically processes such data even when the signal-to-noise ratio is low. This process is demonstrated on several compounds and experimental $^{1}\text{J}_{\text{CC}}$ values obtained at the majority of positions. Similarly, recent advances in theoretical DFT predictions of $^{1}\text{J}_{\text{CC}}$ values now sufficiently accurate to allow a host of candidate structures to be considered.

3.2 Introduction

Studies of molecular structure that involve NMR have long relied upon the measurement of scalar couplings. Such couplings, also referred to as indirect couplings or J-couplings, are perhaps most widely utilized in structural studies involving the three-bond proton-proton coupling (i.e. $^{3}\text{J}_{\text{HH}}$). In this case, the magnitude of $^{3}\text{J}_{\text{HH}}$ allows $\text{H–C–C–H}$ dihedral angles to be predicted from the Karplus relationship. Other J-couplings are also sensitive to structure including $^{1}\text{J}_{\text{CH}}$, $^{2}\text{J}_{\text{CC}}$, $^{3}\text{J}_{\text{CH}}$. The less frequently measured $^{1}\text{J}_{\text{CC}}$ coupling is also sensitive to structure.$^{1-13}$ Here, $^{1}\text{J}_{\text{CC}}$ coupling is emphasized because it has the ability to provide structure in
hydrogen poor compounds and inexpensive commercial software\textsuperscript{14} is now available for automatically interpreting spectra that provide $^{1}\text{J}_{\text{CC}}$ such as the 2D INADEQUATE.

The $^{1}\text{J}_{\text{CC}}$ coupling has demonstrated a sensitivity to dihedral angle in a variety of compounds.\textsuperscript{4, 15-26} Where changes as large as 9 Hz between different conformations have been reported. Stereochemical configuration also influences $^{1}\text{J}_{\text{CC}}$ coupling,\textsuperscript{4, 7, 15, 17, 18, 22, 27-29} where, for example, the $^{1}\text{J}_{\text{CC}}$ coupling in methyl $\alpha$-D-mannopyranoside differs from the $\beta$-D-mannopyranoside by 4 Hz.\textsuperscript{4} Another important contributor to $^{1}\text{J}_{\text{CC}}$ couplings is the presence of adjacent heteroatoms\textsuperscript{8, 30-38} where the heteroatom electronegativity and lone electron pair orientation\textsuperscript{7, 17, 18, 22, 24, 26} have been correlated with the magnitude to the $^{1}\text{J}_{\text{CC}}$ coupling. Ring conformation, ring stain\textsuperscript{11, 39, 40} and percent s-character in a given C–C bond\textsuperscript{41-45} have also been experimentally and theoretically shown to influence $^{1}\text{J}_{\text{CC}}$ couplings.

In early work, $^{1}\text{J}_{\text{CC}}$ couplings were often measured by directly observing the $^{13}\text{C}$ satellite peaks in a proton decoupled 1D $^{13}\text{C}$ spectra. Synthetically $^{13}\text{C}$ enriched products were frequently employed in these studies to enhance the signal-to-noise ratio. Presently, several $^{1}\text{J}_{\text{CC}}$ values can be measured in a single experiment at natural abundance with the carbon-detected 2D INADEQUATE experiment\textsuperscript{8, 18, 20, 24, 27, 29, 31, 33, 34, 36, 37, 46-48} or the proton-detected ADEQUTE experiment.\textsuperscript{12, 15, 49} Theoretical methods from calculating $^{1}\text{J}_{\text{CC}}$ values have also been developed using various approaches.\textsuperscript{1, 10, 16, 22, 26, 50-56} Recent theoretical work often relies upon the SOPPA technique\textsuperscript{7, 10, 11, 18, 20, 22, 28, 29, 37, 54, 57} or DFT computations\textsuperscript{12, 13, 15, 17, 21, 24-26, 31, 34-36, 39, 41, 43, 52, 53, 55, 58} that employ the EPR-III\textsuperscript{59, 60} basis set.\textsuperscript{2, 24, 26, 28, 37, 51, 52} Modern DFT computations are reasonably accurate with recent B3LYP\textsuperscript{61, 62}/EPR-III $^{1}\text{J}_{\text{CC}}$ calculations showing an uncertainty of $\pm 2.92$ Hz relative to experimental data.\textsuperscript{52}
The combination of theoretical predictions and experimental measurement of $^{1}J_{CC}$ values has the potential to solve challenging structural problems, especially in hydrogen poor compounds. This is because only a theoretical structure with the correct stereochemistry, molecular conformation and heteroatom identity is expected to match experimental data. Thus a wide variety of models can be constructed to encompass all likely structural variations and $^{1}J_{CC}$ values computed for each and compared against experimental data. Bifulco et al. recently demonstrated the feasibility of this approach by constructing all possible stereoisomers of strychnine then computing $^{1}J_{CC}$ values. In this case, only the structure having the correct ring conformations and stereochemistry matched experimental data. One aim of the present study is to further evaluate the accuracy of this approach by using a model compound that contains an unusually strong intramolecular hydrogen bond. Such compounds can be a challenge for theoretical methods and verification that such materials can also be accurately modeled is desirable. Another goal of this study is to apply this approach to two natural products where the original characterization left certain structural features unresolved. Specifically, $^{1}J_{CC}$ data are employed to assign configuration about a double bond, to verify heteroatom identity, and to assign conformation at several positions. Three theoretical functionals are evaluated here with B3LYP giving the most accurate $^{1}J_{CC}$ data.

3.3 Experimental

3.3.1 Materials

Samples of $\alpha$-pinene, $\beta$-pinene, and carene were purchased from Sigma Aldrich and used without further purification.
A sample of 5-methylmellein was obtained from an endophytic fungus belonging to the *Biscogniauxia* genus based on analysis of the ribosomal DNA (described below). This fungus was isolated from a small twig of a Sand Pine (*Pinus clausa*) in central Florida using techniques described elsewhere. A culture of this *Biscogniauxia* sp. is maintained at the University of Central Florida on a refrigerated PDA slant and also on a barley seed inoculated with the fungus then cryogenically frozen in glycerol. A total of 8 liters of potato dextrose broth was inoculated with the fungus and the solution cultured without aeration for 30 days. The resulting solution was filtered through cheesecloth and the aqueous solution then extracted with 1:1 (v:v) ethyl acetate (3x). The organic phase was dried by rotary evaporation to yield 1.054 g of material. This crude extract was subjected to column chromatography on silica gel (Fisher, 40–63 µm particles) using a mobile phase of 100% CH$_2$Cl$_2$ to provide 100.6 mg of product. This material was recrystallized in hexanes (HPLC grade, Fisher) to yield 70.42 mg of a crystalline product. Liquid chromatography-mass spectroscopic (LC-MS, Agilent 6230) analysis of this material showed a single peak with a mass of 193.0866, corresponding to ([C$_{11}$H$_{12}$O$_3$]+H)$^+$. An INADEQUATE NMR analysis (see below) unambiguously established the structure as 5-methyl mullein.

Hydroheptilidic acid was isolated from an endophytic fungus isolated from a small twig from a healthy southern live oak (*Quercus virginiana*) in central Florida (28° 36.813’ N 81° 9.537’ W) using a process described elsewhere. A genetic analysis of the ribosomal DNA identified the fungus as a *Xylaria* sp. (see below). Fungal growth conditions were identical to those described above for 5-methylmellein. All cultures were filtered after 30 days then extracted using 1:1 (v:v) ethyl acetate (Fisher) and dried with rotary evaporation to yield several grams of
crude product. This material was subjected to column chromatography using silica gel and a mobile phase of 95:5 CH$_2$Cl$_2$:CH$_3$OH. A TLC analysis of the fractions (silica, 95:5 CH$_2$Cl$_2$:CH$_3$OH) indicate that 7 chemically unique fractions were obtained. Fraction 2 exhibited bioactivity and was thus subjected to further chromatography on silica gel using a mobile phase of hexane/ethyl acetate/methanol/acetic acid (40/50/8/2). The fraction retaining bioactivity was recrystallized in 100% CH$_2$Cl$_2$ to yield 1.48 g of pure material. An INADEQUATE analysis (described below) identified the final material as hydroheptelic acid.

3.3.2 Taxonomic assignment of fungi

Taxonomic assignments of all fungi were made on the basis of genetic barcoding. Genomic DNA was extracted by bead beating. Small samples of freshly grown fungal mycelia was suspended in a lysis buffer (0.0025% SDS, 25 mM Tris, and 6 mM EDTA, pH 8.0) and 1 mm zirconia/silica beads, then vortexed for 5 minutes. DNA was purified by phenol/chloroform protein precipitation followed by isopropanol/ethanol nucleic acid precipitation and stored in TE Buffer. A portion of the ribosomal RNA gene was amplified by routine PCR with Taq polymerase using the primers NS1I and LR3 with annealing temperature of 50°C. Resulting amplicons (~1.5 kbp) were purified by an in-house gel extraction method, then Sanger-sequenced bi-directionally by a commercial third party. After manual inspection, ITSx software was used to extract the complete internally transcribed spacers (ITS1 and ITS2) and enveloped 5.8S sequence. These ~500 bp sequences were used to query the UNITE database utilizing integration with International Nucleotide Sequence Database Collaboration (INSDC).
software\textsuperscript{75} was used to align the trimmed sequences with MUSCLE\textsuperscript{76} and reconstruct phylogenies with the Maximum Likelihood method\textsuperscript{77} using all gap sites in the analysis.\textsuperscript{78}

The fungi included in this study can conservatively be classified to the taxonomic level of genus based on the available genetic information. The use of trimmed ITS sequences in BLAST queries was critical to extracting useful reference data for phylogenetic analysis. The UNITE database was advantageous for taxonomic assignment by clustering reference data into 'species hypotheses' (SH) and providing a curated set of reference sequences. Resolution of phylogenetic reconstruction was improved by inclusion of gap sites in the analysis, which provide valuable taxonomic information. A summary of final taxonomic assignments and other relevant information is given in Table 3–1.

Table 3–1. Taxonomic assignment to the genus level for fungi studied herein.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Inventory #</th>
<th>Taxonomic assignment</th>
<th>Plant source</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-methylmellein</td>
<td>RS 6.2</td>
<td>Biscogniauxia sp.</td>
<td>Serenoa repens</td>
</tr>
<tr>
<td>Hydroheptelic acid</td>
<td>EW 04.01</td>
<td>Xylaria sp.</td>
<td>Quercus virginiana\textsuperscript{a}</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Plant sample deposited in the University of Central Florida herbarium.

3.3.3 Instrumentation

For the \( \text{J}_\text{CC} \) coupling measurements, experiments where performed on a Varian 500 spectrometer operating at 125.694 MHz using the INADEQUATEAD pulse sequence.\textsuperscript{79, 80} All spectra were recorded in CDCl\textsubscript{3} using a 5 mm probe, a 9.2 \( \mu \text{s} \) \(^{13}\text{C} 90^\circ \) pulse width and a 2 s recycle time. The user defined \( \text{J}_\text{CC} \) value was set to 55 Hz in all experiments. Other relevant analysis parameters are summarized in Table 3–2. All analyses were performed using natural
abundance samples. All $^1J_{CC}$ coupling constants were determined by processing the experimental
data with the NMR Analyst software$^{14}$ purchased from ScienceSoft LLC.

Table 3–2. Selected analysis parameters for all 2D INADEQUATE analysis performed in this study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Compound$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Evol. Spectral width (kHz)</td>
<td>18.4</td>
</tr>
<tr>
<td>Acq. Spectral width (kHz)</td>
<td>18.4</td>
</tr>
<tr>
<td>Number of evol. increments</td>
<td>268</td>
</tr>
<tr>
<td>Scans per increment</td>
<td>76</td>
</tr>
<tr>
<td>Acq. digital res. (Hz/point)</td>
<td>0.14</td>
</tr>
<tr>
<td>Evol. digital res. (Hz/point)</td>
<td>233.2</td>
</tr>
<tr>
<td>Experiment Time (h)</td>
<td>62.2</td>
</tr>
</tbody>
</table>

$^a$Compounds 1–5 correspond, respectively, to $\alpha$-pinene, $\beta$-pinene, carene, 5-methylmellein and hydroheptelidic acid

3.3.4 Computational details

All theoretical calculations were performed using Gaussian 09.$^{81}$ A geometry
optimization for each of the molecules was performed at the B3LYP/D95**$^{82}$ level of theory.

Reported $^1J_{CC}$ couplings for the model compounds (i.e. terpenes) were computed using the
functionals B3LYP, B3PW91$^{61,83}$, and PBE0.$^{84,85}$ In each case the EPR-III basis set was
employed. The $^1J_{CC}$ data for 5-methylmellein, hydroheptilidic acid and austrocorin and were
calculated at the B3LYP/EPR-III level of theory.
3.4 Results and discussion

3.4.1 Establishing the accuracy of theoretical methods.

Before employing $^1J_{CC}$ values to resolve actual structural questions, it is necessary to establish the accuracy of predicted values. Previous studies have estimated the error in $^1J_{CC}$ data to be less than $\pm 3$ Hz. New experimental data were acquired herein to further assess accuracy. These analyses involve rigid compounds because such structures provide $^1J_{CC}$ values that do not represent conformational averages. Three bicyclic terpenes, α-pinene, β-pinene and carene (Figure 3–1), were selected because they provide the desired rigidity and exhibit a wide range of $^1J_{CC}$ values. The 2D INADEQUATE was employed to obtain unambiguous shift assignments and $^1J_{CC}$ values at most positions in the model compounds as summarized in Table 3–3.

Figure 3–1. Structures of (A) α-pinene, (B) β-pinene, and (C) carene showing the numbering employed herein.
Table 3–3. The $^{13}\text{C}$ Chemical shift and experimental and theoretical $^{1}J_{\text{CC}}$ couplings for the model terpenes.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Carbon</th>
<th>$\delta^{13}\text{C}$ (ppm)</th>
<th>$^{1}J_{\text{CC}}$ (Hz, exp.)</th>
<th>$^{1}J_{\text{CC}}$ (Hz, theory)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B3LYP</td>
</tr>
<tr>
<td>$\alpha$-pinene</td>
<td>1</td>
<td>41.0</td>
<td>$^{1}J_{1,2} = 33.7$</td>
<td>32.9</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>31.43</td>
<td>$^{1}J_{1,6} = 28.2$</td>
<td>27.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>116.3</td>
<td>$^{1}J_{1,7} = 28.0$</td>
<td>26.9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>144.3</td>
<td>$^{1}J_{2,3} = 39.7$</td>
<td>38.9</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>47.2</td>
<td>$^{1}J_{3,4} = 71.0$</td>
<td>74.0</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>38.1</td>
<td>$^{1}J_{4,5} = 39.8$</td>
<td>39.2</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>31.6</td>
<td>$^{1}J_{4,8} = 43.7$</td>
<td>43.4</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>27.1</td>
<td>$^{1}J_{5,6} = 27.4$</td>
<td>26.4</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>26.5</td>
<td>$^{1}J_{5,7} = 27.5$</td>
<td>26.6</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>20.9</td>
<td>$^{1}J_{6,9} = 37.5$</td>
<td>36.4</td>
</tr>
<tr>
<td>$\beta$-pinene</td>
<td>1</td>
<td>40.5</td>
<td>$^{1}J_{1,2} = 33.4$</td>
<td>32.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>23.7</td>
<td>$^{1}J_{1,7} = 28.1$</td>
<td>27.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>23.6</td>
<td>$^{1}J_{1,7} = 28.1$</td>
<td>27.0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>151.1</td>
<td>$^{1}J_{3,4} = 39.1$</td>
<td>38.2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>51.9</td>
<td>$^{1}J_{4,5} = 41.1$</td>
<td>40.5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>40.6</td>
<td>$^{1}J_{4,8} = 43.7$</td>
<td>76.5</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>27.0</td>
<td>$^{1}J_{5,6} = 26.8$</td>
<td>25.5</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>21.8</td>
<td>$^{1}J_{5,7} = 27.8$</td>
<td>27.2</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>26.1</td>
<td>$^{1}J_{6,8} = 38.5$</td>
<td>37.4</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>106.1</td>
<td>$^{1}J_{6,9} = 37.2$</td>
<td>36.2</td>
</tr>
<tr>
<td>Carene</td>
<td>1</td>
<td>17.0</td>
<td>$^{1}J_{1,2} = 43.5$</td>
<td>42.8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>20.9</td>
<td>$^{1}J_{1,2} = 43.5$</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>119.6</td>
<td>$^{1}J_{1,2} = 43.5$</td>
<td>12.9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>131.2</td>
<td>$^{1}J_{2,3} = 41.8$</td>
<td>41.0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>24.9</td>
<td>$^{1}J_{3,4} = 75.2$</td>
<td>78.6</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>18.8</td>
<td>$^{1}J_{4,5} = 40.4$</td>
<td>39.8</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>16.7</td>
<td>$^{1}J_{4,8} = 44.3$</td>
<td>43.7</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>23.5</td>
<td>$^{1}J_{5,6} = 43.0$</td>
<td>43.3</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>28.4</td>
<td>$^{1}J_{5,10} = 44.7$</td>
<td>43.5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>13.2</td>
<td>$^{1}J_{7,9} = 43.9$</td>
<td>42.9</td>
</tr>
</tbody>
</table>

The $^{1}J_{\text{CC}}$ values were not measured at these sites due to overlap of resonances in the spectra.

The B3LYP, B3PW91 and PBE0 functionals were evaluated using with the EPR-III basis set. The correlation between calculated and experimental $^{1}J_{\text{CC}}$ couplings for the three terpenes is
illustrated in Figure 3–2. The B3LYP functional provides the best agreement with experimental data with a root-mean-square deviation (rmsd) of ± 1.2 Hz. The B3PW91\textsuperscript{61, 83} and PBE0\textsuperscript{84, 85} functionals give larger errors with respective rmsds of ± 3.8 Hz and ± 2.3 Hz (Table 3–4).

Statistically, the B3LYP functional can be said to differ from the B3PW91 and PBE0 functionals with high confidence (i.e. > 99.9%). Thus, for the remainder of this manuscript, all reported $^1$J\textsubscript{cc} couplings are computed at the B3LYP/EPR-III level of theory.

Table 3–4. Uncertainty in theoretical $^1$J\textsubscript{cc} values for three functionals.

<table>
<thead>
<tr>
<th>Functional\textsuperscript{a}</th>
<th>rmsd (Hz)</th>
<th>Maximum absolute deviation (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3LYP</td>
<td>± 1.2</td>
<td>3.4</td>
</tr>
<tr>
<td>B3PW91</td>
<td>± 3.8</td>
<td>4.8</td>
</tr>
<tr>
<td>PBE0</td>
<td>± 2.3</td>
<td>4.0</td>
</tr>
</tbody>
</table>

\textsuperscript{a}The EPR-III basis set was employed for all calculations
3.4.2 Selecting structure with $^{1}J_{CC}$ in the presence of strong hydrogen bonding.

The exclusive use of $^{1}J_{CC}$ to select among theoretical structural models was first successfully employed in 2013.$^{15}$ Here, the accuracy of this approach was further evaluated by a structural analysis of 5-methylmellein. The presence of a very strong low-barrier hydrogen bond in 5-methylmellein presents a challenge to theoretical methods not found in the original study. A total of eight model structures were prepared differing in OH hydrogen orientation, conformation of the methyl group on the lactone ring (i.e axial or equatorial) and the identity of the heteroatom in the lactone ring, with both oxygen and nitrogen atoms considered. Prior work has demonstrated that phenolic OH hydrogens invariably lie in the plane of the aromatic ring.$^{86}$ Thus
only two OH orientations were considered here. Figure 3–3 illustrates all model structures considered.

Figure 3–3. The 8 model structures of 5-methylmellein considered herein. Newman projections viewed along the C3–C4 bond are included for each figure to illustrate the conformations of the C9 methyl group that were evaluated.

A 2D INADEQUATE spectrum of 5-methylmellein provided \(^1J_{CC}\) values for all \(^{13}C\–^{13}C\) bonds together with unambiguous \(^{13}C\) chemical shift assignments (Table 3–5). The calculated \(^1J_{CC}\) couplings exhibit large differences between model structures of > 2.5 Hz (i.e. > 2\(\sigma\)) only at C1-C8a, C8-C8a, C7-C8, and C3-C9. These differences are summarized in Table 3–5 and are
large enough to allow for selection of a single best-fit structure. Model 2 provided the best agreement with experimental data with an rmsd of ±1.3 Hz (Table 3–6). Models 3–6 were rejected as correct structures with high statistical confidence (i.e. ≥92.2% using an F-test). Model 1 differs from model 2 only in the C9 methyl orientation and yet is rejected as a correct structure at a 78.4% confidence level (i.e. ±1.25σ). This analysis demonstrates that the B3LYP/EPR-III level of theory is sufficiently accurate to strongly differentiate between the two hydroxyl orientation and to confidently identify the ring heteroatom. This approach can also distinguish between a methyl group that is axial or equatorial relative to the lactone ring, albeit with less confidence. A comparison of model 2 with the known crystal structure verified that the correct structure was selected by this 1JCC analysis.

Table 3–5. Carbon-13 chemical shifts (ppm) and 1JCC couplings (Hz) for 5-methylmellein.

<table>
<thead>
<tr>
<th>Carbon</th>
<th>δ 13C</th>
<th>1JCC (exp.)</th>
<th>1JCC of model structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>170.1</td>
<td>1J1,8a = 67.2</td>
<td>68.1 68.3 72.3 72.4 62.8 62.9 65.9 65.4</td>
</tr>
<tr>
<td>3</td>
<td>75.3</td>
<td>1J3,4 = 35.7</td>
<td>34.3 34.7 34.0 34.6 34.7 35.0 34.6 34.9</td>
</tr>
<tr>
<td>4</td>
<td>31.6</td>
<td>1J3,9 = 40.2</td>
<td>36.7 40.0 36.8 40.1 34.7 36.9 34.9 36.7</td>
</tr>
<tr>
<td>4a</td>
<td>136.2</td>
<td>1J4,4a = 41.4</td>
<td>40.5 40.8 40.5 40.1 40.6 40.8 40.5 40.9</td>
</tr>
<tr>
<td>5</td>
<td>124.8</td>
<td>1J4a,5 = 61.1</td>
<td>63.0 62.8 61.2 61.1 62.6 62.6 60.8 60.8</td>
</tr>
<tr>
<td>6</td>
<td>137.7</td>
<td>1J4a,8a = 58.3</td>
<td>59.8 60.4 58.7 59.3 59.8 60.3 58.5 59.1</td>
</tr>
<tr>
<td>7</td>
<td>115.3</td>
<td>1J5,6 = 56.9</td>
<td>58.5 58.5 59.4 59.5 58.8 58.8 59.5 59.5</td>
</tr>
<tr>
<td>8</td>
<td>160.2</td>
<td>1J5,10 = 46.1</td>
<td>46.3 46.3 46.2 46.2 46.3 46.3 46.2 46.2</td>
</tr>
<tr>
<td>8a</td>
<td>107.8</td>
<td>1J6,7 = 58.6</td>
<td>60.4 60.4 59.9 60.0 60.4 60.4 59.9 59.9</td>
</tr>
<tr>
<td>9</td>
<td>20.7</td>
<td>1J7,8 = 67.5</td>
<td>68.5 68.6 64.5 64.6 68.3 68.3 64.3 64.3</td>
</tr>
<tr>
<td>10</td>
<td>17.8</td>
<td>1J8,8a = 65.1</td>
<td>65.0 64.9 74.9 74.9 65.1 65.1 74.8 74.8</td>
</tr>
</tbody>
</table>
Table 3–6. Agreement between experiment and calculated $^1J_{CC}$ couplings for the 8 model structures considered for 5-methylmellein.

<table>
<thead>
<tr>
<th>Model</th>
<th>rmsd (Hz)</th>
<th>Maximum absolute deviation (Hz)</th>
<th>Rejection confidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>± 1.7</td>
<td>3.5</td>
<td>78.37</td>
</tr>
<tr>
<td>2</td>
<td>± 1.3</td>
<td>1.9</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>± 3.9</td>
<td>9.8</td>
<td>99.90</td>
</tr>
<tr>
<td>4</td>
<td>± 3.8</td>
<td>9.8</td>
<td>99.87</td>
</tr>
<tr>
<td>5</td>
<td>± 2.5</td>
<td>5.5</td>
<td>97.27</td>
</tr>
<tr>
<td>6</td>
<td>± 2.1</td>
<td>3.3</td>
<td>92.24</td>
</tr>
<tr>
<td>7</td>
<td>± 3.8</td>
<td>9.7</td>
<td>99.87</td>
</tr>
<tr>
<td>8</td>
<td>± 3.6</td>
<td>9.7</td>
<td>99.81</td>
</tr>
</tbody>
</table>

Figure 3–4. Calculated versus experimental $^1J_{CC}$ couplings of the 8 model structures of 5-methylmellein (left). The plot on the right includes only the best-fitting model to more clearly illustrate agreement between experimental and theoretical data. Model structures considered differences in OH hydrogen bonding, ring heteroatom identity and lactone ring conformation.
3.4.3 Characterization of hydroheptelidic acid with $^{1}J_{CC}$ data.

Hydroheptelidic acid (Figure 3–5) represents a natural product that has been investigated by several groups since 1992 and yet retains structural uncertainty at a few sites.\textsuperscript{88-90} For example, the conformation about the C1–C2 bond (i.e. s-cis or s-trans) has never been established. Likewise, the configuration about the C2–C4 double bond was assumed to be $E$ by analogy to xylaric acid B\textsuperscript{89} but not explicitly determined from experimental measurements. Finally, conformation of the moiety bonded to C5 has not been established at C2–C3 and C4–C5.

To resolve these structural questions, experimental $^{1}J_{CC}$ values for 14 positions in hydroheptelidic acid were established from a 2D INADEQUATE (Table 3–7). Four basic model structure types were evaluated to assess the ability of $^{1}J_{CC}$ data to resolve these questions. These models differ in the configuration at the C2–C4 double bond and in conformations about the C1–C2 bond as illustrated in Figure 3–5. For each structure, further sub-variations were also explored involving conformational changes about the C2–C3 and C4–C5 bonds. Overall, a total of 26 models structures were evaluated.
Figure 3–5. The 4 basic model structures considered for hydroheptelidic acid. For each structure, multiple conformational variations were also evaluated involving the C2–C3 and C4–C5 bonds.

Table 3–7. Carbon-13 chemical shifts (ppm) and $^1$JCC couplings (Hz) for hydroheptelidic acid.

<table>
<thead>
<tr>
<th>Carbon</th>
<th>$^{13}$C</th>
<th>$^1$JCC, exp.</th>
<th>$^1$JCC of model structures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$^1$J1,2 = 68.5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>170.1</td>
<td>70.0</td>
<td>70.5</td>
</tr>
<tr>
<td>2</td>
<td>134.9</td>
<td>44.9</td>
<td>47.2</td>
</tr>
<tr>
<td>3</td>
<td>57.4</td>
<td>77.2</td>
<td>74.2</td>
</tr>
<tr>
<td>4</td>
<td>145.4</td>
<td>42.1</td>
<td>42.6</td>
</tr>
<tr>
<td>5</td>
<td>41.6</td>
<td>28.3</td>
<td>28.7</td>
</tr>
<tr>
<td>6</td>
<td>53.8</td>
<td>32.2</td>
<td>32.1</td>
</tr>
<tr>
<td>7</td>
<td>76.4</td>
<td>33.8</td>
<td>33.9</td>
</tr>
<tr>
<td>8</td>
<td>32.9</td>
<td>50.6</td>
<td>50.9</td>
</tr>
<tr>
<td>9</td>
<td>22.0</td>
<td>40.9</td>
<td>40.9</td>
</tr>
<tr>
<td>10</td>
<td>46.5</td>
<td>32.8</td>
<td>32.7</td>
</tr>
<tr>
<td>11</td>
<td>29.4</td>
<td>32.2</td>
<td>32.2</td>
</tr>
<tr>
<td>12</td>
<td>21.9</td>
<td>33.2</td>
<td>33.1</td>
</tr>
<tr>
<td>13</td>
<td>15.8</td>
<td>35.1</td>
<td>35.1</td>
</tr>
<tr>
<td>14</td>
<td>179.3</td>
<td>34.2</td>
<td>34.2</td>
</tr>
<tr>
<td>15</td>
<td>76.2</td>
<td>33.9</td>
<td>33.9</td>
</tr>
</tbody>
</table>

$^a$The $^1$JCC values was not established at this site.

The best agreement between experimental and computed $^1$JCC data for hydroheptelidic acid was obtained with model 2 with an rmsd of ±1.5 Hz. Structures 1 and 3 were rejected as
feasible candidates at $\geq 93.3$ % statistical confidence. This evaluation included numerous conformational variations about the C2–C3 and C4–C5 bonds with none found to agree with experimental data. A similar comparison for structure 4 eliminated it as a feasible structure with $\geq 86.6$ % statistical confidence. A summary of the agreement for all model structures with experimental $^1J_{CC}$ data is provided in Table 3–8.

Table 3–8. Agreement between experiment and calculated $^1J_{CC}$ couplings for the 4 model structures of hydroheptelidic acid.

<table>
<thead>
<tr>
<th>Model</th>
<th>rmsd (Hz)</th>
<th>Maximum absolute deviation (Hz)</th>
<th>Rejection confidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\pm 2.3$</td>
<td>5.2</td>
<td>93.32</td>
</tr>
<tr>
<td>2</td>
<td>$\pm 1.5$</td>
<td>2.3</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>$\pm 2.4$</td>
<td>5.3</td>
<td>94.73</td>
</tr>
<tr>
<td>4</td>
<td>$\pm 2.0$</td>
<td>4.9</td>
<td>86.64</td>
</tr>
</tbody>
</table>

For the best-fit structure (i.e. model 2), feasible conformations about the C2–C3 and C4–C5 bonds were also identified. Two conformations at C4–C5 were retained as highly probable involving the hydrogen bonded to C5 eclipsed with either C2 or the hydrogen bonded to C4. A total of three feasible conformations were found at the C2–C3 bond. An illustration of all feasible conformations is provided in Figure 3–6.
Structures containing all possible combinations of the features established herein were prepared and geometry optimized. Based on an energy calculation (MP2\textsuperscript{91-93}/6-311+G(2d,p)), only three of these conformations were found to be feasible and these are illustrated in Figure 3–7.

In this study of hydroheptilidic acid, the objective was primarily to establish structural details that had been omitted from prior work. Because the C2–C3 and C4–C5 bonds were close to the region of structural ambiguity, it was considered essential to evaluate conformation at
these sites since they were likely to influence $^{1}J_{CC}$ at C1–C2 and C2–C4 bonds. Establishing conformation at other bonds (e.g. C10–C11) was considered beyond the scope of the current study and these sites were not evaluated.

3.5 Conclusion

This study demonstrates that current DFT methods are now sufficiently accurate to allow many organic structures to be characterized solely from $^{1}J_{CC}$ data. This process assumed that the correct structure is contained within the set of candidate structures evaluated by the DFT computation and thus relies upon user input. One of the most striking features of this study is that the best-fit model for all “real compounds” evaluated had errors of $\pm 1.3–1.5$ Hz which are statistically indistinguishable from the error in the benchmark terpene dataset of $\pm 1.2$ Hz. Thus, accurate $^{1}J_{CC}$ values are consistently obtained in all molecules even though solvent effects were neglected. This accuracy is in marked contrast to chemical shifts, which can strongly depend on solvent and lattice effects. This outcome likely results from the fact that the $^{1}J_{CC}$ of lighter element is dominated by the Fermi contact term, which reflects electronic structure at the nucleus. Since solvent effects are small at the nucleus, computed $^{1}J_{CC}$ values for organic compounds are fairly independent of solvent.

Previously, acquisition of $^{1}J_{CC}$ data with the 2D INADEQUATE experiment at natural abundance has been a challenge due to low sensitivity. This problem has been somewhat alleviated by the development of inexpensive software capable of rapidly evaluating low signal-to-noise data. The development of the proton detected ADEQUATE experiment also improves
sensitivity and thus also facilitates acquisition of $^1J_{CC}$ data. It may be anticipated that the ADEQUATE experiment will have less relevance for proton poor compounds.

A final problem with structural analysis with $^1J_{CC}$ data is the presence of heteroatoms. In many INADEQUATE studies the goal is to establish the carbon-carbon bonding arrangement rather than to evaluate $^1J_{CC}$ magnitude. The presence of heteroatoms in a molecular skeleton breaks this connectivity and prevents unambiguous determination of the entire structure. In such cases one obtains a collection of molecular fragments rather than an entire structure. The approach presented here overcomes this limitation by allowing all feasible arrangements of these structural fragments to be considered, including all variations of heteroatoms, and $^1J_{CC}$ values computed for each. The results presented here suggest that only the correct structure will fit the experimental $^1J_{CC}$ values at all positions.
3.6 References


CHAPTER 4: PREDICTING ACCURATE FLUORESCENT SPECTRA FOR HIGH MOLECULAR WEIGHT POLYCYCLIC AROMATIC HYDROCARBONS USING DENSITY FUNCTIONAL THEORY


4.1 Abstract

The ability of density functional theory (DFT) methods to predict accurate fluorescence spectra for polycyclic aromatic hydrocarbons (PAHs) is explored. Two methods, PBE0 and CAM-B3LYP, are evaluated both in the gas phase and in solution. Spectra for several of the most toxic PAHs are predicted and compared to experiment, including three isomers of C_{24}H_{14} and a PAH containing heteroatoms. Unusually high-resolution experimental spectra are obtained for comparison by analyzing each PAH at 4.2 K in an n-alkane matrix. All theoretical spectra visually conform to the profiles of the experimental data but are systematically offset by a small amount. Specifically, when solvent is included the PBE0 functional overestimates peaks by 16.1 ± 6.6 nm while CAM-B3LYP underestimates the same transitions by 14.5 ± 7.6 nm. These calculated spectra can be empirically corrected to decrease the uncertainties to 6.5 ± 5.1 and 5.7 ± 5.1 nm for the PBE0 and CAM-B3LYP methods, respectively. A comparison of computed spectra in the gas phase indicates that the inclusion of n-octane shifts peaks by +11 nm on average and this change is roughly equivalent for PBE0 and CAM-B3LYP. An automated approach for comparing spectra is also described that minimizes residuals between a given theoretical spectrum and all available experimental spectra. This approach identifies the correct
spectrum in all cases and excludes approximately 80% of the incorrect spectra, demonstrating that an automated search of theoretical libraries of spectra may eventually become feasible.

4.2 Introduction

Polycyclic aromatic hydrocarbons (PAHs) are compounds composed of two or more aromatic rings containing only carbon and hydrogen. These products are found in meteors,\(^1\) interplanetary dust particles,\(^2\) and interstellar grains.\(^3\) In our terrestrial environment, PAHs often occur as products of incomplete combustion. PAHs formed by the birth of stars are ubiquitous in nature, and pose great interest in a variety of fields – from measuring the rate of star formation\(^4\) to possible uses as semiconductors.\(^5\) While the presence of PAHs in space inspires studies into the origin of biological molecules and mechanism for PAH growth,\(^6\) in earth’s ecosystem real concerns exist regarding the interaction of PAHs with biomolecules. Epidemiological studies have revealed high toxicity and carcinogenicity for many of these compounds and sixteen are now included on the Environmental Protection Agencies (EPA) list of Priority Pollutants for routine monitoring. Anthropogenic sources of PAHs include wood and coal burning, and their presence in coal and crude oil insures their extraction from the ground and potential for distribution. These factors make PAHs one of the most omnipresent pollutants worldwide. Developing sensitive methods to quantify and identify PAHs in air, water, and soil are therefore of critical importance.

While remote detection of PAHs has revolved around measured and calculated vibrational spectra,\(^7\) work in the Campiglia laboratory and by others has\(^8\) exploited the fluorescence\(^9\) and phosphorescence\(^10\) emission of PAHs to quantify and identify PAHs in
terrestrial environmental samples including water and soil. The use of luminescence provides several advantages over vibrational spectroscopy including improved sensitivity and specificity with fluorescence detection of PAHs routinely providing part-per-trillion limits of detection. Specificity in fluorescence can be further enhanced with time resolved line-narrowing spectroscopy. This process involves diluting PAH samples in solvents with minimal solute interaction (generally n-alkane solvents) then freezing to cryogenic temperatures (liquid nitrogen or helium) to produce vibrationally resolved fluorescence emission spectra with sufficient resolution to identify individual vibronic transitions. This cryogenic technique is generally referred to as Shpol’skii spectroscopy. Additional specificity can be attained with the measurement of fluorescence lifetimes.

In previous work, we have measured Shpol’skii spectra for several of the EPA PAHs and most have been found to exhibit unique spectra. These significant differences between spectra mean that complex mixtures of PAHs (e.g. environmental samples) can potentially be characterized. Moreover, since at least some unique emission frequencies are observed for most PAHs, analysis of mixtures is possible without prior chromatographic separation in analyses requiring only a few minutes. Recently, this methodology has been employed to successfully characterize multiple PAHs in complex mixtures available as “reference materials” from the National Institute of Standards and Technology (NIST). One of the challenges identified in these studies was the discovery that many of the samples contained emission peaks at frequencies not associated with any of the 16 EPA-PAH standards. This is not surprising as the EPA-PAHs are only a small fraction of the total number of PAHs known to exist in environmental samples. Further limiting the monitoring of environmental PAHs is the modest
commercial availability of certain pure standards. Many cases also exist where the cost of pure PAH standards is prohibitive. This is particularly true for dibenzopyrene isomers with molecular weight 302. Several dibenzopyrenes are more toxic than benzo[α]pyrene, which is the most toxic PAH in the EPA priority pollutants list.\textsuperscript{16} Synthesis is always an alternative to overcome the lack of commercial standards. Unfortunately, the existence of numerous isomers with the same molecular weight and very similar molecular structures often challenges the synthesis and purification of individual standards.

Very recently a new alternative to synthesis has become available for identifying PHAs when standards are unavailable. This approach relies on theoretical methods to calculate vibrationally resolved emission spectra for candidate structures.\textsuperscript{17-19} Historically, accurate descriptions of electronic excited states were difficult to obtain because conventional density functional theory (DFT), i.e Hohenberg-Kohn, was limited to ground states.\textsuperscript{20} The development of time-dependent DFT (TDDFT)\textsuperscript{21-23} extends conventional DFT to excited states and allows for studies of emission spectra. A second obstacle arose from the large number of excited vibrational states that must be evaluated in all but the smallest molecules. A key breakthrough in treating these transitions came from the recognition that the majority of the possible vibronic transitions do not, in fact, contribute to the spectrum. A set of rules has recently been developed to identify non-negligible transitions, allowing calculations to be restricted to only these levels.\textsuperscript{24, 25} This development has led to a powerful process for computing DFT emission spectra in larger molecules and a more complete description of this approach is given elsewhere.\textsuperscript{17-19} This methodology allows the influence of temperature and a wide range of solvents to be included. Perhaps most significantly, this computational methodology has now been implemented into the
widely available computational package, Gaussian, and results in a facile prediction process that non-experts can employ to effectively predict spectra in larger molecules. These techniques have the potential to provide a unique path to structure of higher molecular weight PAHs in environmental samples.

At the present time, these computational methods have been employed to predict spectra for several fluorescent molecules, but less has been done to evaluate the accuracy of these methods when computing emission spectra for PAHs. To our knowledge, computational methods have been employed to study a total of 15 PAHs and 3 of these have included solvent effects. Of the predicted spectra, 8 involve PAHs on the EPA list and most have molecular weights of 228 or less. Here, one major aim is to evaluate the ability of these DFT methods to calculate accurate emission spectra for higher molecular weight PAHs in a manner that includes solvent effects. A secondary objective is to develop methods for comparing these predicted spectra to unusually high-resolution experimental data obtained at 4.2 K with the aim of consistently identifying the correct PAHs based on statistical figures-of-merit. Compounds selected for analysis are shown in Figure 4–1 and include benzo[a]pyrene, dibenzo[a,l]pyrene, dibenzo[a,e]pyrene, dibenzo[a,i]pyrene, and (-)7R,8S,9R,10S-7,8,9,10-tetrahydroxy-7,8,9,10-tetrahydrobenzo[a]pyrene (referred to hereinafter as “benzo[a]pyrene tetrol”). Benzo[a]pyrene was selected for analysis because it is one of the most toxic compounds on the EPA list. However, dibenzo[a,l]pyrene and dibenzo[a,i]pyrene are estimated to be roughly 10 times more toxic that benzo[a]pyrene while dibenzo[a,e]pyrene is considered equivalent in toxicity to benzo[a]pyrene. Currently, however, none of these dibenzopyrenes are included on the EPA list. Since these compounds are presently recommended for monitoring in the European Union.
and may be monitored by the EPA in the future, they were also studied here. The inclusion of the three dibenzopyrene isomers is also significant because it allows for a rigorous evaluation of the DFT methods’ ability to distinguish structurally similar isomers. Likewise, benzo[a]pyrene tetrol tests the ability of the theoretical methods to predict spectra of polycyclic aromatic compound with heteroatoms (e.g. O or S) in their molecular structure. Benzo[a]pyrene-7,8,9,10-tetrol is a well-known metabolic product of benzo[a]pyrene often used as a biomarker of human exposure to PAHs.\textsuperscript{36}

In the following discussion, two theoretical methods, PBE\textsuperscript{37,38} and CAM-B3LYP\textsuperscript{39} are evaluated for their ability to accurately reproduce experimental spectra. Because each of these methods is found to exhibit systematic errors, an empirical correction is introduced. These adjusted theoretical spectra are ultimately compared to experimental data using a process that minimizes residuals. The correct experimental spectrum is found to match the predicted spectra in all cases and the majority of the incorrect spectra are eliminated as possible matches. All spectra are computed in an environment that simulates solvent and in all cases the solvent is found to shift the spectrum to higher wavelengths by 10-26 nm.
Figure 4–1. Structures studied herein include benzo[a]pyrene (top left), dibenzo[a,e]pyrene (top right), dibenzo[a,i]pyrene (middle left), dibenzo[a,l]pyrene (middle right) and Benzo[a]pyerene–7,8,9,10-tetrol (bottom).

4.3 Experimental

4.3.1 Materials

Nanopure water from a Barnstead Nanopure Infinity water system was used throughout. HPLC grade *n*-octane was purchased from Acros Organics (Atlanta, GA). Benzo[a]pyrene, dibenzo[a,e]pyrene, dibenzo[a,i]pyrene and dibenzo[a,l]pyrene were purchased from AccuStandard at their highest available purity (100%). Benzo[a]pyerene–7,8,9,10-tetrol was obtained from the NCI Chemical Carcinogen Repository, Midwest Research Institute (Kansas...
Note: use extreme caution when handling PAHs as many are known to be extremely toxic.

Stock solutions of all PAHs were prepared in \( n \)-octane and kept in the dark at 4 °C. Possible PAH degradation was monitored via room-temperature fluorescence (RTF) spectroscopy. Working solutions of PAHs were prepared by serial dilution of their stock solutions with \( n \)-octane.

4.3.2 Instrumentation

All room temperature steady-state excitation and fluorescence spectra were acquired with a commercial spectrofluorimeter (Photon Technology International). The excitation source was a continuous wave 75-W xenon lamp with broadband illumination from 200 to 2000 nm. The excitation and emission monochromator had the same reciprocal linear dispersion (4 nm mm\(^{-1}\)), accuracy (± 1 nm), reproducibility (± 2 nm) and spectral resolution (0.25 nm). Both monochromators have 100 grooves/nm grating were blazed at 300 and 400 nm, respectively. Detection was made with a photomultiplier tube (model 1527) with spectral response from 185 to 650 nm. In the photon counting mode, the maximum count rate was 4 MHz, rise time 20 ns and fall time 100 ns with a 220 ns pulse width. The instrument was computer controlled using commercial software (Felix32) specifically designed for the system. Excitation and emission spectra were corrected for wavelength dependence of excitation light source and detector sensitivity, respectively. Correction was made in the post-acquisition mode using the radiometric correction factors included in Felix32 software. Long pass filters were used when necessary to eliminate second-order emission from the excitation source. Instrumental performance was
monitored with a commercial standard (Photon Technology International) consisting of a single crystal of dysprosium-activated yttrium aluminum garnet mounted in a cuvette-sized holder with a well-characterized quasi-line excitation and emission spectrum. Wavelength accuracy and precision was evaluated periodically by comparing the recorded position of several excitation and emission spectral lines obtained from repetitive scans within 250 – 800 nm to the maximum excitation and emission wavelengths provided by the manufacturer. Room-temperature measurements were made from un-degassed solutions with a standard 600 µL quartz cuvettes with a 1 cm cell path. A 90° excitation/emission configuration was used in all measurements.

The 4.2 K measurements were carried out with the aid of a cryogenic fiber optic probe. The probe consisted of one delivery and six collection fibers. All fibers were 3m-long and 500-mm-core-diameter, silica-clad silica with polyimide buffer coating (Poly-micro Technologies, Inc.). The fibers were fed into a 1.2-m-long section of copper tubing that provided mechanical support for lowering the probe into the liquid helium. At the sample end, the fibers were arranged in a conventional six-around-one configuration with the delivery fiber in the center, bundled with vacuum epoxy (Torr-Seal, Varian Products), fed into a metal sleeve, and aligned with the entrance slit of the spectrometer. The dimensions of the vial were the following: 30-mm length, 5.5-mm inner diameter, and 7-mm outer diameter. Its maximum volume capacity was 750 mL. The measurement procedure was as follows: after transferring a known volume (typically 100-200 mL) of un-degassed sample solution with a pipette into the sample vial of the cryogenic probe, the tip of the fiber-optic bundle was positioned and held constant with the screw cap above the solution surface. Sample freezing was accomplished by lowering the copper tubing into the liquid helium, which was held in a Dewar with 60 L storage capacity. The liquid
helium would typically last three weeks of daily use, averaging 15-20 samples per day. Complete sample freezing took less than 90 s per sample. Replacing the frozen sample involved removing the sample vial from the cryogen container and melting the frozen sample with a heat gun. Because no physical contact between the tip of the fiber-optic bundle and the sample ever occurred during measurements, probe clean up between measurements was not necessary. The entire freeze, thaw, and sample replacement cycle took no longer than 5 min.

All Shpol’skii spectroscopy (LETRSS) Measurements were carried out with a multidimensional luminescence system built in our lab. Its complete description and full measuring capabilities – i.e. for absorption, excitation, fluorescence and phosphorescence measurements - have been reported previously.\textsuperscript{40} The system was operated in the external trigger mode. Data acquisition parameters (gate delay and gate width) were entered on the control computer with Andor software and the appropriate control signals were sent via a GPIB interface to the pulse generator. Once triggered by the laser, the pulse generator used this information to determine when the image intensifier in the detector head was gated on (gate delay) and for how long it was gated on (gate width). When the intensifier was gated off, the acquired data were transferred from the detector head to the controller card (32-bit Intelligent Bus-Mastering PCI card) in the computer. Complete instrument control was carried out with LabView (National Instruments, version 6.0) based software developed in our lab.

Fluorescence spectra were recorded using a minimum delay of 10 ns, which was sufficient to avoid the need to consider convolution of the laser pulse with the analytical signal. The measuring gate was optimized to collect most of PAH fluorescence and still avoid instrumental noise. Unless otherwise noticed, each spectrum corresponds to the accumulation of
100 laser pulses. The limiting resolution for recording excitation spectra was dictated by the minimum scanning rate of the tunable dye laser, namely 0.1 nm/data point. The best resolution for recording fluorescence spectra was dictated by the limiting resolution of the spectrograph/ICCD system, which corresponded to 0.32 – 0.40 nm.  

4.3.3 Computational details

For computations of theoretical spectra, all geometry optimizations and frequency calculations in both the ground state and excited state were calculated using the functionals CAM-B3LYP and PBE0 together with the cc-pVDZ basis set. Fluorescence emission spectra were calculated using the Franck-Condon approximation. The polarizable continuum model (PCM) was used to simulate solvent effects. All calculations were performed using the Gaussian 09 software package.

4.4 Results and discussion

4.4.1 Experimental fluorescence spectra at room temperature and 4.2 K.

Experimental Shpol’skii emission spectra were acquired for each of the 5 PAHs studied herein. Pure standards were diluted in n-octane to 100 ng mL\(^{-1}\) (ppb) then frozen in liquid helium (ca. 4.2 K). For comparison, emission spectra were also acquired for each PAH at room temperature (ca. 278 K) in n-octane using a conventional fluorometer (Figure 4–2). A 3/3 nm excitation/emission band-pass was needed at room temperature to obtain an acceptable signal-to-noise ratio from all the studied compounds. Lowering the temperature to 4.2 K enhanced the fluorescence intensities of the studied PAHs. The low temperature data was then recorded using
a 1/1nm excitation/emission band-pass. The room-temperature fluorescence spectra show typical vibrational structures often observed from PAHs in liquid solutions. The same is true for the room temperature fluorescence spectrum of benzo[a]pyrene tetrol. The 4.2 K spectral features of the PAHs show the quasi-line structure usually observed from Shpol’skii systems. The broad, featureless fluorescence bands we observed were due to molecules freezing out from the crystalline phase into the amorphous phase at the point of matrix solidification. The low solubility of benzo[a]pyrene tetrol in n-octane prevented us from recording quasi-line fluorescence spectra at 4.2 K.

Close examination of 4.2 K data shows characteristic fluorescence profiles for all the studied PAHs. The maximum fluorescence peaks correspond to the $S_{1,0} \rightarrow S_{0,0}$ transition and reveal significant wavelength differences for isomers of the same molecular weight. The same is true for the fluorescence maximum of benzo[a]pyrene. The spectral fingerprints of PAHs in Shpol’skii matrixes make possible their direct determination in complex samples without previous chromatographic separation.\textsuperscript{11, 12, 14}
Figure 4–2. Fluorescence spectra for the PAHs studied here. In each plot, the top spectrum (blue) is obtained in n-octane at room temperature while the bottom spectrum (black) represents the same sample after cryogenic freezing in liquid helium. All Shpol’skii spectra were recorded under site-selective excitation ($\lambda_{\text{exc}}$) at the following wavelengths: $\lambda_{\text{exc}} = 298.0$ nm (benzo[a]pyrene), $\lambda_{\text{exc}} = 306.4$ nm (dibenzo[a,e]pyrene), $\lambda_{\text{exc}} = 312.0$ nm (dibenzo[a,i]pyrene) and $\lambda_{\text{exc}} = 318.0$ nm (dibenzo[a,l]pyrene). No Shpol’skii spectrum was acquired for benzo[a]pyrene tetrol, thus only the solution spectrum in methanol is displayed.

4.4.2 Predicting fluorescence spectra with PBE0 and CAM-B3LYP.

In this study, a key question is whether theoretical methods can accurately reproduce experimental Shpol’skii spectra for PAHs of environmental importance. Here, two methods were evaluated, namely PBE0 and CAM-B3LYP. The PBE0 approach was employed initially because it has been found to give reasonably accurate emission spectra in previous studies.$^{27,44,45}$ Here,
The solvent was included using the polarized continuum model. A more detailed description of the computational process utilized is given in Experimental. It is noteworthy that Shpol’skii spectra are obtained when the solvent is present as a solid, thus one of the aims of this study is to see if a computational solvent model can be used to accurately represent the solid. Predicted spectra for the 5 model PAHs from PBE0 in n-octane are shown in Figure 4–3, together with experimental Shpol’skii data.

Figure 4–3. A comparison of experimental fluorescence spectra (black) in an n-octane matrix and calculated spectra (blue). The experimental spectrum for benzo[α]pyrene tetrall was acquired at room temperature in methanol. Theoretical spectra are calculated at the PBE0/cc-pVDZ level of theory and include n-octane or methanol as a solvent. The wavelength of the S_{1,0} → S_{0,0} transition is marked in both experimental and computed spectra.
In all cases, the computed PBE0 spectra exhibit the majority of the lines observed experimentally. The relative peak intensities are also largely consistent with experiment but there are notable differences, such as the overestimation of peak intensities for the peak corresponding to those near 460 nm in the experimental spectrum of dibenzo[a,i]pyrene. Perhaps more significantly, all predicted spectra also exhibit a systematic error involving an overestimation of the emission wavelengths by an average of 16.1 ± 6.6 nm based upon the position of the \( S_{1,0} \rightarrow S_{0,0} \) transitions. A second type of error is also present in which all PBE0 calculated spectral lines occupy an expanded wavelength range approximately 29% larger, on average, than those in the corresponding experimental spectra.

The errors observed in the PBE0 computations suggest that other theoretical methods may be more suitable for analyzing PAH emission spectra. Recently, the CAM-B3LYP method has been introduced to include long-range corrections and this method has been found to provide accurate emission spectra for some compounds. Accordingly, spectra for the 5 model compounds were calculated at the CAM-B3LYP/cc-pVDZ level of theory. All spectra are illustrated in Figure 4–4.
Figure 4–4. A comparison of experimental emission spectra (black) and calculated spectra (blue). Theoretical spectra are obtained at the CAM-B3LYP/cc-pVDZ level of theory and include n-octane or methanol as a solvent. The wavelength of the $S_{1,0} \rightarrow S_{0,0}$ transition is indicated in both experimental and computed spectra.

Spectra predicted by the CAM-B3LYP method accurately reproduce most experimental lines. The relative peak intensities mimic those observed experimentally, but there are notable regions where over/underestimations of intensities are present. These spectra also exhibit a more significant systematic error in which all emission wavelengths are underestimated by $14.5 \pm 7.6$ nm based upon the position of the $S_{1,0} \rightarrow S_{0,0}$ transitions. Notably, these CAM-B3LYP spectra also contain the second type of error noted for the PBE0 calculated spectra in which the range of
wavelengths predicted was approximately 16% larger on average than that found in experimental spectra. Both kinds of systematic errors are smaller in CAM-B3LYP spectra than in PBE0. One notable difference between CAM-B3LYP and PBE0 is that CAM-B3LYP more accurately predicts the peak intensities for peaks that occur at wavelengths higher than the $S_{1,0} \rightarrow S_{0,0}$ transitions. Taken together, these results indicate that the CAM-B3LYP functional provides more accurate spectra for the PAHs studied here.

4.4.3 Improving comparisons between predicted and experimental spectra.

The ability of the PBE0 and CAM-B3LYP methods to reproduce experimental fluorescence spectra with reasonable accuracy demonstrates that theoretical spectra can potentially be employed to evaluate experimental results. It is, however, desirable to minimize the systematic errors identified above before comparing these DFT spectra. Here, an empirical approach was explored as a way to further improve the match between experimental and computed spectra. It is observed that a plot of the wavelengths of all major experimental lines versus the corresponding wavelengths in computed spectra are linearly correlated (Figure 4–5). A least-squares fit to the data gave the relationships listed in Table 4–1.

Table 4–1. Linear correlations relating computed and experimental emission spectra.

<table>
<thead>
<tr>
<th>Method</th>
<th>Relationship</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBE0</td>
<td>$\lambda_{\text{cm}^{-1}}^{\text{calculated}} = -52.94 \lambda_{\text{nm}}^{\text{Experimental}} + 45,069$</td>
<td>0.925</td>
</tr>
<tr>
<td>CAM-B3LYP</td>
<td>$\lambda_{\text{cm}^{-1}}^{\text{calculated}} = -51.81 \lambda_{\text{nm}}^{\text{Experimental}} + 46,539$</td>
<td>0.932</td>
</tr>
</tbody>
</table>
Figure 4–5. A plot of experimental emission wavelengths for all major peaks in the 5 PAH samples studied versus the corresponding peaks in theoretical spectra. Computed spectra and are strongly correlated with experimental data with respective $R^2$ values of 0.925 and 0.932, respectively, for PBE0 and CAM-B3LYP.

These empirical relationships partially correct the over or underestimation of peak positions and also contract the spectral range over which peaks are predicted to occur to more closely match experimental data. This approach has been used extensively in comparing calculated and experimental NMR data. All computed spectra were adjusted using these relationships and the resulting spectra are shown in Figure 4–6. Because there is a scatter in the data of Figure 4–5, this adjustment is an average correction and the fit in any given individual case, while significantly improved, can be slightly high or low.
Figure 4–6. A comparison of calculated and experimental fluorescence spectra. All computed spectra have been adjusted to correct for systematic over or underestimations in wavelengths and overestimations in the range of wavelengths over which the transitions occur. Computed spectra from the PBE0 and CAM-B3LYP methods are represented, respectively, by blue and orange plots.

After the empirical correction the average difference between theory and experiment is 6.5 ± 5.1 nm and 5.7 ± 5.1 nm, respectively for PBE0 and CAM-B3LYP based upon the position of the $S_{1,0} \rightarrow S_{0,0}$ transitions.

An important consideration in this study is whether calculated spectra are sufficiently accurate to match only the correct experimental spectrum. Here an initial test was performed
comparing a given theoretical spectrum with each of the 5 experimental spectra. For each pair of spectra compared, a residual was computed (see 4.3 Experimental) and this figure-of-merit provided a means to select among the different PAHs. Because any theoretical spectrum can be slightly offset from experimental data, an adjustment is typically needed to obtain the optimal fit.

To make this adjustment process independent of an operator, and thus less biased, a computer controlled process was developed, involving a series of adjustments in which the entire theoretical spectrum was moved in increments of ±0.5 nm and a residual computed at each point. For any given compound, this process creates a group of residuals; one for each offset value. A best fit is taken as the adjustment providing a minimum in the residuals. This process also serves to identify other experimental PAHs fitting a theoretical spectrum. These adjustments are restricted to a small region of the spectrum corresponding to ±3σ (i.e. the uncertainty in the computed spectra) where σ = ±5.7 nm for CAM-B3LYP and ±6.5 nm for PBE0. The Python code used herein for these adjustments is provided in appendix B. An example of the type of output created is illustrated in Figure 4–7 for dibenzo[a,e]pyrene. Ideally, only a single experimental spectrum will match a computed spectrum, allowing unambiguous identification of the PAH. In the case of the five PAH studied here, all theoretical spectra were found to match the correct experimental data. In addition, however, four of the compounds evaluated also matched one other experimental spectrum. Only dibenzo[a,e]pyrene had match exclusively to the correct experimental spectrum and the adjusted spectrum is illustrated in Figure 4–8. A summary of the PAHs matching a given theoretical spectrum is provided in Table 4–2.
Table 4–2. Experimental spectra matching calculated DFT spectra.

<table>
<thead>
<tr>
<th>PAH</th>
<th>Possible matchesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzo[a]pyrene</td>
<td>Benzo[a]pyrene, Dibenzo[a,l]pyrene</td>
</tr>
<tr>
<td>Dibenzo[a,l]pyrene</td>
<td>Dibenzo[a,l]pyrene, Dibenzo[a,i]pyrene</td>
</tr>
<tr>
<td>Dibenzo[a,e]pyrene</td>
<td>Dibenzo[a,e]pyrene</td>
</tr>
<tr>
<td>Dibenzo[a,i]pyrene</td>
<td>Dibenzo[a,i]pyrene, Dibenzo[a,l]pyrene</td>
</tr>
<tr>
<td>Benzo[a]pyrene tetrol</td>
<td>Benzo[a]pyrene tetrol, Dibenzo[a,e]pyrene</td>
</tr>
</tbody>
</table>

aMatches are given at the 95% probability level

Figure 4–7. Residuals derived from a comparison of the theoretical spectrum of dibenzo[a,e]pyrene with experimental data from the 5 PAHs studies herein. Only the experimental spectrum of dibenzo[a,e]pyrene matches this theoretical data, demonstrating that accurate vibronic spectra can be computed using DFT methods.
Overall, the approach of minimizing residuals provided a match to the correct experimental spectra in all cases and eliminated 80% of the incorrect matches with high statistical confidence. This methodology is, admittedly, somewhat limited and was employed only to explore the feasibility of using theoretical spectra to improve analysis of experimental emission spectra. We note that significantly more selective analysis techniques are widely available\textsuperscript{54} and, in future studies, will likely improve the ability to match a given experimental spectrum.

4.4.4 The influence of solvent of calculated spectra

All calculations described here included a solvent in the model with the solvent employed being either methanol (for benzo\[a\]pyrene tetrrol) or \textit{n}-octane (for PAHs containing only C and H). Solvents should be included in fluorescence predictions because the local environment has a strong influence on fluorescence properties. In the present study, this environment involves a solvent that is present as a solid at 4.2 K. Because the solvent in Shpol’skii spectroscopy is
selected to have minimal interactions with the PAH, a solvent model that accurately described
the polarization of the neighboring regions was considered an ideal choice. Accordingly, the
polarized continuum model (PCM) was selected since it correctly describes the average
polarization of the environment without including specific solvent effects such as strongly
hydrogen-bonded moieties. In general, it is possible to include such specific interactions by
including clusters of solvent molecules within the PCM, but in the case of PAHs this was
considered unnecessary.

For the PAHs studied here, the inclusion of n-octane was found to move all peaks to
lower energies (i.e. higher wavelengths) and to influence all peaks in a spectrum equally. A plot
showing computed spectra with and without a solvent model is given in Figure 4–9. Those PAHs
containing only C and H exhibited nearly the same change with CAM-B3LYP shifting the peaks
by + 11.1 ± 2.8 nm and PBE0 altering the peaks by + 10.6 ± 3.1 nm. Because CAM-B3LYP
underestimates the wavelength of the transitions, the addition of solvent significantly improves
the fit to experimental data. In contrast, PBE0 overestimates these wavelengths, thus solvent
inclusion further deteriorates the match with experimental data. It is interesting to note that while
benzo[a]pyrene tetrol also shows a shift of the peaks to larger wavelengths upon solvent
inclusion (MeOH), the effect is roughly 2.6 time larger (i.e. + 26.2 ± 0.2 nm) than that observed
in unsubstituted PAHs. Including solvent also provides a better match between the relative
intensity of the peaks computed for benzo[a]pyrene tetrol and experimental data.
Figure 4–9. A plot showing the influence of solvent on DFT predicted spectra of PAHs. Spectra in the left and right columns show, respectively, data computed with the PBE0 and CAM-B3LYP methods. Green, orange and black spectra denote, respectively, spectra calculated in the gas phase, spectra computed with solvent included and experimental data. Most experimental spectra were acquired at 4.2 K in n-octane, but the benzo[a]pyrene tetrol spectrum was acquired at room temperature in methanol. The computed spectra include either n-octane or methanol to match experimental conditions.
Overall these data comparing gas and solvent effects indicate that the PCM model accurately represents a solvent and that including the influence of solvent becomes more important as the polarity of the PAH increases.

4.5 Conclusion

This study demonstrates that current theoretical methods for calculating fluorescent spectra of PAHs are sufficiently accurate that a given experimental spectrum can be correctly matched to a computed spectrum from a statistical comparison to a small library of computed spectra. The compounds chosen here for comparison provide a challenging test case with three isomers of \(C_{24}H_{14}\) and a PAH containing heteroatoms. In all cases, the best predictions are obtained when the CAM-B3LYP method is utilized and the influence of solvent is included.

Ultimately, these theoretical methods may prove useful in predicting spectra in cases where commercial standards are unavailable. For example, the analysis of higher molecular weight PAHs (i.e. PAHs > 300 g mol\(^{-1}\)) is presently limited by the lack of commercial standards. This is particularly problematic for PAHs with a molecular weight of 302 because such PAHs have been shown to have significant toxicity yet only 23 standards are commercially available for the 88 possible isomers.\(^{55}\) Obviously, for higher molecular weight PAHs the number of structural isomers increases rapidly and the absence of commercial standards becomes even more pronounced. For this ultimate application, however, more sophisticated methods of identification (e.g. pattern recognition) would be required – particularly in the case of PAH identification in complex mixtures.
The ability to theoretically predict accurate spectra also provides for the possibility of generating libraries of fluorescent spectra in a digital format. Presently considerable efforts have been devoted to the development of a database of IR spectra for detecting PAHs in space. The creation of a comparable database of vibrationally resolved fluorescence spectra for PAHs may be of similar value in pollutant identification.
4.6 References


A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.02*, Gaussian, Inc.: Wallingford, CT, USA, 2009.


APPENDIX A: SUPPORTING INFORMATION FOR CHAPTER 2
Assigning $^{13}$C chemical shifts

Assignments of $^{13}$C chemical shifts for lauric acids (phase C) were made by first acquiring a series of $^1$H/$^{13}$C heteronuclear correlation (HETCOR) spectra. A total of three spectra were acquired employing cross-polarization times of 100 µs, 200 µs and 300 µs. These spectra identified the $^1$H → $^{13}$C correlations shown in Figure A-1 and provided assignments for C1 – C4, C11 and C12.

![Figure A-1](image.png)

Figure A–1. The $^1$H → $^{13}$C correlations observed in a HETCOR analysis of lauric acid, phase C. The red, blue and black lines show correlations best observed at contact times of 100 µs, 200 µs and 300 µs. One-bond $^1$H → $^{13}$C correlations were also observed in all spectra but are not included here.

A second approach was employed to assign C5 and C10 that involved calculating $^{13}$C chemical shift tensors for lauric acid using the single crystal x-ray structure and comparing these data to experimental tensors. For this analysis, all atoms of lauric acids was refined using the planewave DFT code CASTEP and then NMR tensors were computed using the GIPAW method. Both computations were conducted at the PBE/ultrafine level of theory. The best fit between theory and experiment provided assignments for C5 and C10.
Carbons 6–9 were nearly degenerate in both the isotropic shifts and in the tensor principal values and thus could not be assigned by either of the methods described above. The shift assignments given in the text for C6–C9 are therefore interchangeable.

**Energy versus O–H separation**

Energies were computed for the lauric acid to evaluate the influence of disorder on hydrogen bond strength. All computations were performed using CASTEP to include lattice effects and employed the PBE functional with the ultrafine pseudopotential. Structures evaluated began with an O–H bond length of 1.00 Å and a O=C-α-Cβ dihedral angle of 180° (i.e. the *trans* conformation) and then increased the bond length in steps ranging in size from 0.04 to 0.1 Å until a bond length of 1.70 Å was achieved. This final structure corresponds to the *cis* structure with a O=C-α-Cβ dihedral angle of 0°.

Table A–1. Energy vs. O–H bond length for lauric acid phase C

<table>
<thead>
<tr>
<th>O–H separation (Å)</th>
<th>Energy$^a$ (kJ mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>4.8</td>
</tr>
<tr>
<td>1.04 (<em>trans</em>)</td>
<td>0.0</td>
</tr>
<tr>
<td>1.10</td>
<td>9.6</td>
</tr>
<tr>
<td>1.20</td>
<td>32.8</td>
</tr>
<tr>
<td>1.30</td>
<td>36.7</td>
</tr>
<tr>
<td>1.40</td>
<td>26.1</td>
</tr>
<tr>
<td>1.50</td>
<td>14.5</td>
</tr>
<tr>
<td>1.60 (<em>cis</em>)</td>
<td>9.6</td>
</tr>
<tr>
<td>1.70</td>
<td>13.5</td>
</tr>
</tbody>
</table>

$^a$All energies are given relative to the lowest energy (r = 1.04 Å), which is arbitrarily given a value of zero.
The Boltzmann populations of the two minima at O–H separations of 1.04 Å and 1.60 Å are 98.0% and 2.0%, respectively.

**Identifying best-fit O–H hydrogen positions using CASTEP**

In all structural comparisons, a butyric acid dimer model hydrogen bonded in an $R_2^2(8)$ arrangement was employed to evaluate the most favorable O–H distances. In order to verify that this model gives comparable results to computational methods that include lattice effects, a study was performed in which the crystal structure for lauric acid phase C was evaluated in an environment that includes lattice effects at the PBE/ultrafine level of theory. A series of 8 structures were prepared that were identical except that the O–H separation varied from 1.0 to 1.7 Å in steps of 0.1 Å. Each structure was optimized using CASTEP while holding the O–H distance unchanged and NMR shift tensor were computed using GIPAW (PBE/ultrafine) for the relaxed structure. The best-fit O–H separations were found at 1.1 Å and 1.4 Å. These values agree favorably with the O–H separations of 1.16 and 1.46 Å obtained using the butyric acid dimer. The NMR agreement versus O–H separation is illustrated in Figure A-2.
Figure A–2. The agreement between experimental and computed $^{13}$C principal values for the COOH in lauric acid versus O–H separation. Here, calculated $^{13}$C shift tensors were computed using GIPAW utilizing the crystal structure of lauric acid phase C to include lattice effects. The best-fit O–H distances of 1.1 Å and 1.4 Å compare favorably with the values of 1.16 Å and 1.46 Å obtained when the butyric acid dimer was used as a model structure.
APPENDIX B: SUPPORTING INFORMATION FOR CHAPTER 4
Python code for comparing theoretical and calculated spectra. All computed spectra were digitally compared to experimental spectra using a computer program that evaluates the residuals. Because the theoretical spectra displayed systematic errors, this program also allowed these spectra to be moved to larger or smaller wavelengths and new residuals to be calculated at each position. In this manner a best fit could be found as that having the lowest residuals. The Python code employed for this comparison is given below.

```
#!/usr/bin/python3

# Modules to be imported for the calculations
import csv
import numpy as np
import scipy.stats as sps
import matplotlib.pyplot as plt

# Inputs to be used in the calculation
speccalc = input('Enter calculated spectrum: ')  
expna = input('Enter experimental spectrum: ')  
ranip = input('Enter standard deviation: ')

# Importing csv of the calculated spectrum as an array
data1 = open(speccalc+'.csv','r')
calc = csv.reader(data1,dialect='excel')
calc = [[eval(row[0]),eval(row[1])] for row in calc]
calc = np.array(calc)
data1.close()

# Importing csv of the experimental spectrum as an array
data2 = open(expna+'_EXP.csv','r')
exp = csv.reader(data2,dialect='excel')
exp = [[eval(row[0]),eval(row[1])] for row in exp]
exp = np.array(exp)
data2.close()

# Creating step size in array
step = 1/(exp[1,0]-exp[0,0])
ranip = 3*eval(ranip)
rnge = (ranip*step)//1
rnge = int(rnge)
```
#Do a zero fill the start of the experimental array
frtar = []
for i in range(-60, 0):
    faa = exp[0, 0] + (i/step)
    frtar.append([faa, 0])

frtar = np.array(frtar)
exp = np.insert(exp, 0, frtar, axis=0)

#Reducing the calculated array to match the experimental array
indx = np.searchsorted(calc[:,0], exp[:,0])
baint = np.take(calc, indx, 0)

#Combine arrays and do a zero fill the end of the array
comb = np.column_stack((exp, baint[:,1]))
endar = []
for i in range(1, 60):
    eaa = comb[len(comb)-1, 0] + (i/step)
    endar.append([eaa, 0, 0])
endar = np.array(endar)
comb = np.append(comb, endar, axis=0)

#Shifting the calculated intensities to best match experimental intensities
shiftmatch = []
for i in range(-range, range):
    shfint = np.roll(comb[:,2], i, axis=0)
    combint = np.column_stack((comb[:,1], shfint))
    errsq = (combint[:,0] - combint[:,1])**2
    sig = np.sqrt(np.sum(errsq)/(len(errsq) - 1))
    shiftmatch.append([i/step, sig])

#Use F-test to print out significant values
shiftmatch = np.array(shiftmatch)
F1 = shiftmatch[:, 1]**2/min(shiftmatch[:, 1])**2
p = sps.f.sf(F1, len(errsq)-1, len(errsq)-1)
alpha = 0.05
F2 = np.where(p > alpha)
F2 = np.array(F2)
F2 = F2.flatten()
shiftWL = shiftmatch[:,0]
finalWL = shiftWL[np.array(F2)]

 pv95 = np.sqrt(1.15*(min(shiftmatch[:,1]))**2)
pv95 = np.ones(len(shiftmatch[:,1]))*pv95

#Save the rsmd array data to a csv file
specrmsd = np.column_stack((shiftmatch, pv95))
p2 = (1-p)*100
specrmsd = np.column_stack((specrmsd, p2))
Columns are wavelength shift (nm), RMSD, 95% confidence limit, Rejection confidence (%)
np.savetxt(speccalc+'_rmsd.csv',specrmsd,delimiter=',

#Plot and save rsmd vs shift with 95% confidence
plt.ion()
plt.figure()
plt.plot(shiftmatch[:,0], shiftmatch[:,1], label='RMSD')
plt.plot(shiftmatch[:,0], pv95, label='95% confidence')
plt.title('RSMD vs Shift')
plt.xlabel('Shift (nm)')
plt.ylabel('RSMD')
plt.legend(loc=1)
plt.savefig(speccalc+'_rmsd.pdf')
plt.clf()

#Calculate the residual between the experimental and calculated data
resishf = shiftmatch[np.argmin(shiftmatch[:,1],axis=0),0]
resishf = (resishf * step)//1
resishf = int(resishf)
sinr = np.roll(comb[:,2],resishf,axis=0)
resib = comb[:,1] - comb[:,2]
resia = comb[:,1] - sinr

#Plot and save residual array to csv
plt.plot(comb[:,0], resib, label='Before')
plt.plot(comb[:,0], resia, label="After")
plt.title('Residual')
plt.xlabel('wavelength (nm)')
plt.ylabel('intensity')
plt.legend(loc=1)
plt.savefig(speccalc+'_resi.pdf')
plt.clf()
specresi = np.column_stack((comb[:,0], resia))
#Columns are wavelength (nm), residue
np.savetxt(speccalc+'_resi.csv',specresi,delimiter="","")

#For debugging code (uncomment to see)
# print(exp)
# print(comb)
# print(shiftmatch)
# print(F1)
# print(p)
# print(p2)
# print(finalWL)
Chapter 2 adapted with permission from:


**DOI:** 10.1039/C6CP00416D
Chapter 4 adapted with permission from: (Page 1 of 6)

ELSEVIER LICENSE TERMS AND CONDITIONS

Jan 11, 2017

This Agreement between Jacob Powell ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number 4025910887996
License date Jan 11, 2017
Licensed Content Publisher Elsevier
Licensed Content Publication Journal of Molecular Spectroscopy
Licensed Content Title Predicting accurate fluorescent spectra for high molecular weight polycyclic aromatic hydrocarbons using density functional theory
Licensed Content Author Jacob Powell, Emily C. Heider, Andres Campiglia, James K. Harper
Licensed Content Date October 2016
Licensed Content Volume Number 328
Licensed Content Issue Number n/a
Licensed Content Pages 9
Start Page 37
End Page 45
Type of Use reuse in a thesis/dissertation
Portion full article
Format both print and electronic
Are you the author of this Elsevier article? Yes
Will you be translating? No
Order reference number
Title of your thesis/dissertation Combining Experimental Data and Accurate Theoretical Values of Nuclear Magnetic Resonance and Fluorescence Spectroscopy in the Determination of Molecular Structures
Expected completion date Apr 2017
Estimated size (number of pages) 130
Elsevier VAT number GB 494 6272 12
Requestor Location Jacob Powell
3952 Stonefield Dr.
ORLANDO, FL 32826
United States
Chapter 4 adapted with permission from: (Page 2 of 6)

Attn: Jacob Powell
Total 0.00 USD

Terms and Conditions

INTRODUCTION
1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS
2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC’s Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC’s Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC’s Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the
Chapter 4 adapted with permission from: (Page 3 of 6)

materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC’s Billing and Payment terms and conditions. These terms and conditions, together with CCC’s Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC’s Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. Translation: This permission is granted for non-exclusive world English rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. Posting licensed content on any Website: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com. All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only.
Chapter 4 adapted with permission from: (Page 4 of 6)

You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

**Preprints:**

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

**Accepted Author Manuscripts:** An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- **immediately**
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- **after the embargo period**
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

**Published journal article (PJA):** A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

**Subscription Articles:** If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect,
Chapter 4 adapted with permission from: (Page 5 of 6)

and so links will help your users to find, access, cite, and use the best available version. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others’ research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
**Gold Open Access Articles:** May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors** the following clauses are applicable in addition to the above:
Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

**Elsevier Open Access Terms and Conditions**
You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

**Terms & Conditions applicable to all Open Access articles published with Elsevier:**
Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated. The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.
If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

**Additional Terms & Conditions applicable to each Creative Commons user license:**
**CC BY:** The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by/4.0.
**CC BY NC SA:** The CC BY-NC-SA license allows users to copy, to create extracts,
Chapter 4 adapted with permission from: (Page 6 of 6)

abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at http://creativecommons.org/licenses/by-nc-sa/4.0.

**CC BY NC ND:** The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0.

Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee. Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions:**

v1.9

*Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.*