Understanding Cellulose Pyrolysis With Simulated Vibrational Spectra

A DISSERTATION PRESENTED
BY
SEAN MCGRATH
TO
THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
BACHELOR OF SCIENCE
IN THE SUBJECT OF
PHYSICS

UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS

MAY 2016

Thesis advisor: Professor Scott Auerbach Sean McGrath

Understanding Cellulose Pyrolysis With Simulated Vibrational Spectra

Abstract

Liquid fuels and plastics are generally made from environmentally unfriendly and nonrenewable sources such as petroleum. These fossil fuels are a major source of carbon dioxide emissions, and their use is the likely cause of phenomena such as acid rain and anthropogenic global warming. Biofuels, made from plants that remove carbon dioxide from the air during their growth, have the potential to replace fossil fuels as a sustainable and carbon-neutral energy source. Attempts at large-scale biofuel production have been made, primarily focusing on bioethanol made from corn and other grains. However, researchers believe that biofuel made from cellulose has the potential to be more efficient and effective.

Cellulose is a polymer that is the main component of plant matter, and it can be broken down by heat into a variety of precursor chemicals that serve as feedstocks for clean fuel synthesis. However, this pyrolytic process is currently poorly understood and hard to control precisely. In this study, we advance a new method for modeling the vibrational behavior and spectra of cellulose, in hopes of gaining a better understanding of the effects of temperature on this chemical. This method employs classical models of cellulose to simulate changes in its structure as it heats, and quantum techniques to calculate the Raman spectra associated with these high-temperature geometries. The data obtained are compared with experimentally observed spectra at the same temperatures and searched for evidence of early steps in the pyrolysis process as well as revised assignments for peaks in the vibrational spectra. Simulation data reveals both high and low-frequency deviations from experiment, indicating non-classical reactive behavior of cellulose at temperatures below the chemical's melting point. This work informs possible routes for improving cellulose pyrolysis and probes the limits of a classical model of cellulose heating.

Contents

For Katelyn, Tito, and my parents; you made this possible.

Acknowledgments

I WOULD LIKE TO THANK Scott Auerbach profusely for his tireless support of this project. His dedication has been an inspiration. Thanks also to Justin Fermann, for serving on the thesis review committee. Thanks to Mike Timko and Geoffrey Tompsett for their technical expertise and for graciously providing experimental reference data. Finally I would like to thank the members of the CRUNCH research group, for their technical and moral support.

1 Introduction

Liquid hydrocarbons are a crucial part of the modern world's infrastructure. These energy-dense chemicals make up the fuels that power our homes and vehicles, and are the basis of plastics, industrial chemicals and synthetic materials[1]. Typically, hydrocarbons are refined from non-renewable, environmentally harmful sources such as natural gas and crude oil. A practical renewable source of hydrocarbons would therefore reduce our dependence on fossil fuels and constitute a major milestone on the path to overall environmental sustainability[2].

A prime candidate for this source is cellulose, the carbohydrate polymer that makes up the majority of plant matter[3]. It is the most abundant organic compound on the planet, and it is capable of being transformed into an array of useful chemicals, including ethanol[4]. However, current conversion methods fail to be cost-effective. Chemical and enzymatic depolymerization approaches are generally slow and expensive, while faster pyrolytic methods lack precise product output and generate significant amounts of solid and gaseous waste[4]. The ideal conversion process would have the speed of pyrolysis and the precise efficiency of enzymatic processing.

To develop such a process, a greater understanding of cellulose pyrolysis is needed. In the interest of developing this understanding, we have studied the effects of increasing temperature on the vibrational spectra of cellobiose, the glucose dimer that makes up cellulose. These spectra reveal the structure and behavior of the molecule, and by investigating them both experimentally and via simulation we have improved our understanding of how heat interacts with cellulose.

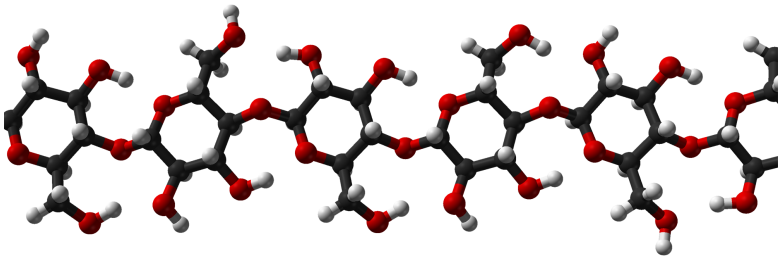


Figure 1.1: A visualization of a single cellulose strand.

1.1 Theory

1.1.1 Vibrational Spectroscopy

As its name suggests, vibrational spectroscopy is predicated on the vibrations of molecules. These vibrations affect a molecule's properties, which in turn affect how it interacts with light. We can exploit these phenomena by illuminating molecules with carefully controlled light and examining the spectrum of frequencies that are reflected. This spectrum encodes information about the underlying molecular vibrations, and thus about the nature of the molecule itself.

A mathematical description of vibration begins with the concept of *degrees of freedom*. The state of any molecule is described by that molecule's degrees of freedom, which enumerate the number of meaningful ways in which the molecule's structure may vary. A single atom has three degrees of freedom, since it may move independently in three spatial directions. A linear molecule composed of two bonded atoms has the same three degrees of freedom, as well as two rotational degrees of freedom, which are axes about which the molecule can spin. Finally, it has one *vibrational* degree of freedom, since the length of the bond may vary, bringing the total number of degrees of freedom to 6. For larger, non-linear molecules, a third axis of rotation becomes available, and the number of vibrational degrees of freedom increases. In general, a molecule with N atoms will have $3N$ degrees of freedom, with three describing overall translations of the molecule and three describing overall rotations, bringing the number of vibrational degrees of freedom to $3N - 6$. For each of these degrees there is an associated quantized vibrational energy. The extent of the vibration is determined by the energy level, and each particular vibration can absorb energy quanta energy to transition to a higher level or emit energy quanta to transition to a lower level.

As an example, vibrational degrees of freedom for water are illustrated in Figure 1.2. They include "stretch" motions, which involve changing bond lengths, and a "bend" which changes bond angle. These motions, also referred to as *vibrational modes,* represent a complete basis for describing the de-

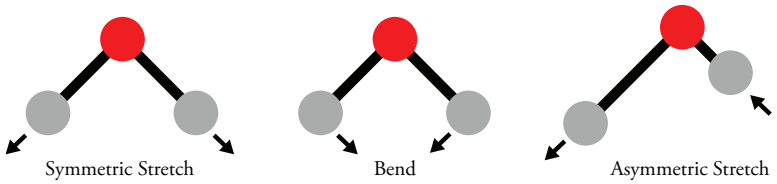


Figure 1.2: The vibrational modes of water.

formations of a water molecule. Any complicated vibration can be described as a linear combination of these motions.

As a molecule vibrates, its physical and chemical properties fluctuate as well. For example, as the water molecule undergoes the symmetric stretch depicted above, the charges held by its atoms go from being relatively concentrated to relatively separated. This amounts to a change in the molecule's dipole moment. Another property that varies in this way is the molecule's polarizability; in general, the shape of the molecule's charge cloud determines how easy the molecule is to polarize, and since vibrations deform this cloud, they change the polarizability.

Before we begin a mathematical treatment of the subject, a note on units: spectroscopists typically measure frequency in wavenumbers, the number of full wave cycles completed per unit distance. Therefore it is common to see frequency reported in inverse centimeters. with this in mind, consider some arbitrary vibrational mode of a cellobiose molecule with a characteristic frequency ν_{vib}. When this mode is in its equilibrium position, the molecule has polarizability α_0, but as the vibration proceeds, the instantaneous polarizability changes. We can write this as

$$\alpha(t) = \alpha_0 + \beta \sin(2\pi\nu_{\text{vib}} t) \tag{1.1}$$

where β is a constant describing the strength of this vibration's effect on polarizability. A similar equation can be written describing the molecule's osciallating dipole moment.

Now consider a beam of light incident on the molecule. This beam is an oscillating electric field E with a strength E_0 and a frequency ν:

$$E = E_0 \sin(2\pi\nu t) \tag{1.2}$$

The molecule is polarizable; therefore, in the prescence of this electric field it will have a time-varying induced dipole μ determined by the equation

$$\mu = \alpha E_0 \sin(2\pi\nu t) \tag{1.3}$$

This oscillating dipole will emit radiation at frequency ν. In other words, shining light of some frequency on the molecule will cause light of the same frequency to be reflected. Physically, this amounts to elastic scattering of photons off of the molecule, a phenomenon known as Rayleigh scattering.

For a static molecule this would be the end of the story. However, since the molecule is vibrating, Equation 1.3 is incomplete. We must include the time-varying polarizability:

$$\mu = [\alpha_0 + \beta \sin(2\pi\nu_{\text{vib}}t)] \cdot E_0 \sin(2\pi\nu t) \tag{1.4}$$

$$= \alpha_0 E_0 \sin(2\pi\nu t) + \beta E_0 \sin(2\pi\nu_{\text{vib}}t) \sin(2\pi\nu t) \tag{1.5}$$

$$= \alpha_0 E_0 \sin(2\pi\nu t) + \frac{1}{2}\beta E_0 [\cos(2\pi(\nu - \nu_{\text{vib}})t) - \cos(2\pi(\nu + \nu_{\text{vib}})t)] \tag{1.6}$$

The result shows that the molecule will radiate at not one but three frequencies: ν, $\nu + \nu_{\text{vib}}$, and $\nu - \nu_{\text{vib}}$. These frequencies will be observed if the polarizability of the molecule changes as it vibrates, and the intensity of the radiation will be proportional to β, the strength of the vibration's effect on the polarizability[5].

This radiation was predicted by Indian physicist C.V. Raman in 1922, and in his honor it is referred to as Raman radiation. Though we have so far framed this discussion by considering light as a wave, on a deeper level the physics are due to the quantum nature of light and energy. We observe Raman radiation because of inelastic scattering of photons, a phenomenon known as the Raman effect.

Most photons incident on an atom are reflected or scattered elastically: they interact with an atom which is in a ground state, and though the atom may increase in energy during this interaction, the atom returns to its ground state as the interaction finishes. By conservation of energy, the photon must exit with an unchanged energy (and hence frequency) as well. As mentioned above, this is Rayleigh scattering. However, in a small fraction of cases (as small as one in ten million), the photon interacts with an atom that returns to an energy state higher or lower than its original state. The energy difference is transferred to the photon, which is scattered with a shifted frequency.

This effect is illustrated graphically in Figure 1.3. When, after being excited by a photon, a system returns to an energy level above its original state, the photon is re-emitted with a lower frequency; this is referred to as Stokes Raman radiation. The opposite effect, in which the system returns to a lower energy state than it started in, results in an increased-frequency photon and is referred to as Anti-Stokes Raman radiation.

It is important to note that the Raman effect operates on so-called "virtual" energy states. This

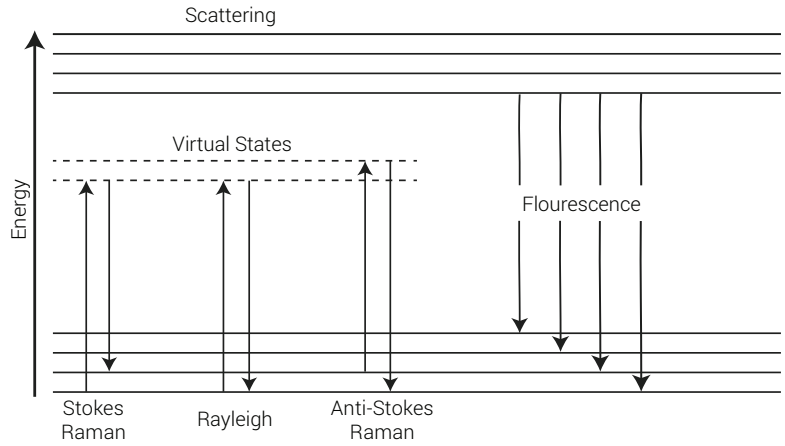


Figure 1.3: Diagram of the Raman effect. Horizontal lines represent energy levels of a system photons are incident upon, while vertical lines represent transitions between these levels. At right, the Raman effect is compared to fluorescence, which occurs when a system is excited to a high energy level and collapses to a lower one, emitting a photon in the process.

nomenclature refers to short-lived states which are not full-fledged quantum states (solutions to the system Hamiltonian). They serve simply as intermediate states during the photon absorption and scattering process, and are generally considered to be more mathematical abstractions than physical entities[5].

To sum up; a molecule vibrates with distinct vibrational modes. Some of these modes affect the molecule's polarizability —these are referred to as Raman-active modes. As a result, light incident on the molecule may be shifted in frequency by the characteristic frequency of vibrational modes. Therefore the spectrum of reflected light will contain a peak for each Raman-active mode, with a frequency that reveals the vibration's ν_{vib} and an intensity that reveals how strongly that vibration affects the polarizability. A generic example of such a spectrum is shown in figure 1.4.

The Raman effect is observed generally with visible frequencies of incident light. However, if the incident frequency is decreased into the infrared region, a different effect is observed; infrared adsorption. This phenomenon is the basis of infrared (IR) spectroscopy.

Infrared light has a frequency in the range of $40-4.000\ \text{cm}^{-1}$. Most molecular vibrational modes have a ν_{vib} in the same range. This coincidence is what makes IR so useful, since IR light is precisely the right energy to excite the vibrational modes of molecules without scattering. When an IR photon in incident on a molecule, if its frequency is the same as that of one of the molecule's vibrational modes, the photon may be absorbed and excite that mode to a higher energy state. In contrast to the Raman effect, no photon is emitted at the end of this process. The result is an absorption, rather

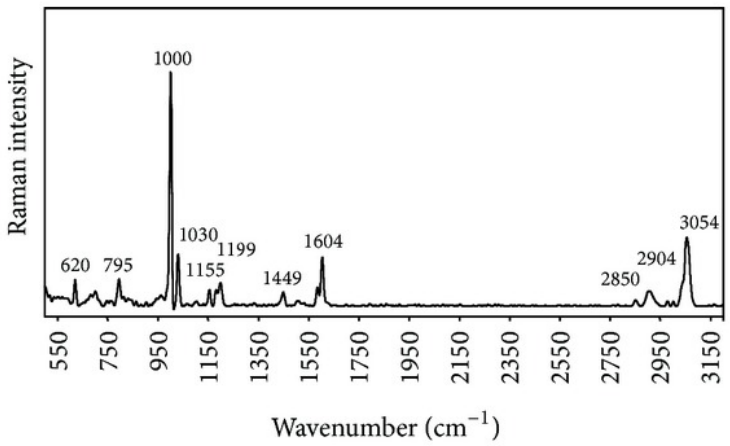


Figure 1.4: Raman spectrum of polystyrene fibers[6]. Peaks indicate the frequency and intensity of Raman-shifted light.

than scattering, spectrum; spectroscopists illuminate samples with a wide range of IR frequencies, and measure decreases in the intensity of transmitted light at the frequencies that match those of the sample's vibrational modes. An example of such a spectrum is shown in Figure 1.5

There is one important restriction on IR adsorption as described here; vibrational modes can only be excited by IR photons if that vibration affects the molecule's dipole moment. In fact, analogously to how polarizability affects the intensity of Raman radiation, the extent to which a vibration affects a molecule's dipole moment dictates how strongly it will absorb incident photons. This restriction makes IR spectroscopy a perfect complement to the Raman variety, since many modes that are IR-active are not Raman-active, and vice versa. Deploying the two techniques in tandem thus allows a full vibrational picture of a material to be developed.

1.2 Pyrolysis

Pyrolysis is defined as the thermal decomposition of materials in the absence of oxygen[7]. In other words, this process involves breaking down materials without oxidizing them, as occurs during combustion. When this technique is applied to biomass, the result is typically a carbon and energy-rich substance, such as a solid char, an oil, or volatile gases. A familiar example of a pyrolytic process is the production of charcoal, which can be done by heating wood to approximately 500 °C in a reduced-oxygen environment.

The wide spectrum of products created by biomass pyrolysis is one of its chief advantages as a method of biofuel production. Typically, pyrolyzed cellulose is converted to levoglucosan, glyco-

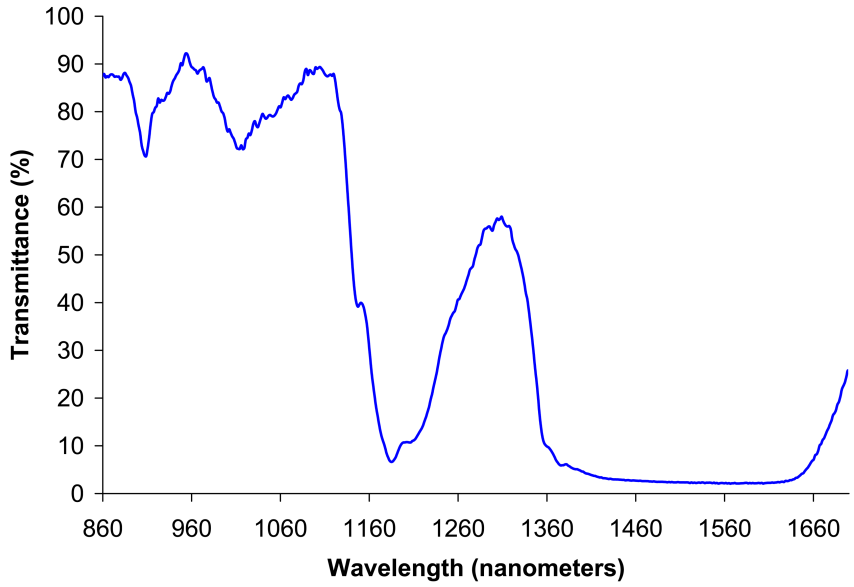
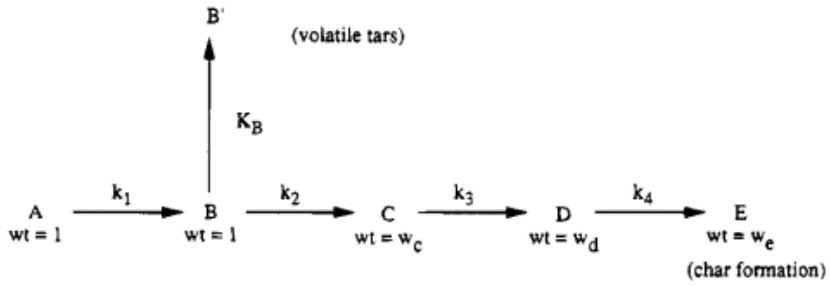
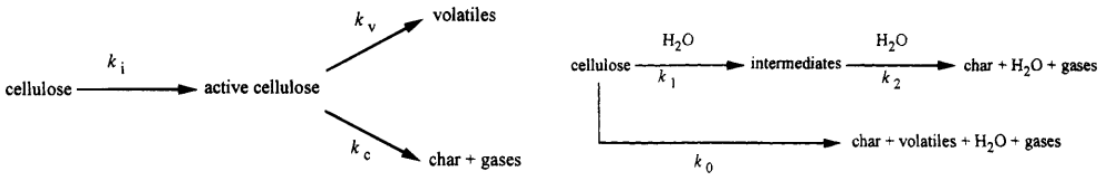


Figure 1.5: Generic IR spectrum, from Wikipedia. In contrast to Raman spectra, troughs are seen at vibrational frequencies, rather than peaks.



(a) Broido (1976) multistep model.



(b) Bradbury, Sakai, and Shafizadeh (1979) active cellulose model.

(c) Varegyi et. al. (1993) high-pressure model.

Figure 1.6: Schematics of various proposed models for cellulose pyrolysis[7].

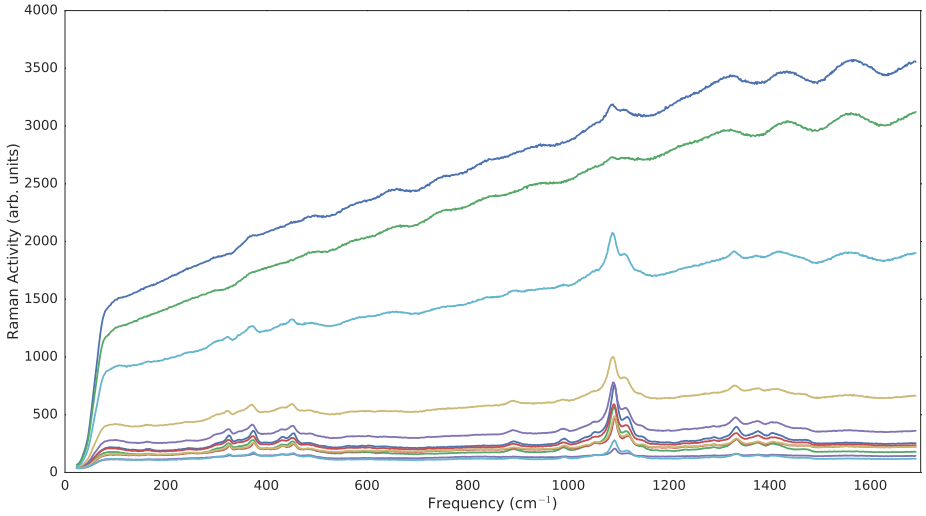


Figure 1.7: Raman spectrum of cellulose at different temperatures. Data courtesy of Michael Timko and Geoffrey Tompsett, Worcester Polytechnical Institute.

laldehyde and anhydrocellulose[7,8]. These chemicals can be broken down further to produce a range of chemicals including furan, hydroxymethylfuran, furfural, tar, and fermentable sugars. Some of these can be processed into a variety of useful products, including ethanol, bio-oil, and other industrial feedstocks. The wide range of pyrolysis products suggests that cellulose could form the basic renewable precursor for a whole host of widely used chemicals, which would make it invaluable in a post-fossil-fuel world.

The distribution of products from cellulose pyrolysis can be coarsely controlled by means of varying the duration of the reaction, the rate of heating, and the presence/concentration of additives such as inorganic salts[9]. However, these techniques are still crude due to our incomplete understanding of how pyrolysis actually occurs in cellulose. Many models of this process have been proposed over the past few decades, several of which can be seen in schematic form in Figure 1.6[10]. These models typically take the form of an initiation step followed by one or more competitive reaction pathways. In some cases, this initiation step involves a transition to a poorly defined state known as "active cellulose," from which further degradation occurs.

A primary goal of the present work is to gain insight into the onset of pyrolysis, including how it begins and at what temperature this occurs. Past research has indicated that pyrolysis is initiated at temperature ranging from 250–400 °C (523-673 K)[10]. However, recent work at Worcester Poly-

technical Institute led by Mike Timko and Geoffrey Tompsett has shown that the Raman spectrum of cellulose actually starts to change dramatically at temperatures as low as 473K, shown in Figure 1.7. This change may indicate heretofore unobserved chemical or structural changes being caused by heating. Identifying the causes of this change is another goal of this thesis. We aim to do this by developing an effective method for simulating the vibrational behavior and thus the Raman spectrum of cellulose at different temperatures. With this model in hand, we can then answer the question, "what causes the Raman spectrum of cellulose to change at 483K?" By answering this question, we will come a bit closer to answering an even more important question: "how do we make biofuels more efficiently?"

2 Methods

The simulations described in this thesis were carried out according to a new protocol for temperature-dependent calculation of vibrational spectra. This protocol consists of three phases: the sampling of molecular configurations, the simulation of molecular wavefunctions, the vibrational analysis of these wavefunctions, and the reconstruction of spectra from the analysis results. The three latter phases are well understood in the computational chemistry community[11,12,13,14,15,16], but the methods outlined here for the first phase are, to our knowledge, novel.

2.1 Sampling Molecular Configurations

Spectral simulation begins with a molecular geometry. This geometry is a set of three-dimensional coordinates that describe the type and spatial location of all the atoms in a molecule. Typically, computational chemistry assumes that the energy of the geometry is minimized—that is, that the atoms in the structure are arranged such that the potential energy stored in the molecule is as small as possible. While this simplifies calculations, it is not realistic. A molecule fixed at its minimum potential energy is at a virtual absolute zero, which is, in effect, the temperature at which most chemical calculations are carried out.

In reality, molecules are always at a nonzero temperature, and move through a large space of possible configurations. The temperature of the system to which a molecule belongs describes the size

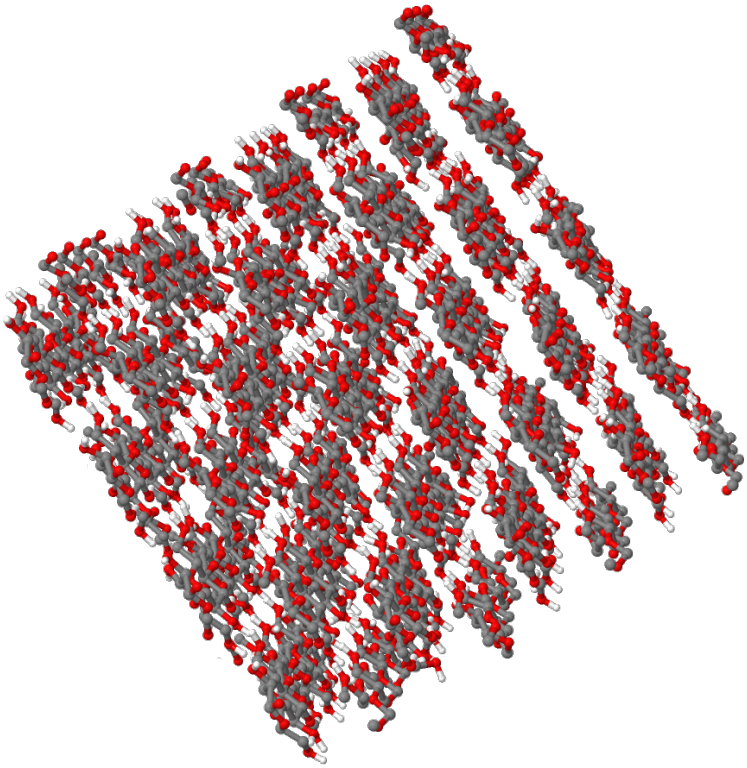


Figure 2.1: Snapshot of cellulose configuration from molecular dynamics simulation at 280K.

of this space; in general, the hotter a system is, the more possible configurations it can assume. Vibrations modulate a molecule through different geometries, and in turn those geometries constrain the ways in which a molecule can vibrate; it follows that molecules in a thermally excited state with access to high-energy geometries will vibrate differently and display different vibrational spectra than those which are frozen at absolute zero.

To address this issue, we drew molecular geometries from the results of large-scale molecular dynamics (MD) simulations of cellulose. MD is a well-understood method for simulating large molecules and molecular systems using classical physics[17]; it works by taking as input the locations and velocities of all the particles in the simulated system, then using a carefully selected potential function to calculate all the forces between the particles. These forces then determine how the particles move according to Newtonian mechanics, which in turn determine how the forces change, and so on. By iteratively evaluating the positions, velocities, and forces at successive discrete time steps separated by as little as a femtosecond, the behavior of molecules can be accurately simulated.

The particular MD approach employed here was designed and implemented by Vishal Agarwal expressly for the purpose of simulating cellulose[18]. This simulation takes in temperature as a parameter, and models the behavior of more than 3,200 atoms in 32 strands of cellulose arranged in 8 parallel sheets for a period of 10 nanoseconds. This simulation is carried out by GROMACS, an open-source software package designed expressly for this type of work[19].

The molecules in this simulation bend, stretch, and warp according to their own thermal energy and the laws of Newtonian mechanics. By taking a snapshot of this system at any given instant, we have access to a cross-section of the configuration space available to a cellulose molecule at a specific temperature. The results of this particular simulation can be examined at intervals of one picosecond, meaning that at any selected temperature, this simulation generates 10,000 points in cellulose state space.

Unfortunately, performing quantum calculations on such large structures is intractable. As a model system, we selected cellobiose units from within the cellulose superstructure, snipping them out between glycosidic oxygens. These oxygens were then capped by manually adding hydrogen atoms at standard O-H bond lengths and angles. All carbon-linked hydrogens had to be added manually as well; the MD simulation protocol used employed a united-atom model, in which hydrogen-carbon pairs were treated as a single atom. These had to be redistinguished to generate complete cellobiose units.

Taken together, these techniques allow us to sample over the realistic configurations of cellulose at any specified temperature and extract ready-to-simulate cellobiose configurations. We conducted the molecular dynamics at 14 temperatures: 280K, 290K, 300K, 310K, 320K, 350K, 400K, 423K, 450K, 473K, 483K, 493K, 500K, and 550K. We discarded the first 5 nanoseconds of each simulation to ensure that the virtual system had reached equilibrium. Then, at each of these temperatures, we randomly selected 24 timesteps from which to extract cellobiose units, resulting in 336 unique molecular geometries, each representing a possible cellulose substructure at a specific temperature.

2.2 Simulation of Wavefunctions

Each cellobiose configuration generated from the MD was input into Gaussian 09, a widely-used computational chemistry program[20]. Gaussian uses density functional theory (DFT)[21] to compute an approximate wavefunction for the molecule. DFT calculations can be conducted at various levels of theory; we used the popular B3LYP hybrid functional, as is standard in carbohydrate simulation literature[22,12].

A basis set, describing the basic functions out of which the computed wavefunctions will com-

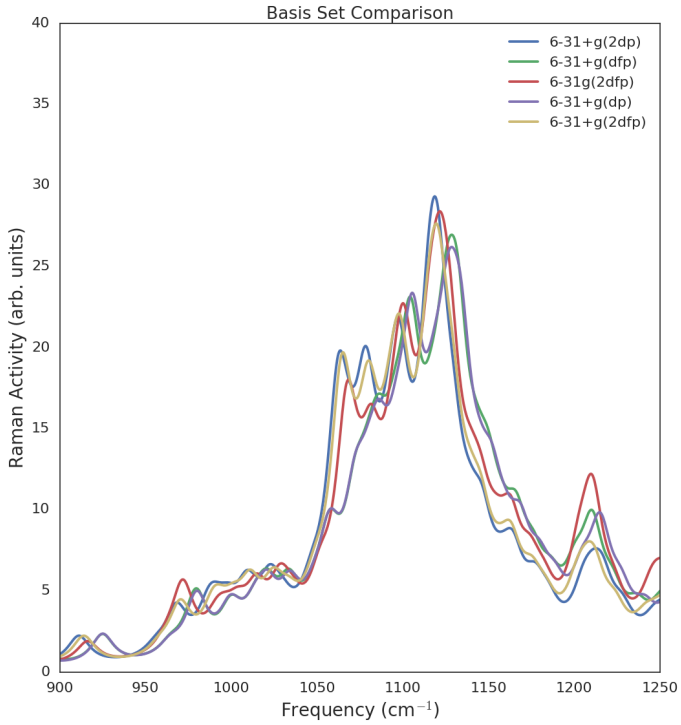


Figure 2.2: Comparison of Raman spectra generated by different candidate basis sets.

posed, is also needed. Past carbohydrate simulation literature suggests 6-31+G(d,p) or 6-31G(2d,f,p), with the former being a relatively simple basis set and the latter being relatively complex[12]. This complexity generally leads to more accurate results, at the cost of much more computationally intense calculations. For example, we found that 6-31+G(2d,f,p) Raman calculations took nearly eight times as long as 6-31+G(d,p) Raman calculations.

One of the first phases of this experiment was to determine which basis set was most efficient, delivering the most accuracy while using the least amount of computer time—a critical factor, given that 336 simulations needed to be performed, and each calculation could take over 90 hours. We assessed the performance of several different basis sets by using each of them to compute the wavefunction and Raman spectrum of the same 280K cellobiose configuration, and comparing the results to those of 6-31+G(2,d,f,p), the standard for carbohydrate simulation.

The results of the comparison are shown in Figure 2.2 and Table 2.1. These indicated that 6-31+G(2d,p) was the superior basis set for cellobiose simulation, delivering spectra nearly identi-

Table 2.1: Times taken to calculate Raman spectrum with different basis sets for an identical cellobiose geometry at 280K.

Basis Set	Computation Time (hours)
6-31+G(2d,f,p)	92
6-31+G(d,f,p)	46
6-31G(2d,f,p)	39
6-31+G(2d,p)	31
6-31+G(d,p)	13

cal to those generated by 6-31+G(2d,f,p) while consuming 60% less computer time. Therefore 6-31+G(2d,p) was used as the basis set for the bulk of the simulations.

2.3 Vibrational Analysis

Raman spectra were calculated directly from the simulation results by Gaussian. For this purpose, Gaussian employs a technique known as normal mode analysis (NMA). NMA depends on the fact that a molecule's vibrational modes are, mathematically speaking, normal to each other. In other words, they are independent, and each one has characteristics such as frequency that are independently calculable.

NMA assumes that the atoms in a molecule behave as a network of coupled three-dimensional harmonic oscillators, in which the movement of each atom (labeled with an index i) is affected by the position and movement of all the other atoms. To frame this mathematically, let each atom have instantaneous position coordinates x_i, y_i, z_i centered about an equilibrium position denoted $x_{i,\mathrm{eq}}, y_{i,\mathrm{eq}}, z_{i,\mathrm{eq}}$. As harmonic oscillators, these atoms experience a force proportional to their displacements from their respective equilibrium positions, governed by a proportionality constant k. In addition, they vibrate with a characteristic frequency ν given by

$$4\pi\nu^2 mx = kx \tag{2.1}$$

where m is the atom's mass. If the atoms were considered to be simple and independent harmonic oscillators, as in the Einstein model of solids[23], k would be simply a constant. However, in NMA, coupling dictates that k is determined by the second derivative of the system's potential energy V with respect to position. V is a function of the overall configuration of the molecule; symbolically, $V = V(x_1, y_1, z_1 \ldots x_N, y_N, z_N)$. Therefore, to calculate all the force constants for all three directions of

the N atoms, we must take $3N$ derivatives of a function of $3N$ variables.

The result is a $3N \times 3N$ matrix of force constants describing the force each atom in the molecule feels, given an overall molecular configuration. This matrix is referred to as the Hessian, and it is of critical importance to a number of processes in computational chemistry, including energy minimization. If we denote the Hessian as H and apply standard mass-weighting techniques, we can write the matrix representation of the molecular configuration as X and use Equation 2.1 to write

$$HX = 4\pi\nu X \tag{2.2}$$

This is an eigenvalue equation. The eigenvalues of the Hessian are the neqative of the squared normal mode frequencies, and the eigenvectors are the mass-weighted normal coordinate displacements. In other words, solving this equation reveals how each normal mode vibrates, as well as the frequency at which it does so. With the eigenvalues and vectors in hand, Gaussian then proceeds to calculate how each vibrational mode affects the molecule's dipole moment and polarizability, producing IR and Raman intensities for each mode as well.

The reason this calculation must follow a wavefunction simulation phase is that the potential energy V is determined by the wavefunction. The wavefunction determines the location and density of the electron cloud, which in turn determines how the atoms in the molecule interact. Therefore it is the combination of DFT and NMA that allows vibrational spectra to be accurately computed.

2.4 Spectral Reconstruction

It is important to note that normal mode analysis does not produce an actual visual spectrum. For each of an input molecule's vibrational modes, NMA generates a frequency, an IR intensity, a Raman activity level, and a set of vectors describing the vibration of itself in terms of atomic displacements. However, there is an accepted method for constructing a continuous spectrum based on this information. In essence, each vibrational mode is assumed to produce a peak (or trough, in the case of IR) in the spectrum with a Lorentzian profile, and the sum of these will reproduce an approximate spectrum [24].

A Lorentzian is a single-variable, three-parameter function similar to a Gaussian. It is described by the equation

$$L(x) = \mathcal{A}\frac{\frac{1}{2}\Gamma}{(x - x_0)^2 + (\frac{1}{2}\Gamma)^2} \tag{2.3}$$

where A is the amplitude, Γ describes the width of the peak, and x_0 is the peak center. This func-

tion generates peaks with a distinctively pointed shape and long wings, illustrated in Figure 2.3. For a mode in the Raman spectrum, for example, peaks are recreated as Lorentzians centered at the mode frequency with an amplitude equal to the Raman activity. The only free parameter is the peak width; typically, this is chosen by investigators to best match the width of experimentally observed peaks. Typically this value is set between 8 to 10 cm^{-1} [11].

A peak reconstruction proceeds in this fashion: let the vectors F and I represent the normal mode frequencies and their associated peak heights, and let F_i and I_i be the particular frequency and intensity for mode i. Assuming fixed peak width Γ_0, the peak L_i is given as

$$L_i(x) = I_i \frac{\frac{1}{2}\Gamma_0}{(x - F_i)^2 + (\frac{1}{2}\Gamma_0)^2} \tag{2.4}$$

Repeating this procedure for all i (that is, all the vibrational modes), we can then derive a function S that gives a continuous spectrum by summing:

$$S(x) = \sum_i I_i \frac{\frac{1}{2}\Gamma_0}{(x - F_i)^2 + (\frac{1}{2}\Gamma_0)^2} \tag{2.5}$$

This function $S(x)$ represents the reconstructed spectrum. It takes a wavenumber in cm^{-1} as input and returns the predicted intensity. In cases where frequencies are close together, the peak amplitudes will sum, and large super-peaks will form, as they do in real spectra.

This method has been proven effective for single sets of NMA output, but for the present work, we have hundreds of such outputs with which grapple. In the interest of isolating the effect of temperature on spectral features, we chose to handle this by averaging over the spectra computed at each temperature, yielding a set of second-order spectral functions that integrate all available information about spectral behavior at a given temperature.

Let Ω_T be the set of all computed spectral functions $S(x)$ at temperature T and $\Omega_{T,i}$ be the ith element in the set. We can then construct the temperature's average spectrum $S_T(x)$ as:

$$S_T(x) = \frac{1}{|\Omega_T|} \sum_i \Omega_{T,t}(x) \tag{2.6}$$

This procedure can be carried out for either Raman intensities or IR intensities. The resulting S_T functions represent our model's prediction of the vibrational spectra of cellobiose at a temperature T.

In a similar fashion, the average vibrational displacement vectors can also be calculated. Let the

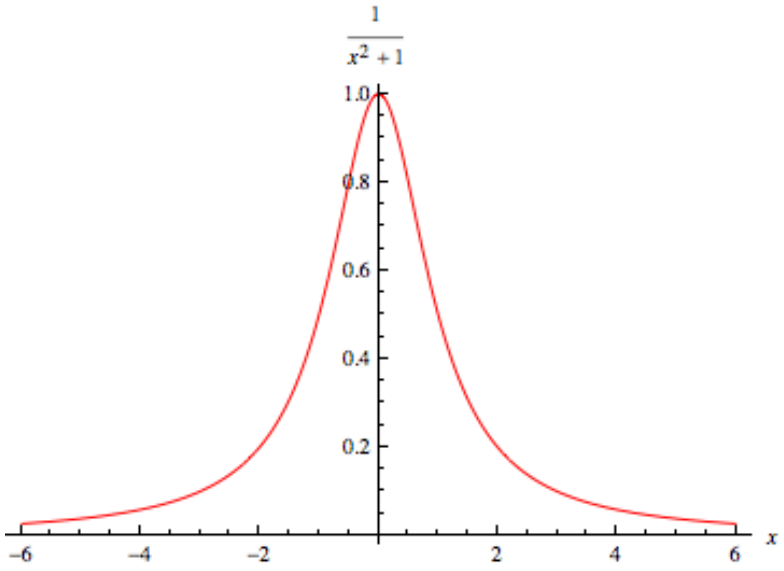


Figure 2.3: Normalized Lorentzian function with center 0, width and amplitude 1.

total number of vibrational modes be $\mathcal{M}$. At a temperature T, the ith vibrational mode of the jth simulation result contains a set of 3D eigenvectors $V_{j,i}$ numbered 1 through N, the number of atoms in the simulated molecule. The eigenvector with index k describes the direction and magnitude of the motion kth atom during the mode's vibration. We wish to compute $\langle V_i \rangle_T$, the average set of atomic displacement vectors for mode i at temperature T, taken over all the simulation results at T. Each entry in this set is given by the following formula:

$$\langle V_{i,k} \rangle_T = \frac{\sum_j^{\mathcal{M}} V_{i,k}}{\mathcal{M}} \tag{2.7}$$

Calculating this quantity for each atom k for each mode i, we can recompose the resulting vectors into a complete $\langle V \rangle_T$, which describes the average displacement vector for each vibrational mode at T.

2.5 Summary

To reiterate, the simulation protocol used for this work proceeds as follows:

1. Molecular dynamics is used to generate cellulose configurations at a given temperature.
2. Cellobiose structures are extracted from the cellulose structure and capped with hydrogens.

3. Density functional theory is used to calculate approximate wavefunctions for the cellobiose configurations.

4. Normal mode analysis is used to compute Raman spectra from the simulated wavefunctions.

5. Averaging techniques are used to compute representative spectra and vibrational modes for each simulated temperature.

2.6 Comparing to Experiment

The spectral data produced by the protocol described above interesting in isolation, but even more so when compared to the equivalent experimental spectra. There are two complicating factors that must be addressed when carrying out such comparisons: slight deviations in predicted mode frequencies, and the arbitrary nature of Raman activity units. The first point refers to the widely observed fact that DFT/NMA calculations of mode frequencies tend to uniformly under or overshoot experimentally observed frequencies by one to two percent. To aid comparison, the standard technique is to multiply each predicted frequency by a constant scaling factor, chosen to align some distinctive spectral feature. In this work, an empirically determined scaling factor of 1.017 was chosen to align the main low-frequency peak at approximately 1100 cm^{-1}.

The second point refers to the fact that the absolute value of a peak's height in a Raman spectrum has no intrinsic meaning. Different calculational and experimental techniques will yield different numbers and use different units to report the Raman activity, but all of these are arbitrary. What matters from the standpoint of chemical insight is the relative height of each peak in the spectrum. To make this feature clear, we have adopted a simple spectral normalization procedure: we set the height of the tallest peak in the spectrum to be one, and take the lowest point in the spectrum to be zero. We then report the height at every other point as a unitless fraction of the maximum peak height. This way we can accurately compare any number of different spectral datasets, regardless of their native units or scales.

A final note is that certain experimental trials produce a spectra that sits on an elevated linear baseline, due to wide-spectrum polychromatic fluorescence (an effect not believed to be relevant to the pyrolysis process). No such baseline is produced in the simulation. Therefore, for these elevated trials a baseline-corrected spectrum is produced by subtracting the baseline value from each point in the spectrum.

The methods described above do have several limitations. Perhaps the most important is that, in this paradigm, the effects of inter-chain hydrogen bonding between strands of cellulose are ignored in the vibrational analysis stage. By sampling thermal configurations, we take into account

the effects of hydrogen bonding on the starting molecular geometry, but certain vibrational modes that may be suppressed by that bonding are calculated to be active with this method, and vice versa. Also, the manual addition of carbon-linked hydrogens makes analysis of modes predominantly involving motion of these atoms unviable, as their positions do not necessarily reflect actual configurations seen at any real temperature.

3
Results

The simulations described in the previous chapter proved to perform significantly better than conventional carbohydrate simulation protocols. A typical off-the-shelf Raman calculation scheme (B3LYP, 6-31+G(dp)), applied to one of the cellobiose configurations sampled at 300 K, produced the spectrum shown in Figure 3.1. However, the application of peak scaling and averaging techniques generated a marked improvement, as shown in Figure 3.2. Applying our new protocol to temperatures approaching 483 K, we produced simulated Raman spectra shown in Figure 3.3, each compared to the corresponding experimental data.

At each temperature, average atomic displacement vectors for each vibrational mode were calculated. From these vectors, overall molecular motions can be reconstructed, and the mode's vibrational type can be assigned. A selection of these assignments for prominent peaks are presented in Table 3.1.

Finally, the effect of changing temperature on the simulated frequency of each vibrational mode was ascertained by averaging over the frequencies predicted for each mode over all the configurations generated at a particular temperature. The results are shown in Table 3. The frequency change for the main peak is shown graphically in Figure 3.4, with a comparison to the behavior of the same peak as observed experimentally. The same comparison is made for the peak at approximately 1330 cm^{-1} in Figure 3.5.

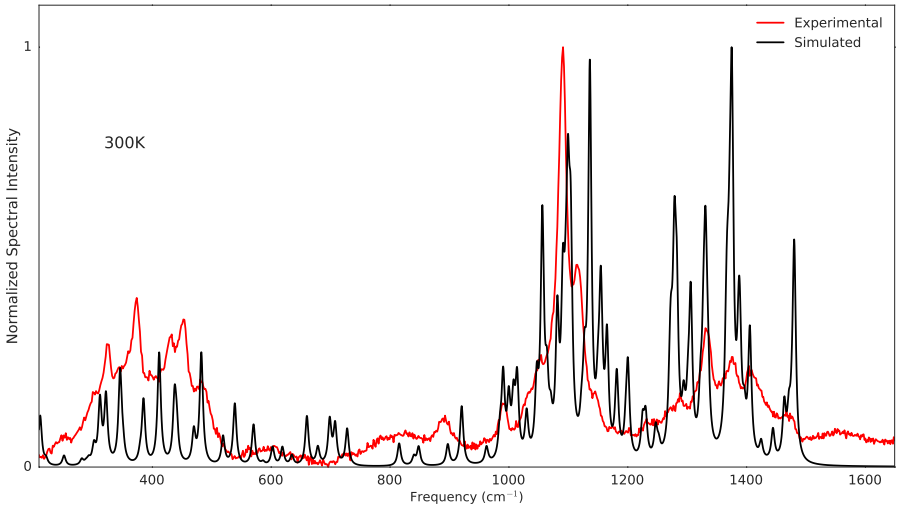


Figure 3.1: Raman spectrum calculated for a single cellobiose unit at 300K using standard protocols.

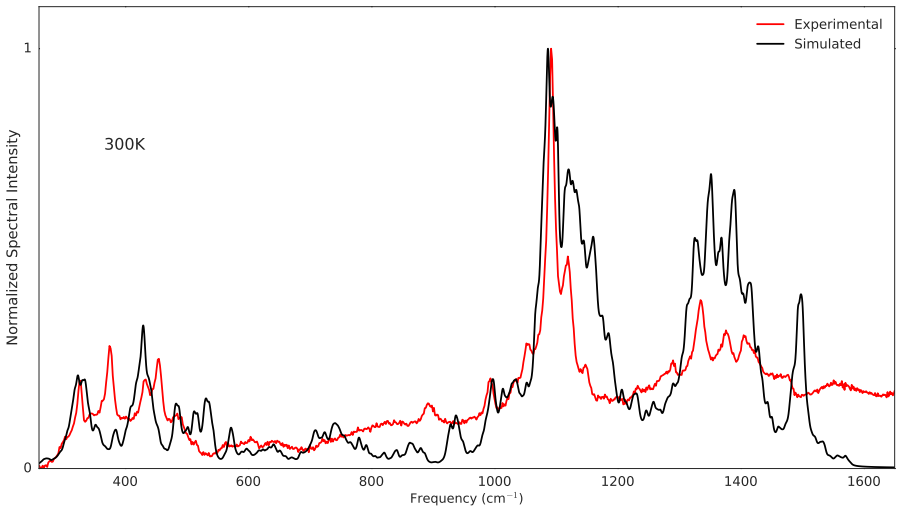


Figure 3.2: Average Raman spectrum calculated at 300K using new simulation protocol.

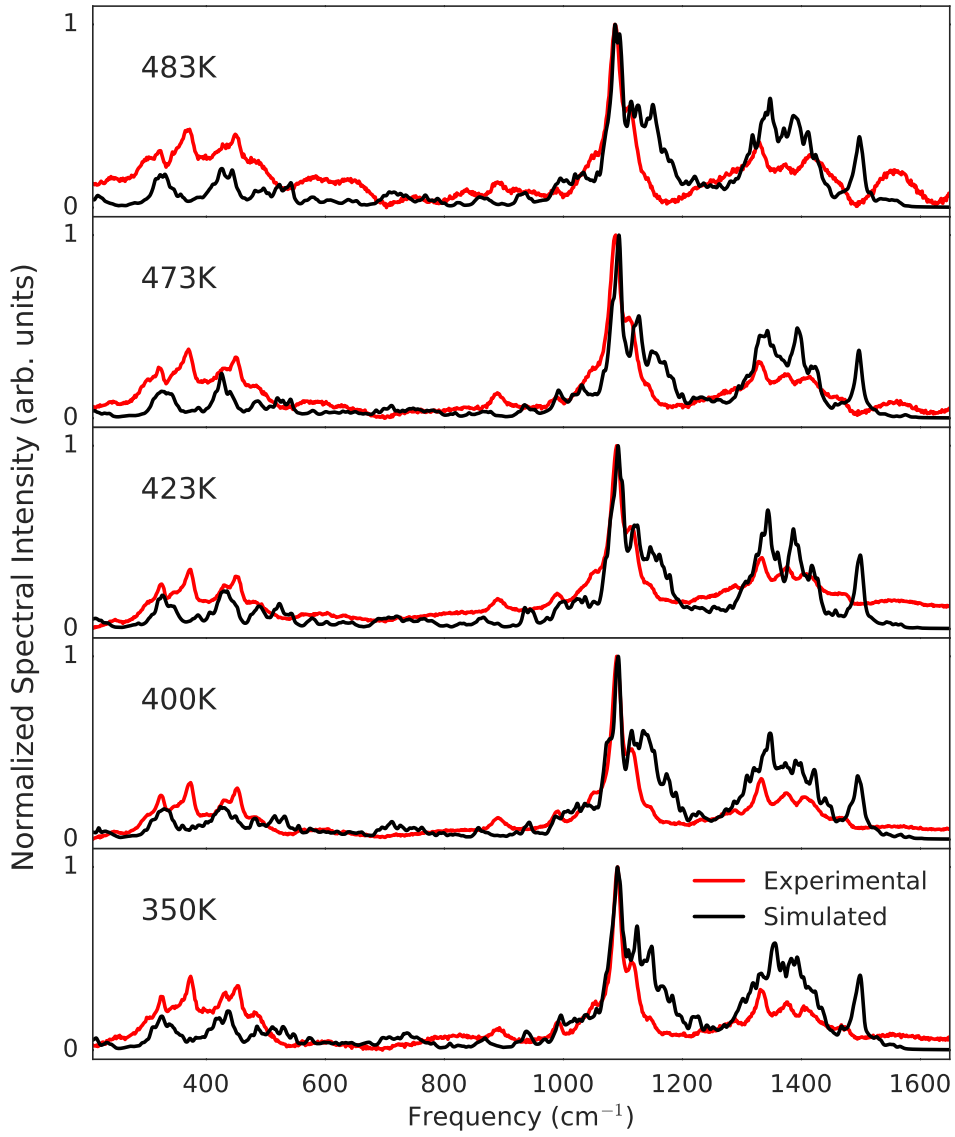


Figure 3.3: Experimental and simulated spectra at a variety of temperatures approaching the divergence temperature of $483\,\mathrm{K}$. Note the emergence of new broad peaks in the experimental data at $483\,\mathrm{K}$.

Table 3.1: Selection of peak assignments generated from simulations.

Peak Wavenumber (cm^{-1})	Vibration Type
432	Ring expansion/contraction
490	Rocking of glycosidic linkage
586	Generalized ring bending/stretching
635	Multiple CO stretches, some bending
666–880	OH wagging
923	one-sided stretch of glycosidic linkage
955	OCO asymmetric stretching
1005	Linkage symmetric stretching
1010–1200	In-plane CC & CO stretching
1200–1500	COH & CCH bending
3230+	OH & CH stretching

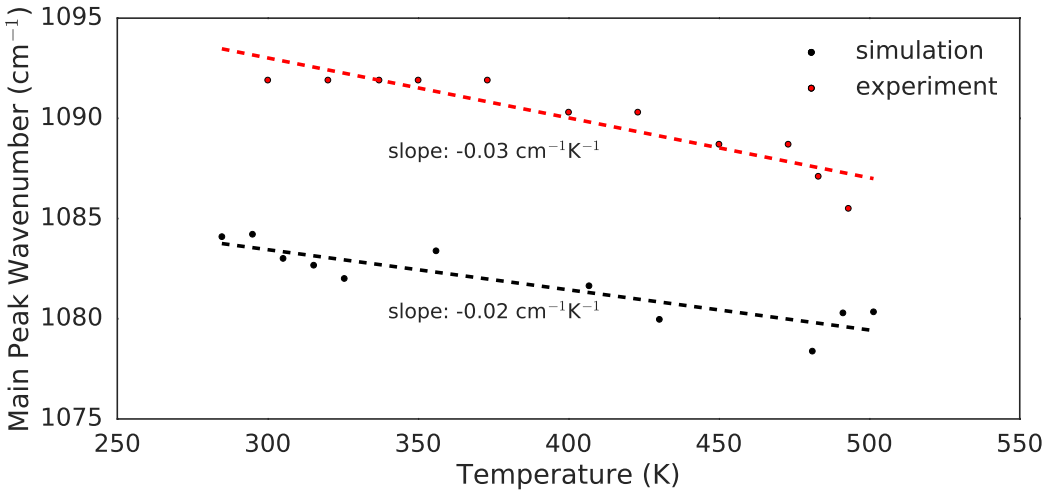


Figure 3.4: Effect of temperature on frequency of main Raman peak in simulation and experiment.

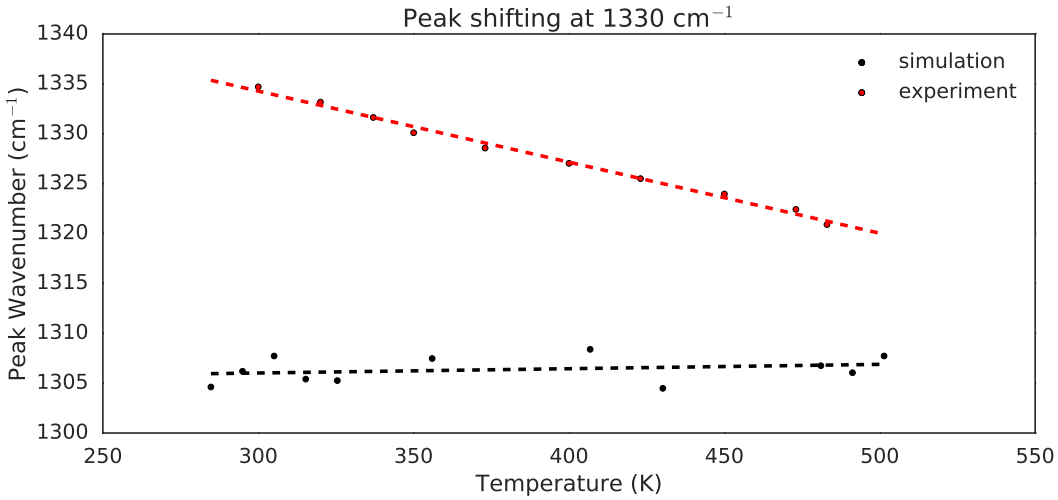


Figure 3.5: Effect of temperature on frequency of peak centered about 1330 cm^{-1} in bending region in both experiment and simulation.

Table 3.2: Simulated temperature-dependent frequencies for each vibrational mode of cellulose. Includes the estimated slope of any linear relationship between temperature and frequency for each mode, as well as the value of the linear correlation between the two quantities.

Mode #	Wavenumber (cm^{-1}) at Temperature													
	280K	290K	300K	310K	320K	350K	400K	423K	473K	483K	493K	500K	Slope ($cm^{-1}\,K^{-1}$)	Absolute Linear Correlation
27	378.6989	379.1851	378.564	376.84083	379.4923	379.7847	381.772	377.2102	378.565	382.4551	378.259	376.2846	0.000429164 2926	0.020315 93813
28	404.936	405.0264	401.6581	399.4409	403.9322	403.2195	403.8744	400.6363	406.7628	402.8635	401.9358	405.9971	0.004704774544	0.1864847032
29	419.0925	418.1785	417.5398	415.5354	418.2121	418.0188	420.3491	419.2244	418.6316	417.9009	416.249	418.1569	0.000416192 6027	0.028154 36945
30	431.5967	430.9698	429.7929	427.7879	431.4348	431.5415	434.1135	430.215	430.2879	433.0093	430.237	433.7397	0.007535508395	0.3656377588
31	442.3573	440.5314	440.1383	440.9496	441.037	444.1456	446.4227	439.3863	439.5433	443.699	441.7239	445.0916	0.006922325514	0.2624017175
32	476.2556	475.6157	474.7457	470.4785	475.0401	477.1622	475.5777	476.1623	476.3716	477.9247	475.9901	473.2233	0.005520855909	0.2452756271
33	490.3402	490.1467	488.2039	486.0022	487.9884	491.9681	492.7459	485.6457	489.8229	490.6783	488.0275	489.6352	0.001675435127	0.066742 10956
34	511.0226	510.3294	507.762	508.5399	508.4107	512.2176	513.9925	507.973	508.8566	512.5564	510.5136	525.4708	0.026152 59881	0.4676096736
35	528.3403	526.8678	527.1704	524.1088	528.7616	533.2863	532.9101	526.9793	528.9913	530.0583	529.8175	542.1466	0.028480 88663	0.5342438983
36	566.2911	565.1127	564.2263	560.9592	563.9382	566.1628	565.4823	564.1527	566.6569	565.2544	564.7332	558.9	-0.004039993597	0.1539861136
37	585.6674	582.3022	582.3461	578.1857	584.2863	584.327	583.0666	582.9446	584.0581	581.9997	583.6345	580.3515	-0.002052005536	0.0892314 3469
38	598.8648	596.0347	596.0097	592.6444	597.2039	597.9472	599.6223	595.9233	596.9261	596.9442	596.5478	593.3302	-0.002711990258	0.1172388519
39	616.1439	615.4952	614.5566	611.2713	616.3306	617.0892	618.2454	614.3097	614.6602	615.1302	618.0048	610.7599	-0.002036846094	0.075665 64978
40	635.2613	630.6773	632.8387	627.652	631.1825	633.6543	635.8533	630.4148	629.5863	632.6532	632.9108	629.0988	-0.004284990497	0.1491966862
41	666.674	670.6042	668.8565	654.6261	662.114	664.7196	663.5405	674.665	666.5249	670.4447	664.5723	645.5942	-0.016096848 04	0.1789938147
42	688.0608	692.2068	689.9225	676.9156	682.455	686.8255	686.8984	690.7561	686.0996	694.9723	683.7822	662.3555	-0.025909099 88	0.2607339398
43	707.5449	709.9102	708.8485	701.2059	699.5396	704.908	703.525	710.5279	709.0461	710.5756	696.2629	677.1041	-0.042256502 49	0.3856488229
44	729.2335	726.3548	723.3977	718.0611	722.118	721.4759	718.7637	728.6333	727.8381	727.0501	716.3916	694.2665	-0.043365732 65	0.3943581339
45	748.8869	746.8289	751.9238	738.5096	749.345	737.5105	736.867	750.6538	755.1583	750.6985	747.818	716.2522	-0.026244 20344	0.2134 01908
46	792.9472	784.4052	787.6062	786.8856	786.4237	778.3762	773.3302	796.6104	785.9247	793.7008	787.5082	744.5595	-0.050787 45355	0.3201729 16
47	835.6727	830.2006	825.5621	822.3207	825.752	834.3713	823.8221	833.5088	822.1924	833.5548	825.1462	798.1973	-0.048746 39518	0.4212908783
48	854.5376	853.9364	853.2124	844.2987	848.6254	857.3562	849.5914	851.6477	851.8267	852.2739	848.1797	838.1896	-0.024 06285423	0.4 0667 98099
49	883.864	877.1797	872.1914	875.5396	869.3102	880.2455	871.8635	868.4026	868.6716	876.4403	871.906	857.9071	-0.045737 96638	0.5924848436
50	923.6555	925.6522	925.3752	923.2177	924.4279	927.469	921.199	919.2716	922.1451	920.2109	918.0012	922.3835	-0.023290 32326	0.7226 469635
51	955.2175	956.5129	953.4362	951.7546	951.0351	954.6315	950.6697	946.9118	952.6433	951.9138	948.6987	946.1075	-0.025508 2199	0.689022846
52	976.0972	975.7356	974.8066	970.7194	972.2404	974.1422	974.1756	967.9629	972.8796	970.2528	967.4493	963.9382	-0.032226183	0.743857356
53	987.2299	984.0033	983.1238	979.0102	982.7165	985.4342	983.7189	977.5943	979.1416	980.0941	976.4778	977.5671	-0.029935 04865	0.7444482294
54	994.6871	992.1686	991.2263	987.4116	989.3381	992.7644	993.3016	985.5403	988.578	986.6473	983.573	987.2646	-0.027269 62094	0.6830370321
55	1005.5338	1003.0008	1002.1345	999.9733	1002.5696	1003.1704	1004.4032	998.8151	998.3377	997.1294	995.9133	995.6443	-0.032502 26813	0.8349313362
56	1014.7258	1014.2083	1012.7069	1011.9758	1012.9156	1012.7911	1012.6866	1006.7244	1010.0572	1009.0073	1007.8159	1008.5271	-0.026083 54764	0.8537703032
57	1024.9747	1024.0414	1024.8521	1023.1992	1022.5573	1025.4501	1024.8343	1018.7794	1019.6054	1020.7996	1019.4729	1021.755	-0.020740 3421	0.7464733003
58	1039.4293	1040.7611	1038.6123	1035.8953	1036.6474	1039.6394	1036.9131	1032.6047	1033.3053	1034.5779	1032.3307	1037.2141	-0.023651 98304	0.7207 466761
59	1053.7521	1055.7221	1053.358	1050.3313	1052.3828	1053.7042	1052.8474	1048.101	1049.4016	1050.3358	1050.5837	1047.9924	-0.021305 49993	0.754 4235824
60	1063.4091	1064.6724	1063.2147	1059.5097	1062.0764	1062.7462	1060.9804	1059.1223	1060.5367	1061.5156	1059.2987	1056.7217	-0.018811 9335	0.7238138688
61	1069.9751	1071.8001	1069.6991	1066.6977	1068.2122	1069.9492	1067.844	1065.2131	1066.8657	1067.1424	1065.109	1064.4532	-0.020775 10202	0.7909370479
62	1077.4693	1077.8849	1075.7682	1074.6143	1074.4031	1076.3297	1075.0709	1072.2063	1072.1078	1073.1625	1072.3616	1071.8378	-0.021452 72018	0.8704 680276
63	1084.097	1084.2178	1083.0119	1082.6728	1082.0106	1083.3908	1081.6459	1079.9704	1078.3844	1080.2936	1080.3466	1079.7938	-0.019647 72761	0.9010323073
64	1095.7022	1097.4517	1096.1753	1095.0477	1093.4943	1094.46	1094.1712	1093.3125	1093.3339	1092.9651	1091.9217	1088.7139	-0.021507 2536	0.8298 00025
65	1103.9375	1103.9733	1104.9116	1103.2924	1101.8424	1101.9391	1101.7819	1100.5783	1101.7848	1101.2655	1100.4186	1096.8962	-0.019799908 42	0.8123162948
66	1111.6253	1110.7598	1110.9824	1109.6209	1109.4246	1110.0712	1110.015	1106.6299	1108.6048	1109.0502	1108.512	1106.4875	-0.014223 56595	0.7737435565
67	1119.0764	1118.7775	1119.4089	1116.5975	1117.3717	1118.1835	1117.425	1114.2847	1116.3372	1116.4595	1115.2885	1115.3375	-0.014725 54008	0.7852315464

68	1125.4612	1128.7572	1127.0464	1123.7215	1124.567	1126.1123	1125.8115	1122.5496	1126.6809	1124.7566	1123.3333	1125.6695	-0.006745241057	0.33836968662
69	1132.725	1136.111	1135.9453	1132.5571	1132.9346	1134.2964	1134.6562	1130.401	1132.9358	1132.2831	1131.9693	1134.246	-0.00782462647	0.4043217832
70	1141.4752	1143.1888	1142.726	1140.9455	1140.1256	1144.4744	1143.4548	1139.0328	1141.5453	1141.4995	1140.9929	1142.4362	-0.002959917655	0.169182211 8
71	1148.4291	1151.6999	1151.6536	1149.3709	1147.6287	1151.8172	1151.3778	1147.0803	1149.8191	1149.7191	1149.3839	1150.445	-0.001801909088	0.097114 22203
72	1156.9261	1158.5131	1159.3297	1155.7453	1155.9568	1158.8837	1158.8758	1155.2888	1157.058	1156.4756	1157.3063	1158.223	-0.001802217476	0.1141644 59
73	1164.7422	1165.5653	1168.2566	1164.0748	1163.2466	1168.0265	1167.05	1163.141	1166.3702	1165.4937	1164.7726	1166.7246	0.000918866 5826	0.04640966413
74	1175.2485	1174.9205	1177.4815	1172.9429	1172.3462	1179.3021	1176.65	1174.3511	1175.7139	1176.0738	1173.5634	1176.4796	0.001568785737	0.06867579595
75	1186.3706	1186.8733	1188.9412	1184.8573	1184.9867	1190.5284	1187.9711	1187.0419	1187.3263	1187.3674	1186.9546	1184.8723	-0.001064134386	0.0548243311 6
76	1200.1767	1200.884	1202.0995	1197.795	1198.6506	1203.1512	1202.8112	1200.4151	1202.6147	1201.3979	1200.8143	1198.7596	0.002802676614	0.1393584846
77	1209.6464	1211.586	1211.809	1209.6773	1210.1923	1212.0674	1212.53	1208.654	1212.7965	1211.8152	1211.4159	1211.3705	0.004754507896	0.3194999565
78	1223.1495	1226.0571	1225.687	1221.3171	1222.8073	1226.096	1223.028	1224.8194	1225.6318	1222.8549	1223.6199	1221.2398	-0.004308395774	0.2121338 82
79	1231.6788	1235.8921	1236.5176	1234.1569	1232.8417	1236.2006	1237.4493	1233.6688	1235.0064	1234.4748	1233.7458	1232.516	-0.002290533986	0.1128668737
80	1243.8372	1246.1098	1245.8574	1244.8698	1241.3098	1246.4531	1249.7893	1243.9508	1244.1333	1242.8505	1245.3385	1242.6424	-0.004212315869	0.1652196864
81	1254.2792	1259.5179	1258.8221	1256.9608	1255.4803	1258.873	1258.651	1254.4417	1258.5405	1256.9226	1256.8249	1255.0672	-0.003135922956	0.1453546813
82	1266.2329	1270.6781	1271.2154	1266.6631	1266.1829	1271.0881	1268.5853	1268.9936	1268.813	1268.9011	1272.0753	1265.1141	0.000988754 7768	0.037588220 79
83	1277.7193	1280.6982	1282.2466	1274.5603	1276.1208	1279.5389	1280.1531	1277.8844	1280.5103	1280.2548	1281.1838	1274.278	0.001489546617	0.048947838 41
84	1284.8354	1286.6595	1288.3281	1282.1278	1283.9398	1287.2789	1287.777	1285.007	1287.2032	1286.9398	1287.4319	1283.8472	0.00407739406	0.1821596165
85	1292.2362	1293.499	1294.6442	1289.0401	1293.3747	1294.1844	1294.7863	1290.5878	1294.409	1292.8963	1293.7883	1292.4861	0.002422075885	0.12156974 19
86	1298.6168	1299.2959	1301.3815	1296.9576	1298.8069	1300.7957	1301.462	1298.7254	1300.4958	1299.2973	1301.0629	1301.5647	0.007486790576	0.4428166417
87	1304.6143	1306.1904	1307.7198	1305.4029	1305.2554	1307.4688	1308.3877	1304.4781	1306.7496	1306.0462	1307.7219	1309.5458	0.008071728594	0.4429663213
88	1313.1796	1313.77	1315.5098	1313.1754	1312.8271	1314.1845	1315.2783	1313.2773	1314.0366	1314.0679	1316.4423	1316.2838	0.007763662078	0.5387275519
89	1320.8451	1322.7952	1324.733	1321.5904	1321.0734	1324.0908	1323.1791	1320.6035	1321.4169	1321.3626	1322.7368	1323.8086	-0.000378265 7325	0.023540 3328
90	1330.0162	1330.719	1331.6652	1330.2273	1329.3705	1332.0048	1330.22	1327.6908	1330.1579	1328.6945	1330.582	1330.4144	-0.004207870177	0.3103739175
91	1337.2573	1340.1665	1340.3225	1336.752	1335.6626	1338.8369	1339.3695	1334.9124	1338.207	1337.5149	1338.2401	1338.2809	-0.002149435297	0.1122416656
92	1345.2703	1346.7034	1346.37	1343.5435	1342.5487	1347.5582	1348.1395	1341.1825	1345.2487	1344.3386	1344.7744	1346.0146	-0.002145645082	0.0914337 7908
93	1351.7654	1353.0072	1354.8922	1352.0282	1349.3944	1355.8099	1355.3145	1351.1049	1353.5964	1351.8348	1352.4075	1352.1141	-0.001062560432	0.0493574 7103
94	1358.907	1359.6241	1362.6845	1358.7795	1358.51	1362.1087	1363.155	1358.4949	1360.7277	1361.0282	1359.2645	1360.5415	0.002115468843	0.1105517603
95	1366.1575	1367.2202	1369.5808	1366.0204	1366.5588	1369.7478	1371.2805	1366.0598	1367.7698	1367.379	1367.5099	1368.0318	0.001941365134	0.1013881826
96	1372.3932	1375.1798	1376.3961	1373.5598	1373.2937	1376.7195	1378.4266	1372.9959	1374.5811	1374.5588	1375.6379	1375.658	0.004091049857	0.2022454045
97	1380.0138	1382.9515	1384.3844	1381.5165	1381.1674	1385.6311	1388.1872	1381.6363	1381.6079	1381.6949	1383.4738	1385.9766	0.006169375 97	0.2204995981
98	1389.6678	1391.1298	1393.4542	1392.3945	1394.6277	1395.7766	1398.4693	1392.265	1392.6479	1391.2516	1393.6485	1396.3424	0.007739571539	0.2690322998
99	1399.2925	1399.0457	1402.4267	1405.2438	1404.4176	1405.716	1408.2866	1403.7606	1402.7434	1401.6817	1407.4763	1407.4076	0.017056734 69	0.4822294224
100	1414.7089	1420.0469	1416.8641	1421.5023	1419.1171	1419.9883	1421.9295	1421.2227	1423.089	1419.4303	1421.687	1419.439	0.01424870809	0.5318708814
101	1436.8722	1438.3596	1440.1853	1438.3368	1433.5148	1440.0278	1438.7948	1443.0473	1438.906	1437.9107	1443.5543	1433.4725	0.005428753237	0.1521334371
102	1453.898	1453.0847	1457.7436	1451.421	1451.9375	1452.2499	1450.8607	1460.8545	1456.1165	1456.3698	1456.4656	1445.6531	0.001297027763	0.0284165547 3
103	1465.2378	1464.0347	1468.04	1463.9469	1464.6049	1464.1479	1462.3125	1468.1756	1465.6171	1465.4643	1466.6213	1456.0116	-0.008930848128	0.2433847037
104	1471.8289	1472.6315	1475.1093	1473.2725	1473.2537	1472.5595	1470.5069	1474.9605	1472.0881	1471.5902	1474.2709	1463.4142	-0.01443658107	0.4070656607
105	1479.1814	1480.3919	1481.8489	1479.1994	1480.4772	1480.4891	1478.9703	1480.4929	1477.4099	1480.1625	1481.6284	1472.9069	-0.011 47572852	0.4164764743
106	1493.4843	1487.6052	1495.3883	1498.887	1494.228	1492.7035	1489.4379	1500.6258	1488.6033	1495.6095	1496.5685	1484.9282	-0.008392760 58	0.1533287146
107	1513.6522	1515.9775	1531.0903	1525.1316	1520.375	1513.334	1515.9784	1519.5869	1516.528	1520.6742	1525.4567	1520.4219	0.003072138292	0.0501304 9118
108	3237.1647	3246.2815	3245.2197	3233.2049	3237.5232	3243.263	3245.135	3246.1913	3247.8872	3234.1426	3243.2991	3239.9274	0.007182471724	0.1233356438
109	3258.085	3257.9095	3256.6516	3253.0253	3255.2441	3255.3172	3255.0541	3254.291	3255.571	3254.296	3253.7725	3256.736	-0.007274589593	0.3966736409
110	3263.4044	3261.9603	3261.9275	3260.7517	3261.9956	3261.6967	3260.5886	3260.3794	3260.2076	3259.0021	3259.9477	3262.4041	-0.009006524981	0.6278457972
111	3267.8601	3266.6126	3266.545	3266.706	3267.1446	3267.0231	3264.808	3264.5534	3264.7157	3263.5765	3263.7377	3266.9107	-0.01246915917	0.7346032804
112	3271.9646	3271.256	3272.0998	3271.4905	3272.889	3271.1856	3270.8374	3269.3577	3268.6254	3267.564	3268.4632	3272.9806	-0.012659935 08	0.6078959154
113	3276.9469	3276.3635	3275.6581	3275.7021	3278.822	3275.2763	3275.676	3273.0348	3273.1622	3271.3254	3272.3133	3277.7141	-0.01538377721	0.5858800356
114	3281.3097	3280.3604	3280.2819	3280.5178	3282.8064	3280.4597	3280.9805	3277.748	3277.1426	3275.5246	3276.8413	3282.9731	-0.014 99980954	0.5258289111
115	3286.0633	3284.9864	3284.2544	3285.2445	3287.1657	3284.9341	3285.2081	3281.7948	3282.1201	3280.7231	3281.611	3287.8086	-0.012603373 02	0.4776523254
116	3289.7845	3289.284	3288.2396	3289.483	3289.9958	3289.6768	3290.0799	3285.6383	3285.6311	3283.8351	3285.4816	3291.4664	-0.014 32499341	0.5113907047

117	3294.4073	3292.8258	3293.095	3293.2227	3293.9701	3294.4628	3294.5414	3288.9096	3289.731	3288.1556	3289.2482	3296.3971	-0.01432536644	0.4596197706
118	3297.9652	3296.9029	3297.3931	3298.4315	3299.1356	3299.0817	3298.7767	3292.7462	3294.2983	3292.7274	3293.7273	3301.5305	-0.01298706229	0.3927434762
119	3302.8973	3303.0977	3302.6137	3303.0006	3303.5372	3303.7785	3303.595	3298.6331	3299.6584	3298.3514	3298.8775	3306.8299	-0.01139678746	0.3784283976
120	3307.9281	3306.8086	3307.7288	3307.5136	3308.5122	3308.4752	3308.015	3303.1147	3304.6013	3302.9978	3303.1263	3313.2327	-0.009320365973	0.2705911376
121	3312.2303	3312.6614	3312.4667	3312.1344	3314.112	3313.8612	3312.8232	3308.1346	3309.1491	3308.0633	3308.623	3319.7005	-0.007167296883	0.188756398
122	3318.0377	3319.7035	3320.1816	3318.4652	3322.1885	3321.4905	3317.4753	3314.3052	3314.8495	3313.6835	3313.6189	3334.2652	-0.002842063161	0.04348840179
123	3329.1125	3331.1623	3333.5724	3331.0414	3331.03	3337.4354	3332.9768	3323.6862	3325.2882	3325.5476	3322.0643	3364.3775	0.01604002318	0.1254006611
124	3365.188	3368.4942	3365.6134	3367.7098	3367.8106	3366.7929	3365.6733	3360.8396	3361.778	3360.4451	3360.3317	3371.1923	-0.01851287389	0.4490343 25
125	3370.5441	3372.6811	3370.845	3372.3926	3372.3787	3371.8878	3371.0549	3366.6408	3367.3493	3366.7253	3365.0584	3374.0926	-0.0183953369	0.5434853762
126	3375.4508	3377.5315	3375.2569	3377.1524	3376.4047	3375.8792	3375.0084	3371.2828	3372.2094	3370.0267	3370.1796	3378.1898	-0.01969622194	0.5910278786
127	3381.281	3382.3111	3381.781	3383.3258	3383.5411	3380.3129	3379.5648	3376.8046	3376.3705	3374.5969	3376.5911	3384.8904	-0.02162585637	0.5615365956
128	3895.9596	3897.8234	3896.8244	3896.1263	3896.496	3897.0018	3894.4439	3889.8807	3889.3618	3889.5921	3889.3092	3893.5911	-0.03378741262	0.8623719956
129	3898.3405	3899.4715	3898.7322	3898.4441	3898.7951	3899.164	3897.6492	3891.518	3891.0535	3891.9396	3891.1842	3897.5052	-0.031725725	0.7752794109

4

Discussion

The preceding results point clearly to the relative efficacy of our new temperature-dependent cellulose simulation protocol. Figure 3.1 shows the inadequacy of conventional techniques when applied to a single cellobiose unit; the computed spectrum, shown in black, fails to agree with the experimental spectrum in red. The simulation places the main (1100 cm^{-1}) peak at the wrong frequency, and predicts a large peak at 1400 cm^{-1} that is not observed. Generally, the simulated spectrum is much sharper with more distinct peaks than the experimental one.

The application of the new protocol produces a simulated spectrum much closer to that of real cellulose, as shown in Figure 3.2. Here, the main peak is located properly, and has the correct shape, with two distinct shoulders to the right of the main peak. The spectrum appears much smoother, and the peaks are generally confined to low (0–600 cm^{-1}), middle (1000–1200 cm^{-1}), and high (1300–1500 cm^{-1}) frequency regions, as is seen in the cellulose spectrum. There are some notable deviations between the two data sets, particularly a distinct peak at approximately 1500 cm^{-1} that appears in the simulation but not in the experiment. We suspect such deviations arise from the fact that our isolated cellobiose units do not capture the effects of interchain hydrogen bonding that occurs in real cellulose, and presumably suppresses certain vibrational modes.

The peak assignments listed in Table 3.1 suggest that the three frequency regions discussed above map clearly onto three distinct types of vibrations. The low frequency region is comprised of large-scale vibrations distributed over many atoms and bonds in the glucose rings. The middle frequen-

cies about the main peak are caused by heavy atom stretching, or changes in the bond lengths between carbon and oxygen atoms. Finally, the high frequency region corresponds to bending modes, where bond angles involving both heavy atoms and hydrogen atoms change. There is also another vibrational region, consisting of hydrogen-oxygen and hydrogen-carbon stretching modes, that occurs above 3200 cm^{-1}; however, due to a lack of comparable experimental data in this region, it will be neglected from further analysis until such data becomes available. It is worth noting that these rough assignments agree with the conclusions reached in past work on cellulose's vibrational spectra[15].

Figures 3.4 and 3.5 show that the simulation is capable of accurately predicting the effect of changing temperature on mode frequency in the stretching region, but not in the bending region; for the main stretching peak, the simulation predicts that the frequency of the peak should shift by $-0.02\ cm^{-1}$ per Kelvin, in excellent agreement with an observed redshift of $-0.03\ cm^{-1}$ per Kelvin. We attribute this redshift to the anharmonicity of the Lennard-Jones potential describing atom-to-atom repulsion and attraction; essentially, at higher temperatures, the atoms in the molecule vibrate more vigorously, and go both closer and farther away from each other as stretching vibrations occur. For a perfectly harmonic potential, this would not affect the frequency, but for a Lennard-Jones potential, larger separation distances lead to a restoring force that is weaker than that of a simple quadratic potential. Therefore atoms that travel farther from the equilibrium distance spend more time in the relatively weak anharmonic region, feeling a weaker average restoring force and vibrating at a lower frequency.

The inability of the simulations to capture the behavior of the bending modes is attributed to the fact that the majority of the bending motions describe the behavior of hydrogen atoms attached to in-ring carbons. These are the precisely the hydrogens that had to be added manually to the otherwise temperature-dependent cellobiose structures. Therefore these hydrogens are essentially unaffected by the simulated temperature, and their bending motions do not change as the simulation temperature increases. This is the cause of the slope mismatch evident in Figure 3.5. It is a distinct shortcoming of this computational approach that it does not permit quantitative analysis of modes in the bending region.

The data in Table 3 represent our most comprehensive attempt to understand the effect of increasing temperature on different types of vibrations. It lists the computed frequency of all 129 vibrational modes of cellobiose at each examined temperature up to 500K. For each mode, a linear regression has been performed, and the slope of the best-fit line describing the linear relationship between temperature and frequency is listed. This is accompanied by the absolute value of the Pearson product-moment correlation function, which measures the linearity of the temperature-frequency

relationship. A value of 1 for this function implies a perfect linear relationship between the two variables, and a value of 0 implies no linear relationship.

The results indicate that the vibrations most clearly affected by increasing temperature are the stretching vibrations. No non-stretching mode has a linear correlation function value higher than 0.43; carbon-oxygen and hydrogen-oxygen stretches have correlations as high as 0.9. Low-frequency skeletal modes display little linear dependence on temperature, with correlations below 0.25 and typically below 0.1. The bending modes also have very weak linear correlations, though for the reasons described above, simulation data for these modes are probably not useful.

Vibrations associated with movement of the glycosidic linkage are affected by temperature. Modes 50 and 55, asymmetric and symmetric stretches of the linkage respectively, both display a redshift with respective linear correlations of 0.78 and 0.80 respectively.

5 Conclusion

The results discussed in the preceding section suggest that the work presented here has brought us to two main conclusions. The first is that we now have a robust technique for predicting and calculating the Raman spectrum of complex polymers at elevated temperatures. The polymer investigated here was cellulose, but we see no reason that similar polymers could not be treated in the same way. This finding is significant because, to our knowledge, no technique for such temperature-dependent spectral analysis currently exists in the chemical literature.

The second finding is the emergence of two broad peaks in the experimental Raman spectrum of cellulose at 483K that are not predicted by our non-reactive model. These peaks indicate not just that the structural/vibrational behavior of cellulose begins to change at that temperature, they may potentially indicate exactly how that change is occurring. Identifying the nature of those peaks is the primary goal as this research continues.

Though we cannot yet confidently say what is causing new peak emergence, we have produced and are in the midst of expanding on a powerful data set that attempts to quantify the effects of temperature on the different kinds of vibrational modes. Our first attempt at this analysis is seen in Table 3. We note that this particular type of data is not attainable with any other method known to date; it is common for experimental spectroscopists to analyze the shifting of observed peaks, but in this way it is impossible to identify the exact underlying vibration that is being shifted. Our simulation makes such conclusions possible, and thus represents a significant increase in the analytic

power available to physical chemists.

We regret that, due to time constraints on the production of this thesis, this further analysis is incomplete. However, we do believe that it will prove to provide powerful new insights into cellulose thermochemistry. Some key areas for new work include tabulating the Raman intensity for each vibrational mode at each temperature, and carrying out peak assignments at every simulated temperature. In this way we can quantify the possible effects of temperature on spectral intensity and the nature of the vibrations themselves, in addition to simply the vibrational frequency.

All of this work, of course, is in service to an overarching goal of improving the efficiency of pyrolytic biofuel production. Understanding changes in the Raman spectrum of cellulose is simply another step towards more efficient energy production and a more sustainable society.

A

Online Resources

In the interest of promoting openness and cooperation in the scientific community, all data produced during this work has been made publicly available online. Also included are various scripts and other bits of technical tooling used to facilitate the gathering and analysis of spectral data.

A.1 Main Repository

The majority of the published data reside in a Git repository hosted at the following URL:

https://github.com/SeanMcGrath/McGrath396Data.

This repository includes the raw output of all the Gaussian simulations, numerous utility scripts, experimental reference spectra, all the figures used in this thesis, and more. It is documented with an accompanying README file.

A.2 Peak Shifting Data

A key piece of data not included in the above repository is the raw data used to construct Tables 3.1 and 2.1. This is kept in the form of a Google spreadsheet publicly available at the following URL: https://drive.google.com/open?id=1CNW38XCJoEd2ucuxoX7hFKjDLFzx_5maocOOE_bjVKc

A.3 gParse

The analysis techniques used in this paper require the simultaneous manipulation and recombination of dozens of sets of Gaussian output. To make this process as simple and robust as possible, a Python library named gParse has been created. This library provides several simple classes that consume Gaussian output files and convert the structural and spatial data therein into powerful data structures. Along with these data structures come many useful methods for carrying out the types of analysis presented in this thesis. gParse is a key part of nearly all the analysis scripts contained in the main repository.

The complete, installable source for gParse is hosted in its own Git repository, available here:

https://github.com/SeanMcGrath/gparse

References

[1] U. E. I. Administration, "What are the products and uses of petroleum?," 2015.

[2] R. C. Brown and T. R. Brown, *Biorenewable Resources: Engineering New Products from Agriculture*. Ames, Iowa: Iowa State Press, 2003.

[3] United States Department of Energy, "Biomass program: Feedstock composition and property database," 2015.

[4] Y. Sun and J. Chen, "Hydrolysis of lignocellulosic materials for ethanol production: a review," *Bioresource Technology*, vol. 83, 2002.

[5] J. M. Chalmers and P. R. Griffiths, *The Handbook of Vibrational Spectroscopy, Volume 1: Theory and Instrumentation.* Wiley, 2002.

[6] A. Serafim, R. Mallet, F. Pascaretti-Grizon, I.-C. Stancu, and D. Chappard, "Osteoblast-like cell behavior on porous scaffolds based on poly (styrene) fibers," *BioMed research international*, vol. 2014, 2014.

[7] F. Shafizadeh, "Introduction to pyrolysis of biomass," *Journal of Analytical and Applied Pyrolysis*, vol. 3, no. 4, pp. 283–305, 1982.

[8] H. Yang, R. Yan, H. Chen, D. H. Lee, and C. Zheng, "Characteristics of hemicellulose, cellulose and lignin pyrolysis," *Fuel*, vol. 86, no. 12, pp. 1781–1788, 2007.

[9] D. Mohan, C. U. Pittman, and P. H. Steele, "Pyrolysis of wood/biomass for bio-oil: a critical review," *Energy & fuels*, vol. 20, no. 3, pp. 848–889, 2006.

[10] M. J. J. Antal and G. Varhegyi, "Cellulose pyrolysis kinetics: the current state of knowledge," *Industrial & Engineering Chemistry Research*, vol. 34, no. 3, pp. 703–717, 1995.

[11] J. R. Cheeseman, M. S. Shaik, P. L. A. Popelier, and E. W. Blanch, "Calculation of raman optical activity spectra of methyl-β-d-glucose incorporating a full molecular dynamics simulation of hydration effects," *Journal of the American Chemical Society*, vol. 133, p. 4991, 2011.

[12] T. Kim, R. S. Assary, L. A. Curtiss, C. L. Marshall, and P. C. Stair, "Vibrational properties of levulinic acid and furan derivatives: Raman spectroscopy and theoretical calculations.," *Journal of Raman Spectroscopy*, vol. 42, p. 2069, 2011.

[13] S. Cael, J. Koenig, and J. Blackwell, "Infrared and raman spectroscopy of carbohydrates: Part iii: raman spectra of the polymorphic forms of amylose," *Carbohydrate research*, vol. 29, no. 1, pp. 123–134, 1973.

[14] J. J. Cael, J. L. Koenig, and J. Blackwell, "Infrared and raman spectroscopy of carbohydrates: Part iv. identification of configuration-and conformation-sensitive modes for d-glucose by normal coordinate analysis," *Carbohydrate research*, vol. 32, no. 1, pp. 79–91, 1974.

[15] J. Cael, K. Gardner, J. Koenig, and J. Blackwell, "Infrared and raman spectroscopy of carbohydrates. paper v. normal coordinate analysis of cellulose i," *The Journal of chemical physics*, vol. 62, no. 3, pp. 1145–1153, 1975.

[16] J. J. Cael, J. L. Koenig, and J. Blackwell, "Infrared and raman spectroscopy of carbohydrates. part vi: Normal coordinate analysis of v-amylose," *Biopolymers*, vol. 14, no. 9, pp. 1885–1903, 1975.

[17] M. E. Tuckerman and G. J. Martyna, "Understanding modern molecular dynamics: Techniques and applications," *The Journal of Physical Chemistry B*, vol. 104, pp. 159–178, 01/01 2000. doi: 10.1021/jp992433y.

[18] V. Agarwal, G. W. Huber, W. Conner, and S. Auerbach, "Simulating infrared spectra and hydrogen bonding in cellulose iβ at elevated temperatures," *The Journal of Chemical Physics*, vol. 135, p. 134506, 2011.

[19] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "Gromacs 4.5: a high throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*, vol. 29, no. 7, p. 845, 2013.

[20] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, M. A. Jr.J., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N.

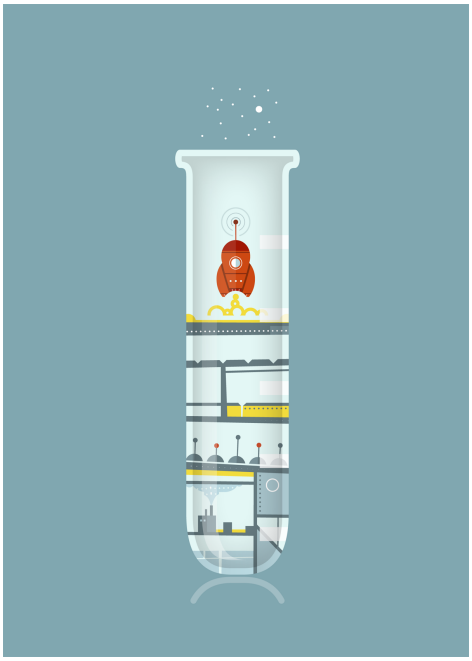
Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, . Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, "Gaussian 09 revision e.01." Gaussian Inc. Wallingford CT 2009.

[21] R. G. Parr, "Density functional theory," *Annual Review of Physical Chemistry*, vol. 34, no. 1, pp. 631–656, 1983.

[22] K. Kim and K. Jordan, "Comparison of density functional and mp2 calculations on the water monomer and dimer," *The Journal of physical chemistry*, vol. 98, no. 40, pp. 10089–10094, 1994.

[23] D. V. Schroeder, *An introduction to thermal physics*, vol. 60. Addison Wesley New York, 2000.

[24] M. Bradley, "Curve fitting in raman and ir spectroscopy: Basic theory of line shapes and applications," Tech. Rep. 50733, Fisher Scientific, 2007.



This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, *Science Experiment 02*, was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD dissertation with this look & feel has been released under the permissive AGPL license, and can be found online at github.com/suchow/Dissertate or from its lead author, Jordan Suchow, at suchow@post.harvard.edu.