2011

A low-frequency Raman study of six nucleosides

Craig Koontz
The University of Toledo

Follow this and additional works at: http://utdr.utoledo.edu/theses-dissertations

Recommended Citation
http://utdr.utoledo.edu/theses-dissertations/612

This Thesis is brought to you for free and open access by The University of Toledo Digital Repository. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of The University of Toledo Digital Repository. For more information, please see the repository’s About page.
A Thesis

entitled

A Low-Frequency Raman Study of Six Nucleosides

by

Craig Koontz

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the
Master of Science Degree in Physics

Dr. Scott Lee, Committee Chair

Dr. Brian Bagley, Committee Member

Dr. Bo Gao, Committee Member

Dr. Patricia Komuniecki, Dean
College of Graduate Studies

The University of Toledo
August 2011
An Abstract of
A Low-Frequency Raman Study of Six Nucleosides

by

Craig Koontz

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Master of Science Degree in Physics

The University of Toledo
August 2011

Low-frequency modes in nucleosides are important to study due to the molecules' similarities to DNA and RNA, and these modes' involvement in replication, transcription, and other processes which involve large molecular units. This paper will report on Raman spectroscopic studies of six nucleosides from 20-1600 cm\(^{-1}\), with a focus on the low-frequency (20-200 cm\(^{-1}\)) region. These data will be used to analyze the nature of bonding in the nucleosides. It will also be compared to low-frequency theoretical predictions on nucleosides from Shishkin et al.
# Contents

Abstract iii

Contents iv

List of Figures v

1 Introduction 1

1.1 DNA, RNA, Nucleoside Overview ................................. 1

1.2 Raman Spectroscopy ............................................... 17

1.3 Previous Research .................................................. 18

1.4 Purpose ............................................................. 21

2 Experimental 23

2.1 Experimental Details ................................................. 23

3 Results and Discussion 24

3.1 Individual Nucleosides ............................................ 28

3.2 Comparisons Between Groups of Nucleosides .................... 32

3.3 Theoretical Predictions ............................................ 33

3.4 Basestacking Character of Each Nucleoside ..................... 34

4 Conclusions 37

References 38
List of Figures

1-1 DNA Double-helix. .................................................................1
1-2 Composition of nucleotides and nucleosides.................................2
1-3 Chemical Structure Composition.............................................3
1-4 Ribose and deoxyribose.........................................................4
1-5 Sugar puckering of ribose and deoxyribose sugars in DNA and RNA.........5
1-6 Bases found in DNA and RNA................................................6
1-7 DNA in the A conformation and B conformation ............................9
1-8 Assembly of nucleosides..........................................................13-15
1-9 View of nucleoside crystal assembly of multiple unit cells viewed along axis of base stacking........................................................................................................................................16-17
4-1 Raman spectra from 200 to 1700 cm\(^{-1}\) for 6 nucleosides..................25
4-2 Raman spectra from 20 to 200 cm\(^{-1}\) for 6 nucleosides. Included is overlays of predicted modes for deoxyadenosine with C3’-endo sugar pucker, deoxythymidine with C3’-endo sugar pucker, and deoxycytidine with C3’-endo and C2’-endo sugar pucker. 26
4-3 Graph of intensity of S mode to intensity of rest of peaks lower than 200 cm\(^{-1}\) for six nucleosides........................................................................................................................................................................35
Chapter 1

Introduction

1.1 DNA, RNA, Nucleoside Overview

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are central in the storage and transmission of genetic information in life. They are important in transcription (the process by which gene expression begins) and replication (which allows for biological inheritance). DNA and RNA play a role in interactions with proteins for regulation of cell growth\(^1\), as well as interactions with drugs. Because of the importance of DNA and RNA in life, it is crucial to learn the nature of these molecules and their components.

Figure 1-1. DNA Double-helix. 
(http://commons.wikimedia.org/wiki/File:A-DNA,_B-DNA_and_Z-DNA.png)
DNA consists of long, polymeric molecules which form a double-stranded, helical structure (See Figure 1-1). Nucleotides form the simplest unit of DNA and are composed of three parts, a phosphate, a sugar, and a base (Fig. 1-2). The double-stranded, helical structure can be viewed as a twisted ladder. In this twisted ladder, phosphates bonded with sugars form the two risers while the bases, which are bonded to the sugars, are paired with each other forming the rungs (Fig. 1-3). So, for each step of this ladder, there are two nucleotides, which are bonded to each other by their bases through hydrogen bonding. As the ladder climbs higher, nucleotides are stacked on one another by covalent bonding of one nucleotides sugar to the next nucleotide’s phosphate group.
The formation of DNA in this way allows for the bases, which contain the genetic information, to be sheltered from chemical modification by the “protection” of the “monotonous and uniform” \(^1\) phosphate-sugar backbone. Only specific proteins and other molecules can unwind the DNA to obtain access to the genetic information contained in the bases.

RNA itself is a similar molecule to DNA, made up of nucleotides connected through sugar-phosphate bonds. RNA has several differences from DNA. First, the sugar in RNA is different from DNA. Second, in RNA the base uracil replaces thymine (the other three bases remain the same). Finally, RNA tends to be single stranded, contrasting DNA’s double-stranded structure.
The simplest component of DNA and RNA are nucleotides, which are made of three parts, a phosphate group, a sugar, and a base. The sugar for both DNA and RNA is a furanose ring which has 4 carbons and one oxygen atom. There is a numbering method for these sugars which is shown in figure 1-4. In RNA this sugar is a ribose, whereas in DNA, this sugar is a deoxyribose, which has a hydrogen attached to the 2’ position in place of the hydroxyl (-OH) found in ribose.
Planar furanose is not energetically favorable for DNA or RNA. The carbons in the furanose ring are in sp$^3$ hybridization, which results in four orbitals being 109.5° with respect to each other pointing toward the four corners of a tetrahedron. This mismatches the 108° planar angle that the pentagonal furanose would exist at. To accommodate for this mismatch, the furanose puckers$^2$. This pucker comes in two forms: envelope, which has 4 in-plane atoms and 1 out-of-plane, and twist, which has 3 in-plane atoms and 2 adjacent out-of-plane atoms. Atoms out-of-plane on the side of C5’ (as in figure 1-5) are called endo, while atoms on the other side are called exo. The carbons that are energetically favorable to adopt a puckered position are C2’ and C3’ because of the nonbonding interactions of the furanose ring substituents.
A phosphate group is bonded to a sugar at C5’. These phosphate groups bond to the next nucleotide at the C3’ position through a phosphodiester bond (See Fig. 1-2). This pattern, C3’-phosphate’-C5’, continues through the DNA or RNA macromolecule.

![Diagram of Purines and Pyrimidines](http://commons.wikimedia.org/wiki/File:Nucleotides_1.svg)

Figure 1-6. Bases found in DNA and RNA.

Bases are bonded to the sugar at the C1’ position by a glycosidic bond. There are two types of bases, purines and pyrimidines. These two types are differentiated by their base organic compound. Purine is a double ring structure while pyrimidine is single-ringed. As such, more atoms are contained in purines than pyrimidines. There are two types of bases in DNA and RNA that have purine as their base organic compound, adenine and guanine (Fig. 1-6). There are three types of pyrimidines found in DNA and RNA, thymine, cytosine, and uracil. In DNA, only cytosine and thymine exist, while in RNA, thymine is replaced by closely related uracil. Thymine and uracil are only differentiated by uracil’s lack of a methyl group at the C5 position.

In DNA, pairs of bases are bound to each other through hydrogen bonds to form the aforementioned rungs of the ladder. Certain bases are paired with other certain bases. In general a certain purine base will only bond to another certain pyrimidine base. For
example, cytosine will bond to guanine, and thymine to adenine, but not cytosine to adenine or thymine to guanine. This phenomenon of favorable matching of bases is called base-pairing.

The sequence of bases in DNA is crucial to the storage of genetic information. Base pairing facilitates the transfer of this information to gene expression. In transcription, a strand of RNA is formed based on the sequence of bases in DNA (which has been matched by base pairing). This strand of RNA is then used in the production of proteins. In replication, DNA is unwound into two single strands. The bases on each single strand are matched to their corresponding base partner. The consequence of this is two strands of DNA identical to the original strand of DNA. These processes emphasize the differences in function that DNA and RNA serve. DNA serves as the storage of genetic information, with the instructions for protein production and cell growth. RNA sometimes plays a role in storing this genetic information as well, but its main role is in transferring the genetic information to different parts of the cell and regulating protein production.

As mentioned before, nucleotides are the simplest component of DNA and RNA. They are made of three parts, a phosphate, sugar, and base. Nucleosides are composed of only the sugar and base. They lack the phosphate contained in nucleotides. Nucleosides contain the same two sugar structures and five base structures as in nucleotides. As such, there are a total of eight nucleosides related to DNA and RNA: adenosine (having an adenine base and ribose sugar), deoxyadenosine (adenine base, deoxyribose sugar), cytosine, deoxycytosine, uridine, deoxythymisine, guanosine, and deoxyguanosine. These can form as a crystal, contrasting the previously discussed polymeric DNA and
RNA. Nucleosides discussed in this paper will be isolated molecules in a crystal state. They are not in polymeric form. Nucleosides can aid the study of DNA and RNA due to their similar composition and bonding nature. This paper will focus on the first six of the eight nucleosides mentioned above and neglect guanosine and deoxyguanosine.

Bonding in DNA, RNA, and, consequently, nucleosides is important in getting a full understanding of the nature of these molecules. In nucleoside crystals, there are several categories of bonds which vary in strength, and therefore, frequency of vibrational modes. Strongest of these involve covalent bonds. The stretching modes involving covalent bonding of hydrogen will be at the highest frequency (~3000 cm\(^{-1}\)), while covalent stretching modes involving heavier atoms will be at a lower frequency (400-1700 cm\(^{-1}\)). Bending and torsional forces involving three- and four-member systems will be weaker and at a lower frequency (20-600 cm\(^{-1}\)). The weakest and lowest frequency modes (< 200 cm\(^{-1}\)) involve motions of entire molecular units bound by intermolecular bonds (such as hydrogen bonds and van der Waals interactions).

Intermolecular interactions are the focus of this study. These bonds have the weakest restoring force. Of particular interest in this study are the low frequency vibrational modes. Such modes involve large portions of the molecule and weak restoring forces. Such modes are of biological interest since large numbers of atoms must move in coherent manners during transcription and replication. As said before, these interactions originate from hydrogen bonding and van der Waals interactions. In nucleoside crystals, although the nature of the bonding is different from DNA, similarities exist such as base stacking interactions.
Figure 1-7. DNA in the A conformation (left) and B conformation (right). A-DNA has bases at an angle of about 22° relative to the helix axis while B-DNA has bases nearly perpendicular to the helix axis.

The intermolecular interactions focused on for this study are base stacking, column formation due to this base stacking, and hydrogen bonding between neighboring nucleosides. The distinction between base stacking and column formation due to base stacking is significant. Bases may stack without forming a column if the base stacking is not continuous throughout the polymer or crystal. For example, pairs (or perhaps more) of bases may stack without a following base, leaving a structure with base stacking, but without a continuous column of bases. Base stacking occurs due to van der Waals interactions between neighboring molecules in DNA, RNA, and in nucleoside crystals. In DNA, bases may stack nearly perpendicular to the axis of the helix (as in figure 1-7 in B-DNA), or, more likely, at an angle relative to the axis of the helix (as in figure 1-7 in A-DNA). Figure 1-8 shows the crystal structure\textsuperscript{3-10} of the six nucleosides studied. Significant base-stacking occurs in five of the nucleosides. Of these 5, 2 nucleosides (deoxythymidine and cytidine) exhibit bases stacking directly on top of each other, with the other 3 (deoxyadenosine, adenosine, and uridine) stacking with an offset in one
direction. This continuous base-stacking leads to the formation of columns of nucleosides that repeat in the same orientation continuing “up” the crystal. The remaining nucleoside, deoxycytidine, exhibits base stacking in pairs of bases, with one nucleoside oriented in the opposite direction as its base-stacking partner (See fig. 8(d)). Because of the lack of continuous base-stacking, no column formation occurs.

Hydrogen bonding between neighboring nucleosides occurs in all 6 nucleosides. The specific bonding that occurs, i.e. which atom the hydrogen is donated to, is different depending on the crystal structure. The neighboring pattern is shown for the six nucleosides in figure 1-9. Adenine based nucleosides form with nucleoside bases facing each other, forming “ribbons” of bases (in blue), beside ribbons of sugars (primarily red). (In the figure, these are oriented from top to bottom and emphasized with black lines in the Adenine figure.) In cytidine and deoxythymidine, bases and sugars face each other. Ribbons also form in these two nucleosides, although thinner than in the adenine base nucleosides. In cytidine the ribbons are oriented from left to right and in deoxythymidine the ribbons are oriented from top to bottom. Lines of nucleosides (oriented left to right) form in uridine with alternating base-sugar orientation, and with neighboring lines offset by half of a nucleoside. Deoxycytidine forms no columns and so cannot be analyzed in the same way as the five other nucleosides.

Table 1-1 lists the nucleosides and several characteristics which are important in base-stacking and intermolecular interactions. The base composition affects the mass of the nucleoside with purines being heavier than pyrimidines. This could lead to lower frequency modes in heavier nucleosides. Crystal parameters are listed. They determine the distance between individual nucleosides, which affects the strength of intermolecular
interactions. Stacked base to base distances and angles bases make relative to the stacking axis are listed specifically because of the proposed importance of base stacking to the modes that are being examined.
Table 1-1. Characteristics of the six nucleosides being studied.

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>&quot;S-mode&quot; frequency (cm⁻¹)</th>
<th>Molar Mass (g/mol)</th>
<th>Unit Cell</th>
<th>Space Group</th>
<th>a (Å)</th>
<th>b(Å)</th>
<th>c(Å)</th>
<th>alpha (°)</th>
<th>beta (°)</th>
<th>gamma (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uridine</td>
<td>28</td>
<td>244.2</td>
<td>Monoclinic</td>
<td>P2₁</td>
<td>4.981</td>
<td>14.649</td>
<td>13.964</td>
<td>90.000</td>
<td>95.500</td>
<td>90.000</td>
</tr>
<tr>
<td>Deoxythymidine</td>
<td>26</td>
<td>242.229</td>
<td>Orthorhombic</td>
<td>P2₁2₁2₁</td>
<td>4.860</td>
<td>13.810</td>
<td>16.320</td>
<td>90.000</td>
<td>90.000</td>
<td>90.000</td>
</tr>
<tr>
<td>Cytidine</td>
<td>30</td>
<td>243.22</td>
<td>Orthorhombic</td>
<td>P2₁2₁2₁</td>
<td>13.990</td>
<td>14.790</td>
<td>5.116</td>
<td>90.000</td>
<td>90.000</td>
<td>90.000</td>
</tr>
<tr>
<td>Deoxycytidine</td>
<td>29</td>
<td>244.217</td>
<td>Triclinic</td>
<td>P1</td>
<td>7.285</td>
<td>6.866</td>
<td>11.074</td>
<td>104.300</td>
<td>84.900</td>
<td>72.400</td>
</tr>
<tr>
<td>Adenosine</td>
<td>36</td>
<td>267</td>
<td>Monoclinic</td>
<td>P1</td>
<td>4.825</td>
<td>10.282</td>
<td>11.823</td>
<td>90.000</td>
<td>99.300</td>
<td>90.000</td>
</tr>
<tr>
<td>Deoxyadenosine</td>
<td>35</td>
<td>251.24</td>
<td>Monoclinic</td>
<td>P2₁</td>
<td>11.298</td>
<td>10.393</td>
<td>4.819</td>
<td>90.000</td>
<td>101.510</td>
<td>90.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>Base</th>
<th>Sugar conformation</th>
<th>Base-base distance (stacked) (Å)</th>
<th>Base-base Offset (Å)</th>
<th>No. of H bonds/Nucleoside</th>
<th>Orientation of Base w.r.t Stacking Axis (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uridine</td>
<td>Pyrimidine</td>
<td>C(3')-endo</td>
<td>4.981</td>
<td>0.477</td>
<td>6</td>
<td>12.7</td>
</tr>
<tr>
<td>Deoxythymidine</td>
<td>Pyrimidine</td>
<td>C(3')-exo</td>
<td>4.86</td>
<td>0</td>
<td>4</td>
<td>43.7</td>
</tr>
<tr>
<td>Cytidine</td>
<td>Pyrimidine</td>
<td>C(3')-endo</td>
<td>5.12</td>
<td>0</td>
<td>5</td>
<td>40.1</td>
</tr>
<tr>
<td>Deoxycytidine</td>
<td>Pyrimidine</td>
<td>C(3')-endo/C(2')-endo</td>
<td>3.95</td>
<td>2.69</td>
<td>5, 5</td>
<td>N/A</td>
</tr>
<tr>
<td>Adenosine</td>
<td>Purine</td>
<td>C(3')-endo</td>
<td>4.825</td>
<td>0.78</td>
<td>6</td>
<td>34.2</td>
</tr>
<tr>
<td>Deoxyadenosine</td>
<td>Purine</td>
<td>C(3')-endo</td>
<td>4.7</td>
<td>0.937</td>
<td>4</td>
<td>37.5</td>
</tr>
</tbody>
</table>
(a) Adenosine

(b) Deoxyadenosine
Figure 1-8. Assembly of nucleosides. (a) Adenosine. (b) Deoxyadenosine. (c) Cytidine. (d) Deoxycytidine. (e) Uridine. (f) Deoxythymidine. Images created with Jmol\textsuperscript{11}.
**Adenosine** (Black lines emphasizing “ribbon”-like structure.)

**Deoxyadenosine**

**Cytosine**

**Deoxycytidine**
Uridine

Figure 1-9. View of nucleoside crystal assembly of multiple unit cells viewed along axis of base stacking. Images created with Jmol\textsuperscript{11}.

1.2 Raman Spectroscopy

Raman spectroscopy is a technique that can study the vibrational modes in a molecule. This technique requires a monochromatic light to be scattered off a sample. Most of this scattering will be elastic and will have no change in frequency from the original beam. Elastic scattering is termed Raleigh scattering. Some of the light will be inelastically scattered. This phenomenon occurs by the incoming photon exciting a molecule from the ground vibrational state to a higher energy virtual state. The molecule then relaxes to an energy state different from the original, emitting a photon with an energy different from the original photon corresponding to the energy difference from the original state to the final. This scattering is termed Raman scattering. The frequency change from the incident light is calculated with the following equation:
\[ \Delta w = \left( \frac{1}{\lambda_0} - \frac{1}{\lambda_1} \right), \]

with \( \lambda_0 \) being the wavelength of the incident light, \( \lambda_1 \) being the wavelength of the Raman scattered light, and \( \Delta w \) being the wavenumber shift in inverse length units. The frequency shift is given in units of inverse centimeters, known as wavenumbers. This unit is used simply because of convenience as wavenumbers of molecular bonds will be less than 3000 cm\(^{-1}\). The frequency shift corresponds to the strength of the bond. For example, a large frequency change (~3000 cm\(^{-1}\)) originates from relatively strong covalent bonding involving hydrogen atoms.

Raman spectroscopy at low frequencies will study modes involving large portions of molecules with low restoring forces. At low frequencies the Raman scattered light will be close in proximity to the relatively strong Raleigh scattered light. This creates a problem of obscuring the Raman scattered light. Careful experimental procedures must be taken to prevent this from happening. Obscurity may also exist in Raman spectra because of overlapping modes and fluorescence from substances in the sample. This problem makes low frequency Raman spectroscopy extremely difficult. The study reported in this thesis is the first systematic study of the 20 to 200 cm\(^{-1}\) region of the Raman spectra of the nucleosides.

Raman spectroscopy can be used to study the vibrational modes of nucleosides to explore the nature of their bonds.

### 1.3 Previous Research

Raman spectroscopic research on DNA was first published by R. C. Lord in 1967\(^1\). The focus of Raman spectroscopic research has been to assign modes to
certain constituents of DNA, or closely related molecules, which can be indicators of “local structure, global conformation, intermolecular interaction, or molecular dynamics”\(^{13}\). Early studies were focused on higher frequencies than 200 cm\(^{-1}\) due to the complications of Raman spectroscopy at low frequencies. These studies identified high frequency modes for DNA and RNA conformation\(^{14,15}\).

As refined methods of Raman spectroscopic experimentation were put into use, the low-frequency region became available for inspection. Various conditions were altered, such as water content\(^{17,18}\), counterion species\(^{18,19}\), base sequences\(^{20}\), and temperature\(^{19,21}\), to examine the modes important in the collective motions in DNA and related molecules. Most modes in this region (20-100 cm\(^{-1}\)) were found to be insensitive to changes in water content (which changes the distance DNA bases stack from each other as well as the distance DNA helices neighbor each other)\(^{17}\). These modes strongly depend on DNA conformation and not intermolecular or base stacking interactions. Hydrogen bonds in bases were found to be necessary for these modes\(^{22}\) and base sequence (which affects the hydrogen bonding of the bases) was found to affect the spectra of these modes\(^{18}\). Interhelical interactions were ruled out by studying Raman spectra of molecules similar to DNA which either had alternate helices\(^{20,23}\), or no helix at all\(^{24,25,26}\).

The lowest-frequency mode was found to be sensitive to changes in water content, contrasting the other modes in the low-frequency region. This mode has since been dubbed the S-mode by Urabe and co-workers\(^{21}\). It was determined to be intermolecular in nature, depending mainly on the interactions between adjacent molecules, not interactions within a molecule\(^{17,27,28,29}\). This mode had various theorized
causes such as the DNA backbone chain\textsuperscript{17,30}, water content of the sample\textsuperscript{31}, and adjacent, hydrogen bound nucleotides\textsuperscript{23}. The mode changed frequency with changes in DNA conformation (which alters intermolecular and base stacking interactions)\textsuperscript{23,28}. Changing the counterion in DNA also showed a change in frequency in the S mode, which suggested a dependence on intermolecular and/or base-stacking interactions. The S-mode, along with other modes below 200 cm\textsuperscript{-1}, also changed frequency with changing temperature\textsuperscript{16,32,33,34}. As temperature decreased, the frequency increased, indicating a strengthening of intermolecular forces.

The origin of the S mode was still mysterious, so related molecules were used to study the importance of interhelical and intrahelical interactions on this mode. The nucleotide, guanosine monophosphate, was used to establish the common origin of the lowest mode in nucleotides to the S mode in DNA\textsuperscript{23}. This origin was also found to be common in nucleosides polymers and crystals\textsuperscript{26}, ruling out the necessity of phosphate-sugar backbone chains. Also, hydrogen bonding was ruled out as the origin of this mode due to differences in hydrogen bonding between different guanine-based molecules. Similarities of the S-mode to the lowest-frequency mode in methyluracil crystal, lead to the proposed origin of libration of bases in an accordion-like fashion. The S mode was suggested to be affected by base stacking, column formation due to this base stacking, and hydrogen bonding between bases, with the exact frequency being determined by intra- and intercolumn interactions.

As well as experimental research, theoretical research has been performed on DNA and its relatives to predict what frequency modes exist and assign origins to these modes. Theoretical predictions by Shishkin and co-workers have been carried out mainly
using density functional theory\textsuperscript{35}. Theoretical predictions have been compared to experimental results in order to assist in assigning origins of modes, as well as for the purpose of refining theoretical procedures. Such work is greatly complicated by the fact that low-frequency modes, which involve many distant atoms, are affected by long-range interactions. The Coulombic interactions between the partial charges in the molecule is particularly important and difficult to model accurately, mainly because the exact local partial charge is poorly understood.

Much research has been done to predict and assign modes at higher wavenumbers\textsuperscript{36, 38, 37}. Simpler molecules related to DNA have been researched as well, including in the low-wavenumber region \textsuperscript{39}. Recently, low-wavenumber modes have been predicted for isolated nucleosides with assignments given\textsuperscript{40}.

It should be noted that the theoretical work of Shishkin \textit{et al.} has been performed for the C2’ and C3’ puckers. Theoretical calculations have the advantage of not being constrained by physical reality. Our experimental results are for samples with specific sugar puckers. Further elaboration on this topic is given later.

1.4 Purpose

Low-frequency modes in nucleosides are important to study due to the molecules similarities to DNA and RNA, and these modes’ involvement in replication, transcription, and other processes which involve large molecular units. This paper will report on Raman spectroscopic studies of six nucleosides from 20-1600 cm\textsuperscript{-1}, with a focus on the low-frequency (20-200 cm\textsuperscript{-1}) region. These data will be used to analyze the nature of
bonding in the nucleosides. It will also be compared to low-frequency theoretical predictions on nucleosides from Shishkin et al.
Chapter 2

Experimental

2.1 Experimental Details

Nucleosides were obtained from ISIS Pharmaceuticals and used without any further purification. The nucleosides were prepared as a finely ground powder. In preparation for Raman spectroscopy, a powder of a single type of nucleoside was placed between two microscope slides held together by adhesive tape. After Raman spectra were obtained for one type of nucleoside, the process was repeated with another nucleoside powder until spectra for all six nucleosides were recorded.

Raman spectra were excited with an argon ion laser operating at a wavelength of 514.5 nm and at powers up to 20 mW. Radiation scattered at 90° was collected by a camera lens and focused on the entrance slit of a double monochromator. The detector was a photomultiplier tube, cooled by a thermoelectric cooler to reduce dark current, and coupled to photon counting electronics. Spectrometer scans, data collection and processing were controlled by a personal computer.
Chapter 3

Results and Discussion

Figure 3-1 shows Raman spectra from six nucleosides in the region of 200-1700 cm\(^{-1}\). Figure 3-2 shows Raman spectra from six nucleosides in the low-frequency region, 20-200 cm\(^{-1}\). This figure also shows the frequencies of predicted modes for deoxyadenosine, deoxycytidine, and deoxythymidine from Shishkin et al. Table 3.1 lists frequencies of observed peaks from experiment and predicted peaks with assignments from Shishkin et al.
Figure 3-1. Raman spectra from 200 to 1700 cm$^{-1}$ for 6 nucleosides.
Figure 3-2. Raman spectra from 20 to 200 cm\(^{-1}\) for 6 nucleosides. Included is overlays of predicted modes for deoxyadenosine with C3’-endo sugar pucker, deoxythymidine with C3’-endo sugar pucker, and deoxycytidine with C3’-endo and C2’-endo sugar pucker.
Table 3.1. Observed peaks for six nucleosides with predicted peaks and assignments from Shishkin et al.

<table>
<thead>
<tr>
<th>Sugar Modes</th>
<th>Base Modes</th>
<th>Sugar Modes</th>
<th>Base Modes</th>
</tr>
</thead>
<tbody>
<tr>
<td>vs1  Ring out-of-plane</td>
<td>Thymidines</td>
<td>vs1  Ring out-of-plane</td>
<td>N(1) out-of-plane</td>
</tr>
<tr>
<td>vs2  Rotation around C(4')-C(5'), Ring out-of-plane</td>
<td>vs2 methylrotation</td>
<td>vs2  Rotation around C(4')-C(5'), Ring out-of-plane</td>
<td>methylrotation</td>
</tr>
<tr>
<td>vs3</td>
<td>N(3) out-of-plane</td>
<td>vs3</td>
<td>N(3) out-of-plane</td>
</tr>
</tbody>
</table>

**Sugar Modes**
- vs1: Ring out-of-plane
- vs2: Rotation around C(4')-C(5'), Ring out-of-plane
- vs3: Rotation around C(4')-C(5'), Ring out-of-plane

**Base Modes**
- Thymidines
  - vs1: N(1) out-of-plane
  - vs2: methylrotation
  - vs3: N(3) out-of-plane
- Cytosine
  - vs1: N(1) out-of-plane
- Adenine
  - vs1: C(8) out-of-plane
3.1 Individual Nucleosides

Deoxyadenosine has 39 distinct modes in the 20-1700 cm\(^{-1}\) region. Table 3.1 presents the predicted frequencies and assignments from Shishkin et al., which are given for 0-200 cm\(^{-1}\), for the C(3')-endo sugar pucker only (and not the C(2')-endo) because deoxyadenosine is in the C(3')-endo sugar pucker conformation in crystal. Eleven distinct modes exist below 200 cm\(^{-1}\). The lowest mode exists at 35 cm\(^{-1}\), which is then followed by two slightly weaker modes at 50 and 62 cm\(^{-1}\). These compare well with predicted modes at 26, 50 and 58 cm\(^{-1}\) and are assigned as modes involving the rotation around C(1')-N bond, a bending mode in the base unit plane, and a bending in the sugar unit plane respectively. Experimentally observed modes exist at 72, 79, 89, and 99 cm\(^{-1}\). These four modes are relatively weak in comparison to their neighbors. A strong mode is then observed at 115 cm\(^{-1}\) followed by several very weak modes at 136, 150, and 162 cm\(^{-1}\). Modes are predicted to exist at 97, 120, 144, and 171 cm\(^{-1}\). Several of these modes match well, however only a total of seven modes are predicted below 200 cm\(^{-1}\), where 11 are observed to exist. Several of predicted modes correspond to observed modes (97 to 99, 120 to 115, 144 to 150, and 171 to 162 cm\(^{-1}\)). The 58-97 cm\(^{-1}\) region is devoid of predicted peaks where several peaks clearly exist. Modes in this region have been proposed to originate from base bonds and are sensitive to structural difference (Urabe). Shishkin et al. have predicted modes only for single molecules and not in repeating crystal structure. This suggests that these modes are intermolecular in nature.

Adenosine has 56 distinct modes in the 20-1700 cm\(^{-1}\) region. No theoretical predictions currently exit for adenosine in the low-frequency region. 10 distinct modes
are observed below 200 cm\textsuperscript{-1}. The lowest distinct mode exists at 37 cm\textsuperscript{-1} and is very strong. This mode exists at the same frequency as in deoxyadenosine. This suggests a strong correlation between base and lowest-frequency mode. Medium strength modes are observed at 46 and 54 cm\textsuperscript{-1}. These are similar to in strength and relative position to deoxyadenosine’s modes at 49 and 61 cm\textsuperscript{-1}, which suggests similar origin for these modes. Therefore, I propose that the 45 and 52 cm\textsuperscript{-1} modes in adenosine originate from bending in the base unit plane and bending in the sugar unit plane respectively. The base unit plane mode in both nucleosides are closer in frequency due to the molecules’ identical base, while the sugar unit plane mode has a greater difference due to the different, but very similar sugars in both molecules. Continuing to higher frequencies, modes are observed at 67 and 76 cm\textsuperscript{-1}. The mode at 67 cm\textsuperscript{-1} is broader and could be two modes obscured by their proximity in frequency to each other. A strong mode exists at 99 cm\textsuperscript{-1} in adenosine which matches closely in relative strength and frequency to the mode at 99 cm\textsuperscript{-1} in deoxyadenosine. This suggests a similar origin, which is \(v^s_2(\text{ring}) + v^A_1\) from Shishkin et al. Several weaker modes exist at higher frequencies, 109, 140, and 152 cm\textsuperscript{-1}.

Deoxythymidine has 61 distinct modes in the 20-1700 cm\textsuperscript{-1} region. Table 3.1 presents the predicted frequencies for the C(3\textsuperscript{\prime})-endo sugar pucker only because that is the only conformation deoxythymidine takes in the crystalline state. Eleven distinct modes are observed below 200 cm\textsuperscript{-1}. The lowest mode is observed at 24 cm\textsuperscript{-1}, near the predicted frequency of 29 cm\textsuperscript{-1} for the rotational mode around C(1\textsuperscript{\prime})-N bond. This strong mode is followed by 4 weaker modes at 38, 47, 55, and 63 cm\textsuperscript{-1}. Two modes are predicted in this region at 46 and 47 cm\textsuperscript{-1}, which are assigned to be bending in the base
unit plane and bending in the sugar unit plane, respectively. Judging by the similarity in spectral pattern to deoxyadenosine, the two observed lower frequency peaks at 38 and 44 cm\(^{-1}\) should correspond to the predicted modes at 46 and 47 cm\(^{-1}\). There is a strong peak at 77 cm\(^{-1}\), which does not have a closely matching predicted peak. Broad peaks are observed at 95 and 119 cm\(^{-1}\), followed by several relatively weak peaks at 151, 166, and 180 cm\(^{-1}\). There are six modes predicted above 95 cm\(^{-1}\), most of which closely match observed modes. No peaks are predicted from 47 to 95 cm\(^{-1}\) where several strong peaks are observed. This is similar to the situation in deoxyadenosine.

Uridine has 64 distinct modes in the 20-1700 cm\(^{-1}\) region, with 10 observed below 200 cm\(^{-1}\). There are no theoretical predictions for modes in the low-frequency region for uridine at this time. The spectrum for uridine is obscured by the Raleigh scattered light below 50 cm\(^{-1}\). The lowest distinct peak is observed at 29 cm\(^{-1}\). This is followed, in increasing frequency, by a strong peak at 40 cm\(^{-1}\), a weak peak at 47 cm\(^{-1}\), and strong peak at 57 cm\(^{-1}\) with a shoulder at 61 cm\(^{-1}\). Above this frequency several weak modes exist at 75, 92, and 114 cm\(^{-1}\). Few similarities exist in the spectral pattern between uridine and deoxythymidine despite similarities in their base composition. Notably, the lowest frequency mode is higher in uridine than in deoxythymidine, contrasting the previously mentioned adenine-based nucleosides.

Deoxycytidine has 73 distinct observed modes. Table 3.1 presents predicted modes for both the C(3')-endo and C(2')-endo conformation as deoxycytidine exists in both conformations in crystal. Thirteen modes are observed below 200 cm\(^{-1}\). Some Raleigh scattered light is observed to obscure the spectrum below 35 cm\(^{-1}\). The lowest mode is observed at 30 cm\(^{-1}\) which is within both predicted values for the mode involving
the rotation around C(1′)-N. This mode is followed, in increasing frequency, by two weak modes at 36 and 45 cm\(^{-1}\). Predicted frequencies for bending in the base plane and bending in the sugar plane are 50 and 61 cm\(^{-1}\) and 57 and 44 cm\(^{-1}\). Both of these groups of predicted modes overestimate the frequency assuming the observed modes at 33 and 44 cm\(^{-1}\) originate from the bending of these two planes. Strong modes are observed at 62, 74, and 86 cm\(^{-1}\) with no corresponding predicted modes. The observed mode at 62 cm\(^{-1}\) matches well with the predicted mode at 61 cm\(^{-1}\). This could indicate the presence of two sugar plane modes at 45 and 62 cm\(^{-1}\) originating from the two, differently puckered sugars. These are followed by modes at 103, 111, 114, 124, 127, 134, and 148, some of which have well-matching predicted modes (95, 105, 113, 124, and 149). From 150 cm\(^{-1}\), 3 predicted modes exist, but no strong modes are observed. From 60-150 cm\(^{-1}\), deoxycytidine has several broad peaks or ones being obscured by other peaks. This could be due to the presence of different sugar puckers within the deoxycytidine crystal.

Cytidine has 56 distinct observed modes from 20-1700 cm\(^{-1}\), with 10 being below 200 cm\(^{-1}\). The lowest mode is observed at 31 cm\(^{-1}\), similar to deoxycytidine mode at 30 cm\(^{-1}\). The observed mode at 38 cm\(^{-1}\) corresponds well with the deoxycytidine’s mode at 33 cm\(^{-1}\) due to its similarity in relative position. I propose that this mode originates from bending in the base plan in both cytidine and deoxycytidine due to their proximity in frequency, their similarity in relative position, and existence of identical bases. The next highest mode in cytidine is observed at 50 cm\(^{-1}\), which is higher than deoxycytidine’s third highest mode at 44 cm\(^{-1}\). I propose that due to their similarity in relative position and intensity, this mode originates from bending in the sugar plane. A shoulder is observed at 57 cm\(^{-1}\), followed by another at 70 attached to a peak at 74 cm\(^{-1}\). Several
strong peaks are observed at 85, 102 and 110 cm\(^{-1}\) followed by a weak peak at 155 cm\(^{-1}\). Similarities exist between the spectra of deoxycytidine and cytidine from 70-120 cm\(^{-1}\), notable at 71, 84, and 108 cm\(^{-1}\). This suggests similar origin of modes, for example from bonds in the bases.

### 3.2 Comparisons Between Groups of Nucleosides

Comparisons between nucleosides containing the same bases have been discussed in the above section. In general, the lowest frequency modes match well between like-base nucleosides. In adenine- and cytidine-based nucleosides, the lowest frequency mode matches between the deoxy- and ribonucleosides. The pair that had different bases, deoxythymidine and uridine, had different low frequency modes. This suggests that the bases greatly affect the lowest frequency mode.

Same-base nucleosides also had similarities in frequency and intensity for the second highest mode. This observed data supports theoretical predictions. Shishkin et al. predicted for most nucleosides that this mode originates from bending in the base plane. This suggests that this mode is mainly unaffected by composition of the sugar unit.

Similarities exist in like-based nucleosides above the base mode peak, however not all modes correspond to their different-sugared counterpart. Modes from 40-100 cm\(^{-1}\) have been attributed to base interactions (Urabe, or at least cited in Urabe). Modes which are affected by base composition or similar sugar composition (not affected by the presence or lack of oxygen at 2\(^{'}\)) are present (i.e. 100 cm\(^{-1}\) in the adenine-based nucleosides, 84 cm\(^{-1}\) in the cytidine-based nucleosides). Some modes are clearly affected
by differences in sugar composition, base stacking, or other low-frequency interactions, which is apparent in the differences in this region between like-based nucleosides.

Comparison between deoxynucleosides and nucleosides lack substantial similarities. The ribose-sugared nucleosides each have a broader peak or peak with a shoulder above the mode which originates from bending the base plane each near the same frequency (56 cm\(^{-1}\)). Deoxynucleosides have distinct, singular peaks above the base plane mode. The lowest frequency mode is unaffected by the composition of the sugar in the nucleoside.

The purine nucleosides (adenosine, deoxyadenosine) had the lowest frequency mode at higher frequencies when compared to the pyrimidines. This occurs despite the existence of more mass in the purines than the pyrimidines, which should mean lower frequency modes. This suggests that mass is not the main contributing factor in the frequency of the lowest mode. Purines should have more modes than pyrimidines due to the larger number of atoms in purines when compared to pyrimidines. This is not found in the low-frequency region. Modes involving individual atoms will be at higher frequencies due to their stronger bonding and lower mass.

### 3.3 Theoretical Predictions

Theoretical predictions and assignments have been compared. Predictions show different levels of success in each nucleoside, which have been discussed above. In general, many predicted modes are with 10 cm\(^{-1}\) of an observed peak and have similar pattern to the observed spectra. There is a lack of predicted modes from \(~60-100\) cm\(^{-1}\) where many peaks are observed. This could be a consequence of the predictions being
for single molecules of the nucleosides. However, modes in this region have been predicted to originate from bonds in the bases and not intermolecular interactions. At frequencies above 100 cm$^{-1}$, more modes are predicted than are observed. This could be due to overlapping observed modes or overestimated predicted modes.

3.4 Basestacking Character of Each Nucleoside

Column structure of stacked bases has been proposed as the origin of the S mode with intra- and intercolumn interactions affecting its frequency$^{30}$. Intermolecular structures have been shown in Figure 7 for the six nucleosides being studied. Analysis of the intermolecular structure and characteristics can be used to test the effect of base stacking, column structure, and other characteristics on the S-mode.

Urabe et al. proposed that bases stacking to form a column is essential to the S mode. Five of the nucleosides in this study have bases stacked to form columns. Two of these, thymidine and cytidine, form straight columns while the other three, uridine, adenosine, and deoxyadenosine, form tilted columns with stacked bases moving perpendicular to the direction of stacking, making the bases not stack directly on top of one another. The last nucleoside, deoxycytidine, does not have bases stacked in columns. Rather, only pairs of bases stack without a clear column forming. Contrary to the proposal, all of these nucleosides have a low-frequency mode with characteristics of the S mode despite the lack of base stacked columns in deoxycytidine.

It was also suggested that hydrogen bonds do not affect the S mode. The data on these six nucleosides supports that claim. Adenosine and deoxyadenosine have a similar S mode, but do not have similar hydrogen bonds with neighboring molecules. The same
is true for cytidine and deoxycytidine. Although their S modes are similar, the similarities in the hydrogen bonding are limited.

It has been shown previously that covalent bonds from the phosphate group are unnecessary for the S mode. This is clearly true as all six of these nucleosides have a low-frequency mode characteristic of the S mode without having phosphate groups.

![Is/Iave vs. Nucleosides](image)

Figure 4-3. Graph of intensity of S mode (Is) to intensity of rest of peaks (Iave) lower than 200 cm\(^{-1}\) for six nucleosides. Higher values are theorized to have a higher degree of base stacking. Error in these measurements were less than 1%.

The intensity of the S mode has been proposed to be determined by the stacking degree of the bases. Figure 4-3 shows the ratio of the intensity of the lowest-frequency mode to the rest of the modes below 200 cm\(^{-1}\) for the six nucleosides. High values indicate high intensity S-modes and are predicted to have good base stacking.

Deoxycytidine is poorly stacked and lacks columns of bases. It has a low intensity S mode relative to the rest of its spectrum with an I\(_s\)/I\(_{ave}\) value of 1.408. Thymidine and cytidine are well stacked and have strong relative peaks with I\(_s\)/I\(_{ave}\) values of 1.881 and 2.273, respectively. Deoxyadenosine is less well stacked than thymidine and cytidine.
and has the lowest ratio of 1.198. This would support the proposition that S mode intensity is affected by stacking degree. However, this is not conclusively clear. Deoxyadenosine has similar stacking to adenosine but has a more intense peak. Thymidine is well stacked and has a relatively strong peak, however it has a more intense peak at 78 cm$^{-1}$.

The frequency of the S mode has been proposed to be determined by intra- and intercolumn interactions. Urabe et al. contend that the S mode is not entirely dependent on base stacking. This appears to be true as there is different base stacking but same frequency S modes in same-based nucleosides (deoxyadenosine, adenosine and deoxycytidine, cytidine). It is unclear to us which intra- and intercolumn interactions are important. Base composition strongly affects the frequency of the S mode. However differences in same-based nucleosides such as base-stacking, orientation of base to sugar, crystal structure, molecular mass, and angle of bases with respect to stacking direction appear to be ruled out as key contributors in S mode frequency, without there being any strong pattern in these characteristics to the S mode frequency. Hydrogen bonding from bases may be a key contributor to the frequency of the S mode. Differences in the specific hydrogen bonds leaves some doubt, but could explain the 1-2 cm$^{-1}$ difference in S mode frequency in same-based nucleosides although these differences could be attributed to error.
Chapter 4

Conclusions

Raman spectroscopic data were taken for six nucleosides from 20-1700 cm\(^{-1}\). This data allowed for analysis of the low-frequency modes which are involved in collective motions of the nucleosides. The data were compared to theoretical predictions, which was found to match well with some observed modes but had regions with observed modes with no predictions. This data can be used in future work by theoreticians to refine models of nucleosides. The data were also examined to analyze the lowest frequency mode dubbed the S-mode. Column structure due to base-stacking was found not to be necessary for the S-mode, but was found to increase the intensity of the S-mode peak. Base composition was found to affect the frequency of the S mode, while other characteristics such as base orientation, crystal structure, molecular mass, were found not to greatly affect the S mode.
References

References

11. Jmol: an open-source Java viewer for chemical structures in 3D.