I, Kemberly G. Patteson, hereby submit this as part of the requirements for the degree of:

Doctor of Philosophy

in:

Chemistry

It is entitled:

Analysis of the Mass Silent Post-transcriptional Modification Pseudouridine in RNA by Mass Spectrometry

Approved by:

Dr. Patrick A. Limbach

Dr. Apryll M. Stalcup

Dr. Albert M. Bobst
Analysis of the Mass Silent Post-transcriptional Modification Pseudouridine in RNA by Mass Spectrometry

A dissertation submitted to the
Division of Research and Advanced Studies
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of
DOCTORATE OF PHILOSOPHY (Ph.D.)
in the Department of Chemistry
of the College of Arts and Sciences
2003
by
Kemberly G. Patteson
B.S., Central Washington University, 1997

Committee Chair: Patrick A. Limbach Ph.D
ABSTRACT

Pseudouridine has been of considerable interest in the last decade for several reasons. Of all post-transcriptional modifications found in RNA, it is the most common modification found. Because it is a structural isomer of uridine, it has the potential for more hydrogen bonding than uridine. It is found in conserved regions of RNA, therefore it may play some role in the translation process or in the ability of the ribosome to take its final, tertiary structure.

Although pseudouridine has been known for quite some time, its sequence location in RNA has not been studied until recently. This is because, as a structural isomer of uridine, its analytical properties including UV absorbance, electrophoretic mobility, interaction with stationary phases and mass are similar to those of uridine. It is extremely difficult to place pseudouridine in an oligonucleotide sequence when, for all intents and purposes, it behaves as uridine.

Mass spectrometric techniques alone are unable to distinguish pseudouridine. As it has the same mass as uridine, pseudouridine is a “mass-silent” modification. An interest of this lab has been to develop analytical methods to characterize pseudouridine. N-cyclohexyl-N’-β-(4-methylmorpholinium)ethylcarbodiimide selectively derivatizes pseudouridine. The focus of this work is to use this selective derivatization to develop an analytical method that can determine the location of pseudouridine within an RNA, no matter the size of the RNA. This method was initially tested using E. coli. RNAs, which have been sequenced, including modifications, so it is a good organism to validate the techniques developed. Initial studies focused on tRNA, which is the smallest of all RNAs, then moved to 16S rRNA, and then finally to 23S rRNA.
Copyright notice
ACKNOWLEDGEMENTS

I would like to provide a great deal of thanks for the patience and understanding that my major professor, Dr. Patrick A. Limbach, has shown me over the past few years. Despite some rather unusual, and difficult, situations that have presented themselves during my time in graduate school; he has always provided me with wonderful advice and guidance, although at times I seemed too stubborn to listen to it. I would also like to thank both my committee members at LSU: Dr. Stephen Watkins, and Dr. Robert Gale; and my UC committee members: Dr. Apryl Stalcup, and Dr. Albert Bobst, for their help and guidance throughout the years. Special thanks go to Dr. William Daly for his continued support and presence on my committee and to him and his wife for support during difficult times. And a very special thanks to Dr. Thomas Ridgway for all he has done for me.

I would also like to thank both the Departments of Chemistry at Louisiana State University and University of Cincinnati for their financial support and for the valuable teaching experience.

I would like to thank the following for their personal support and guidance, both in my research, and in my life: Sean Troxclair, Garrett Doucet, Gary Lyon, Patricia Anne Beck, Glenna Sowa, Michele Haun, and Shirish Deorukhkar. I would also like to acknowledge all of my research group members and fellow graduate students but especially: Stephen Macha, Beniam Berhane, Rama Tummala, Chad Robbins, Justin Mecomber and David Monk. To all of those who have befriended me but are not mentioned here, you have my thanks.
Finally I would like to dedicate this work to Dr. Timothy Sean Evenson, for all of his love and patience. I wish he could have been here to see this day. All my love goes to my family and puppies, without them, I could never have accomplished so much, and I am forever grateful for them in my life.
# Table of Contents

**Abstract** .......................................................................................................................... ii  
**Acknowledgements** ....................................................................................................... iv  
**List of Figures** ................................................................................................................ 4  
**List of Tables** .................................................................................................................. 7  
**List of Schemes** ............................................................................................................... 8  
**List of Abbreviations** .................................................................................................... 9  
**Chapter 1: Pseudouridine Introduction** ........................................................................ 11  
  1.1 Molecular Biology ....................................................................................................... 11  
  1.2 Mass Spectrometry ..................................................................................................... 15  
  1.3 Previous Studies of Pseudouridine ............................................................................. 20  
**Chapter 2: Initial Pseudouridine Studies** ................................................................. 24  
  2.1 Introduction ............................................................................................................... 24  
  2.2 Methods and Materials ............................................................................................. 25  
  2.3 Results ..................................................................................................................... 26  
  2.4 Conclusions.............................................................................................................. 28  
**Chapter 3: Development of a New Pseudouridine Identification Method on tRNA from *E. coli*** ................................................................. 30  
  3.1 Introduction ............................................................................................................... 30  
  3.2 Materials and Methods ............................................................................................. 31  
  3.3 Results ..................................................................................................................... 34  
**Chapter 4: Reactions of CMC with dT_{10}** ................................................................. 48  
  4.1 Introduction ............................................................................................................... 48
4.2 METHODS AND MATERIALS ........................................................................................................... 50
4.3 RESULTS ......................................................................................................................................... 53
4.4 CONCLUSIONS ................................................................................................................................. 62

CHAPTER 5: CULTURING OF E. COLI MRE-600 .................................................................................. 64

5.1. INTRODUCTION ......................................................................................................................... 64
5.2 METHODS AND MATERIALS ....................................................................................................... 66
5.3 RESULTS ......................................................................................................................................... 70

CHAPTER 6: APPLICATION OF A METHOD TO ANALYZE PSEUDOURIDINE IN 16S RRNA FROM E. COLI ........................................................................................................... 76

6.1 INTRODUCTION ........................................................................................................................... 76
6.2 METHODS AND MATERIALS ......................................................................................................... 78
6.3 RESULTS ......................................................................................................................................... 79

CHAPTER 7: DEVELOPMENT OF NEW RNA CLEAVAGE TECHNIQUES USING RNASE H ................................................................................................................................. 84

7.1 INTRODUCTION .......................................................................................................................... 84
7.2 MATERIALS AND METHODS ....................................................................................................... 85
7.3 RESULTS ......................................................................................................................................... 87
7.4 CONCLUSIONS ............................................................................................................................... 92

CHAPTER 8: ANALYSIS OF PSEUDOURIDINE IN 23S RRNA FROM E. COLI .............................................................. 93

8.1. INTRODUCTION ........................................................................................................................ 93
8.2. METHODS AND MATERIALS ..................................................................................................... 96
8.3. RESULTS ....................................................................................................................................... 98
8.4. CONCLUSIONS ........................................................................................................................... 107

CHAPTER 9: CONCLUSIONS ........................................................................................................... 109
9.1 Final Conclusions to Pseudouridine Project .................................................. 109

9.2 Future Work ....................................................................................................... 110

BIBLIOGRAPHY ......................................................................................................... 114
LIST OF FIGURES

Figure 1.1. The components of the prokaryotic ribosome include 23S, 16S, 5S rRNA and ribosomal proteins. .......................................................... 13

Figure 1.2. Structures of uridine and pseudouridine. ............................................. 14

Figure 1.3. General illustration of the fundamentals of matrix-assisted laser/desorption ionization time-of-flight mass spectrometry [27]. ................................. 16

Figure 1.4. General illustrations of: A) fundamentals of an electrospray ionization (ESI) source and B) charged droplet formation [28]. .................................................. 18

Figure 1.5. Illustration of a time-of-flight mass analyzer. ........................................ 18

Figure 2.1. Separation of the nucleoside test mix by HPLC........................................ 27

Figure 2.2. Mass spectra of 50 ppm adenosine using atmospheric pressure chemical ionization (APCI) showing (M+H)+ and (B+2H)+ ions. ........................................... 28

Figure 3.1. Comparison of: A) tRNAVal (m/z 24705) and B) tRNAVal after reaction with CMC (m/z 24950) providing a difference of 250 Da, showing the addition of 1 CMC. ......................................................................................................................... 35

Figure 3.2. Comparison of the RNase T1 digestion of tRNA^Phe for the A) control and B) CMC-reacted species.................................................................................. 39

Figure 3.3. HPLC separation of RNase T1 digestion products for A) the control and B) CMC-reacted species................................................................. 41

Figure 3.4. UV-Vis analysis of control and CMC modified oligonucleotides. ........... 42

Figure 3.5. MALDI-MS spectra of the HPLC fractions eluted at A) 13.4, B) 16.4 and C) 30.0 minutes. .............................................................. 43

Figure 3.6. Snake-venom phosphodiesterase digestion from the 30.0 minute HPLC fraction analyzed by MALDI-MS. ......................................................... 46

Figure 4.1. Theoretical structures of the CMC-nucleoside bond. A) Suggested by Ofengand, and B) suggested by protein literature................................. 49

Figure 4.2. Comparison of dT10 A) before and B) after reaction with CMC under initial conditions. ........................................................................ 53
Figure 4.3. Effects of CMC concentration on the CMC reaction. A) 50 µg/µL CMC, B) 100 µg/µL, C) 500 µg/µL

Figure 4.4. Effects of buffer concentration on the CMC reaction. A) 50 mM Tris, B) 200 mM Tris

Figure 4.5. Effects of reaction temperature. A) 37 °C, B) 55 °C

Figure 4.6. HPLC separation of dpT and dpT+CMC

Figure 4.7. UV-Vis comparison of A) dpT-CMC reaction and B) control, after separation with chloroform

Figure 4.8. 10-mer containing Ψ A) after addition reaction of CMC. B) After addition of NH₄OH and reaction for 4 hours with no cleaning in between

Figure 5.1. E. coli growth curve showing the UV absorbance at 600 nm over time

Figure 5.2. UV-Vis calibration curve at 260 nm for RNA

Figure 5.3. Sucrose gradient separation of 23S, 16S, and 5S rRNA from E. coli

Figure 6.1. 16S rRNA and the 510 loop containing pseudouridine from 16S rRNA from E. coli

Figure 6.2. Anion-exchange separation of RNase T₁ digestion of A) Control and B) CMC modified fragments from 16S rRNA

Figure 6.3. Reversed-phase HPLC separation of: A) anion-exchange peak I. Peak labeled A is suspected to be the CMC peak. B) Anion exchange peak II. Peak labeled B is suspected ΨGp+CMC fragment

Figure 6.4. Comparison of CMC RP-HPLC fraction B A) with matrix and B) control fraction B

Figure 7.1. PAGE separation of RNase H cleavage products. Lanes 1-4 and 6-9 are cleavage products D-I from 23S rRNA of E. coli. Lane 5 is RNA size marker from 100-300 nt in length

Figure 7.2. HPLC separation of RNase T₁ digestion of Fragment H from the gel. Peak labeled with (*) at 17.4 minutes is suspected to have several oligonucleotide fragments

Figure 7.3. MALDI-MS spectrum of 17.4 minute peak from the HPLC separation. Peaks labeled with (*) correspond to RNase T₁ fragments of the 136-mer
Figure 7.4. UV scan comparing a viable and non-viable RNA extracted from gel... 91

Figure 8.1. Comparison of RNase H products before and after PAGE separation... 98

Figure 8.2. UV comparison scans between the control and the CMC modified 136-mer. 99

Figure 8.3. UV scan comparison between the 114 control sample and CMC sample... 101

Figure 8.4. HPLC chromatograms from control and CMC species from the 114 band. Peaks labeled with (*) indicate UV increases between control and CMC samples. 101

Figure 8.5. UV scan comparison of the 143-mer showing a tenfold increase between control and CMC samples. ................................................................. 102

Figure 8.6. HPLC chromatogram from the 146-mer. Peak labeled (*) at 14.0 minutes in the CMC sample shows an eight-fold increase in absorbance. .................. 103

Figure 8.7. UV scan comparison between the control and CMC reaction for the 179-mer. ..................................................................................................... 104

Figure 8.8. HPLC separation of the 179-mer. Peak labeled (*) at 24.0 minutes shows a large increase between the CMC and control samples................................. 105

Figure 8.9. UV comparison scan between the control and CMC of the 207-mer showing a four-fold increase at 260 nm.............................................................. 106

Figure 8.10. HPLC separation of RNase U2 products from the 207-mer CMC and control reaction. The peak at 22.0 minutes labeled with (*) denotes the fragment suspected to have CMC attachment................................................................. 106

Figure 8.11. UV scan comparison between the control and CMC reaction from the 235-mer................................................................. 107

Figure 9.1. Flow chart describing the steps needed to fully characterize pseudouridine in any size of RNA................................................................. 110
LIST OF TABLES

Table 2.1. Names, common symbols, concentrations and retention times of nucleosides found in the nucleoside test mix used to optimize the HPLC methodology. ............ 26

Table 2.2. Nucleoside names, symbols, molecular and base fragment masses and detected amounts. ....................................................................................................................... 29

Table 3.1. Predicted and experimentally determined masses from the RNase T₁ fragments generated from tRNA^{Phe} . ........................................................................................................ 38

Table 3.2. Predicted and experimentally determined masses from the three pseudouridine-containing RNase T₁ fragments after digestion with snake venom phosphodiesterase. ................................................................................................. 44

Table 4.1. Conditions for the reaction of CMC with dT₁₀ that may affect the number of CMC adducts that can be added to the oligomer. ................................................................. 49

Table 4.2. List of reactions I-VIII, and their reaction variables. ........................................... 51

Table 4.3. Number of CMC adducts observed with reactions I-VIII. ............................... 55

Table 5.1. Cultured cellular yield of E. coli MRE-600................................................................. 71

Table 5.2. Calculated amount of total RNA generated from culturing by UV-Vis analysis. ....................................................................................................................... 72

Table 7.1. Sequence, melting temperature, % guanine/cytidine composition and molecular weight information for the selected DNA oligonucleotides......................... 88

Table 7.2. RNA lengths generated from DNA/RNA hybrids and where in 23S rRNA sequence they occur. ........................................................................................................... 88

Table 7.3. RNase T₁ fragments generated from the 134-mer and their expected and experimental m/z values. ................................................................................................. 90

Table 8.1. Predicted RNase T₁ and U₂ fragments containing pseudouridine, molecular weights before and after CMC addition. ................................................................. 100

Table 8.2. Comparing the increase between UV absorbance due to scan of the intact oligomers vs. the absorbance seen due to HPLC separation of fragments. ............ 108
LIST OF SCHEMES

Scheme 1.1. Proposed reaction mechanism for the: A) addition and
B) removal of CMC from guanine, thymine and uracil………………………….20
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ψ</td>
<td>pseudouridine</td>
</tr>
<tr>
<td>β</td>
<td>beta</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>μg</td>
<td>microgram</td>
</tr>
<tr>
<td>μL</td>
<td>microliter</td>
</tr>
<tr>
<td>Δm</td>
<td>change in mass</td>
</tr>
<tr>
<td>μm</td>
<td>micrometer</td>
</tr>
<tr>
<td>µM</td>
<td>micromolar</td>
</tr>
<tr>
<td>2,4,6 THAP</td>
<td>2,4,6 trihydroxyacetophenone</td>
</tr>
<tr>
<td>A</td>
<td>adenosine</td>
</tr>
<tr>
<td>ACN</td>
<td>acetonitrile</td>
</tr>
<tr>
<td>AE-HPLC</td>
<td>anion exchange HPLC</td>
</tr>
<tr>
<td>APCI</td>
<td>Atmospheric pressure chemical ionization</td>
</tr>
<tr>
<td>AU</td>
<td>absorbance units</td>
</tr>
<tr>
<td>BH₂⁺</td>
<td>protonated base fragment ion</td>
</tr>
<tr>
<td>cDNA</td>
<td>complimentary DNA</td>
</tr>
<tr>
<td>CI</td>
<td>chemical ionization</td>
</tr>
<tr>
<td>CMC</td>
<td>N-cyclohexyl-N’-β-(4-methylmorpholinium)ethylcarbodiimide</td>
</tr>
<tr>
<td>Da</td>
<td>Daltons</td>
</tr>
<tr>
<td>DAC</td>
<td>Diammonium citrate</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EI</td>
<td>Electron impact</td>
</tr>
<tr>
<td>ESI</td>
<td>fast atom bombardment</td>
</tr>
<tr>
<td>FAB</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>FTICR</td>
<td>Fourier-Transform Ion Cyclotron Resonance</td>
</tr>
<tr>
<td>G</td>
<td>guanine</td>
</tr>
<tr>
<td>GC-MS</td>
<td>gas chromatography-mass spectrometry</td>
</tr>
<tr>
<td>HPLC</td>
<td>High Performance Liquid Chromatography</td>
</tr>
<tr>
<td>ICP</td>
<td>inductively coupled plasma</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>K</td>
<td>Degrees Kelvin</td>
</tr>
<tr>
<td>KDa</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>kV</td>
<td>kilo volts</td>
</tr>
<tr>
<td>LC-MS</td>
<td>Liquid Chromatography-Mass Spectrometry</td>
</tr>
<tr>
<td>LD</td>
<td>laser desorption</td>
</tr>
<tr>
<td>M</td>
<td>mass</td>
</tr>
<tr>
<td>M</td>
<td>molarity</td>
</tr>
<tr>
<td>m/z</td>
<td>mass-to-charge ratio</td>
</tr>
<tr>
<td>M¹acp³Ψ</td>
<td>1-methyl 3-(3-amino-3-carboxypropyl)pseudouridine</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix-assisted laser desorption/ionization</td>
</tr>
<tr>
<td>MH⁺</td>
<td>protonated molecular ion</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>mL</td>
<td>milliliters</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem mass spectrometry</td>
</tr>
<tr>
<td>nm</td>
<td>nanometers</td>
</tr>
<tr>
<td>nmol</td>
<td>nanomole</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotides</td>
</tr>
<tr>
<td>ODU</td>
<td>optical density units</td>
</tr>
<tr>
<td>PD</td>
<td>plasma desorption</td>
</tr>
<tr>
<td>PMT</td>
<td>photomultiplier tube</td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>R</td>
<td>resolution</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>RP-HPLC</td>
<td>reversed phase HPLC</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>snRNA</td>
<td>small nuclear RNA</td>
</tr>
<tr>
<td>SVP</td>
<td>snake venom phosphodiesterase</td>
</tr>
<tr>
<td>T</td>
<td>thymidine</td>
</tr>
<tr>
<td>TOF</td>
<td>time-of-flight</td>
</tr>
<tr>
<td>TRIS</td>
<td>2-Amino-2-(hydroxymethyl)-1,3-propanediol</td>
</tr>
<tr>
<td>tRNA\textsuperscript{\textit{Phe}}</td>
<td>Transfer RNA, phenylalanine specific</td>
</tr>
<tr>
<td>tRNA\textsuperscript{\textit{Val}}</td>
<td>transfer RNA, valine specific</td>
</tr>
<tr>
<td>U</td>
<td>uridine</td>
</tr>
<tr>
<td>UV-VIS</td>
<td>Ultraviolet/Visible</td>
</tr>
</tbody>
</table>
CHAPTER 1: Pseudouridine introduction

1.1 Molecular Biology

Since the beginning of the 20th century, advances in medicine, molecular biology, biochemistry, and chemistry have developed to such a point where we, as scientists, are now able to start looking into the complex biochemical mechanisms of the human body. The variety of biomedical, biological and chemical topics introduced in the past century are nearly limitless, and as of now, very few of these topics have been explored to the point where we have a working understanding of what happens in these systems. Medical researchers tend to focus their research in the area of diseases, trying to cure them, as they become apparent in the patient. Molecular biologists look at the human cell and try to understand the structures and functions that go on in the cell. Biochemists tend to look at the chemical reactions that go on in the body, not limited to the cell, to provide an understanding of how molecules interact with each other, and how introducing one chemical causes a cascade of reactions in the body. Chemists, more specifically analytical chemists, look for new techniques to analyze the processes that interest medical, biological, and biochemical researchers.

Previously the boundaries between these fields were clear-cut, with researchers having a limited understanding of the background and techniques used in the other fields. However, a recent surge in research has removed those barriers and brought forth a new generation of researchers who are now able to look at a general topic, and all of its ramifications. In the 1990’s, the human genome became a topic of significant research as
researchers from many disciplines came together to sequence it. In about a decade, they completed a task that was first projected to take several decades, thanks to this merging and cooperation between disciplines. Many new topics still need to be addressed now that the genome has been sequenced. We still need to identify protein sequences, for instance, and determine the function of the ribosome.

Our group has decided to pursue the goal of ribosome characterization. We are not confining ourselves to single topics such as how to identify RNA sequences [1-4], or how RNA combines to form the ribosome [5-14], or to determining how mRNA binds in the ribosome [7], or how amino acids join together to form proteins [15-17]. The long-term goal of this group is to answer as many of these questions as possible.

Obviously all of these questions could not even begin to be answered by an individual researcher, so many must come together to achieve the overall goal. To this end, individual researchers must identify specific topics that are feasible for them. In order to look at functions inside the ribosome, we must know what the components of the ribosome are on a molecular scale. As of now, it is known that the ribosome is made up of three ribosomal RNAs which make up a large and small subunit (Figure 1.1).

The sizes of these subunits depend upon the type of organism. The remainder of this discussion will be focusing on the bacterium Escherichia coli, which, at this point, is the only organism to have its entire DNA, RNA, and proteins sequenced [18-21].

While E. coli may be fully sequenced, it was no simple chore. Traditional methods have been slow and laborious, and they have required large amounts of sample in order to yield viable sequence information. Additional challenges have made RNA
particularly difficult to analyze. This work deals specifically with the area of post transcriptional modifications [21-25], those that are purposefully added to an RNA sequence after the transcription of RNA from DNA. These modifications are very common; approximately 5% of all nucleoside bases in RNA are modified. Various types of modifications occur including methylation, thiolation, and isomerization. To date, approximately 250 different modifications have been characterized. Different separation methods, and most recently mass differences, can identify most of these modifications. Early identification was done with gels [26]. While gels are still used, new techniques have been developed to aid in detecting modified nucleotides, most notably the separation technique high performance liquid chromatography (HPLC), and the detection technique mass spectrometry. However, one post-transcriptional modification cannot be easily differentiated from other nucleosides. This nucleotide, pseudouridine, is a structural isomer of the major nucleoside, uridine.
On the surface, the differences between uridine and pseudouridine seem trivial. However, there are several characteristics of pseudouridine that make it unique from all other nucleosides. As shown in Figure 1.2, the difference between uridine and pseudouridine involves rotation of the cyclic ring from an N-1 to C-5 attachment to the ribose sugar. While such a rotation does not seem significantly different, it produces a carbon-carbon glycosidic bond, which tends to be more stable than a nitrogen-carbon glycosidic bond, and two nitrogens with labile protons are on the cyclic ring, which theoretically, allow for more hydrogen bonding. At this time, it is not known what this extra bond may be used for, but this modification is clearly important. Out of the 5% of all nucleosides that are modified, pseudouridine is the most common one. Yet ironically, this significant modification is nearly undetectable by current characterization methods.
The main focus of my work has been to develop a fast and accurate method for determining the sequence locations of this modification in RNA sequences.

1.2 Mass spectrometry

Mass spectrometry has provided biological analysis a very powerful tool. The amounts of information that can be generated from large biomolecules has grown with each passing year. Mass spectrometry has been around for quite some time; however, Hillenkamp did not discover matrix-assisted laser desorption ionization (MALDI) until 1985, and Fenn did not develop electrospray ionization (ESI) until 1988. While this may not seem significant at the outset, it is important to note that until the 1980s, it was not possible to generate molecular ions from large biomolecules. Previously, when electron impact and chemical ionization techniques were the only ones that were easy to implement, molecules either had to be volatile or chemically modified to be made volatile. Until thermospray, fast atom bombardment, MALDI, and ESI, ionization techniques were not capable of vaporizing large biomolecules into the gas phase. For the purpose of this discussion, only MALDI-TOF and ESI-FTICR mass spectrometry will be discussed in detail.

When choosing between the use of MALDI-TOF and ESI-FTICR, several considerations have to be taken into account. First is the polarity of the analyte that determines the ease by which the analyte becomes an ion. MALDI, due to the matrix involved, can handle a wide range of molecules. On the other hand, ESI requires that the molecules are pre-charged in solution, so it is necessary to have highly polar molecules. The second consideration is the complexity of the analyte mixture. If there are a large number of analytes in solution that must be detected, MALDI-TOF is capable of
producing a spectrum that contains only single peaks for each analyte, and therefore spectral interpretation is not as complicated. However, because the ionization efficiency in MALDI-TOF is not the same for all molecules, ESI-FTICR is more adaptable to quantitative analysis. A third consideration when choosing between MALDI-TOF and ESI-FTICR is the amount of available sample. Because biological samples can be difficult to isolate and purify, it is sometimes necessary to analyze picomole and femtomole amounts of material. Because ESI-FTICR can be scaled down to micro- or even nano-ESI-FTICR, it has the ability to analyze molecules that are in that range. MALDI and ESI are both theoretically unlimited as to the molecular weights of compounds that they can ionize, however practical limitations are between 100 and 200 kDa. MALDI is more tolerant of salts and biological buffers than ESI is.

Figure 1.3. General illustration of the fundamentals of matrix-assisted laser/desorption ionization time-of-flight mass spectrometry [27].

MALDI-TOF analysis is done by mixing the sample with a matrix, which is usually a small, polar organic acid that absorbs light at the wavelength of the laser that is used. The matrix must be soluble in a common solvent as the sample to insure that the
interaction between matrix and sample is optimum after the solvent evaporates. After evaporation, there should be homogenous crystals on the surface of the MALDI plate. A diagram of MALDI-TOF MS is shown in Figure 1.3.

ESI-MS, shown in Figure 1.4, is a solution-based technique. ESI is capable of creating highly charged analytes, so that a 10000 Da protein may give peaks from $m/z$ 5000.5 down to 1001.0, equating to charge states $2^+$ to $10^+$. The advantage of this is that large biomolecules can be readily analyzed with mass analyzers that have limited $m/z$ ranges; however the drawback is that the spectra become more complicated, as the same analyte may have many different yet related peaks.

There are many different types of mass analyzers, and the choice of mass analyzer depends upon several factors such as choice of ionization method, mass range needed, mass accuracy required, and the resolution needed. However, as stated earlier, I will be focusing exclusively on time-of-flight (TOF) and Fourier transform ion cyclotron resonance (FTICR) mass analyzers. The purpose of a mass analyzer is to separate ions based upon mass and focus them onto the detectors.

A TOF mass analyzer is among the least complicated type of mass analyzer available. As shown in Figure 1.5, ions are accelerated from the source region with the same kinetic energy into a field-free region. In this field-free region, analytes are separated based upon their flight time, with heavier analytes moving slower than lighter analytes. Despite the simplicity in the design of this mass analyzer, there are some inherent problems which have been overcome by recent improvements. In the design of the analyzer, it is assumed that all ions will be given the same initial kinetic energy.
Figure 1.4. General illustrations of: A) fundamentals of an electrospray ionization (ESI) source and B) charged droplet formation [28].

However, depending on the initial position of the ions in the source region, this may not be the case. Ions closer to the acceleration plate may be given less energy than the ones farther away, so ions with the same mass may not move at exactly the same speed, reaching the detector with a larger time distribution than would be expected. This effect

Figure 1.5. Illustration of a time-of-flight mass analyzer.
causes a loss in resolution; however, this problem can now be corrected by the addition of a reflectron and a second flight tube. At the end of the first flight tube, ions with more kinetic energy can pass farther into the set of voltage plates that make up the reflectron. Ions of the same mass, but with less kinetic energy, cannot penetrate so deeply; so that when both ions are accelerated towards the detector, they will arrive at the same time, thereby narrowing the packet of ions. This type of mass analyzer is ideally suited for MALDI ionization, as both are pulsed techniques. However, while ESI can be used with TOF, there is another mass analyzer which is more ideally suited for ESI analysis.

FTICR mass analyzers require a large magnetic field. An ion’s radial motion is confined by the magnetic field provided its mass is not too high; this is the basis of ion cyclotron resonance. The smaller the velocity of the ion, the smaller the radius of the circle that ion will follow. After resonant excitation, the radius of the ion will increase. Because an ion trap is used to confine the ions within the magnetic field, if the radius of an ion becomes too great, it will eventually crash against the side of the trap.

Measurements of mass with this type of system are based upon the frequency of the ions in their circular trajectories. Typically when using an FTICR instrument, all ions are simultaneously excited within the ion trap and a large frequency range is detected. The ions are excited so that their trajectory is close to the wall of the trap, where they are held for a time, and the instrument can then transform a time-dependent function into a frequency-dependent function. Like any technique that depends upon the Fourier transform, the resolution depends upon the observation time. In the case of FTICR, the measurement time is based upon the relaxation time of the ions, as they go from their excited orbits down to their smallest orbits. Practical limits of this type of analysis allow
for spectra to be taken at the interval of one spectra per second, which is compatible with most separation techniques. For low \( m/z \) ions, resolutions exceeding one million can be achieved. This makes FTICR mass analyzers ideal for coupling with ESI sources, as the multiple charging effect from the ESI would bring large mass ions down to an \( m/z \) range where the resolution is optimum.

1.3 Previous Studies of Pseudouridine

There have been previous studies that have focused on pseudouridine. In the early 1970’s, Ho & Gillham explored the reaction of pseudouridine with N-cyclohexyl-N’-β-(4-methylmorpholinium)ethylcarbodiimide (CMC) [29]. They observed that CMC was able to form a bond with the nucleosides thymidine, uracil, guanine, and pseudouridine. The reaction that adds CMC to nucleosides is outlined in Scheme 1.1a, while the reaction that removes CMC from guanine, thymine, and uracil is shown in Scheme 1.1b.

![Scheme 1.1. Proposed reaction mechanism for the: A) addition and B) removal of CMC from guanine, thymine and uracil.](image)

In 1971, Ho & Gilham did initial studies of pseudouridine that focused on the interaction of CMC with various single nucleotides. Initially they saw that the addition
of CMC to guanine, thymine and uracil was reversible upon changes in pH, while adenine and cytosine did not appear to react at all. Later Ho & Gilham looked into the reaction of CMC with pseudouridine and inosine, two modified bases found in RNA. Their results suggested that while inosine showed similar results to guanine, thymine, and uracil, pseudouridine did show attachment of CMC, but when CMC was removed from other bases, it remained on pseudouridine. This ability of CMC to selectively modify pseudouridine provided a way to distinguish pseudouridine from any other nucleoside. Ho & Gilham theorized that the CMC group bonded with the nitrogen groups on the ring, and that only the N-3 position was stable under basic conditions. The analysis of the linkage between various nucleobases and CMC was made primarily based upon UV analysis, elemental analysis and IR. However, UV and elemental analysis do not aid in the determination of the bonding. Only with IR did a small band appear in the CMC-nucleoside bond that was assigned as a new carbon-nitrogen bond. However, I do not believe that there is sufficient evidence to arrive at a definitive conclusion based upon protein literature [30-32] and UV absorbance data (shown in later chapters) The UV data will be discussed in detail in Chapter 4.

In the method developed by Ofengand and coworkers [33-35], they also modify pseudouridine with CMC. However, after modification they use reverse transcriptase to create a copy of the RNA, one base at a time. When there is a base modification, like a CMC-modified pseudouridine, the enzyme is unable to complete its transcription and there is a “stop” point. These points where the transcription fails can be mapped out with gel electrophoresis, and stops that occur due to pseudouridine will differ in the CMC-derivatized sample when it is compared to the non-derivatized control. However there
were inherent limitations in the method that were apparent due to the enzyme and technique for analysis. Because reverse transcriptase will add one base consecutively, detection requires a different “stop” point from one that was observed in the control. Reverse transcriptase will not continue past a base modification, so if there is a base modification exactly one nucleotide past the previous one -- two consecutive pseudouridines modified with CMC -- reverse transcriptase will not be able to transcribe past the first pseudouridine, and therefore the second one cannot be detected. Second, due to the limitation of gel electrophoresis, if pseudouridine is not completely transcribed in all copies of the RNA, reverse transcriptase will be able to transcribe through some of the copies, showing a band that may be fainter than others, but still detectable. Therefore, the analysis would miss a pseudouridine present, even if all others were detected.

While the method of Bakin and Ofengand is extremely powerful, the underlying approach is based upon dideoxy chain-termination sequencing using reverse transcriptase. Some of the potential difficulties associated with the Bakin and Ofengand approach can include: (1) sequencing oligonucleotides that contain $\Psi$ in a run of U residues [35]; difficulties in interpreting data arising from weak, strong or stutter bands [3, 36]; and the inability to identify modified pseudouridine residues directly [34]. All of these limitations occur because the entire method hinges the fact that pseudouridine detection is done by the lack of enzyme transcription.

ESI-MS and MALDI-MS now permit the analysis of oligonucleotides and intact nucleic acids [37, 38]. McCloskey has pioneered the use of mass spectrometry for the analysis of modified nucleosides from nucleic acids [39]. McCloskey and co-workers
developed a method for determining the sequence locations of modifications in RNA using mass spectrometry [40], and have applied this method to the analysis of a variety of RNAs [21, 41-45]. The basic method has two components. First, the RNA is digested down to the single nucleoside level, and then analyzed by LC-MS. This procedure allows for the quantitation of pseudouridine, which helps alleviate the problem of failure to identify partially transcribed pseudouridines that were present in a particular RNA. Then, mass spectrometric analysis of the pseudouridine peak eluted from the HPLC is done. As the glycosidic bond of pseudouridine (Figure 1.1a) is a carbon-carbon bond, it will not be readily cleaved under ionization conditions in the mass spectrometer, as most nucleosides are. All other nucleosides, due to their N-C glycosidic bond, will yield two ions: the molecular ion, designated MH+, and an ion that is a result of glycosidic bond cleavage, designated BH2+. Thus, uridine will have both a MH+ ion (m/z 245), and BH2+ ion (m/z 113), while pseudouridine will only have a MH+ ion (m/z 245). Therefore, pseudouridine identification is made based upon a new HPLC peak and the lack of a BH2+ peak, and this does not pinpoint the locations of the pseudouridines in a particular RNA.
CHAPTER 2: Initial Pseudouridine Studies

2.1 Introduction

As stated earlier, previous methods to analyze pseudouridine were done by taking an enzymatically digested oligonucleotide and separating it into individual nucleosides and then confirmed the identity of each nucleoside by the use mass spectrometry. This allows researchers to have three pieces of evidence to confirm the presence of pseudouridine: 1) retention time in the HPLC, 2) mass of the molecular ion from the intact nucleoside, and 3) mass from the nucleobase ion that is formed from all nucleosides except for pseudouridine and its derivatives [46-49].

Originally this work was done by McCloskey and co-workers [42, 44] using HPLC coupled with thermospray mass spectrometry. I wanted to confirm whether this approach would work with atmospheric pressure chemical ionization mass spectrometry (APCI/MS). To do this analysis, I decided to use a nucleoside test mix to optimize HPLC and APCI/MS conditions. In order to use this method with an intact RNA, it is necessary to first digest the oligonucleotide with a combination of alkaline phosphatase [50-52], nuclease P₁ [53-55], and venom phosphodiesterase [56-58]. The alkaline phosphatase removes all phosphate groups. Nuclease P₁ hydrolyzes phosphate bonds between nucleotides, and venom phosphodiesterase hydrolyzes oligonucleotides in a sequential manner. Once the RNA is exposed to these three enzymes [59], all that should remain are single nucleosides, which can be easily separated by HPLC and analyzed by APCI/MS in positive-ion mode.
2.2 Methods and Materials

2.2.1 Materials

Alkaline phosphatase, nuclease P₁ and snake venom phosphodiesterase (SVP) were purchased from Worthington Biochemical Corp (Lakewood, NJ). The nucleoside test mix, molecular biology grade ammonium acetate, Tris-HCl, and HPLC grade acetonitrile were obtained from Sigma-Aldrich (St. Louis, Mo). HPLC was done on a Beckman System Gold HPLC with detection at 260 nm, (Fullerton, CA). APCI/MS was done with a Hitachi M-1000 APCI/quadrupole MS system (San Jose, CA).

2.2.2 HPLC separation of nucleosides

A linear gradient from 100% of 25 mM ammonium acetate, pH 6.0, to 100% of 40% acetonitrile/H₂O in 100 minutes at a flow rate of 1 mL/minute was used. Separation was done on a Supelco (St. Louis, MO) supelcosil C-18 reversed phase column (4.6x125 mm) and detection was done at 260 nm. HPLC eluents were infused into the APCI/MS spectrometer for immediate analysis.

2.2.3 APCI/MS analysis of nucleosides

Optimal conditions for analysis of nucleosides by APCI/MS were a drift voltage of 33 Volts, a focusing voltage of 72 Volts and a resolution voltage of 11 Volts. The desolvation temperature was held constant at 375 degrees Celsius and the vaporization temperature was held constant at 330 degrees Celsius. Nitrogen was used as a drying gas, and the flow rate was set to provide the most intense analyte signal. For all APCI
experiments, the flow rate was 1 mL/minute. Analysis was done in selected ion monitoring mode.

2.3 Results

2.3.1 HPLC separation of nucleoside test mix

To optimize the separation conditions, I first analyzed a set of standards that have more modified nucleosides than tRNA, so that I would know if there would be adequate time to separate and analyze these peaks by LC-MS.

A list of the compounds, concentrations, and retention times is found in Table 2.1. Except for the separations of: 1-methyladenosine (m\(^1\)A) and 2’-O-methylcytidine (Cm), and 7-methylguanosine (m\(^7\)G) and guanosine (G); all peaks showed near baseline resolution. The separation of m\(^7\)G and G were similar in retention time, showing a difference of only 0.266 minutes. While these two retention times were fairly close, there

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Symbol</th>
<th>Concentration (µg/mL)</th>
<th>Retention Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudouridine</td>
<td>Ψ</td>
<td>16.7</td>
<td>6.463</td>
</tr>
<tr>
<td>Cytidine</td>
<td>C</td>
<td>33.3</td>
<td>9.347</td>
</tr>
<tr>
<td>Uridine</td>
<td>U</td>
<td>16.7</td>
<td>13.126</td>
</tr>
<tr>
<td>2-thiocytidine</td>
<td>S(^2)C</td>
<td>6.7</td>
<td>14.746</td>
</tr>
<tr>
<td>5-methylcytidine</td>
<td>M(^5)C</td>
<td>66.7</td>
<td>16.903</td>
</tr>
<tr>
<td>1-methyl adenosine</td>
<td>m(^1)A</td>
<td>16.7</td>
<td>17.467</td>
</tr>
<tr>
<td>2’-O-methylcytidine</td>
<td>Cm</td>
<td>13.3</td>
<td>17.998</td>
</tr>
<tr>
<td>Inosine</td>
<td>I</td>
<td>16.7</td>
<td>19.253</td>
</tr>
<tr>
<td>5-methyl uridine</td>
<td>M(^5)U</td>
<td>33.3</td>
<td>20.444</td>
</tr>
<tr>
<td>7-methylguanosine</td>
<td>M(^7)G</td>
<td>16.7</td>
<td>21.050</td>
</tr>
<tr>
<td>Guanosine</td>
<td>G</td>
<td>16.7</td>
<td>21.316</td>
</tr>
<tr>
<td>Adenosine</td>
<td>A</td>
<td>33.3</td>
<td>24.162</td>
</tr>
</tbody>
</table>
was enough separation between the peaks to distinguish them from each other. A representative chromatogram of the separation is found in Figure 2.1.

2.3.2 *APCI/MS analysis of standard nucleosides*

Before coupling the HPLC system with the APCI/MS, I had to analyze nucleosides solely by MS. In order to accomplish this, a syringe pump was directly connected to the APCI/MS, and ~50 ppm solutions of the standard nucleosides guanosine, cytidine, uridine, and adenosine were analyzed. A spectrum of 50 ppm adenosine is shown in Figure 2.2.

![Figure 2.1. Separation of the nucleoside test mix by HPLC.](image)

The masses of the molecular ion (MH\(^+\)) and base fragment (BH\(_2\)^+) are shown at \(m/z\) values of 269 and 136 respectively. Once these standards were run and used to calibrate the instrument, the nucleoside test mix was then injected into the LC-MS system. A list of the compounds, molecular ions and base fragment masses, as well as detection limits, is shown in Table 2.2. For all analytes, signals were observed.
2.4 Conclusions

The analysis of an oligonucleotide by nuclease P₁ digestion allows us to determine what types of modifications are present, including pseudouridine, and their relative concentration in the oligonucleotide. In an optimized system, we can identify nucleosides by measuring: 1) the retention time in the HPLC, 2) the MH⁺ peak, and 3) the BH₂⁺ peak. Uridine and pseudouridine has the same $m/z$ value for the MH⁺ peak. However, since the carbon-carbon glycosidic bond of pseudouridine will not fragment when the nucleoside is protonated, pseudouridine identification is done based upon the different retention time and the lack of a BH₂⁺ ion. These differences provide a relatively quick way to analyze the number and types of modifications found in a given oligonucleotide, regardless of the presence of pseudouridine. However, while this method may be useful in measuring the number of pseudouridine residues that are
present, it will not tell us where the pseudouridines are, and so it must be used in conjunction with other techniques.

Table 2.2. Nucleoside names, symbols, molecular and base fragment masses and detected amounts.

<table>
<thead>
<tr>
<th>Analyte</th>
<th>Symbol</th>
<th>Molecular Ion MH⁺ m/z</th>
<th>Base Fragment BH₂⁺ m/z</th>
<th>Detected Amount Ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudouridine</td>
<td>Ψ</td>
<td>245</td>
<td>n/a</td>
<td>16.7</td>
</tr>
<tr>
<td>Cytidine</td>
<td>C</td>
<td>244</td>
<td>112</td>
<td>33.3</td>
</tr>
<tr>
<td>Uridine</td>
<td>U</td>
<td>245</td>
<td>113</td>
<td>16.7</td>
</tr>
<tr>
<td>2-thiocytidine</td>
<td>s²C</td>
<td>260</td>
<td>128</td>
<td>6.7</td>
</tr>
<tr>
<td>5-methylcytidine</td>
<td>m¹C</td>
<td>258</td>
<td>126</td>
<td>66.7</td>
</tr>
<tr>
<td>1-methyl adenosine</td>
<td>m¹A</td>
<td>282</td>
<td>150</td>
<td>16.7</td>
</tr>
<tr>
<td>2-O-methylcytidine</td>
<td>Cm</td>
<td>258</td>
<td>112</td>
<td>13.3</td>
</tr>
<tr>
<td>Inosine</td>
<td>I</td>
<td>269</td>
<td>137</td>
<td>16.7</td>
</tr>
<tr>
<td>5-methyl uridine</td>
<td>m⁵U</td>
<td>259</td>
<td>127</td>
<td>33.3</td>
</tr>
<tr>
<td>7-methylguanosine</td>
<td>m⁷G</td>
<td>298</td>
<td>152</td>
<td>16.7</td>
</tr>
<tr>
<td>Guanosine</td>
<td>G</td>
<td>284</td>
<td>166</td>
<td>16.7</td>
</tr>
<tr>
<td>Adenosine</td>
<td>A</td>
<td>268</td>
<td>136</td>
<td>33.3</td>
</tr>
</tbody>
</table>
3.1. Introduction

The determination of pseudouridine using mass spectrometric methods is problematic in that this modification is the only known mass-silent modification. In this chapter, I describe the development of a mass spectrometric approach for the direct determination of pseudouridine in RNA. My approach uses the same chemical derivatization approach of Bakin and Ofengand [35] that converts the originally mass-silent modification to one that contains a unique mass tag easily identified by mass spectrometric analysis. As in their derivatization step, pseudouridine is chemically modified using CMC metho-p-toluenesulfonate and alkaline buffer. Reaction with CMC modifies all G and U-like residues. However, the CMC adduct is cleaved under alkaline conditions from all residues except at the N-3 position of pseudouridine [29].

After derivatization all pseudouridine residues except for m1acp3Ψ will exhibit a mass shift of 252.21 u (due to the CMC adduct). A comparison of the masses of the unreacted and CMC-reacted oligonucleotides will determine the number of pseudouridine residues in the oligonucleotide. The sequence location of pseudouridine can be determined using MALDI-MS by mapping of nucleotide-specific RNase digestion products [24] or by the monitoring of exonuclease digestion products [60, 61]. I demonstrate this new approach to pseudouridine identification by analyzing two tRNAs from *Escherichia coli*.
3.2. Materials and Methods

3.2.1 Materials

*E. coli* tRNA\textsuperscript{Val} (UAC) was obtained from Subriden RNA (Rolling Bay, WA) and used without further purification. *E. coli* tRNA\textsuperscript{Phe} was obtained from Sigma (St. Louis, MO) and used without further purification. CMC metho-p-toluenesulfonate, Tris HCl, urea, EDTA, and ammonium bicarbonate were obtained from Sigma. RNase T\textsubscript{1}, SVP and alkaline phosphatase were obtained from Worthington Biochemical Corp (Lakewood, NJ). Molecular weight centrifuge filters were obtained from Millipore Corp. (Bedford, MA). MALDI quality 2,4,6-trihydroxyacetophenone (THAP) and diammonium citrate (DAC) were obtained from Fluka (St. Louis, MO).

3.2.2 Reaction of CMC with tRNA.

Reaction mixtures consisting of 34 nmol (10 A\textsubscript{260} units) of tRNA, 1.2 mmol CMC in 3 mL of 50 mM Tris (pH 8.3), 4 mM EDTA and 7 M urea were incubated at 37 °C for 20 to 30 minutes. The tRNA was then separated from the reaction mixture by use of centrifugation filters. The resulting tRNA solution was then mixed with 6 mL of 50 mM ammonium acetate (pH 10.4) and 0.1 mM EDTA solution and allowed to react for 2 to 4 hours at 37 °C. As before, the tRNA was then separated from the reaction mixture by use of centrifugation filters, and the final volume was adjusted to 10 µL using nanopure H\textsubscript{2}O. Control samples of tRNA were prepared following the same protocol described above except there was no addition of CMC to the initial reaction mixture. UV analysis (254 nm) of the control showed that the recovery of tRNA from the centrifugation filters was approximately 70%.
3.2.3 RNase T₁ digestion of derivatized and control tRNA.

A 5 µL (~12 nmol) aliquot of the tRNA obtained above was reacted in 50 µL of a 50 mM Tris-HCl/1mM EDTA solution containing 10,000 units of RNase T₁ at 37 °C for 30 minutes. After digestion, half of the reaction mixture was separated by reverse-phase HPLC (described below) and each fraction was dried down and reconstituted in 10 µL of nanopure H₂O. The other half of the reaction mixture was analyzed directly using MALDI-MS (described below).

3.2.4 Exonuclease digestion by SVP.

A 3 µL (~3.5 nmol) aliquot of the fractions shown by MALDI-MS to contain the CMC modification were incubated for 30 minutes at 37 °C with 1 µL of a 0.5 units/µL solution of alkaline phosphatase. That solution was then reacted with 5 µL of 0.05 units/µL solution of SVP, 5 µL of 100 mM ammonium acetate and 5 µL nanopure H₂O. The reaction was allowed to proceed at 37 °C for 60 minutes with 3 µL aliquots removed every fifteen minutes and placed upon ice. 1 µL of the reaction sample was mixed with 1 µL of matrix for immediate MALDI-MS analysis.

3.2.5 HPLC purification of RNase T₁ Digests.

HPLC separations were done on a Hitachi (San Jose, CA) Model 7400 instrument with the UV detector set to monitor at 254 nm. The column used was a 4.6x 125 mm reversed phase Supelco (Bellefonte, PA) Nucleosil C-18 column. Buffer A was composed of 25 mM triethylammonium bicarbonate, pH 6.5, and buffer B consisted of 40% aqueous acetonitrile. Gradient elution from 0 to 100 %B at 2 %B/min at a flow rate of 1 mL/min was used. The eluted fractions were dried down and reconstituted in 10 µL nanopure...
H₂O. It should be noted that the conditions stated here have been changed in Chapter 8, and the conditions listed in that chapter provide for better separation and elution of compounds and should be used in the future for this type of analysis.

3.2.6 Mass Spectrometry.

All analyses, except for that of the intact tRNA, were done using a Bruker (Billerica, MA) ProFlex III MALDI-TOF instrument equipped with an N₂ laser operated in negative-ion mode. Intact tRNAs were analyzed on a PerSeptive Biosystems (Framingham, MA) Voyager DE MALDI-TOF instrument equipped with an N₂ laser operated in negative-ion mode. For all experiments, the accelerating voltage was held at 20 kV, and the laser power was set to the minimum level necessary to generate a reasonable signal (threshold). A two point calibration using dT₄ and dT₁₀ was used for all analyses. The matrix solution used for all experiments, except where noted, consisted of 0.325-M 2,4,6-THAP and 0.225-M DAC.

3.2.7 Analysis of intact tRNA by MALDI-MS.

A 1 µL (~2.5 nmol) aliquot of tRNA was mixed with 1 µL of matrix. 0.5 µL of this mixture was spotted on the sample target and another 0.5 µL of matrix were spotted on top of the previously dried layer. MALDI measurements were obtained as described above with the instrument operating in linear mode.

3.2.8 Analysis of RNase T₁ digest by MALDI-MS before HPLC.

A 1 µL (~1.2 nmol) aliquot of the RNase T₁ digest was mixed with 1 µL of matrix. 0.5 µL of this mixture was spotted on the sample target and MALDI measurements were obtained as described above with the instrument operating in reflectron mode.
3.2.9 MALDI-MS analysis of HPLC separated fractions.

A 1 µL (~ 0.6 nmol) aliquot of the HPLC fraction was mixed with 1 µL of matrix. 0.5 µL of this mixture was spotted on the sample target and MALDI measurements were obtained as described above with the instrument operating in reflectron mode.

3.2.10 MALDI-MS analysis of exonuclease reaction.

Reaction aliquots were mixed with 1 µL of matrix. 0.5 µL of this mixture was spotted on the sample target and MALDI measurements were obtained as described above with the instrument operating in reflectron mode.

3.3. Results

Derivatization of oligoribonucleotides or ribonucleic acids with CMC results in the modification of only pseudouridine or modified pseudouridine residues [29]. Thus, molecular weight measurements of the underivatized and derivatized sample will allow for the direct determination of the number of pseudouridine residues present. To demonstrate that molecular weight measurements of nucleic acids containing pseudouridine with MALDI-MS can be used to determine the number of pseudouridine residues present in the sample, my co-worker first analyzed *E. coli* tRNA\(^{\text{Val}}\). *E. coli* tRNA\(^{\text{Val}}\) contains a single pseudouridine residue which resides in the conserved TΨC stem-loop of the molecule. The molecular mass of underivatized tRNA\(^{\text{Val}}\) is calculated to be 24,643 Da. Based upon the known chemistry of the CMC derivatization reaction, each pseudouridine residue should increase in mass by 252 Da. Thus, the overall molecular mass of tRNA\(^{\text{Val}}\) after derivatization is predicted to be 24,895 Da.
Figure 3.1 shows representative mass spectra obtained on tRNA\textsuperscript{Val1} before derivatization (Fig. 3.1a) and after the CMC derivatization reaction (Fig. 3.1b). Suitable high-quality MALDI mass spectra of underivatized tRNA\textsuperscript{Val1} are readily obtained with mass measurement errors of less than 0.5%, which is typical for the instrument used for these experiments. Furthermore, because the determination of the number of pseudouridine residues present in the molecule is made by taking the difference in two mass values, the requirements for ultrahigh mass accuracy are reduced. As seen in Figure 3.1, the mass shift arising due to CMC derivatization is approximately that expected for the addition of a single CMC group to the molecule. Replicate experiments found that

![Figure 3.1. Comparison of: A) tRNA\textsuperscript{Val1} (m/z 24705) and B) tRNA\textsuperscript{Val1} after reaction with CMC (m/z 24950) providing a difference of 250 Da, showing the addition of 1 CMC.](image)
the mass shift was 250 u ± 6 u, which is within the experimental error for the modification of a single pseudouridine residue. Thus, the direct determination of the number of pseudouridine residues by mass measurement of underivatized and derivatized RNA samples is feasible using MALDI-MS.

It should be noted that the representative data shown in Figure 3.1 were obtained after extensive sample purification and required salt-reducing co-matrices [62]. High molecular weight oligonucleotides readily adduct metal ions present in the sample solutions, thereby reducing the sensitivity of the measurement [38, 63]. The underivatized tRNA was easier to analyze than the derivatized tRNA. The signal-to-noise ratio for the underivatized tRNA is approximately three times that of the derivatized sample. Upon spotting the derivatized tRNA with the matrix on the MALDI sample plate, co-crystallization of the sample and matrix often did not occur. As the quality of MALDI mass spectral data depends upon the co-crystallization of the analyte with the matrix [64], I believe that the lack of co-crystallization is responsible for the poorer mass spectral data obtained for the CMC-derivatized tRNA.

3.3.1 RNase T1 mapping of tRNA

While mass spectrometric analysis of intact nucleic acids provides an indication of the number of pseudouridine residues present in the molecule, this approach is not suitable for identifying the sequence location of such modifications. My approach to placing pseudouridine residues in specific sequence locations is based on the mass...
spectral analysis of RNase T₁ fragments. The use of the highly specific endonuclease, RNase T₁, simplifies the analysis of the mass spectral data. 

*E. coli* tRNA<sub>Phe</sub> was cleaved *in silico* to generate the predicted mass values reported in Table 3.1. *E. coli* tRNA<sub>Phe</sub> contains three pseudouridine residues, and only the three resulting RNase T₁ fragments containing pseudouridine should increase in mass after derivatization with CMC.

The further advantage benefits of using a specific endonuclease that generates lower molecular weight fragments has been elucidated in detail by McCloskey and coworkers [21, 44]. As noted by that group, two constraints are introduced in determination of the base composition of the oligonucleotide. The first constraint is the mass value obtained from the mass spectrum. The second constraint is the knowledge that the oligonucleotide will terminate in a specific residue (here, guanosine 3′-monophosphate). With these two constraints, accurate mass measurement (within ± 0.10%) of the resulting RNase T₁ fractions can limit the number of base compositions to a single possibility for any RNase T₁ products less than a 7-mer. Once the base composition is established, this base composition can be matched against predicted RNase T₁ digest products to place the oligonucleotide in the overall sequence. In the example demonstrated here, each pseudouridine residue resides in an oligonucleotide whose base composition occurs only once in the overall sequence, thereby simplifying assignment.

Figure 3.2 shows representative mass spectral data obtained for (a) a control with no CMC present during the derivatization reaction and (b) RNase T₁ digest of CMC-reacted
Table 3.1. Predicted and experimentally determined masses from the RNase T$_1$ fragments generated from tRNA$^{\text{Phe}}$.

<table>
<thead>
<tr>
<th>RNase T$_1$ Fraction</th>
<th>Predicted Mass</th>
<th>Experimental Mass</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gp</td>
<td>362.05</td>
<td>n.d.</td>
<td>--------</td>
</tr>
<tr>
<td>pGp</td>
<td>442.02</td>
<td>n.d.</td>
<td>--------</td>
</tr>
<tr>
<td>I AG&gt;p</td>
<td>673.09</td>
<td>n.d.</td>
<td>--------</td>
</tr>
<tr>
<td>II UDG&gt;p</td>
<td>958.11</td>
<td>958.03 ± 2</td>
<td>0.01</td>
</tr>
<tr>
<td>III CAG&gt;p</td>
<td>978.13</td>
<td>979.99 ± 1</td>
<td>0.19</td>
</tr>
<tr>
<td>IV DAG&gt;p</td>
<td>981.13</td>
<td>982.37 ± 2</td>
<td>0.13</td>
</tr>
<tr>
<td>V CCCC&gt;g&gt;p</td>
<td>1259.2</td>
<td>1260.4 ± 1</td>
<td>0.10</td>
</tr>
<tr>
<td>VI UCCG&gt;p</td>
<td>1260.2</td>
<td>1262.5 ± 1</td>
<td>0.19</td>
</tr>
<tr>
<td>VII m$^5$UΨCG&gt;p</td>
<td>1275.2</td>
<td>1276.3 ± 1</td>
<td>0.09</td>
</tr>
<tr>
<td>VIII AΨUG&gt;p</td>
<td>1285.2</td>
<td>1285.4 ± 0</td>
<td>0.02</td>
</tr>
<tr>
<td>IX As$^8$UAG&gt;p</td>
<td>1324.2</td>
<td>1325.6 ± 2</td>
<td>0.11</td>
</tr>
<tr>
<td>X CACCA(OH)</td>
<td>1510.3</td>
<td>n.d.</td>
<td>--------</td>
</tr>
<tr>
<td>VIIa m$^5$UΨCG&gt;p + CMC</td>
<td><strong>1527.4</strong></td>
<td><strong>1527.1 ± 2</strong></td>
<td><strong>0.02</strong></td>
</tr>
<tr>
<td>VIIIa AΨUG&gt;p + CMC</td>
<td><strong>1537.4</strong></td>
<td><strong>1535.8 ± 2</strong></td>
<td><strong>0.10</strong></td>
</tr>
<tr>
<td>XI CUCAG&gt;p</td>
<td>1589.2</td>
<td>1591.4 ± 1</td>
<td>0.14</td>
</tr>
<tr>
<td>XII AUUCCG&gt;p</td>
<td>1895.2</td>
<td>1896.1 ± 1</td>
<td>0.05</td>
</tr>
<tr>
<td>XIII Um$^7$Gacp$^3$UCCUUG&gt;p</td>
<td>2640.6</td>
<td>2641.1 ± 0</td>
<td>0.01</td>
</tr>
<tr>
<td>XIV AAms$^{\text{i6}}$AAΨCCCCCG&gt;p</td>
<td>3303.1</td>
<td>3304.9 ± 1</td>
<td>0.05</td>
</tr>
<tr>
<td>XIVa AAms$^{\text{i6}}$AAΨCCCCCG&gt;p + CMC</td>
<td><strong>3555.1</strong></td>
<td><strong>3556.3 ± 1</strong></td>
<td><strong>0.05</strong></td>
</tr>
</tbody>
</table>

tRNA$^{\text{Phe}}$. A comparison of the $m/z$ values obtained in these two mass spectra finds the presence of several new peaks in the CMC-reacted data. Three of these peaks are detected at $m/z$ 1524, 1536 and 3555. If one assumes that these three peaks contain CMC modified nucleosides, their unmodified counterparts should be present in the control data at $m/z$ 1277, 1287 and 3303, which is indeed the case. Based upon the known sequence of tRNA$^{\text{Phe}}$ and the predicted RNase T$_1$ digestion products, these three ions can be
assigned to the following oligonucleotides (Table 3.1): m$^5$UΨCG>p, AΨUG>p and AAms$^{2,6}$AAΨCCCCG>p.

As expected from the known chemistry of the CMC-derivatization reaction [29], each RNase T$_1$ fraction containing a pseudouridine residue undergoes a 252 Da mass shift due to the addition of the CMC group to the pseudouridine nucleoside. The complete RNase T$_1$ map of tRNA$^{\text{Phe}}$ was obtained in both cases with the exception of oligonucleotide X, the 3’-terminus of the tRNA (5’-CACCA(OH)-3’). It is not clear why this particular oligonucleotide was not detected at measurable abundances above the background. As seen in the mass spectra in Figure 3.2, several other ions are also present in the MALDI data. Many of these ions are sodium adducts of the RNase T$_1$ digestion products. The

![Figure 3.2. Comparison of the RNase T$_1$ digestion of tRNA$^{\text{Phe}}$ for the A) control and B) CMC-reacted species.](image)

39
CMC-reaction spectrum contains more sodium adducts than does the control sample, although both samples underwent the same purification prior to analysis. More extensive sample cleanup should reduce these adducts[65]. The data obtained here is of similar quality to that presented by Kirpekar et. al for the MALDI RNase mapping of 5S rRNAs [24].

3.3.2 Exonuclease digestion of CMC-modified fractions by snake venom phosphodiesterase

While the MALDI RNase T1 mapping data discussed above demonstrates that CMC-derivatization can be used in combination with mass spectrometry to identify the sequence locations of pseudouridine residues in nucleic acids, the sequence assignments of the pseudouridine residues were only inferred based upon the known sequence of tRNA^Phe and the predicted mass values of the RNase T1 fractions. Accurate assignment of pseudouridine requires that the appropriate T1 fractions be sequenced to confirm that the CMC mass tag resides on the pseudouridine residue. Further, application of this method to larger nucleic acids (e.g., rRNAs) or for nucleic acids whose posttranscriptional modifications are unknown requires that the sequence location of pseudouridine be identified within each modified RNase T1 fraction.

The approach I have chosen for sequence placement of pseudouridine residues includes RP-HPLC separation of the RNase T1 fractions followed by exonuclease digestion of the CMC-modified fractions. Representative HPLC chromatograms of (a) a control and (b) CMC-derivatized T1 digest are shown in Figure 3.3. A comparison between the control and CMC-derivatized chromatograms found that the absorbance of
three peaks was from 4 to 10 times greater in the CMC-derivatized digest than in the control digest.

Figure 3.3. HPLC separation of RNase T₁ digestion products for A) the control and B) CMC-reacted species.

Other UV studies of CMC-derivatized oligonucleotides, shown in Figure 3.4, have also demonstrated that the UV absorbance of CMC-derivatized oligonucleotides is much greater at 254 nm than those of underivatized oligonucleotides. Thus, the increase in UV absorbance allows one to determine which fractions contain CMC-derivatized
oligonucleotides that should be further digested for sequence placement of the pseudouridine residues.

Figure 3.4. UV-Vis analysis of control and CMC modified oligonucleotides.

MALDI-MS was first used to confirm that the three peaks selected from the chromatogram of the CMC-derivatized RNase T1 fractions actually contained the CMC-modified oligonucleotides (Figure 3.5). Figure 3.5a is the MALDI mass spectrum of the peak eluting at 13.4 minutes in the chromatogram in Figure 3.3b. Four ions are detected which correspond to AΨUG>p, AΨUGp and their CMC-derivatized counterparts. The CMC-modified T1 products are detected nearly exclusively in this chromatographic fraction.

Figure 3.5b is the MALDI mass spectrum of the peak eluting at 16.4 minutes in the chromatogram in Figure 3.3b. Here, unlike the case found in Figures 3.5a and 3.5c, this
chromatographic fraction contains a number of oligonucleotides. The two most abundant ions, at $m/z$ 1274 and 1284, can be assigned to $T\Psi CG>\text{p}$ and $A\Psi UG\text{p}$, respectively. Their CMC-derivatized counterparts are also found in this chromatographic fraction, at $m/z$ 1527 and 1537. Figure 3.5c is the MALDI mass spectrum of the peak eluting at 30.0 minutes in the chromatogram in Figure 3.3b. The most abundant ion in this spectrum can be assigned to CMC-derivatized $A\text{Am}^2\text{i6}
\Psi \text{CCCCG}\text{p}$. Its underivatized counterpart is also detected at $m/z$ 3306.

Once the presence of CMC-modified oligonucleotides was confirmed in the selected chromatographic fractions, exonuclease digestion was done using SVP after removal of the 3’-phosphate group by alkaline phosphatase. SVP cleaves sequentially from the 3’
terminus of the oligonucleotide. MALDI-MS can be used to monitor the progress of this digestion process [60, 61].

Table 3.2. Predicted and experimentally determined masses from the three pseudouridine-containing RNase T1 fragments after digestion with snake venom phosphodiesterase.

<table>
<thead>
<tr>
<th>RNase T1 Fraction</th>
<th>Predicted Mass</th>
<th>Experimental Mass</th>
<th>% Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.4 min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m^5UΨCG + CMC</td>
<td>1465.4</td>
<td>1465.1</td>
<td>0.02</td>
</tr>
<tr>
<td>m^5UΨC + CMC</td>
<td>1120.4</td>
<td>1124.6</td>
<td>0.37</td>
</tr>
<tr>
<td>m^5UΨ + CMC</td>
<td>815.31</td>
<td>814.2</td>
<td>0.14</td>
</tr>
<tr>
<td>m^5U</td>
<td>257.08</td>
<td>n.d.</td>
<td>n/a</td>
</tr>
<tr>
<td>13.4 min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AΨUG + CMC</td>
<td>1475.4</td>
<td>1473.3</td>
<td>0.14</td>
</tr>
<tr>
<td>AΨU + CMC</td>
<td>1130.4</td>
<td>1135.7</td>
<td>0.47</td>
</tr>
<tr>
<td>AΨ + CMC</td>
<td>824.32</td>
<td>827.0</td>
<td>0.33</td>
</tr>
<tr>
<td>A</td>
<td>266.09</td>
<td>n.d.</td>
<td>n/a</td>
</tr>
<tr>
<td>30.0 min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAm^2iαAAΨC + CMC</td>
<td>2230.6</td>
<td>2228.5</td>
<td>0.09</td>
</tr>
<tr>
<td>AAm^2iαAAΨ + CMC</td>
<td>1925.5</td>
<td>1925.0</td>
<td>0.02</td>
</tr>
<tr>
<td>AAm^2iαA</td>
<td>1367.3</td>
<td>1369.9</td>
<td>0.19</td>
</tr>
<tr>
<td>AAm^2iαAΨ</td>
<td>1038.2</td>
<td>1041.7</td>
<td>0.34</td>
</tr>
<tr>
<td>AA</td>
<td>595.14</td>
<td>598.65</td>
<td>0.59</td>
</tr>
<tr>
<td>AAm^2iαAAΨC</td>
<td>1978.4</td>
<td>1978.7</td>
<td>0.02</td>
</tr>
<tr>
<td>AAm^2iαAAΨ</td>
<td>1673.3</td>
<td>1674.8</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Analysis of the exonuclease digestion products of TΨCG and AΨUG is complicated because the 5’-terminal nucleosides (T and A, respectively) are not detected in negative-ion mode MALDI-MS. Furthermore, matrix ion interferences below 500 Da further complicate interpretation of the data. However, exonuclease digestion of these two T1 fractions does demonstrate that the CMC modification must reside on either the T or Ψ residues for TΨCG and on either the A or Ψ residues for AΨUG (Table 3.2). Based on
the known chemistry of the CMC derivatization reaction and based upon the results found for the larger pseudouridine containing T_1 fraction (see below), it is most likely that the Ψ residues do contain the CMC modification. If rigorous confirmation of Ψ sequence placement were necessary, calf spleen exonuclease, a 5’ to 3’ exonuclease, could also be used. Figure 3.6 is the MALDI mass spectrum of the 30.0 min chromatographic fraction in Figure 3.3b digested for 15 minutes with SVP. As discussed above, this fraction was found to contain the pseudouridine containing 10-mer oligonucleotide AAm^6^AAΨCCCCG>p with one CMC group. Assignment of the peaks in the mass spectrum conclusively shows that the CMC group resides on the Ψ residue (Table 3.2). In addition, two additional peaks are present in the mass spectrum that can be assigned to the 10-mer without a CMC group. It is not clear whether the CMC group is lost during the exonuclease digestion reaction or whether these ions are due to the original underivatized 10-mer present in the chromatographic fraction. In any event, the use of exonuclease digestion serves to confirm that the CMC mass tag resides on the pseudouridine residues in all three oligonucleotides, and there is no evidence that CMC derivatized any other nucleoside except for pseudouridine.

The use of mass spectrometry for the analysis of CMC-derivatized samples allows for the direct determination of pseudouridine in a single experiment. An additional advantage of this method is its ability to identify modified pseudouridine residues. Modified pseudouridine residues will exhibit an anomalous mass shift during exonuclease digestion that can be used to assign the modification [61, 66]. Further refinements to the derivatization reactions will be necessary to identify undermodified
pseudouridine residues as the present reaction conditions do not result in a 100% conversion to CMC-derivatized pseudouridine residues. This protocol can also be used within a general MALDI RNase T1 [24] mapping or ESI-LC-MS approach [40] for posttranscriptional modification identification.

In summary, the method presented in this chapter allows for the unambiguous determination of the number of pseudouridine residues, as well as their sequence location, in RNAs. The number of pseudouridine residues in an RNA can be determined...
simply from mass measurement of intact underivatized and derivatized samples. Mass spectrometric analysis of enzymatic digestion products can be used to identify the sequence location of pseudouridine. While the method presented here has been demonstrated using tRNAs, it should be generally applicable to any RNAs (e.g., snRNAs or rRNAs).
Chapter 4: Reactions of CMC with dT₁₀

4.1. Introduction

From the studies of tRNA it was clear that the reaction that modifies pseudouridine with CMC does not go to completion. This could cause problems later in the analysis of more complex molecules, such as rRNA, that contain more pseudouridines. There are several specific problems that a more complete, or complete reaction would prevent. First, there are cases where the conversion of uridine to pseudouridine does not always take place in every copy of RNA [35], which means that the number of detectable pseudouridines in some places may be lower than in others. Therefore, if the derivatization is not efficient, some pseudouridines may not be modified to an extent that can be measured. Back-to-back pseudouridines [35] may also cause problems. If there are two pseudouridine modifications adjacent to each other, then both pseudouridines may not be modified, so one or both pseudouridines may not be identified.

To avoid these complications, it seems reasonable to study the reaction of CMC with oligonucleotides to determine what factors may be important in the reaction and to try optimizing these factors to provide the best conditions for pseudouridine analysis. Some of the possible factors that could be involved include analyte concentration, CMC concentration, buffer type, buffer concentration, pH, and temperature. Other factors include comparative rates of the derivatization and neutralization reactions, and steric interference. While some of these factors could be directly involved in improving the efficiency of the reaction, others could be directly involved in the quality of the data generated by the mass spectrometer. Therefore, one must balance these factors to
generate a salt-free spectrum that has the most CMC additions possible. The goal here is not to generate a rate constant or a rate equation; it would simply be useful to examine each of the factors to determine which is most important in the generation of spectra. The variables that will be modified are shown in Table 4.1.

Table 4.1. Conditions for the reaction of CMC with dT10 that may affect the number of CMC adducts that can be added to the oligomer.

<table>
<thead>
<tr>
<th>CMC Concentration (µg/µL)</th>
<th>TRIS Concentration (mM)</th>
<th>Temperature (°C)</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>50</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>500</td>
<td>200</td>
<td>90</td>
<td>120</td>
</tr>
</tbody>
</table>

My aim is to determine which combination of experimental conditions provides the best spectra with the most complete derivatization. In order to accurately test these variables, I must have a set of conditions that are standard and will not change with each experiment. The initial conditions I used were 0.67 µmols of dT10, 25µg/µL of CMC in

![Figure 4.1. Theoretical structures of the CMC-nucleoside bond. A) Suggested by Ofengand, and B) suggested by protein literature.](image)
100 µL of 50 mM TRIS, pH 8.5 and the reaction was allowed to proceed at 37 °C for 30 minutes.

I also wanted to investigate the increase in absorbance that has been observed when the CMC modification was added. According to the literature [29], the modification to the base is done through the N-3 position of pseudouridine. However, in proteins, CMC is said to modify the carbonyl of the amino acid and not the amine, suggesting that the modification would be done through oxygen instead of nitrogen [30-32]. The two possible structures of attachment of CMC to pseudouridine are shown in Figure 4.1.

The initial assignment of the structure from Figure 4.1a was based primarily upon a small peak observed in the IR spectra. Thus, UV spectra will be analyzed to see if an increase in intensity, a shift in the maximum wavelength, or both occur after derivatization.

4.2 Methods and Materials

4.2.1 Materials

The dT_{10} and Ψ 10-mer were synthesized by standard phosphoramidite chemistry using an 8909 DNA synthesizer (PerSeptive Biosystems, Framingham, Ma) and were purified by reverse-phase cartridge purification. All solutions were made in nanopure, sterilized water. dTp, CMC and Tris were purchased from Sigma-Aldrich (St. Louis Mo). THAP and DAC were purchased from Fluka (St. Louis Mo). All mass spectral experiments were done using a ProFlex III MALDI-TOF (Bruker Daltonics,
Bellerica, Ma), and all UV-Vis studies were done using a Lambda 40 UV-Vis Spectrophotometer, Perkin-Elmer (Shelton, Ct).

4.2.2 Derivatization of oligonucleotides

Using the conditions listed in Table 4.2, reactions I-VIII were carried out in 1.5 mL centrifuge tubes. The CMC solution was made fresh daily for all derivatizations in 50 or 200 mM Tris, pH 8.5. This solution was then added to 0.67 μmoles of dT₁₀ and allowed to react at the specified temperature for the specified time. Derivatization of the 10-mer containing pseudouridine was done using 50 mM Tris, pH 8.5, with 100 μg/μL CMC and allowed to react at 37 degrees for 30 minutes.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>[CMC] μg/μL</th>
<th>[TRIS] mM</th>
<th>Temp °C</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>100</td>
<td>200</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>II</td>
<td>100</td>
<td>50</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>III</td>
<td>500</td>
<td>200</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>IV</td>
<td>500</td>
<td>50</td>
<td>37</td>
<td>30</td>
</tr>
<tr>
<td>V</td>
<td>100</td>
<td>200</td>
<td>55</td>
<td>30</td>
</tr>
<tr>
<td>VI</td>
<td>500</td>
<td>50</td>
<td>55</td>
<td>30</td>
</tr>
<tr>
<td>VII</td>
<td>100</td>
<td>50</td>
<td>37</td>
<td>60</td>
</tr>
<tr>
<td>VIII</td>
<td>100</td>
<td>50</td>
<td>37</td>
<td>120</td>
</tr>
</tbody>
</table>

4.2.3 Removal of CMC from Ψ₁₀-mer

Removal of CMC was done without any attempt to remove salts or excess CMC. 100 μL of 50 mM NH₄HCO₃, pH 10.4, was added to the sample, and the reaction was allowed to proceed for 4 hours at 37 degrees Celsius.
4.2.4 Derivatization of single nucleotides

A 0.2 mmol sample of dTp and 0.4 g of CMC were dissolved in 1 mL of sterile water and the pH was adjusted to 8.5 for the reactions and allowed to react at room temperature for 72 hours. The reaction mixture was dissolved in 5 mL methanol and the CMC-dTp product was precipitated out of solution with 50 mL chloroform. The solution was spun in a centrifuge at 5000 rpm for 30 minutes and the supernatant decanted. The resultant precipitate was then dissolved in 1 mL of sterile water for UV analysis. A control experiment was also done using the same protocol as listed above, but without adding CMC to the reaction mixture.

4.2.5 Mass Spectrometry

All reacted samples were dried and then re-dissolved in 20 µL of sterile water. A microliter of reaction mixture was mixed with a microliter of matrix that consists of 0.325-M 2,4,6 THAP and 0.225-M DAC in 50/50 ACN:H₂O. The mixture was then spotted on the sample plate and allowed to dry, and then another microliter of matrix was spotted on top of the sample and allowed to dry. The instrument was set to negative-ion mode and the instrument was equipped with a N₂ laser. For all experiments, the accelerating voltage was held at 20 kV and the laser power was set to the minimal level necessary to generate a reasonable signal threshold. A two-point external calibration was done with dT₁₀ and dT₂₀ as calibrants.
4.3 Results

4.3.1 Initial Conditions

The initial conditions for this reaction had been modified from previous protocols [22]. Under these conditions, I found that out of a possible ten residues, there were up to three additions of CMC observed; however, there was still a large peak relating to starting material. Figure 4.2 confirms what was seen in tRNA, that addition of CMC to oligonucleotides does not go to completion.

The relative abundance between the starting material at 2983 Da, and the first modified peak, at 3234 Da, was approximately 78% of the underivatized peak. This shows that a significant fraction of dT\textsubscript{10} was derivatized by a single CMC. However, the second and third additions of CMC, at 3486 and 3737 Da respectively, show only 17 and

![Comparison of dT\textsubscript{10} A) before and B) after reaction with CMC under initial conditions.](image)

Figure 4.2. Comparison of dT\textsubscript{10} A) before and B) after reaction with CMC under initial conditions.
1% conversions respectively. Based upon these results, it appears that while one modification will readily attach to a linear oligonucleotide, if there are two or three modifications, problems may arise. However, in this situation it is not clear if the problem is based on thermodynamics or on steric factors. Clearly, a large amount of salt is adducting to the oligonucleotide [67], making the actual conversion amount unclear. Moreover, because CMC has a quaternary amine, it has a permanent positive charge, which makes it difficult to use standard purification techniques. Because of the small size of the oligonucleotide, it will not precipitate in ethanol, nor can it be trapped by a molecular weight centrifuge filter. Because of the positive charge on the CMC group, cation exchange beads cannot be used and elution from a C18 tip will be difficult. Therefore, there are several options for oligonucleotide analysis. First, and probably the easiest, is to perform the derivatization on molecules of a larger size so that centrifuge filters can be employed. Second, smaller oligonucleotides can be separated with HPLC which should remove much of the salt that is present. According to the literature [29], single nucleotides can be separated from ones modified with CMC by dissolving them in methanol and then precipitating out both modified and unmodified nucleotides with chloroform. However, while this is good for removing excess CMC, it may still be necessary to desalt the sample.

4.3.2 Optimization Studies

Table 4.3 shows the number of CMC additions to dT10 for reactions I-VIII. Under the conditions investigated here, as many as 6 out of 10 possible modifications can be added to dT10. However, the yields decrease dramatically the more modifications one adds. A few noticeable effects are seen when changing the variables listed in Table 4.2.
Most noticeable was the effect of CMC concentration. As seen in Figure 4.3, as the concentration of CMC added to the reaction is increased, the amount of added CMC modifications increases from three to five. The solubility of CMC is approximately 500 µg/µL, so after that point I cannot determine if increasing the concentration beyond this point would be useful. Interestingly, at a concentration of 100 µg/µL (Figure 4.3b), the ion containing one CMC addition is more abundant than the unmodified starting material, suggesting that one modification was the most abundant species present in the mixture. This effect was not present at 500 µg/µL of CMC (Figure 4.3c), suggesting that adding too much CMC can actually inhibit the reaction, although this is not understood at this time.

Figure 4.4 shows the effect of changing the buffer concentration. In all cases, I used a Tris buffer at pH 8.5, but varied the concentration between 50 and 200 mM. There are several interesting things to note about the differences between the spectra generated by changing the buffer concentration. There is a higher relative abundance of the first

Table 4.3. Number of CMC adducts observed with reactions I-VIII.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Number of Modifications Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>4</td>
</tr>
<tr>
<td>II</td>
<td>6</td>
</tr>
<tr>
<td>III</td>
<td>5</td>
</tr>
<tr>
<td>IV</td>
<td>6</td>
</tr>
<tr>
<td>V</td>
<td>6</td>
</tr>
<tr>
<td>VI</td>
<td>6</td>
</tr>
<tr>
<td>VII</td>
<td>6</td>
</tr>
<tr>
<td>VIII</td>
<td>6</td>
</tr>
</tbody>
</table>
Figure 4.3. Effects of CMC concentration on the CMC reaction. A) 50 µg/µL CMC, B) 100 µg/µL, C) 500 µg/µL.

Modification compared to the unmodified species at both concentrations. However, only 4 CMC adducts are detected from the 200 mM Tris buffer reaction as compared to 5 CMC adducts for the 50 mM Tris buffer reaction.

Figure 4.4. Effects of buffer concentration on the CMC reaction. A) 50 mM Tris, B) 200 mM Tris.
There is also a loss of ion abundance from the 200 mM Tris buffer reaction. This is most likely a result of excessive sample buffer present in the MALDI sample crystal. It is not known whether this is also limiting the detection of the more modified forms of dT_{10}.

A third, less noticeable factor of the CMC-dT_{10} reaction is temperature. As shown in Figure 4.5, the two spectra are very similar in terms of relative abundance and number of CMC additions observed. However there is a large difference in the signal intensity between the two spectra; with the spectrum resulting from the higher temperature showing a higher background and baseline. It is difficult to precisely determine the cause of the differences, as MALDI is not a quantitative technique, and ion abundances are not always consistent from sample to sample. There is considerably more degradation of the

![Figure 4.5. Effects of reaction temperature. A) 37 °C, B) 55 °C.](image)
oligonucleotides in Figure 4.5b. It cannot be determined at this point whether this degradation is due to mass spectrometry conditions or to excessive heating of the sample during the derivatization.

The final variable considered was time, but this seemed to have no appreciable effect on CMC addition. In fact, no differences between spectra of reaction at different times were seen, which suggests that the reaction happens relatively quickly, and that once it reaches a threshold, no more CMC adducts can be added to dT₁₀.

4.3.3 UV analysis of the CMC-nucleoside reaction

As demonstrated in earlier chapters, after an oligonucleotide is modified with CMC, it exhibits a large increase in its UV absorbance at 260 nm. In order to fully characterize the reaction between CMC and a nucleoside, I decided to try to look at that reaction with methods that were better able to make structural determinations. However, before trying to analyze the bonds in the molecule, I first had to determine if a reaction had been completed on a single nucleotide. To check for modification, I compared the product’s retention time to a control by reversed phase HPLC. As shown in Figure 4.6, a small difference in the retention time between the reaction and control species is noted along with an indicative increase in the UV absorbance of the reacted species. This comparison was made by taking the water phase of both the control and reaction species after precipitation with chloroform.

After it had been confirmed that a reaction had taken place, three comparisons were made in the UV. All UV scans were made from 300 to 200 nm under the same instrumental conditions. The first comparison was between the water and chloroform phases of the control. This comparison showed almost no difference except for the absorbance, which
Figure 4.6. HPLC separation of dpT and dpT+CMC.

is higher in the chloroform phase than in the water phase, suggesting that less than half of the nucleotide fell out of solution upon precipitation with chloroform. The second comparison was between the reaction chloroform and water phases. In the chloroform phase, some fine band structure is seen, Figure 4.7a, which is attributed to the anion (p-toluenesulfonate) that came with the CMC. No trace of anion in the water phase, possibly due to solvent dampening, and only a small absorbance in the chloroform phase were seen.

Figure 4.7. UV-Vis comparison of A) dpT-CMC reaction and B) control, after separation with chloroform.
In Figure 4.7b, the comparison between the reaction water phase and the control water phase shows that the absorbance spectra does not shift in wavelength, but just increases dramatically. In fact, when the absorbance from both control phases are added together, they do not add to half of the signal shown from the water phase of the reaction, suggesting that the extinction coefficient rises dramatically. While both the HPLC and the UV scan suggest that a reaction does in fact occur, they do not tell us what the new molecule looks like, or how CMC is bound.

ESI-MS was attempted to determine the structure of the new molecule [68-71]. However, in all attempts to generate a molecular ion from the new species, the only discernable peak detected was at $m/z$ 280, which based upon the isotope pattern, could be the mass of the unmodified CMC molecule plus a water molecule. Because it was not possible to generate a molecular ion of a nucleotide-CMC product in either positive or negative ion mode, it was impossible to do tandem MS experiments. At present, I do not know why spectra could not be generated from the reaction product.

It was suggested to try NMR spectroscopy to probe the structure of the molecule [72-74]. However, as the structure to be analyzed had no protons, a large amount of purified material would have to be generated to use $^{13}$C NMR. Also, I would need standard molecules with structures similar to the possible target areas for comparison to determine which of the structures was correct. Due to the lack of commercial standards, and the difficulty in producing large amounts of purified products, I did not attempt further analysis of this bond.

Despite the problems of analysis, one thing was clear when dealing with the reaction of CMC with nucleotides and nucleosides: sample preparation and cleaning are
essential before analysis by mass spectrometry. It is also not clear whether the limit of addition of CMC was based upon thermodynamics or steric, but looking at various factors has provided some insights on what the correct reaction conditions should be in order to maximize the number of additions onto an oligonucleotide.

Figure 4.8. 10-mer containing Ψ A) after addition reaction of CMC. B) After addition of NH₄OH and reaction for 4 hours with no cleaning in between.

The final topic which needs to be addressed is the removal of CMC from nucleotides other than pseudouridine upon exposure to a highly basic aqueous system. Under normal conditions, CMC modifications are stable on all uracil, guanine, thymidine, and pseudouridine bases. Once the pH of solution becomes very basic (>pH 10), then the CMC additions should be removed from all bases except for pseudouridine. Figure 4.8 shows a 10-mer containing pseudouridine as one of its bases and several guanine and thymine bases. After the first step of the reaction, several CMC additions to the 10-mer are observed. However, upon exposure to basic conditions for 4 hours, no
detectable change in the number of CMC adducts is observed. This suggests that sample purification is required to remove excess CMC from the reaction mixture because the equilibrium reaction is not favorable for the removal of these modifications.

4.4 Conclusions

While the study of the CMC-nucleoside reaction did not optimize the conditions for complete addition of CMC to any oligonucleotide, several insights will be helpful for further investigations of RNA. First, it is difficult to add CMC to all possible reaction sites under a variety of conditions. Second, without a good separation technique to remove excess CMC from the initial reaction, it is difficult to irreversibly remove CMC additions from molecules that are not pseudouridine. These findings suggest that while the bond is not stable under highly basic conditions, if there is a large amount of CMC available, removal is not apparent. Third, a delicate balance seems to exist between optimum conditions for CMC addition and optimum conditions for detection by mass spectrometry. Possibly, fewer CMCs are added to the molecule than are detected, but to force them on to the oligonucleotide creates buffer and salt conditions that may overwhelm the mass spectrometric detection and the observed peak abundance is reduced.

To fully test these ideas, it would be necessary to try a longer, mixed based oligonucleotide, preferably one that contains the pseudouridine modification. A longer oligonucleotide would allow for better desalting and CMC removal in two ways. First, ethanol precipitation could be employed, and second, molecular weight centrifuge filters could be employed. Both of these could be used to remove the excess CMC and buffers out of the sample, which would allow for better mass spectrometric data, as well as
analysis of the reverse reaction. The oligonucleotide would still have to be limited to 15-20 bases in length for optimum mass spectrometric detection.
CHAPTER 5: Culturing of E. coli MRE-600

5.1. Introduction

To have purified 16S and 23S rRNA to work with, one must culture *E. coli* and separate the rRNA from all of the other cellular material. This can be accomplished in four basic steps. These steps are: 1) bacterial growth, 2) ribosome isolation, 3) RNA isolation, and 4) RNA separation. When these steps are completed, one should have pure samples of all three types of rRNA that is found in the organism [75, 76].

Bacterial growth is a process during which stock bacteria, in this case MRE-600, are inoculated into a flask of media and allowed to grow as fast as possible with minimal organism death to maximize the number of live, viable cells that can be harvested. In a typical harvest, which will be described later in this chapter, 10 grams of cells can be harvested from 4 liters of starting media. Once the cells are harvested and collected, it is then necessary to break open, or lyse, the cells to obtain intact ribosomes. Several methods can be used to break open cells; in this case, I use a purely mechanical device called a French press [77-80], which can put cells under intense pressure until the cell walls break. Typically several repetitions with the French press are required to completely destroy cellular membranes. After lysing the cells, it is then necessary to use DNase I to hydrolyze any intact DNA that is found in the cell so that the intact ribosome can be recovered by centrifugation.

Once I recover the intact ribosome, I must break apart the ribosome and remove the proteins associated with the ribosome. This is done by several separation steps [81]. The first step uses phenol saturated into an aqueous solution. This will remove low molecular
weight material as they will be able to stay in the non-polar environment while RNA and other large biomolecules will stay in the aqueous phase. The second step is a phenol(aq)/chloroform:isoamyl alcohol extraction step, which is responsible for allowing insoluble material to stay in the interface of the aqueous/organic later, thereby removing them from the RNA. The final step is to lower the polarity of the organic phase by extracting samples with the chloroform/isoamyl alcohol. This allows the ribosomal proteins to stay in the organic layer, while the polar nucleic acids will stay in the aqueous layer, thereby completing the separation of the nucleic acids from the other cellular material. After precipitation with ethanol, there should be pure RNA, since the DNase I enzyme cleaves DNA into small pieces that will not precipitate out of solution.

After these steps, it is then necessary to separate the various types of RNA from each other. In MRE-600, there are mainly 23S rRNA, 16S rRNA, and 5S rRNA. The separation of these three compounds can be accomplished by several methods, including gel electrophoresis. However, the most efficient way to separate large amounts of RNA quickly is to perform sucrose gradients [82-84]. A sucrose gradient is a density gradient, and when RNA is added and centrifuged, the RNA molecules will stay in the area of the gradient where their density closely matches that of the sucrose. For bacterial RNA, a gradient from 40% to 15% sucrose is sufficient to effect separation of rRNA. The RNA is initially loaded into a tube that has a premade gradient and then centrifuged at high speeds, typically 27,000 rpm, for 18 hours. The sucrose is then eluted by putting a small hole in the bottom of the gradient tube and allowing the solution to drip slowly from the bottom of the tube. This allows for separation to occur as the various RNAs are typically collected in different tubes.
5.2 Methods and Materials

5.2.1 Materials

Bacto-trypton and bacto-yeast extract were purchased from Difco, (Sparks, Md). Sodium chloride, sodium hydroxide, Tris-HCl, magnesium acetate, ammonium chloride, sodium acetate, magnesium chloride and sucrose were purchased from Fisher Scientific (Hampton, NH). Equilibrated phenol, phenol/chloroform:isoamyl alcohol, and chloroform:isoamyl alcohol solutions and sodium dodecyl sulfate (SDS) were purchased from Sigma-Aldrich (St. Louis, MO). All solutions were made from nanopure water that was purified and autoclaved in house. Incubation of *E. coli* was done in an Innova 4000 incubator, New Brunswick Scientific (Edison, NJ). An American Instrument Co. French press was used to lyse cells, a GM-100 gradient maker, CSB Scientific Co (Del Mar, Ca), Beckman ultracentrifuge XL-80 (Fullerton, Ca), and a Sorval RC5C centrifuge (Newtown, Ct) were used to separate and purify RNA. All concentration measurements were made on a Shimadzu Biospec 1601, (Columbia, Md) UV/Vis spectrometer at 600 and 260 nm.

5.2.2 MRE-600 Cell Growth

For this procedure, all materials and chemicals should be kept sterile to preserve RNA. After four in the afternoon, 100 mL of Luria-Bertani (LB) media is inoculated with stock bacteria, and allowed to incubate at 37 degrees Celsius and 300 rpm overnight. After approximately 16 hours, 25 mL from the started flask is transferred to a 2 L media flask containing 500 mL of fresh LB media. After the transfer, an aliquot is removed from the media flask, to be used as a blank, and the UV absorbance is measured at 600
nm. The media flask is then allowed to incubate at 37 degrees Celsius and 300 rpm and an aliquot is removed every hour and the absorbance is checked at 600 nm to monitor the growth process. The media flask is allowed to incubate until the UV absorbance stops doubling. Once this has occurred, the cultures are divided into 12-250 mL centrifuge bottles and then centrifuged at 4 degrees Celsius and 5000 rpm for 15 minutes. After the liquid is decanted the bottles are separated into 2 6-bottle batches. A volume of less than 10 mL of 0.9% NaCl is added to one bottle per batch. This volume is used to dissolve the pellet containing the cells and then that volume is transferred to each successive bottle in the batch. After all pellets are dissolved, then the first was rinsed with 10 mL of NaCl, then the solution was transferred to the each subsequent bottle, providing a final volume of < 20 mL per batch. One then transfers the contents of the final bottles to 2-50 mL centrifuge tubes that were previously weighed, centrifuge the tubes at 4 degrees Celsius at 9500 rpm for 15 minutes, decant the liquid and weigh the tubes to determine the cell yield.

5.2.3 Ribosome Isolation

The cell homogenizer is prepared for use by rinsing with ethanol and sterile water and then chilled on ice. Cells that were thawed on ice are dissolved in 7.5 mL of RB buffer (0.01 M Tris-HCl, 0.01 M MgOAc, 0.06 M NH₄OAc, and 7 M beta-mercaptoethanol). The contents of both tubes (~15 mL) were then added to the homogenizer and homogenized with 20 strokes. The sample is then divided into 3-50 mL tubes. The homogenizer is rinsed twice with 12.5 mL RB buffer, and equal amounts are added to the three tubes. The pressure on the French press is set to 1200 psi and the cylinder is inserted securely into the apparatus. Once the pressure builds to 1200 psi, the
valve is slowly opened so sample slowly pours into the receptacle tube. This step is then repeated. Then 20 µL of DNase I (10 U/µL) per gram of cells is added to each tube and pressed 2 additional times. Tubes are then centrifuged at 4 degrees Celsius for 15 minutes at 9500 rpm. At this point, the RNA is contained in the aqueous supernatant and is transferred to 10 mL tubes. The pellet is then dissolved in a minimum volume three more times to remove additional RNA, and these subsequent aqueous rinses are also transferred to the 10 mL tubes. All 10 mL centrifuge tubes are spun at 48000 rpm for 15 hours at 4 degrees Celsius. The RNA is now contained in the pellet, so the supernatant is discarded.

5.2.4 RNA isolation

Ribosome pellets are dissolved in a total of 16 mL of 10 mM Tris-HCl. Solutions are transferred to a clean 50 mL centrifuge tube and each 10 mL tube is rinsed with 1 mL of Tris-HCl and added to a 50 mL tube. The approximate final volume should be about 25 mL. Using the saturated aqueous phenol, the following extractions are done: one phenol extraction, three phenol/chloroform:isoamyl alcohol extractions, and two chloroform:isoamyl alcohol extractions. Each extraction should be done by adding 25 mL of each organic solution to each aqueous solution. All extracted aqueous phases are kept separate. To separate the aqueous phase from the organic phases, centrifugation at 3500 rpm for 3 minutes at room temperature is done. When decanting the aqueous phase, avoid removing white protein interface material. Once all aqueous phases have been collected, 0.53 mL of 3 M NH₄OAc for each 8 mL of aqueous phase collected is added and then three times the volume of ice cold ethanol is added to precipitate RNA. The bottles with the precipitate are centrifuged at 10500 rpm at 4 degrees Celsius for 1 hour.
Pellets are then dissolved in a minimum volume of water and then the bottle is rinsed with 3 mL sterile water, combining all solutions for UV absorbance measurement at 260 nm to determine concentration of total RNA.

5.2.5 Sucrose gradients

In the gradient maker, 17.5 mL of 40% sucrose is added to the left reservoir and 17.5 mL 15% sucrose to the right reservoir. The gradient maker is then placed on a stir plate and a stir bar is added to the right reservoir. The stopcock and flow valve are opened together to allow flow to an empty centrifuge tube. The outlet tube is removed slowly to avoid disturbing the gradient. Once the gradients are complete, the tubes are stored in the refrigerator overnight. 750 µg of total RNA is added to the top of each tube and then the tubes are spun at 4 degrees Celsius for 18 hours at 27000 rpm in a SW-28 rotor. Once finished, gradients must be collected immediately by punching a small hole in the bottom of the tube with a 18-20 gauge needle and allowing the solution to flow into approximately 65-1.5 mL centrifuge tubes. For best resolution, approximately 0.75 mL of gradient should flow into each tube at a rate of about 1 drop per second. If the flow rate exceeds 2 drops per second, the resolution will be severely compromised. 20 µL aliquots from each collected fraction are diluted with 80 µL of sterile water and the absorbance is measured at 260 nm for each diluted fraction.
5.3 Results

5.3.1 Cell Culturing

Four 2 L media flasks were inoculated and the cells were collected from each. After 19 hours, the absorbance at 600 nm was consistently around 1.6 AU, which corresponded to a total cell yield of between 10-12 grams of cells. The total amount of cultured cells is shown in Table 5.1. To fully characterize pseudouridine in later studies with 23S rRNA, it was necessary to generate at least 6 mg of 23S rRNA. Based upon the growth of the bacteria, I determined that the optimal growing time was approximately 18-19 hours. A typical growth curve for MRE-600 is shown in Figure 5.1. It shows that the UV absorbance starts to slow at approximately 3 hours after transfer into the 2L flasks, which means a total growth time of 18 hours.

![Growth Curve](image)

Figure 5.1. *E. coli* growth curve showing the UV absorbance at 600 nm over time.
5.3.2 RNA isolation

Once the cells were destroyed and the RNA separated from the ribosome, it was necessary to quantitate the amount of RNA that was collected from the cells. UV analysis indicated that approximately 80 mg of RNA had been recovered. Table 5.2, shows the amount of cells generated during culturing. To ensure that the sample contains pure RNA, there are several calibration steps required. First, the accuracy of the UV must be measured with a known RNA standard. This must generate a calibration curve (Figure 5.2) that has a high $r^2$ value.

Once it is determined that the UV-Vis spectrometer is working correctly, one must then consider whether there are contaminants that contribute to the measured UV absorbance. Some of the possible types of contaminants that can be found after RNA isolation [85, 86] are: 1) proteins, 2) phenol, 3) carbohydrates, and 4) DNA.

Table 5.1. Cultured cellular yield of *E. coli* MRE-600.

<table>
<thead>
<tr>
<th>UV abs (max)</th>
<th>Initial Mass (g)</th>
<th>Final Mass (g)</th>
<th>Cell Mass (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tube 1</td>
<td>1.58</td>
<td>17.8</td>
<td>23.7</td>
</tr>
<tr>
<td>Tube 2</td>
<td>1.61</td>
<td>18.1</td>
<td>23.8</td>
</tr>
<tr>
<td>Tube 3</td>
<td>1.60</td>
<td>17.7</td>
<td>24.6</td>
</tr>
<tr>
<td>Tube 4</td>
<td>1.60</td>
<td>17.9</td>
<td>22.9</td>
</tr>
<tr>
<td>Tube 5</td>
<td>1.54</td>
<td>18.0</td>
<td>23.6</td>
</tr>
<tr>
<td>Tube 6</td>
<td>1.64</td>
<td>17.8</td>
<td>23.7</td>
</tr>
<tr>
<td>Tube 7</td>
<td>1.56</td>
<td>17.9</td>
<td>22.9</td>
</tr>
<tr>
<td>Tube 8</td>
<td>1.59</td>
<td>17.6</td>
<td>24.9</td>
</tr>
<tr>
<td>Tube 9</td>
<td>1.61</td>
<td>17.7</td>
<td>23.8</td>
</tr>
<tr>
<td>Tube 10</td>
<td>1.60</td>
<td>17.8</td>
<td>22.9</td>
</tr>
</tbody>
</table>

Total Cell Mass 58.5 g
Table 5.2. Calculated amount of total RNA generated from culturing by UV-Vis analysis.

<table>
<thead>
<tr>
<th>TUBE</th>
<th>ODU</th>
<th>ODU/stock</th>
<th>µg/mL RNA</th>
<th>mg RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.854</td>
<td>854</td>
<td>17080</td>
<td>17.08</td>
</tr>
<tr>
<td>2</td>
<td>1.299</td>
<td>1299</td>
<td>25980</td>
<td>25.98</td>
</tr>
<tr>
<td>3</td>
<td>0.597</td>
<td>597</td>
<td>11940</td>
<td>11.94</td>
</tr>
<tr>
<td>4</td>
<td>0.832</td>
<td>832</td>
<td>16640</td>
<td>16.64</td>
</tr>
<tr>
<td>5</td>
<td>0.495</td>
<td>495</td>
<td>9906</td>
<td>9.906</td>
</tr>
<tr>
<td>6</td>
<td>0.025</td>
<td>25</td>
<td>500</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Total mg RNA 82.046

Protein contamination can be attributed to the phenol extractions and is due to poor technique when decanting the aqueous phase from the protein interface. However, by taking a ratio of absorbances at 260 and 280 nm [87] it is easy to determine protein contamination. If the ratio is above 1.8, then there is no significant contamination by protein. The second possible contaminant, phenol, can also be a result of poor pipetting during extraction. It shows up as a large shoulder at a 276 nm.

Figure 5.2. UV-Vis calibration curve at 260 nm for RNA.
The third contaminant, carbohydrates, is usually found in the organic layer of the phenol extraction, and when carbohydrates are detected, then phenol is usually also a contaminant. The extent of contamination is determined by measuring a ratio of 260 to 230 nm and needs to be greater than 2 to show no appreciable carbohydrate contamination. The final possibility, DNA, is the only contaminant that cannot be determined by UV analysis alone, as the optimum wavelength for DNA is the same as for RNA. DNA can be a contaminant if there is insufficient DNase I added when the cells are lysed. If insufficient DNase I is added, then the nuclear DNA will not be cleaved and will be of sufficient length to precipitate with the RNA after it is isolated from the ribosome. The only way to determine whether there is DNA interference is to run gel electrophoresis, preferably using a 1% agarose gel. This will determine two things: 1) if there is DNA contamination, and 2) if the RNA bands show intact RNA.

Upon analysis of the total RNA sample, the 260/280 ratio indicating proteins was 1.8, which shows no substantial protein interference. The 260/230 ratio indicating carbohydrates was 2.1, also indicating no substantial interference from carbohydrates, and there was no shoulder in the UV scan at 276, indicating that phenol concentrations were minimal. An agarose gel was run from which it was determined that the RNA was intact and there was no DNA interference.

5.3.3 RNA separation

From those analyses, it appears that the total RNA sample is pure. It is then necessary to separate the three individual rRNAs by sucrose gradient. As stated earlier, the goal was to generate approximately 6 mg of 23S rRNA for later studies. Approximately half of the RNA in the cell is 23S rRNA [87], so if 12 mg of total RNA is
separated, there should be enough to do further studies. This should also produce a yield of 16S rRNA that is approximately 3 mg, which should be more than adequate for pseudouridine analysis. Optimal separation of RNA was previously determined by this researcher with the assistance of Martine Marcotte to be approximately 750 µg of total RNA per tube. This suggests that in order to separate the correct amount of RNA, 16 tubes need to be run. A typical example of a histogram generated from a sucrose gradient is shown in Figure 5.3. The first large peak generated is the 23S rRNA; the second is 16S rRNA, and the final peak is 5S rRNA.

The separation between the 23S and the 16S rRNA is not complete, so to generate a pure sample of RNA, care must be taken. Typically, that last faction that is included in the 23S peak is when absorbance starts to fall. Usually, the first 16S fraction used is either the maximum absorbance peak, or the fraction immediately preceding it, depending upon the quality of the separation.

Figure 5.3. Sucrose gradient separation of 23S, 16S, and 5S rRNA from *E. coli*. 
Based upon the relationship that the number of optical density units (ODU) measured by UV, is equivalent to 20 µg/mL of RNA, it was determined that after 25 sucrose gradients, the amount of 23S rRNA generated was 15.8 mg; the amount of 16S rRNA generated was 6.1 mg, and the amount of 5S rRNA was not quantitated. Note that a large amount of unseparated total RNA remaining, so if the amount of pure 16S or 23S rRNA is not sufficient for later experiments, it will be a short process to generate more of these pure rRNAs.
Chapter 6: Application of a Method to Analyze Pseudouridine in 16S rRNA from E. coli

6.1 Introduction

In Chapter 3, I presented a method to analyze pseudouridine in tRNA from E. coli. The newly developed method has several advantages over previous methods used to detect pseudouridine including the ability to: 1) directly sequence RNA, 2) sequence RNA quickly, 3) determine the presence of adjacent pseudouridines, 4) detect pseudouridines that are not present in all copies, and 5) detect pseudouridine by generation of a new signal peak. These are all advantages that allow for the rapid and reliable detection of pseudouridine in small RNAs.

However, with larger pieces of RNA, such as 16S and 23S rRNA, the problem of pseudouridine detection becomes much more complicated. In typical bacterial tRNA, there are approximately 75 bases in the sequence, which will generate approximately 15 RNase T₁ fragments. Fifteen fragments is a reasonable number of fragments that can be easily separated by reversed-phase HPLC and therefore it is fairly straightforward to get pure fragments that can easily be sequenced. With a larger piece of RNA, like 16S rRNA from E. coli [22, 88, 89], there are approximately 1600 bases in the sequence, which can generate as many as 300 RNase T₁ fragments many which can be identical. This increases the difficulty of separation by at least a factor of twenty, and so in these cases, reversed phase HPLC is not sufficient to generate pure RNase T₁ fragments that can be sequenced.

Another separation technique, anion exchange HPLC [90-92] separates primarily based upon the size of oligonucleotides. The anion exchange resin initially attracts the
negatively charged backbone of the oligonucleotide and during the separation, as the conditions change, it progressively releases molecules sequentially based upon the number of charges that are present. The first to elute off of the column will be single nucleotides, then dinucleotides, and so forth until all sizes have been eluted. This allows for the collection of oligonucleotides that are similar in size and that can subsequently be

Figure 6.1. 16S rRNA and the 510 loop containing pseudouridine from 16S rRNA from *E. coli*.

Figure 6.1. 16S rRNA and the 510 loop containing pseudouridine from 16S rRNA from *E. coli*.

separated by reversed phase HPLC. Using this two dimensional approach, efficient separation of a large number of RNase fragments can be achieved.

In 16S rRNA from *E. coli*, there is only one pseudouridine residue present at position 516 [42], and it is found in the hairpin loop shown in Figure 6.1. The highlighted residue is at position 516. Based upon its position in the sequence, RNase T₁
will generate the fragment $\Psi$Gp. The mass of this fragment before derivatization is 668.38 Da, and after derivatization, the mass will be 920.45 Da.

6.2 Methods and Materials

6.2.1 Materials

16S rRNA from *E. coli* was generated from bacterial cells as described in Chapter 5. All buffers were made from Nanopure, sterilized water. CMC, Tris, EDTA, urea, NH$_4$OAc, NH$_4$HCO$_3$, triethylamine and acetonitrile were purchased at molecular biology grade from Sigma-Aldrich (St. Louis, Mo). MALDI quality THAP and DAC were obtained from Fluka (St. Louis, Mo). RNase T$_1$ was purchased from Worthington Biochemical Corporation (Lakewood, NJ). Centricon Plus-20 centrifuge filters with a 3000 MWCO were obtained from Millipore Co. (Billerica, Ma).

6.2.2 HPLC separation

The anion exchange column was a nucleogen DEAE 60-7 4x125 mm column, Macherey-Nagel (Easton, Pa), with a matching guard column, and the reversed phase column was a Supelco (St. Louis, Mo) Supelcosil LC-18 4.6x125 mm column. All separations were done on a Beckman System Gold HPLC with gradient elution at a flow rate of 1 mL/min and UV detection at 254 nm. For anion exchange HPLC, a linear gradient of 1% per minute of buffer A (25 mM TEAB (pH 6.4) in 20% ACN in water) to buffer B (1 M TEAB (pH 7.8) in 20% ACN in water) for 50 minutes. For reversed phase HPLC, a linear gradient from buffer A (25 mM TEAB (pH 6.4) in 20% ACN in water) to buffer B (40% ACN in water) for 50 minutes was used [59].
6.2.3 Mass spectrometry

All analyses were done using a Bruker (Billerica, MA) ReFlex IV MALDI-TOF instrument equipped with a N₂ laser operated in negative-ion mode. For all experiments, the accelerating voltage was held at 20 kV and the laser power was set to the minimum level necessary to generate a reasonable signal. A two point calibration was done using dT₅ and dT₁₀ for all analyses. The matrix solution used for all experiments consisted of 0.325-M THAP and 0.225-M DAC.

6.2.4 Analysis of Pseudouridine in 16S rRNA from E. coli

The reaction of CMC with 16S rRNA and RNase T₁ reactions were done as described in Chapter 3, with the reaction scaled up by a factor of 15. Centricron Plus-20 MWCO filters were used in place of the Microcon filters listed in Chapter 3.

6.3 Results

MALDI-TOF cannot be used to analyze intact 16S rRNA for CMC addition. Instead after the reaction with CMC, 16S rRNA was digested with RNase T₁ and separated by HPLC. According to the sequence of 16S rRNA [89, 93-95], Figure 6.1, there should be one pseudouridine modification causing one addition of CMC that results in a fragment of 920.45 Da. A comparison of the control (Figure 6.2a) and CMC modified (Figure 6.2b) RNase T₁ fragments shows, as expected, that early in the separation there is a large jump in UV absorbance in the first two peaks (labeled I & II), with a small shift in retention time. The supposition is that the first peak eluted from the
column (6.9 minutes) is excess CMC, while the second peak (13.5 minutes) will be the one that contains the pseudouridine containing fragment.

Upon isolation of the mononucleotide and dinucleotide-containing HPLC fractions, it was necessary to separate these fragments by reversed phase HPLC. For the anion exchange peak I, shown in Figure 6.3a, the control only showed evidence of one peak, indicative of Gp. In the CMC-modified sample, a second, high intensity peak (labeled A) was isolated and analyzed by MALDI-MS. Because of signals due to the matrix, it is very difficult to detect signal relating to both CMC and Gp [96].

Figure 6.3b shows the control and CMC-modified reversed phase HPLC separation of the anion exchange separation, peak II. Several peaks in both the control
Figure 6.3. Reversed-phase HPLC separation of: A) anion-exchange peak I. Peak labeled A is suspected to be the CMC peak. B) Anion exchange peak II. Peak labeled B is suspected $\Psi Gp^{+}CMC$ fragment.

and the CMC modified separation would be attributed to CGp, UGp, and AGp. A peak that would be due to $\Psi Gp$ would be difficult to detect because there is only one of these fragments, as compared to between 25 and 35 fragments of each of the other three. In the CMC-modified anion exchange chromatogram, there was evidence of a high intensity peak at a new retention time, which is presumably $\Psi^{CMC}Gp$. This new fraction (labeled B) was then analyzed by MALDI-MS (Figure 6.4) and compared to the spectrum of the control and the spectrum generated just by matrix to show that any peak generated at 920.45 Da is a result of the CMC-modified peak and not the matrix or the control.
Because the pseudouridine containing fraction generated by RNase T₁ is a dinucleotide, exonuclease digestion is not helpful. However, the chromatograms and mass spectral data support the fact that the pseudouridine containing fragment is identified and isolated. This process has some limitations, because, as it stands, it cannot identify where in a large RNA a pseudouridine is located. As the mass of the RNase T₁ fraction generated is exactly the same as 32 other fragments, it is impossible to prove that the identified fragment comes from position 516.

One way to alleviate this problem would be to use different RNases. Enzymatic digestion using RNase U₂ [97, 98] cleaves RNA at unmodified adenine residues. With this enzyme a fragment from 16S rRNA having the sequence CUCCGΨGCCAp, (\(m/z\ 3157.88\) Da) would be generated. This fragment is large enough to provide the sequence information needed, not only to identify pseudouridine, but to identify the position of pseudouridine. This would easily be the most quick and efficient way to analyze pseudouridine in 16S rRNA. However, the ultimate goal of this project is to analyze
pseudouridine, not just in 16S rRNA, but in all sizes of rRNA. Therefore analysis of the fractions generated by RNase U2 alone will not be sufficient. To accommodate the additional size and complexity of other RNAs, another enzyme called RNase H can be employed. RNase H can be used to cleave RNA in a position that is determined by the researcher by using single stranded DNA that is complimentary to the RNA sequence that is to be cleaved. The first successful implementation of this technique and the details of the method development using RNase H will be described in Chapter 7.
Chapter 7: Development of new RNA cleavage techniques using RNase H

7.1 Introduction

It is difficult to obtain RNA in sizes small enough to determine sequence information, but still large enough to place sequences into the larger scheme. Cutting DNA to a specific size is relatively simple with use of restriction enzymes [99-103] that will cleave DNA based upon specific sequences. However there are limited ways to generally cut DNA. With RNA, the situation is reversed. Many enzymes can cut RNA in general areas [97, 98, 104, 105]. For example, RNase T₁ is capable of cutting RNA at every unmodified guanine residue. This is useful for sequencing RNA, but sometimes we want to determine where an RNase T₁ fragment is located in a larger strand of RNA. To do that, an approach that can cut RNA based upon the RNA sequence, thereby allowing control of the size of RNA generated, must be used.

The method presented here is based upon use of RNase H [106-115], an enzyme that can cleave RNA wherever there is a DNA/RNA hybrid. This allows for the researcher to decide the sequence of DNA to be annealed and therefore control where the RNA is hydrolyzed. After the RNA is cleaved, the resulting fragment can be separated using poly-acrylamide gel electrophoresis (PAGE). To confirm the identity of each band, the RNA from each band can be cleaved by RNase T₁ and a map of the generated fragments can be compared with the predicted fragments. In this work I will apply this approach to 23S rRNA from *E. coli*. 
7.2 Materials and Methods

7.2.1 Reagents

*E. coli* was cultured and 23S rRNA was separated, purified by precipitation and the concentration was determined by UV absorbance. DNA oligonucleotides were synthesized on an Expidite 8909 DNA synthesizer (PerSeptive Biosystems, Framingham, Ma), purified by cartridge purification, and the concentration was determined by UV absorbance. Tris, MgCl₂, Stains-All, NH₄OAc, urea, gel loading solution, RNA markers and EDTA were purchased from Sigma-Aldrich (St. Louis, Mo). MALDI-quality THAP and DAC were obtained from Fluka (St. Louis, Mo). The 10x TRIS/Borate/EDTA (TBE) buffer was purchased from Promega (Madison, Wi); acrylamide and bis-acrylamide were purchased from Amersco (Solon, OH) and RNase H and RNase T₁ were purchased from USB (Cleveland, OH). All water was deionized and autoclaved prior to use.

7.2.2 RNase H reaction

Each reaction contained 100 µL of 0.7 µM purified 23S rRNA, which is then heated to 95 °C for 10 minutes. Immediately after being removed from heat, selected DNA oligonucleotides, 20 µL of 37.5 µM for each, were added to the RNA and the solution was allowed to cool slowly to room temperature. Once the RNA had annealed to the DNA, 200 µL of 40 mM Tris, pH 8.0 and 350 µL of 10 mM MgCl₂ were added to the reaction. After vortexing the solution, 50 µL of 5 unit/µL RNase H enzyme was added and the solution was vortexed again. The reaction was then allowed to proceed at 20 °C for 3-4 days.
7.2.3 Polyacrylamide Gel Electrophoresis

To separate fragments of 23S rRNA between 100 and 300 bases in length, a 7 M urea, 7 % acrylamide/bis-acrylamide denaturing gel was run in a Hoeffer 660 vertical gel electrophoresis unit (Amersham Bioscience, Piscataway, NJ) with dimensions 160 mm x 120 mm x 3 mm. The gel was run with 1X TBE at 300 V, 30 mA for 3 hours. Bands were then visualized using Stains-All dye and scanned with a desktop scanner.

7.2.4 RNA removal from Gel

Gel bands were excised from the gel with a clean, sharp razor blade and then crushed. 5 mL of elution buffer, 20 mM Tris (pH 7.6), 1 M NH₄OAc, and 2 mM EDTA was added and the gel was allowed to soak at 37 °C for 17 hours. The gel slurry was transferred to 0.22 μm pore size Steri-flip vacuum filtration units (Millipore, Bedford, Ma). The filtrate was collected and precipitated with ethanol and NH₄OAc overnight. Dried RNA was dissolved in 100 μL of sterile H₂O and saved for analysis. UV analysis of standard RNAs showed that recovery from the gel was approximately 70%.

7.2.5 RNase T₁ digestion, HPLC separation, and analysis by MALDI-MS

A 70 μL (~0.3 nmol) aliquot of RNA obtained above was reacted in 50 μL of a 50 mM Tris-HCl, 1 mM EDTA solution containing 25 U RNase T₁ at 37 °C for 30 min. 100 μL of this solution was then injected onto a Supelcosil C-18 reversed phase HPLC 4.6x125 mm column (Supelco, St. Louis, Mo) run on a Beckman Gold Nouveau HPLC system (Fullerton, Ca) with a 2% per minute gradient from 100% 25 mM TEAB (pH 6.0) to 100% of 40:60 acetonitrile/water with UV detection at 260 nm. The peaks were collected and dried down, and then reconstituted in 3 μL of sterile, deionized water for
analysis by MALDI-MS. A 1 μL aliquot of digested RNA was mixed with 1 μL of matrix (0.325 M THAP and 0.225 M DAC) and spotted onto the MALDI plate. MALDI analysis was done using a Bruker (Billerica, MA) ReFlex IV MALDI-TOF instrument equipped with an N₂ laser operated in negative-ion mode and reflectron TOF mode. Accelerating voltage was set at 20 kV and laser power was set to the minimum level necessary to generate a reasonable signal (threshold). A three point calibration using dT₅, dT₁₅, and dT₂₀ was used for all analyses.

7.3 Results

To selectively cut RNA in a specific region using RNase H, it was essential to choose DNA oligonucleotides which can effectively anneal with RNA and will stay annealed during the reaction. According to the literature [106-115], oligonucleotides that are 10 bases in length are optimum for annealing to RNA. Other considerations include the % GC content of the oligonucleotide and the melting temperature of the RNA/DNA hybrid. Also, it is important to note that with a DNA/RNA hybrid, the RNase H enzyme starts cutting after the fourth DNA base and then continues until the tenth. Thus the ends of the cleavage products are not exact. Therefore, it is important to choose DNA oligonucleotides that are far enough away from the region of interest so that the RNA edges will not complicate the analysis. The DNA oligonucleotides used in this study and their characteristics are shown in Table 7.1.

The addition of the DNA oligomers from Table 7.1 allows for 23S rRNA to be cut into nine fragments that differ in length by at least 5 nucleotides (nt) to ease separation by PAGE. The expected fragments are shown in Table 7.2. Only the fragments between 100 and 300 nt long, fragments D-I, were analyzed by RNase T₁.
Table 7.1. Sequence, melting temperature, % guanine/cytidine composition and molecular weight information for the selected DNA oligonucleotides.

<table>
<thead>
<tr>
<th>ID</th>
<th>5' --&gt; 3' DNA sequence</th>
<th>Tm (°C)</th>
<th>% GC</th>
<th>MW (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>ATATCTGGGC</td>
<td>25.6</td>
<td>50</td>
<td>3043.0</td>
</tr>
<tr>
<td>II</td>
<td>AAGAGGGGGCT</td>
<td>33.0</td>
<td>50</td>
<td>3117.1</td>
</tr>
<tr>
<td>III</td>
<td>CCAATTCACC</td>
<td>30.0</td>
<td>50</td>
<td>2932.0</td>
</tr>
<tr>
<td>IV</td>
<td>TCACCTGCAT</td>
<td>30.2</td>
<td>50</td>
<td>2963.0</td>
</tr>
<tr>
<td>V</td>
<td>GACAGAGGTG</td>
<td>28.1</td>
<td>60</td>
<td>3117.1</td>
</tr>
<tr>
<td>VI</td>
<td>TCCCGGTTAGC</td>
<td>35.3</td>
<td>70</td>
<td>3004.0</td>
</tr>
<tr>
<td>VII</td>
<td>GACTTCTTACC</td>
<td>22.4</td>
<td>50</td>
<td>2963.0</td>
</tr>
<tr>
<td>VIII</td>
<td>CTTTGAACG</td>
<td>26.6</td>
<td>50</td>
<td>3043.0</td>
</tr>
</tbody>
</table>

PAGE separation of the RNase H fragments of lengths 100 to 300 nt is shown in Figure 7.1. It should be noted that in all cases, no evidence of intact 23S rRNA appear in the gel, so it seems that RNase H reaction went to completion in the time frame allowed. One of the problems in developing this approach was to optimize the PAGE conditions and the elution conditions. This will be discussed later in the chapter.

Table 7.2. RNA lengths generated from DNA/RNA hybrids and where in 23S rRNA sequence they occur.

<table>
<thead>
<tr>
<th>RNA ID</th>
<th>RNA Length</th>
<th>DNA ID</th>
<th>5' Start Position</th>
<th>3' End Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>792</td>
<td>III, IV</td>
<td>1022</td>
<td>1814</td>
</tr>
<tr>
<td>B</td>
<td>668</td>
<td>I</td>
<td>1</td>
<td>669</td>
</tr>
<tr>
<td>C</td>
<td>421</td>
<td>V, VI</td>
<td>1995</td>
<td>2416</td>
</tr>
<tr>
<td>D</td>
<td>235</td>
<td>VII, VIII</td>
<td>2532</td>
<td>2767</td>
</tr>
<tr>
<td>E</td>
<td>207</td>
<td>II, III</td>
<td>814</td>
<td>1021</td>
</tr>
<tr>
<td>F</td>
<td>179</td>
<td>IV, V</td>
<td>1815</td>
<td>1994</td>
</tr>
<tr>
<td>G</td>
<td>143</td>
<td>I, II</td>
<td>670</td>
<td>813</td>
</tr>
<tr>
<td>H</td>
<td>136</td>
<td>VIII</td>
<td>2768</td>
<td>2904</td>
</tr>
<tr>
<td>I</td>
<td>114</td>
<td>VI, VII</td>
<td>2417</td>
<td>2531</td>
</tr>
</tbody>
</table>
Figure 7.1. PAGE separation of RNase H cleavage products. Lanes 1-4 and 6-9 are cleavage products D-I from 23S rRNA of *E. coli*. Lane 5 is RNA size marker from 100-300 nt in length.

Once the RNase H cleavage products are isolated from the gel, they are digested using RNase T\textsubscript{1}. RNase T\textsubscript{1} digested oligonucleotides were separated by HPLC and collected for analysis by MALDI-MS. Figure 7.2 shows the separation of the RNase T\textsubscript{1} digestion products generated from the 136 nt rRNA fragment. While the separation does not show baseline resolution, a number of peaks are seen in the HPLC. It is apparent

![HPLC separation of RNase T1 digestion of Fragment H from the gel. Peak labeled with (*) at 17.4 minutes is suspected to have several oligonucleotide fragments.](image)

Figure 7.2. HPLC separation of RNase T\textsubscript{1} digestion of Fragment H from the gel. Peak labeled with (*) at 17.4 minutes is suspected to have several oligonucleotide fragments.
from the MS analysis of the 17.4 minute peak, that there is co-elution of many of the oligonucleotide fragments (Figure 7.3). However, a large number of peaks corresponding to oligonucleotide fragments from the 136-mer are observed. Based upon the RNase T1 digestion products that were predicted (Table 7.3), over 70% of the possible oligonucleotide fragments were seen in the analysis, several of which are specific only to the 136-mer. Oligonucleotide RNase T1 fragments that had a one to one molar ratio with the parent 136-mer were not detected by MALDI-MS. Further optimization of the conditions for separation by gel electrophoresis and HPLC may improve sample recovery.

<table>
<thead>
<tr>
<th>RNase T1 fragment</th>
<th>Expected m/z</th>
<th>Experimental m/z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gp</td>
<td>363.22</td>
<td>n.d.</td>
</tr>
<tr>
<td>CG&gt;p</td>
<td>650.39</td>
<td>650.01</td>
</tr>
<tr>
<td>UG&gt;p</td>
<td>651.38</td>
<td>651.10</td>
</tr>
<tr>
<td>AG&gt;p</td>
<td>674.42</td>
<td>672.99</td>
</tr>
<tr>
<td>CCGp</td>
<td>973.59</td>
<td>n.d.</td>
</tr>
<tr>
<td>UUG&gt;p</td>
<td>957.55</td>
<td>958.21</td>
</tr>
<tr>
<td>CAG&gt;p</td>
<td>979.60</td>
<td>978.91</td>
</tr>
<tr>
<td>ACGp</td>
<td>979.60</td>
<td>978.91</td>
</tr>
<tr>
<td>AUG&gt;p</td>
<td>980.59</td>
<td>979.25</td>
</tr>
<tr>
<td>AAG&gt;p</td>
<td>1003.63</td>
<td>1004.29</td>
</tr>
<tr>
<td>CUUG&gt;p</td>
<td>1262.73</td>
<td>1263.07</td>
</tr>
<tr>
<td>AACGp</td>
<td>1326.82</td>
<td>1325.60</td>
</tr>
<tr>
<td>UAAAGp</td>
<td>1327.81</td>
<td>1326.74</td>
</tr>
<tr>
<td>AUAGp</td>
<td>1327.81</td>
<td>1326.74</td>
</tr>
<tr>
<td>CCCCCGp</td>
<td>1583.96</td>
<td>n.d.</td>
</tr>
<tr>
<td>UCCUGp</td>
<td>1585.93</td>
<td>1587.22</td>
</tr>
<tr>
<td>AACCgp</td>
<td>1632.01</td>
<td>1634.10</td>
</tr>
<tr>
<td>CUAACCG&gt;p</td>
<td>2225.35</td>
<td>2224.01</td>
</tr>
<tr>
<td>ACUCCUUGp</td>
<td>2526.49</td>
<td>n.d.</td>
</tr>
<tr>
<td>UACUAAUGp</td>
<td>2574.54</td>
<td>n.d.</td>
</tr>
<tr>
<td>UUCUCCCUGp</td>
<td>2808.63</td>
<td>n.d.</td>
</tr>
<tr>
<td>CUUAAACCUU&gt;p</td>
<td>2798.65</td>
<td>2799.10</td>
</tr>
</tbody>
</table>
Figure 7.3. MALDI-MS spectrum of 17.4 minute peak from the HPLC separation. Peaks labeled with (*) correspond to RNase T1 fragments of the 136-mer.

While developing this approach, several attempts were unsuccessful. Reasons for this include: poor separation in the gel, poor elution from the gel, inability to separate the gel components from the RNA, and poor gel polymerization. However, it was discovered

Figure 7.4. UV scan comparing a viable and non-viable RNA extracted from gel.
that the RNA could be efficiently monitored before and after the separation by a UV scan from 300 to 200 nm. A comparison between clean, viable RNA removed from a gel and RNA that can not be used is shown in Figure 7.4. It was also found that if the RNA removed from the gel is not viable, numerous ethanol precipitations, centrifuge filtrations, and C18 filtrations would not be able to separate the RNA from its contaminants. However, if the gel and elution were successful, this would be immediately apparent by UV.

There are several other important considerations to note for the RNase H digestion. First, there was no non-specific binding of DNA oligomers. Second, DNA specific binding occurred regardless of RNA secondary structure. Finally, DNA oligomers bind independently of the presence of other oligomers.

7.4 Conclusions

While it is evident that further optimization is needed, RNase H can be used to specifically cleave large RNA. These cleavage products can then be separated and prepared for mass spectrometric analysis. Now that the RNase H cleavages have been made, it is possible to use the other generated fragments for the analysis of pseudouridine in 23S rRNA.
CHAPTER 8: Analysis of pseudouridine in 23s rRNA from *E. coli*

8.1. Introduction

Before studies into the number and placement of pseudouridine residues in 23S rRNA could go forward, it was necessary to make sure I could successfully analyze pseudouridine in smaller RNAs, such as tRNA and 16S rRNA. Then, because of the very large size of 23S rRNA (2900 bases, compared to either tRNA or 16S rRNA, 76 and 1600 bases respectively), I had to develop an approach to accommodate rRNA of this size.

Previously for tRNA and 16S rRNA, analysis depended upon cleavage of the intact RNA into a number of small oligonucleotide fragments using an endonuclease such as RNase T₁ or U₂ [104, 105]. This method works quite well for tRNA, because it generates approximately 15-20 manageable and detectable oligonucleotide fragments. For 16S rRNA, this method could be used as long as a second separation technique was added to handle the large number of oligonucleotide fragments, approximately 300 generated by RNase T₁. However, even using this additional separation method, it was clear that localization of a pseudouridine in a large RNA would be problematic without a better way to cleave the RNA. It was essential to have a method to cleave RNA using a set of specific cleavage sites, sites that ideally would be selected by the researcher.

RNase H provides the best way to attack this problem. As noted in the previous chapter, RNase H can cleave RNA wherever there is a DNA/RNA hybrid present [106-115]. In the previous chapter, an RNase H approach suitable for use with mass spectrometry was shown with cleavage sites that were specifically designed to cleave 23S
rRNA derived from *E. coli* [20, 21, 26, 116, 117]. These sites were specifically chosen to cleave 23S rRNA into several small (100-300 residues) fragments that were readily amendable to previous methods for analyzing pseudouridine. Cleavage sites were chosen to create oligonucleotides, no longer than 300 bases, that encompass the known modification-rich areas of 23S rRNA.

There are ten known pseudouridine residues in 23S rRNA [33-35]. The 5' half of 23S rRNA has 2 pseudouridine residues located at positions 746 and 955. To analyze the modification at 746, DNA oligomers are added to cleave RNA at positions 669 and 813, which generates an oligonucleotide that is 143 bases long and includes pseudouridine at position 68 of the oligonucleotide. For pseudouridine 955, the DNA oligomer that cleaves at 813 was used with another oligomer that cleaves at position 1019 to generate RNA that is 207 bases in length. The pseudouridine residue in the 207-mer is located at position 142.

There are eight pseudouridine residues located in the 3' half of the 23S rRNA. These pseudouridine modifications are found in three relatively small groups. In the first group, the pseudouridines are located at positions 1911, 1915, and 1917 and can be handled together by using DNA oligomers which cleave at positions 1814 and 1994 of the 23S rRNA. This generates a 179-mer with the pseudouridines located at positions 77, 79, and 83. The second group of pseudouridines is located at positions 2457 and 2504, and is handled with RNase H cleavage at 23S positions 2416 and 2531, which create a fragment that is 114 bases long and contains pseudouridine at positions 27 and 74 of the new oligonucleotide. The final group has pseudouridines is located at positions 2580, 2639, and 2640 and also uses the cleavage at 23S position 2531 and another cleavage
point at 2767. This places the pseudouridines in positions 162, 163 and 187 of a 235-mer. Previous analysis methods have had difficulty looking at modifications 2639 and 2640 using reverse transcriptase, as it is difficult to develop transcription primers that start specifically at position 2640. The first group of pseudouridine residues is handled with DNA oligomers which cleave at 23s rRNA positions 1814 and 1994 with the pseudouridines located at positions 77, 79, and 83. The second group of pseudouridine residues are handled with RNase H cleavage at 23s positions 2416 and 2531 leaving a 115-mer with pseudouridines at positions 27 and 74. The final group of pseudouridines also uses the 23S rRNA cleavage at 2531 and another at 2767. This places the pseudouridines in positions 162, 163, and 187 of a 235-mer. One reasonably sized cleavage product also exists at the very 3’ end of the molecule, a 136-mer that contains no pseudouridines, but can easily be analyzed as a control to determine that the cleavages were made in the correct location.

Once RNase H cleavages products are formed, they are separated by polyacrylamide gel electrophoresis (PAGE) and eluted from the gel. These RNase H cleavage fragments are then handled in the same manner as previously reported for tRNA. From this point, the only difference between this procedure and the tRNA procedure is that these cleavage products are still too large to be analyzed before and after CMC addition while they are intact, so after the CMC reaction, RNase T1 or RNase U2 will be applied followed by RP-HPLC separation and MALDI confirmation of the products. Several fragments contain pseudouridine residues but not uridine residues after cleavage by RNase T1 and RNase U2, and these will not need testing with exonucleases
such as snake venom phosphodiesterase, as their positions will be known based upon the known sequence.

8.2. Methods and Materials

8.2.1 Materials

_E. coli_ MRE-600 was cultured in the laboratory according to procedures discussed in Chapter 5. All water used was distilled, deionized and sterilized. All buffers, solvents, and matrices were purchased from either Sigma-Aldrich (St. Louis, MO) or Fisher Scientific (Pittsburg, PA). RNase H was purchased from USB (Cleveland, OH); RNase U2 was from Industrial Research Limited, (Auckland, New Zealand). RNase T1 was purchased from Roche Pharmaceuticals, (Nutley, NJ). Acrylamide and bis-acrylamide was purchased from Amersco (Solon, OH).

8.2.2 RNase H digestion and PAGE separation

RNase H digestion, PAGE separation and sample elution from PAGE bands were done according to the procedure outlined in Chapter 7 with no changes.

8.2.3 CMC and Control reaction

CMC and control studies were done as described in Chapter 3 for RNA of similar size to tRNA. 40% of the available sample was used for the control reaction and the remaining 60% of the sample was used for the CMC derivatized sample.

8.2.4 RNase T1 and U2 digestions, HPLC separation, and MALDI analysis

RNase T1 reactions were done as described in Chapter 3. RNase U2 reactions were done by diluting the CMC and control sample into 10 µL of sterile water, taking 8
µL of the sample and diluting it with 10 µL of a sodium citrate buffer composed of 20 mM sodium citrate (pH 5.0±0.2), 1 mM EDTA, and 7 M urea. The pH of the sodium citrate solution should be checked daily, and if pH differs by more than 0.2 units, it should be remade so that cleavage will occur only at adenine residues. After dilution with the sodium citrate buffer, 50 units of RNase U₂ are added and the reaction is allowed to proceed for 15 minutes at 50 °C. Samples are then ready for injection into HPLC for separation.

HPLC conditions are done as outlined in Chapter 3; however, the composition of the buffers has been changed to promote better separation on the HPLC column. Buffer A is unchanged (25 mM TEAB (pH 6.0) in 100% H₂O); however, buffer B is 25 mM TEAB (pH 6.0) in 50% acetonitrile, 50% H₂O. Both buffers are prepared under sterile conditions and are run using a linear gradient from 0 to 100% B buffer in 100 minutes at a flow rate of 1 mL/min with UV detection at 260 nm.

MALDI conditions are as discussed in Chapters 3 and 7. Fractions collected from the HPLC that are not needed for exonuclease reaction are dissolved in 0.5 µL of sterile water and mixed with 0.5 µL of matrix and allowed to dry. Samples needed for exonuclease digestion are dissolved in 1.0 µL of sterile water and 75% of the sample is used for MALDI-MS analysis.
8.3. Results

8.3.1 RNase H reaction and PAGE analysis

As shown in Chapter 7, 6 RNase H fragments were expected to be between 100 and 250 bases in length. The fragment lengths were 115, 134, 146, 180, 207 and 237 bases and are easily resolved using the gel conditions outlined in the previous chapter.

As stated earlier, five of the six bands that were observed contained at least one pseudouridine modification. The one band that does not have pseudouridine, the 134-mer, was used as a control to make sure that the RNase H digestion had worked as expected. First, the bands were extracted from the gel; then separated and purified by ethanol precipitation. After purification, the bands were separately analyzed for their

Figure 8.1. Comparison of RNase H products before and after PAGE separation.
concentrations. A comparison scan of the RNase H mixture before the gel and the different gel bands by UV analysis from 300 to 200 nm is shown in Figure 8.1.

8.3.2 CMC and control reactions

Each band was then reacted using conditions for tRNA with CMC. As previously stated, 40% of the sample was used for a control, while the remaining 60% was used for the CMC reactions, as it is likely that the samples will not be completely modified. After addition of base, the pseudouridine-containing fragments were then precipitated and analyzed by UV to determine if the fragments that were supposed to have CMC attachment showed the increase in UV absorbance expected when CMC is attached to a nucleotide. As shown in Figure 8.2, the 134-mer did not show any increase in UV absorbance between the control and the CMC-modified sample, as expected.

Figure 8.2. UV comparison scans between the control and the CMC modified 136-mer.
Table 8.1. Predicted RNase T₁ and U₂ fragments containing pseudouridine, molecular weights before and after CMC addition.

<table>
<thead>
<tr>
<th>Band</th>
<th>Number</th>
<th>23S rRNA Positions</th>
<th>RNase</th>
<th>fragment</th>
<th>Predicted mass (Da)</th>
<th>Mass after CMC (Da)</th>
</tr>
</thead>
<tbody>
<tr>
<td>114</td>
<td>2</td>
<td>2457, 2504</td>
<td>RNase U₂</td>
<td>GΨCGGAp</td>
<td>1976.19</td>
<td>2228.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RNase U₂</td>
<td>CUCGGCU₂ΨAGC₉UmGC₉Ap</td>
<td>5177.14</td>
<td>5429.61</td>
</tr>
<tr>
<td>143</td>
<td>1</td>
<td>746</td>
<td>RNase T₁</td>
<td>ACU₉Um₁GΨTGp</td>
<td>3235.96</td>
<td>3488.43</td>
</tr>
<tr>
<td>179</td>
<td>3</td>
<td>1911, 1915, 1917</td>
<td>RNase T₁</td>
<td>CAA₉ΨA₉ΨCAA₉Gp</td>
<td>3520.13</td>
<td>4277.54</td>
</tr>
<tr>
<td>207</td>
<td>1</td>
<td>955</td>
<td>RNase U₂</td>
<td>CGGC₉GGG₉ΨGCU₉Ap</td>
<td>3928.33</td>
<td>4180.48</td>
</tr>
<tr>
<td>236</td>
<td>3</td>
<td>2580, 2639, 2640</td>
<td>RNase U₂</td>
<td>UUUGGG₉ΨCGAp</td>
<td>3239.91</td>
<td>3492.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RNase U₂</td>
<td>UCCCU₉GGG₉ΨGAp</td>
<td>3810.25</td>
<td>4315.19</td>
</tr>
</tbody>
</table>

For the remaining bands, there are 10 pseudouridine residues. Several of the bands contain two or three pseudouridines per band, while some bands only have one. Table 8.1 lists the length of the band, along with the number of pseudouridines, and the position of the pseudouridines from 23S rRNA and the projected RNase T₁ and U₂ digestion products for each of the pseudouridine containing fragments.

The 114-mer band was expected to have two pseudouridine residues (originally, positions 2457 and 2504) and therefore have two CMC modifications. As seen in Figure 8.3, there is a six-fold increase in absorbance at 260 nm after CMC derivatization. There is also the existence of some fine band structure from 270 nm to 250 nm that was previously attributed to the CMC counter ion p-toluenesulfonate. This feature has been discussed in Chapter 4. It is evident that the cleaning was not complete by ethanol precipitation at there is, in addition to the anion, a high absorbance from 240 nm to 200
Figure 8.3. UV scan comparison between the 114 control sample and CMC sample.

nm that is presumed to be the result of buffers which contain EDTA and urea, as they are not present in samples that have not been exposed to these buffers.

Figure 8.4. HPLC chromatograms from control and CMC species from the 114 band. Peaks labeled with (*) indicate UV increases between control and CMC samples.
Figure 8.4 shows the comparison control and CMC chromatograms after RNase U₂ digestion of the 114 band. Two peaks in the CMC chromatogram are greatly enhanced from the control chromatogram. The peak that elutes at 17.4 minutes shows a nine-fold increase compared to the control, and which may corresponds to the GΨCGGAp fragment. The 23.8 minute peak shows a seven-fold increase from the control. However, due to limited sample quantities, the peak at 23.8 minutes could not be observed by MALDI-MS.

The 143-mer was expected to contain one pseudouridine residue from position 746 of 23S rRNA. As with the 114 band, 40% was used for the control and 60% for the 143 control reaction

143 CMC reaction

Figure 8.5. UV scan comparison of the 143-mer showing a tenfold increase between control and CMC samples.
CMC reaction. As seen in Figure 8.5, there is a ten-fold increase in absorbance at 260 nm after CMC derivatization. There was a similar pattern due to the anion counter ion p-toluenesulfonate and a large absorbance from 240 to 200 nm, again showing that cleaning procedures need to be modified to better remove these contaminants.

Figure 8.6. HPLC chromatogram from the 146-mer. Peak labeled (*) at 14.0 minutes in the CMC sample shows an eight-fold increase in absorbance.

Figure 8.6 shows the chromatogram for the RP-HPLC separation of both the control and the CMC-modified species after RNase T₁ digestion. Only one observed peak increases dramatically (0.12 AU vs. 0.015 AU), as predicted, at 14.0 minutes, which should correspond to ACUAAUm¹GΨTGp.
The 179-mer was expected to contain three pseudouridine residues from positions 1911, 1915 and 1917 of 23S rRNA. As seen in Figure 8.7, there is a three-fold increase in absorbance at 260 nm after CMC derivatization.

Figure 8.7. UV scan comparison between the control and CMC reaction for the 179-mer.

The 179 band was digested with RNase T1 and one HPLC peak at 25.2 minutes increased ten-fold between the control and the CMC reacted species as seen in Figure 8.8. There are three pseudouridines, which are contained within the same RNase T1 fragment, CAAΨAΨCAAΨGp.

The band that corresponds to the 207-mer has three pseudouridines, and Figure 8.9 shows the comparison of the UV scans of the control and CMC reacted species before RNase U2 digestion. The absorbance at 260 nm observed in the control species is 0.500
Figure 8.8. HPLC separation of the 179-mer. Peak labeled (*) at 24.0 minutes shows a large increase between the CMC and control samples.

AU, where the absorbance is 2.50 AU for the CMC modified species, showing a fivefold increase. When the 207-mer is digested with RNase U₂ there are 25 fragments that are generated, one of which contains pseudouridine. The sequence of this fragment after RNase U₂ digestion is CGGCGGGΨGCUAp.

The control and CMC-reacted digest product separations from the 207-mer are shown in Figure 8.10. One peak is expected to significantly increase, and it is observed eluting at 22.2 minutes, while the control does not show a corresponding peak and given the low intensity peaks, the sample maybe so dilute that the peak is undetectable by UV. The 235-mer contains three pseudouridines that upon digestion with RNase U₂ produce two digestion products that containing pseudouridine. The sequence of these fragments is UUUUGGGΨCGAp and UCCCCUGGCΨΨGAp, the latter containing a fragment that
Figure 8.9. UV comparison scan between the control and CMC of the 207-mer showing a four-fold increase at 260 nm.

The UV scan comparison between the control and CMC derivatized species before RNase U2 digested is shown in Figure 8.11. The control reaction shows an absorbance of 0.600 AU while the CMC-derivatized sample shows an absorbance of 0.500 AU.

Figure 8.10. HPLC separation of RNase U2 products from the 207-mer CMC and control reaction. The peak at 22.0 minutes labeled with (*) denotes the fragment suspected to have CMC attachment.
absorbance of 2.00 AU at 260 nm, a 3.3-fold increase. Insufficient sample was available for HPLC analysis.

Figure 8.11. UV scan comparison between the control and CMC reaction from the 235-mer.

8.4. Conclusions

The approach presented in this chapter demonstrates that RNase H cleavage of 23S rRNA followed by CMC-derivatization, RNase T1 or U2 digestion, and HPLC analysis is suitable for preparing pseudouridine containing oligonucleotides for mass spectrometric analysis. While these results are encouraging, more optimization is required to fully characterize pseudouridines in 23S rRNA. More specifically, the separation of the larger oligonucleotides must be optimized to allow the most amount of sample to be available for analysis.

At this time, there is not enough evidence to determine the relationship between the addition of CMC and the increase in absorbance. Table 8.2 shows the comparison between the number of CMC molecules in the oligonucleotide, the increase before RNase
Table 8.2. Comparing the increase between UV absorbance due to scan of the intact oligomers vs. the absorbance seen due to HPLC separation of fragments.

<table>
<thead>
<tr>
<th>Band</th>
<th>Ψ</th>
<th>Position</th>
<th>UV scan increase</th>
<th>Chromatographic Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>115</td>
<td>2</td>
<td>2457</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2504</td>
<td>&quot;</td>
<td>7</td>
</tr>
<tr>
<td>150</td>
<td>1</td>
<td>746</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>180</td>
<td>3</td>
<td>1911, 1915, 1917</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>207</td>
<td>1</td>
<td>955</td>
<td>5</td>
<td>IND</td>
</tr>
<tr>
<td>235</td>
<td>3</td>
<td>2580</td>
<td>3.3</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2639, 2640</td>
<td>&quot;</td>
<td>n/a</td>
</tr>
</tbody>
</table>

T₁ or U₂ cleavage and the increase of the individual peaks during separation. While there is not a direct relationship between the number of pseudouridines and increases in absorbance at 260 nm, for all of the oligonucleotides analyzed in this work significant increases were found.
Chapter 9: Conclusions

9.1 Final Conclusions to Pseudouridine Project

This body of work presented here has been done to try to develop a method to quickly and efficiently analyze pseudouridine in rRNA of any size, without the problems that are experienced in many of the current methods. At this time, there are a few things which need to be examined before this work can be done efficiently on an unknown sample. The efficiency of the RNase H method needs to be optimized to ensure: 1) complete digestion, and 2) optimal separation of fragments. These limitations do not prevent the current analysis of pseudouridine.

This method does not suffer from the constraints that hinder many other available methods. The detection in this method is based upon the generation of a new band, as opposed to other techniques that are based upon the lack of a band, either electrophoretic or mass spectrometric. This method should not suffer from the problem of back-to-back base determination, because even if a single exonuclease cannot cleave between the pseudouridine residues once they have been modified with CMC, two exonucleases can be used to determine the sequences from both sides, leaving two bases in the middle that correspond to the mass of two pseudouridines and two CMC additions. The final advantage of this method is the speed and accuracy that this method can be performed with. This method can be completed, even in the case of 23S rRNA, in one week for all of the pseudouridines present given sufficient sample concentration. Mass accuracies are shown to be within 1 Da and are highly reproducible. A flow chart of this method is shown in Figure 9.1.
9.2 Future Work

There are several things that can be done to increase the efficiency and the use of this method. The main issues lie in the RNase H reaction. Improvement of the separation and recovery of RNase H fragments are the highest priority. When the RNase H reaction is performed with a pure DNA oligomer, cleavage occurs at several points along the RNA sequence. In fact, it occurs at every residue after the fourth DNA base.
This leaves RNA with various lengths that differ by one base that can blur separation bands by gel electrophoresis, and decrease separation resolution. One way to avoid this problem is through the use of a chimera [118]. A chimera is a molecule which incorporates both DNA and RNA bases. The sequence of chimeras includes four DNA bases followed by 3 RNA bases followed by another three DNA bases. The RNase H enzyme cleaves after the fourth DNA base, but cannot cleave between an RNA/RNA hybrid, so cleavage occurs in only one position.

Another way that the method could be improved is by changing the separation technique. Antibodies selectively for CMC-modified DNA have been raised [119], and after the reaction of CMC with an RNA that contains pseudouridine, these antibodies could be employed to clean and preconcentrate the solution by removing the CMC-oligonucleotides. The combination of chimeras and antibodies could create RNase H fragments of similar lengths and pure CMC-modified samples. This would enable separation techniques to have sharper, clearer bands that can be separated with more efficiency. Other techniques for separation could be employed to provide a higher percentage of recovery and a faster separation.

Another possible separation technique could be gel permeation chromatography. Gel permeation chromatography experts have developed columns to deal specifically with nucleic acids; however, resolution is not as good as gels. This can be overcome by choosing DNA or chimeras that provide a larger difference in length between RNase H cleavage fragments. Another method that could help lower the limit of detection and reduce the need for more efficient separation would be the use of nano-ESI FTICR/MS. This type of mass spectrometry has the ability to analyze RNA samples at a level of pico-
to femtomole levels, which can enhance the detection of pseudouridine-CMC adducts even with peaks that are difficult to detect by UV. Separation and analysis of RNase T₁ and U₂ fragments could be done with either a capillary-LC/MS system, or on a microchip that is coupled with a mass spectrometer. With a nano-ESI FTICR/MS system, the use of exonucleases may not be necessary, as the FTICR has the capability of tandem mass spectrometric analysis, and so structural and sequence information could all be obtained at the same time.

Finally, there are some questions about the CMC-nucleoside reaction that could be addressed. While these questions are not essential to the development of this method, addressing these questions may provide researchers a way of enhancing the method. Among those questions is how the bond between CMC and a nucleotide is made, through nitrogen or oxygen. Discovering whether the CMC reaction is dominated by thermodynamics, kinetics, steric; or if reaction is governed by a combination of all of these factors. Another aspect that could be explored is relationship between the CMC-nucleotide bond and the UV absorbance. My data has shown that there is no observable pattern at this point between the increase in absorbance and the number of CMC attachments that has been made. Determination of the extinction coefficient and other UV studies may help provide insight for the detection of this bond. Another avenue that could be addressed would be the investigation of other carbodiimide compounds that could be used to specifically modify pseudouridine, but do not have permanent charges. Optimally, there would be no charge on the carbodiimide, but a negative charge would coincide with the negative charge of an oligonucleotide backbone. It is the opinion of this researcher, that the positive charge on the CMC has made the analysis of these
adducted oligonucleotides by mass spectrometry much more difficult than unmodified oligonucleotides. The removal of excess salts, buffers and CMC has been more difficult as a result of the positive charge. Also, separation of fragments has been difficult because of the presence of both a permanent positive charge from the CMC adduct, and a negative charge resultant from the phosphate backbone. With a different modifying agent, the analysis of these compounds may be easier. However, this carbodiimide is water-soluble, while most other compounds of this type are not, so identification of a suitable carbodiimide would most likely need to be designed and synthesized in the lab.


86. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J., **Isolation and Biologically Active Ribonucleic Acid from Sources Enriched in Ribonuclease.** Biochemistry, 1979. **18**: p. 5294-5299.


95. Manz, W., Amann, R., Ludwig, W., Vancanneyt, M., and Schleifer, K.H., **Application of a suite of 16S rRNA-specific oligonucleotide probes designed to**


