I, C. Callie Coombs, hereby submit this work as part of the requirements for the degree of:

Master of Science

in:

Chemistry

It is entitled:

Developing Liquid Chromatography and Mass Spectrometry Approaches for Studying Posttranscriptional Modifications in Small RNAs

This work and its defense approved by:

Chair: Dr. Patrick Limbach

Dr. Albert Bobst

Dr. Joseph Caruso
Developing Liquid Chromatography and Mass Spectrometry Approaches for Studying Posttranscriptional Modifications in Small RNAs

A thesis submitted to the
Division of Research and Advanced Studies of the University of Cincinnati
In partial fulfillment of the Requirements for the degree of

Master of Science (M.S.)
in the Department of Chemistry of the College of Arts and Sciences
May 2006
By
C. Callie Coombs
B.A. St. Louis University, 2004

Committee Chair: Patrick A. Limbach
Abstract

The objective of this work is to develop mass-spectrometry based approaches for analyzing modified nucleosides in RNA. A technique for isolating and analyzing miRNA and siRNA is described. While the method succeeds at isolating these RNAs, the amount of time required to isolate enough sample to analyze by MS is unreasonable. Therefore, the method presented here serves to guide others who aim to develop approaches for isolating miRNAs and siRNAs.

A method for analyzing mixtures of small ribonucleic acids for the presence of modified nucleosides using direct infusion electrospray ionization mass spectrometry has been developed. This method is compared to a liquid chromatography mass spectrometry based approach. It has been found that direct-infusion mass spectrometry of nucleoside digests allows for detection of all mass-shifting modifications present within a standard tRNA molecule enzymatically digested to nucleosides. However, the sensitivity and dynamic range are insufficient to characterize more complex RNA mixtures. This can instead be accomplished by performing LC-MS experiments on RNA hydrolyzates.
Acknowledgments

I would like to thank my advisor, Dr. Patrick A. Limbach, for his unwavering support during my time here at the University of Cincinnati. During my two years here, I feel that I have grown as a person and as a scientist, and much of this due to his guidance. He has been amazingly patient and understanding with me, and for that I am eternally grateful.

I am thankful to my committee members, Dr. Joe Caruso and Dr. Albert Bobst, for mentoring me during my time at UC. I appreciate their kindness and willingness to help me along my way. Dr. Tom Ridgway has also proved to be an enormous help, always willing to provide advice or candy (peppermint patties) when I am having rough days.

I would also like to thank my parents, Ann and Bill Coombs, for their continual support in regard to my education and my well being. I would also like to thank the rest of my family and friends for being there for me during these years.

Finally, my fellow lab mates in the Limbach group also deserve a big “thank you” because they have helped to provide an enjoyable work environment and have also become great friends of mine over the past couple years.
Table of Contents

Abstract........................................................................................................................................... ii
Acknowledgments............................................................................................................................... iv
List of Figures........................................................................................................................................ 3
List of Tables......................................................................................................................................... 5
List of Abbreviations........................................................................................................................... 6

Chapter 1: The Significance of Ribonucleic Acid Studies................................................................. 8
1.1 Introduction...................................................................................................................................... 8
1.2 Examples of Past Studies Regarding Nucleoside Modifications................................................. 11
   1.2.1 MELAS, a Neurogenenerative Disorder.............................................................................. 12
   1.2.2 MERRF Syndrome, a Muscular Disorder......................................................................... 13
   1.2.3 MLASA, an Oxidative Phosphorylation Disorder............................................................ 14
1.3 The Emerging Role of miRNA and siRNA in Molecular Biology............................................. 15
   1.3.1 The Difference between miRNAs and siRNAs................................................................. 16
   1.3.2 Experimental Techniques for Studying miRNAs and siRNAs......................................... 17
   1.3.3 Methods for Isolating miRNAs and siRNAs.................................................................... 19
1.4 Mass Spectrometry as a Tool for Studying RNA...................................................................... 21

Chapter 2: Developing an Approach for Analyzing Heterogeneous Mixtures of MicroRNAs by LC-UV................................................................. 25
2.1 Introduction...................................................................................................................................... 25
2.2 Experimental.................................................................................................................................... 26
   2.2.1 Materials............................................................................................................................... 26
   2.2.2 Methods............................................................................................................................... 27
       2.2.2.1 RNA Isolation............................................................................................................... 27
       2.2.2.2 Polyacrylamide Gel Electrophoresis........................................................................... 28
           2.2.2.2.1 Preparation of Gel................................................................................................. 28
           2.2.2.2.2 Running and Visualizing the Gel.......................................................................... 28
           2.2.2.2.3 Using a Fluorescently-Labeled Oligonucleotide as a Marker........................... 29
           2.2.2.2.4 Extraction of Bands............................................................................................ 30
       2.2.2.3 Analyzing the Gel Extract............................................................................................ 31
           2.2.2.3.1 Nucleoside Digestion.......................................................................................... 31
           2.2.2.3.2 Verification of a Complete Digestion.................................................................. 32
           2.2.2.3.3 Capillary-LC Analysis........................................................................................ 32
2.3 Results and Discussion................................................................................................................ 33
   2.3.1 RNA Isolation....................................................................................................................... 33
   2.3.2 Polyacrylamide Gel Electrophoresis.................................................................................... 34
   2.3.3 Analyzing the Gel Extract.................................................................................................. 36
2.4 Conclusions..................................................................................................................................... 39
Chapter 3: Comparing Direct ESI-MS to LC-MS for the Analysis of Modified Nucleosides

3.1 Introduction
3.2 Experimental
  3.2.1 Materials
  3.2.2 Methods
    3.2.2.1 Enzymatic Digestion
    3.2.2.2 HPLC-UV of Nucleoside Digests
    3.2.2.3 Direct Infusion Electrospray Mass Spectrometry
    3.2.2.4 Microbore and Capillary-LC Separation of Nucleoside Digests and LC-MS
3.3 Results and Discussion
  3.3.1 Enzymatic Digest
  3.3.2 HPLC-UV of Nucleoside Digests
  3.3.3 Direct Infusion Analysis of Modified Nucleosides
  3.3.4 LC-MS of Modified Nucleosides
  3.3.5 Comparison of Direct Infusion MS to LC-MS
3.4 Conclusion

Chapter 4: Conclusions and Future Perspectives

4.1 Summary
4.2 Future Work

References

References
List of Figures

Figure 1.1 Structures of nucleosides and nucleotides present in RNA [1]......................9
Figure 1.2 Structures of various modified nucleosides..............................................9
Figure 1.3 A description of the numbering system used in the naming of nucleosides…10
Figure 1.4 Structure of the missing nucleoside modification (τm5U) that causes MELAS.................................................................13
Figure 1.5 Structure of τm5s2U, the nucleoside modification absent in mitochondrial tRNA1Lys in patients with MERRF.................................................................14
Figure 1.6 Structure of pseudouridine, the only mass-silent modified nucleoside, which is often not present in tRNA in MLASA patients.......................................................14
Figure 1.7 Basic schematic representations for four possible pathways of a gene: A. is for a non-silenced gene, B. is a gene silenced by miRNA, C. is a gene silenced by siRNA at the mRNA level, and D. is a gene silenced by siRNA at the DNA level.................15
Figure 1.8 Structure of a microRNA hairpin structure, prior to processing..............16
Figure 1.9 A portion of microarray data from miRNAs expressed in different tissues, taken from [35]..............................................................................................................18
Figure 1.10 An example of a Northern blot experiment, taken from [57]..............19
Figure 1.11 An electrospray mass spectrum of a multiply-charged oligonucleotide [49].........................................................................................................................21
Figure 1.12 Diagram of the generation of desolvated ions in the electrospray process…22
Figure 2.1 Structure of the 6-FAM NHS ester, taken from http://www.idtdna.com........29
Figure 2.2 PAGE gel using the fluorescently labeled oligonucleotide as a marker for siRNAs and miRNAs.................................................................30
Figure 2.3 A representative gel, using the preparative comb, containing 5 µg of E.coli small RNAs in the marker lane and 33 µg of Soleracea small RNAs in the preparative lane...........................................................................................................31
Figure 2.4 A gel showing that the fluorescent marker, and subsequently other smaller RNAs, moves the same distance on the gel whether in a mixture of RNAs or by itself...35
Figure 2.5 HPLC chromatogram depicting the separation of nucleosides from a digested tRNA molecule extracted from a PAGE gel…………………………………………………………………………….36

Figure 2.6 A representative calibration plot used in estimating the percent recovery from gels………………………………………………………………………………………………………………36

Figure 2.7 HPLC chromatogram showing the nucleosides recovered from the gel segment thought to contain microRNA………………………………………………………………………..38

Figure 3.1 Schematic of enzymatic hydrolysis of RNA……………………………………………………………46

Figure 3.2 A representative chromatogram from the 4.6mm x 25.0cm LC-18S column, preceded by a 4.6mm guard column [61. Sample was 46 µg of the nucleoside digestion product of tRNA Val, corresponding to approximately 10 µg of each major nucleoside……………………………………………………………………………………………………….47

Figure 3.3 Direct infusion ESI-MS of 8.5 µg/mL of tRNA Tyr digestion……………………………………………………………………………………………………48

Figure 3.4 The linear response of the nucleosides with higher signal-to-noise values….49

Figure 3.5 The linear response of nucleosides with lower signal-to-noise values……..50

Figure 3.6 A. 1-methyladenosine, B. 2-methyladenosine, and C. 2'-O-methyladenosine. All have the molecular weight of 281.27 Da……………………………………………………………………..50

Figure 3.7 Chromatogram from 1.0 mm column, separating components from Supelco’s nucleoside test mix. Each peak contains between 1-10 ng of a nucleoside…………….51

Figure 3.8 Separation of nucleoside test mix using 0.15 mm column, containing 1-10 ng of each nucleoside…………………………………………………………………………51

Figure 3.9 Chromatogram depicting the separation of 0.2 ng of each of the four major nucleosides……………………………………………………………………………………52

Figure 4.1 Example of experimental methodology currently in use for detecting a methylated ribose sugar. In the lane labeled henl-1 –β, this shows the regular miRNA, before chemical treatment. In the lane labeled henl-1 +β, this corresponds to a miRNA that was sensitive to chemical reaction and shifted movement due to the lack of a 2’-OH methyl group. If the sugar is methylated, the result would look like the Ler lanes, with no shift in migration upon treatment……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………
List of Tables

Table 1.1 Table listing the various types of RNAs, including the known number of modified nucleosides presently identified in various types from the database at http://medlib.med.utah.edu/RNAmods/................................................................................................................8

Table 1.2 pKa values for four major nucleobases found in RNA. *pKa values were calculated using Advanced Chemistry Development Software, Solaris V 4.67, except for uridine which was calculated according to Poulter and Frederick’s estimation [50,51]................................................................................................................22

Table 2.1 Gradient used for separating nucleosides on 4.6 mm column......................32

Table 2.2 Gradient used to separate nucleosides on 0.15 mm x 10 cm Targa column.....33

Table 3.1 Nucleosides found in tRNA^Val, the number of times they occur, and relative mole ratios.................................................................42

Table 3.2 Nucleosides found in tRNA^Tyr1, the number of times they occur, and relative mole ratios.................................................................42

Table 3.3 Instrumental parameters used during direct infusion mass spectrometry experiments.................................................................................44

Table 3.4 Gradient used on microbore column.......................................................45

Table 3.5 Origin of peaks within mass spectrum that are not attributable to nucleosides, confirmed by control experiments...............................................48

Table 3.6 Limits of detection for various modified nucleosides from tRNA^Tyr1 .........49

Table 3.7 Modified nucleosides in E. coli 16S rRNA and mole abundances, relative to guanosine.........................................................................................54
List of Abbreviations

A  adenosine
ATCC  American type culture collection
C  cytidine
Da  dalton
DNA  deoxyribonucleic acid
DNase  deoxyribonuclease
dsRNA  double-stranded RNA
E.coli  Escherichia coli
EDTA  ethylenediaminetetraacetic acid
ESI  electrospray ionization
6-FAM NHS  6-Carboxy-Fluorescein N-hydroxysuccinimyl
G  guanosine
Gm  2’-O-methylguanosine
HIV  human immunodeficiency virus
HPLC  high performance liquid chromatography
i.d.  inner diameter
LC  liquid chromatography
LTQ  linear ion trap quadropole
m5C  5-methylcytosine
mLASA  mitochondrial myopathy and sideroblastic anemia
MS  mass spectrometry
ms2i6A  2-methylthio-\textit{N}^6-isopentenyladenosine
m5U  5-methyluridine
m/z  mass-to-charge ratio
PAGE  polyacrylamide gel electrophoresis
\( P_i \)  phosphate
Pus1p pseudouridine synthase 1
Q queuosine
RNA ribonucleic acid
RNase ribonuclease
rRNA ribosomal RNA
RT-PCR reverse transcription-polymerase chain reaction
SDS sodium dodecyl sulfate or sodium lauryl sulfate
siRISC siRNA induced silencing complex
siRNA short (or small) interfering RNA
S/N signal-to-noise
snoRNA small nucleolar RNA
snRNA small nuclear RNA
\( S. oleracea \)  \( Spinach oleracea \)
s\(^4\)U 4-thiouridine
T ribosylthymine
t\(^6\)A  \( \Lambda^6 \)-threonylcarbamoyladenosine
TBE Tris-borate EDTA
TEMED N,N,N',N'-tetramethylethylenediamine
tmRNA transfer/messenger RNA
TOF time of flight
\( \text{tRNA}^{\text{Leu(UUR)}} \) transfer RNA, leucine specific, for UUR codon
\( \text{tRNA}^{\text{Lys}} \) transfer RNA, lysine specific
\( \text{tRNA}^{\text{Tyr}} \) transfer RNA, tyrosine 1 specific
\( \text{tRNA}^{\text{Val}} \) transfer RNA, valine specific
\( \Psi \) pseudouridine
U uridine
UV ultraviolet
UTR untranslated region
v/v volume/volume percent
w/w weight/weight percent
Chapter 1: The Significance of Ribonucleic Acid Studies

1.1 Introduction

Ribonucleic acid (RNA) is a ubiquitous molecule found in every cell of every organism. There are many different types of RNA, listed in Table 1.1, which serve various roles in the cell. RNA is involved in the process of protein synthesis and regulation of the genome.

Table 1.1 Table listing the various types of RNAs, including the known number of modified nucleosides presently identified in various types from the database at http://medlib.med.utah.edu/RNAmods/

<table>
<thead>
<tr>
<th>Type of RNA</th>
<th>Number of modified nucleosides</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA</td>
<td>91</td>
</tr>
<tr>
<td>rRNA</td>
<td>31</td>
</tr>
<tr>
<td>mRNA</td>
<td>13</td>
</tr>
<tr>
<td>tmRNA</td>
<td>2</td>
</tr>
<tr>
<td>snRNA and snoRNA</td>
<td>11</td>
</tr>
<tr>
<td>Chromosomal RNA</td>
<td>2</td>
</tr>
<tr>
<td>Other small RNAs, including miRNAs, siRNAs</td>
<td>1</td>
</tr>
</tbody>
</table>

RNA, most simply explained, is a biopolymer, comprised of repeating nucleotides. Each nucleotide consists of a nitrogenous base, a ribose sugar, and a phosphate group. Further terminology includes nucleosides, which are composed of the base and the sugar. There are four major bases present in RNA: adenine, guanine, cytosine, and uracil. Adenine and guanine are referred to as purines, while cytosine and uracil are identified as pyrimidines.
Figure 1.1 Structures of nucleosides and nucleotides present in RNA [1]

While RNA is minimally explained as repeating units of four different nucleotides, it is in reality often more complicated. RNA has long been known to contain in addition to the four major bases, modified nucleosides. The different modifications were summarized by Limbach et al. in 1994 and a continuously updated database can be found at [http://medstat.med.utah.edu/RNAmods/][2]. Figure 1.2 shows that modifications can be as simple as a methyl group added to the base or the sugar, but they can also be much more complicated, such as the addition of amino acid functionalities.

Figure 1.2 Structures of various modified nucleosides
The nomenclature is quite simple. If a modification is present on the nitrogenous base, a letter representing the modification is placed before the letter representing the nucleoside, followed by the number corresponding to its location on the ring. Hence, m^5C is so named because there is a methyl group on the fifth position on the ring. If the modification occurs on the sugar, the letter representing the modification is presented after the nucleoside letter. Therefore, Gm means a guanosine nucleoside has a methyl group on the sugar. No number is necessary because sugars can only be methylated at the 2’-OH position [2].

**Figure 1.3** A description of the numbering system used in the naming of nucleosides

Between 3 and 11% of the proteins encoded in the genome are involved in the metabolism of RNA, which includes modification, regulation and degradation [3]. Nucleosides are modified at precise positions during a posttranscriptional step by the action of RNA-modifying enzymes [4]. A number of these enzymes have been identified, and their mechanisms of action have often been characterized [5-6]. Many modifying enzymes in different organisms share conserved sequences, showing evolutionary significance [4]. The mechanism of modification varies significantly between prokaryotic and eukaryotic organisms. Prokaryotic organisms typically have enzymes which are site-specific, while eukaryotic organisms rely on snoRNAs, small nucleolar RNAs that guide enzymes to specific sites [7-8]. The biological significance of
modified nucleosides is still being examined. While a precise role for each modification has not yet been found, there have been numerous reports regarding the functional significance of specific modified nucleosides in the RNA in which they are found [9-12]. For example, Ofengand hypothesized that pseudouridine, the most abundant modified nucleoside, serves as “molecular glue” that stabilizes RNA conformations through an additional hydrogen bond donor [12].

The goal of this chapter is to summarize the importance of modified nucleosides in the structure and function of RNA molecules by illustrating past studies regarding RNA modifications and the consequence of their absence. Further, the emerging role of newly discovered types of RNA, miRNA and siRNA, will be examined, in addition to an explanation of how mass spectrometry has been used to learn more about RNAs. Chapter 2 will describe a method by which miRNAs and siRNAs can be isolated from plant cells and further studied using LC-UV. Chapter 3 will present a method by which to directly analyze a mixture of nucleosides arising from the nuclease digestion of RNA for the presence of a modified nucleoside and will compare such a technique to LC-MS techniques. Chapter 4 will conclude these studies and suggest future directions in this area of research.

1.2 Examples of Past Studies Concerning Nucleoside Modifications

The importance of developing reliable methods for studying nucleoside modifications is best stressed by providing examples of what will happen if a modification is not present. There have been numerous studies regarding a phenotypic change in an organism based solely on a difference in the modification of a single
nucleoside. The examples given in this chapter are only a fraction of the cases described in the literature. Often modifications cease to arise when the base from which they originate from has been mutated to something else. Such mutations are called point mutations, which transform one base to a different base. The nomenclature lists the name of the regular base and the position at which it is found in the genome, followed by the base that replaced it. For example, a mutation changing an A to a C at position 6500 in the genome is named A6500C.

Interestingly, the majority of nucleoside-related diseases are caused by a missing modified nucleoside in mitochondrial tRNAs. The function of a tRNA is to deliver an amino acid to a nascent polypeptide chain during the process of protein synthesis. Specificity is achieved because each tRNA contains a different anticodon that pairs to the mRNA molecule being translated. These tRNAs specify different amino acids, and according to standard nomenclature, the amino acid is listed after the name of the tRNA. For example, the tRNA that carries valine is called tRNA$^{\text{Val}}$. Due to the degeneracy of the genetic code, a number of different tRNAs can transport the same amino acid, and these tRNAs are differentiated by including the name of the anticodon after the name of the amino acid, such as tRNA$^{\text{Leu(UUR)}}$.

1.2.1 MELAS, a Neurodegenerative Disorder

MELAS is an acronym with letters representing different symptoms of the disease: mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke. Essentially, this disease has negative effects on the muscles and the nervous system, and often results in death. It has been known to be a disease passed along by the mother through a mutation in mitochondrial DNA [13].
Specifically, this disease is generated by a point mutation in the DNA encoding tRNA\textsuperscript{Leu(UUR)}. There are a number of point mutations that result in the disease, such as A3243G and G3244A. However, other mutations in the tRNA, such as G3242A, can cause other mitochondrial diseases, but do not cause MELAS [13-14]. This information seems perplexing but can be explained by studying the mature tRNA molecules. It was found that MELAS patients are missing a modified nucleoside (5-taurinomethyluridine) at the wobble position of the tRNA, while the non-MELAS cohort has this modification [13]. The tRNAs lacking the modification have been found to be unable to decode UUG codons [15]. Therefore, the entire basis of a human disease could be explained by the lack of a single modified nucleoside.

![Figure 1.4](structure.png)

**Figure 1.4** Structure of the missing nucleoside modification (τm\textsuperscript{5}U) that causes MELAS

### 1.2.2 MERRF Syndrome, a Muscular Disorder

MERRF, or myoclonic epilepsy and ragged-red fiber disease, results when a mutation occurs in the DNA coding for mitochondrial tRNA\textsuperscript{Lys}. Specifically, it is a point mutation (A8344G), which corresponds to the wobble position of the mature tRNA [16]. As a result, instead of uridine, cytidine is present, which is unable to be modified into the 5-taurinomethyl-2-thiouridine found in wild-type cells [17]. This has been found to disturb the codon-anticodon interaction necessary for the tRNA’s function, reducing the
efficiency of mitochondrial translation [18]. Once again, the absence of a nucleoside modification leads to the onset of a disease.

**Figure 1.5** Structure of $\tau m^5s^2U$, the nucleoside modification absent in mitochondrial tRNA$^{Lys}$ in patients with MERRF

### 1.2.2 MLASA, an Oxidative Phosphorylation Disorder

MLASA, or mitochondrial myopathy and sideroblastic anemia, is a disease that is generated when a missense mutation occurs in the $PUS1$ gene. $PUS1$ encodes pseudouridine synthase 1 (Pus1p), an enzyme responsible for catalyzing the conversion of uridine into pseudouridine on numerous tRNAs. Particularly, the missense mutation, occurring near the active site of the enzyme, changes an arginine into a tryptophan, making Pus1p nonfunctional. It has previously been found that certain pseudouridines are necessary for the folding of the characteristic cloverleaf structure of tRNAs [19]. The exact consequence of these missing pseudouridines is the subject of future studies, but it is apparent that another missing modification results in a severe disease [20].

**Figure 1.6** Structure of pseudouridine, the only mass-silent modified nucleoside, which is often not present in tRNA in MLASA patients.
1.3 The Emerging Role of miRNA and siRNA in Molecular Biology

It is extraordinarily rare for a major player in molecular biology to go unnoticed for so many years. However, such was the case with miRNAs and siRNAs, small RNA molecules with the capability of regulating gene expression. Discovered in 1993 by Victor Ambros while studying *Caenorhabditis elegans*, miRNAs, also called microRNAs, are small RNA molecules that regulate gene expression [21]. By binding to a target mRNA in its 3’ UTR (untranslated region), miRNAs arrest the mRNA’s translation into a protein. siRNAs, also known as short interfering RNAs, are essentially identical in structure to miRNAs but serve a slightly different role in the silencing of genes. They either bind to a target mRNA sequence and signal for its degradation, or they methylate cytosines in genomic DNA, preventing their transcription into mRNAs (Figure 1.7).

![Figure 1.7](image.png)

**Figure 1.7** Basic schematic representations for four possible pathways of a gene: A. is for a non-silenced gene, B. is a gene silenced by miRNA, C. is a gene silenced by siRNA at the mRNA level, and D. is a gene silenced by siRNA at the DNA level
1.3.1 Differences between miRNAs and siRNAs

miRNAs and siRNAs are difficult to distinguish from one another based on appearance. Instead, they are differentiated based upon where they originate. Often produced from distinct places in the genome, miRNAs are made from a single-strand RNA, which folds into a hairpin structure. This hairpin structure is enzymatically processed by a nuclear enzyme called Drosha, leaving a mature single-stranded miRNA [22].

\[
5' \text{ UGUCG66UAGCUUAUC A GACUG A UGUUG A CUGU U G A A U} \\
3' \text{ AGUCUGUCGGGUAG---CUGAC C ACAAAC---GUA---C U C}
\]

**Figure 1.8** Structure of a microRNA hairpin structure, prior to processing

siRNAs can originate from a number of different sources, but all are composed of double stranded RNA [23]. Such sources include viruses and transposons, although siRNA has also been shown to originate within the mRNA of the gene it affects [24]. Another significant difference between these RNAs regards how well they match their target mRNA. siRNAs typically base pair to their target mRNA perfectly, while miRNAs bind with only partial complementarity [25]. However, an exception occurs in plants, where miRNAs tend to base pair with no mismatches. Both miRNA and siRNA form an RNA-induced silencing complex, either miRISC or siRISC for short. This complex is comprised of a number of different proteins, some of which are included in both miRISCs and siRISCs [26].

Numerous other proteins are involved in the processing of these RNAs to their functional form. Of particular interest is *HEN1*, a protein that was known to be involved
in the metabolism of miRNA [27-28]. Only recently was the precise function of HEN1 discovered; it methylates the 2’-OH of the terminal nucleotide in both miRNA and siRNA [29]. This methyl group is necessary for the miRNA or siRNA to be fully functional.

1.3.2 Experimental Techniques for Studying miRNAs and siRNAs

These RNAs, theoretically capable of turning off essentially any gene in the genome, seem to be ideal candidates for new therapeutics. In fact, they are currently being studied as potential therapeutics for several diseases, such as poliovirus and HIV-1 [24]. However, as with any therapeutic, as much should be learned about the drug candidate as possible to determine whether it is safe and effective [30]. One topic of current research regards how the RNA would be delivered to the cell. Other studies have been conducted to determine the possibility of off-site targets, such as those that would result from incomplete base pairing [31-32]. Off-target effects result in the silencing of an unintended gene. In general, these studies require the use of synthetically produced siRNAs.

There are many valuable experimental techniques being utilized to study the function of endogenous miRNAs and siRNAs. Microarrays have been gaining popularity as a biochemical tool since they were introduced in the 1980s. Originally designed for studying gene expression, now they have been developed to examine the expression of different miRNAs in different tissues [33-34]. In fact, such microarrays are now commercially available. The microarray platform is composed of different spots containing immobilized complementary sequences that specifically bind to miRNAs. A
color change signals that hybridization has occurred, and therefore the miRNA is present. The intensity of the color shows a relative amount of how much miRNA is there. They require only a small amount of total RNA (1 µg in most recently developed microarrays), which in turn displays the differential expression of the miRNAs present [35].

![Figure 1.9](image)

**Figure 1.9** A portion of microarray data from miRNAs expressed in different tissues, taken from [35]

Another method used in the analysis of miRNAs and siRNAs is RT-PCR, or reverse transcription-polymerase chain reaction. PCR is a method of amplifying DNA segments, so RT-PCR instead amplifies RNA segments by first converting them into DNA segments, and amplifying these. Recent RT-PCR techniques have allowed for both the detection and quantitation of small RNAs [36]. Unfortunately, because RNA is reverted to its DNA foundation, valuable information concerning posttranscriptional nucleoside modifications is lost when using this technique.

Northern blotting is another common technique used to detect small RNAs. First, RNA must be fractionated using gel electrophoresis. The bands from the gel are blotted
onto a nitrocellulose sheet, which is incubated with labeled DNA sequences complementary to the sequence to be detected. After a period of time, the sheet is washed to remove unhybridized probes and subsequently examined using autoradiography or fluorescence detection, depending on the nature of the DNA probe. The resulting data shows whether the target sequence was present, and the intensity of the band estimates the relative amount of the sequence. If using a size marker, the location of the band will also give an estimation of the size of the RNA segment containing the sequence complementary to the probe [37].

**Figure 1.10** An example of a Northern blot experiment, taken from [57]

### 1.3.3 Methods for Isolating miRNAs and siRNAs

To further examine the characteristics of these small RNAs, an efficient method of isolating endogenous RNAs is necessary. The extraction of RNA from cells is a well-established method, and there are commercially-available kits that simplify the process. Many experimental protocols are based on phenol extractions, although there are hundreds of other methods described in the literature [38]. However, once RNA is extracted, the remaining issue concerns how to separate the miRNAs and siRNAs from all the other cellular RNAs.
One possible technique for isolating small RNAs is by using chromatography. Anion-exchange chromatography, which requires a charged stationary phase to separate analytes according to charge, has been used in the past to separate synthetic oligonucleotides (between one to three nucleotides) that vary in base composition [39]. Further, ion-pair chromatography, which uses a reversed phase with ion-pairing reagents in the stationary phase, has previously been used to separate synthetic oligonucleotides ranging from 12 to 60 nucleotides [40]. Finally, a protocol utilizing size-exclusion chromatography, a method that separates based on the size of analytes, was recently described that allows for large-scale purification of RNA, which results in a more pure sample than could be achieved using a polyacrylamide gel [41]. While this specific protocol admittedly is designed for purification, it would be reasonable to adapt the method for separation.

Many methods currently used require running polyacrylamide gels and using radioactive labels as a means of labeling which bands should be extracted. Not every laboratory is licensed to experiment with radioactive chemicals, which is why seeking other types of methods for isolating miRNAs and siRNAs is desirable. A recent paper by Ruth Martin described a new method for isolating these small RNAs, which does not rely on radioactive probes. Instead it uses a miniblot system, which blots the gel containing the sample to another gel containing miRNA probes [42]. The main problem with the technique is that the extraction conditions used vary from sample to sample, meaning that optimization is required if one is moving from one organism to another.
1.4 Mass Spectrometry as a Tool for Studying RNA

Over the years, RNA has been studied by many different chemical and biochemical methods. However, mass spectrometry (MS) has emerged in recent years as an indispensable tool for studying RNA. Mass spectrometry is an advantageous technique because it provides a measurement of an intrinsic parameter: mass. As many RNAs have been shown to contain modifications, mass spectrometry is well suited to detect the modifications that result in a mass shift [43]. While RNA can be analyzed using numerous types of ionization methods, the focus of this work is to utilize electrospray ionization mass spectrometry (ESI-MS) in the development of novel techniques for characterizing modified nucleosides [44-47].

Electrospray ionization was introduced in 1988 by J. Fenn [48]. Electrospray ionization is extremely valuable because it allows for the analysis of high molecular weight compounds that possess numerous charge states. In regards to RNA, oligonucleotides are easily analyzed in negative-ion mode, because the molecule carries a number of different charge states (Figure 1.11). Electrospray ionization allows for the measurement of the mass-to-charge (m/z) ratios, providing multiple measurements of mass on a single molecule [49].

![Electrospray mass spectrum of a multiply-charged oligonucleotide](image.png)

**Figure 1.11** Electrospray mass spectrum of a multiply-charged oligonucleotide [49]
The generation of charged ions in ESI-MS is a relatively simple concept. In electrospray ionization, all molecules ionized in solution should be ionized in the mass spectrometer. Charged droplets are generated at a high voltage capillary tip [49]. Droplets shrink when coulombic repulsions overcome the surface tension of the droplet. Eventually, all solution will have evaporated and a single ion in the gas phase remains, which can be detected by the mass analyzer [47,49].

![Figure 1.12 Diagram of the generation of desolvated ions in the electrospray process](image)

The type of ions drawn into the mass spectrometer can be selected by changing the voltage applied at the capillary needle. While oligonucleotides are best analyzed in negative-ion mode, nucleosides, studied in this work, are analyzed in positive-ion mode, because the nitrogenous bases are capable of picking up a proton leading to positively charged species. The ability of individual ions to be ionized in solution, and therefore in the mass spectrometer, is dependent on their pKa values.

<table>
<thead>
<tr>
<th>Nucleobase</th>
<th>pKa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytosine</td>
<td>4.60 ± 0.20</td>
</tr>
<tr>
<td>Uracil</td>
<td>-2.98 ± 0.20*</td>
</tr>
<tr>
<td>Adenine</td>
<td>2.95 ± 0.20</td>
</tr>
<tr>
<td>Guanine</td>
<td>3.33 ± 0.20</td>
</tr>
</tbody>
</table>

*Table 1.2 pKa values for four major nucleobases found in RNA. *pKa values were calculated using Advanced Chemistry Development Software, Solaris V 4.67, except for uridine which was calculated according to Poulter and Frederick’s estimation [50,51].
The pKa value of a nucleobase generally does not significantly change upon posttranscriptional modification. For example, the pKa for 2’-O-methylguanosine is 3.33, while the pKa for N⁶-(3-methyl-2-butenyl)-2-methylthioadenosine is 2.87 [50]. Preferred electrospray solutions for analyzing RNA are made of water and an organic solvent, such as acetonitrile or methanol [52]. The ionization of nucleosides can be favored by adding to solutions a small percent of an acid, either formic acid or acetic acid.

Performing a separation, usually using liquid chromatography, prior to mass spectrometry is the preferred technique in the analysis of modified nucleosides. During LC-MS, nucleosides are generally separated on a reversed-phase column, and the column eluent is directed to a mass spectrometer. Initially, thermospray was the most common method of ionization, because it was the first ionization technique that was easily interfaced to liquid chromatography systems [53]. One problem with using thermospray ionization is that the mass spectra it generates are not very reproducible [61]. Other ionization techniques that have been used in LC-MS analysis of nucleosides and nucleotides are continuous flow fast atom bombardment and atmospheric pressure chemical ionization, reviewed in [54].

Not long after thermospray techniques for analyzing nucleosides were well refined, electrospray ionization was introduced to the scientific community. Originally, LC-ESI-MS analysis of nucleosides was difficult because of problems with solvent compatibility. Often mobile phases for separating nucleosides contained salts that would form unfavorable adducts with nucleosides. Volatile salts provided a solution; for example, a widely-used salt during LC-ESI-MS separations of nucleosides is ammonium
acetate, because it is able to donate a proton to the nucleoside, and the leftover ammonia molecule evaporates in the electrospray source. Other problems involved the fact that aqueous solutions used at the beginning of gradients have high surface tension and are difficult to desolvate [52]. Eventually, conditions for LC-ESI-MS were optimized, and electrospray became a popular technique for analyzing nucleosides [55].
Chapter Two: Developing an Approach for Analyzing Heterogeneous Mixtures of MicroRNAs by LC-UV

2.1 Introduction

In many multicellular organisms, there are two classes of small RNAs that serve to regulate gene expression. MicroRNAs (miRNAs) bind to target messenger RNAs and thus interfere with their translation into proteins. Short interfering RNAs (siRNAs) also bind to messenger RNA; however, in doing so, they signal for the mRNA’s degradation \[28\]. Another class of siRNAs is involved in gene silencing at the transcriptional level. In a process called RNA-directed DNA methylation that takes place in plant cells, siRNAs bind to nuclear DNA, methylating cytosines, which causes that gene to no longer be transcribed into RNA \[8\].

Since these small RNAs have such a vital role in the expression of genes, it would be beneficial to have a reliable method by which to study them. Mass spectrometry has been used in the past to identify a modified nucleoside present in a plant miRNA. In *Arabidopsis thaliana*, the Chen group at Rutgers found that *HEN1* methylates the ribose sugar of the 3’ terminal nucleotide from an miRNA duplex. This methylation is vital for the function of the miRNA. The methylated miRNA was isolated using affinity purification with a biotinylated complementary DNA sequence and detected by performing electrospray ionization mass spectrometry \[57\].

Such experiments require that the sequence of the RNA to be studied be known. However, what if the sequence was unknown? For example, there could be an organism that has a phenotypic difference from the wild type that suggested a defect in a gene
silencing pathway. If there was a way to nonspecifically study the modified nucleosides in the mutant versus the wild type organism, the lack of a modified nucleoside could be responsible for the defect. This project seeks to develop a method to study the modified nucleosides present in microRNAs and short interfering RNAs. Primarily, the process of isolating and analyzing microRNAs from a plant source will be discussed in this chapter.

2.2 Experimental

2.2.3 Materials

*Escherichia coli* ATCC MRE 600 was cultured according to standard protocols [58]. *Spinach oleracea* was purchased from Kroger as whole organic leaves. Ultra-pure grade bisacrylamide (catalog # 0172) and acrylamide (catalog # 0341) were purchased from Amresco (Solon, OH, USA). Biotechnology-grade boric acid (catalog # B1934), molecular-biology grade ethanol (catalog # E7023), Trizma base (catalog # T6066), tRNA<sup>Val</sup> (catalog # R2645), nuclease P1 (catalog # N8630), venom phosphodiesterase I (catalog # P6877), ammonium bicarbonate (catalog # A6141), adenosine (catalog # A9251), cytidine (catalog # C4654), guanosine (catalog # G6264), and uridine (catalog # U3003) were purchased from Sigma (St. Louis, MO, USA). The electrophoresis-grade EDTA disodium salt (catalog # BP120), buffer-saturated phenol:chloroform:isoamyl alcohol solution (25:24:1) (catalog # BP1752), sodium acetate (catalog # S210), SDS (catalog # S529), TEMED (catalog # BP150), and ammonium acetate (catalog # A639) were purchased from Fisher Scientific (Hampton, NH, USA). Alkaline phosphatase (catalog # LS006124) was purchased through Worthington (Lakewood, NJ, USA). Glacial acetic acid (catalog # 281000ACS) was purchased from Pharmco (Brookfield,
Molecular biology grade urea (catalog # 51456) was purchased from Fluka (Milwaukee, WI, USA). Ammonium persulfate (DNase and RNase free, catalog # 327081000) was purchased from Acros Organics (Fairlawn, NJ, USA). Denaturing loading buffer (catalog # 8546G) was purchased from Ambion (Austin, TX, USA). Syber gold (catalog # S-11494) was obtained from Molecular Probes (Eugene, OR, USA).

2.2.2 Methods

2.2.2.1 RNA Isolation

RNA was isolated from cells using Ambion’s mirVana kit (catalog # 1560). When isolating RNA from plant cells, Ambion’s plant isolation aid was used (catalog # 9690). Spinach tissue was homogenized using a mortar and pestle. Liquid nitrogen may be added to aid in the homogenation process, although doing so is not essential. The centrifuge used in these experiments was a Hettich Mikro 22 R, purchased through Helmer (Noblesville, IN, USA). The amounts of RNA isolated were estimated by measuring absorbance at 260 nm using a Biospec-1601 Shimadzu UV-visible spectrophotometer (Addison, IL, USA). The concentration of the mixture is estimated using Equation 2.1:

\[
\text{Concentration (in } \mu\text{g/mL}) \approx 33 \times A_{260} \tag{2.1}
\]

RNA purity was determined from the \( A_{260}/A_{280} \) ratio. A ratio between 1.8 and 2.1 indicates RNA not contaminated by protein.
2.2.2.2 Polyacrylamide Gel Electrophoresis

2.2.2.2.1 Preparation of Gel

A 10X TBE buffer was prepared by combining 120 g of Trizma base with 27.5 g of boric acid and 9.3 g of disodium EDTA, brought to 1 L with autoclaved nanopure water and sonicated. The acrylamide stock was made with 30 g of acrylamide and 0.8 g of bisacrylamide, brought to 100 mL with autoclaved nanopure water and heated.

15% denaturing polyacrylamide gels were prepared by first mixing 21.02 g of urea with 22.1 mL of water. This mixture was sonicated to aid in dissolving the urea. 22.1 mL of this mixture was mixed with 2.5 mL of 10X TBE, 25 mL of an acrylamide stock solution, 350 µL of ammonium persulfate, and 50 µL of TEMED. The solution was immediately added between the gel plates and allowed to set for 1.5 to 2 hours after diagonal insertion of the comb. Electrophoresis was performed on a Hoefer SE 600 standard vertical electrophoresis unit, purchased through Amersham Biosciences (Piscataway, NJ, USA).

2.2.2.2.2. Running and Visualizing the Gel

Gels were conditioned by placing the gel into the apparatus, and then pouring 1X TBE buffer into the top and bottom portion of the apparatus. Enough solution was added to cover the bottom of the gel. The electrophoresis unit was hooked up to an electrophoresis power supply (EPS 301) and run in the refrigerator for 20 minutes at 300 volts and 20 mAmps.

In 7 mm gel lanes, up to 10 µL of solution can be loaded. If not using a marker, 7.5 µL of a sample, 2.0 µL of denaturing loading buffer and 0.5 µL of Syber gold was added. When using a marker, the volume of marker combined with 2 µL of denaturing
loading buffer was brought to 10 µL using autoclaved nanopure water. In the preparative 121 mm lane, approximately 174 µL may be added, which consists of 130 µL sample, 35 µL denaturing loading buffer, and 9 µL of Syber gold. The gel is run in the refrigerator at 300 volts, 20 mAmperes for approximately 2 hours.

To visualize the gel, an Alpha Innotech multi-image light cabinet is used with the Alpha Imager program (Leandro, CA, USA).

2.2.2.2.3 Using a Fluorescently-Labeled Oligonucleotide as a Marker

To verify the location of the miRNAs and siRNAs on the gel, a fluorescently-labeled oligonucleotide was used. This was specially made by IDT DNA (Coralville, IA, USA). The sequence is 5’-UUGGCAUUCUGUCCACCUCC-3’, and the fluorophore, 6-FAM NHS ester was added at the 5’ end (Figure 2.1). The product was purified using HPLC.

![Figure 2.1](http://www.idtdna.com) Structure of the 6-FAM NHS ester, taken from http://www.idtdna.com

Bands were extracted according to where the 20-mer labeled oligonucleotide showed up on the gel. A range of approximately 1 cm in each direction of the fluorescence was excised to include the entire possible range of miRNA and siRNA sizes (18-26 nucleotides).
2.2.2.4 Extraction of Bands

Bands were extracted from the excised preparative lane according to a previously described protocol [59]. After excision, bands were placed in Corning sterile plastic tubes with a negligible amount of 1X TBE buffer. Gels were crushed using an autoclaved glass stir rod. The 3 M sodium acetate solution was made by mixing 2.46 g of sodium acetate with enough water to make 10 mL, and the pH was adjusted to 5.6 using glacial acetic acid. The 0.3 M solution of sodium acetate with 0.2% SDS was prepared by adding 0.246 grams of sodium acetate and 0.020 grams of SDS to enough water to make 10 mL, with the pH adjusted to 6.0 using glacial acetic acid. The amount of solution listed in the protocol was added for every 1 mL volume of solution from the gel, estimated from the markers on tubes.
Figure 2.3 A representative gel, using the preparative comb, containing 5 µg of *E.coli* small RNAs in the marker lane and 33 µg of *S.oleracea* small RNAs in the preparative lane.

2.2.2.4 Analyzing the Gel Extract

2.2.2.4.1 Nucleoside Digestion

RNA samples were digested to nucleosides according to a previously described protocol [60]. Concentration of RNA was estimated by measuring its wavelength at 260 nm using a Shimadzu UV-visible spectrophotometer. Samples were heated using a Thermolyne type 17600 dri-bath. Subsequent incubations with different enzymes (nuclease P1, venom phosphodiesterase, and alkaline phosphatase) were performed using either an Innova 4000 incubator shaker at 100 RPM or the Thermolyne dri-bath. Solutions used during the digestion (0.1 M ammonium acetate, pH 5.3, and 1.0 M ammonium bicarbonate) were filter-sterilized using 0.20 µm filters. Once the digestion was complete, RNA samples were stored in a -80°C freezer.
2.2.2.4.2 Verification of a Complete Digestion

The completeness of a digestion was verified using a previously described procedure, with minor variations to the gradient listed in Table 2.1 [61]. The column used was a 4.6 mm x 25.0 cm Supelcosil LC-18S, with 5µm particle size and 120 Å pore size (catalog # 58928-U). A nucleoside test mix, also purchased through Supelco (catalog # 47310-U, Bellefonte, PA, USA) was used to verify the column’s performance. The HPLC used was a Hitachi D-7000, with an L-7100 pump and an L-7400 UV detector, with a flow rate of 2.0 mL/min and UV detection at 254 nm.

Table 2.1 Gradient used for separating nucleosides on 4.6 mm column

<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>Percent A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>5.8</td>
<td>99</td>
</tr>
<tr>
<td>7.2</td>
<td>98</td>
</tr>
<tr>
<td>8.6</td>
<td>97</td>
</tr>
<tr>
<td>10.0</td>
<td>95</td>
</tr>
<tr>
<td>25.0</td>
<td>75</td>
</tr>
<tr>
<td>30.0</td>
<td>50</td>
</tr>
<tr>
<td>34.0</td>
<td>25</td>
</tr>
<tr>
<td>37.0</td>
<td>25</td>
</tr>
<tr>
<td>45.0</td>
<td>1</td>
</tr>
<tr>
<td>45.5</td>
<td>1</td>
</tr>
<tr>
<td>47.5</td>
<td>99</td>
</tr>
<tr>
<td>55.0</td>
<td>99</td>
</tr>
</tbody>
</table>

2.2.2.4.3 Capillary-LC Analysis

The column used to perform LC-MS was a 0.15 mm X 10 cm Targa C18 column, with 5 µm particle size and 120 Å pore size (Higgins Analytical, purchased through the Nest Group (Southboro, MA, USA), catalog number TC-1015-C185, sorbent lot T28938). A scaled-down version of gradient, shown in Table 2.2, used on the 4.6 mm column was calculated using Equation 2.2:
Time 2 is the calculated time, adapted from the gradient listed in Table 2.1, with Time 1. Diameter 2 is 0.075 mm, and diameter 1 is 4.6 mm. Flow 2 is 1.06 µL/min, and flow 1 is 2000 µL/min. Length 2 is 10 cm, and length 1 is 25 cm.

The HPLC used was a Dionex/LC Packings Ultimate system, equipped with a LC Packings Famos autosampler (Sunnyvale, CA, USA). Flow rate was 1.06 µL/min, with UV detection at 254 nm.

<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>Percent A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>0.10</td>
<td>99</td>
</tr>
<tr>
<td>1.76</td>
<td>98</td>
</tr>
<tr>
<td>2.88</td>
<td>97</td>
</tr>
<tr>
<td>4.00</td>
<td>95</td>
</tr>
<tr>
<td>16.0</td>
<td>75</td>
</tr>
<tr>
<td>20.0</td>
<td>50</td>
</tr>
<tr>
<td>23.2</td>
<td>25</td>
</tr>
<tr>
<td>25.6</td>
<td>25</td>
</tr>
<tr>
<td>32.0</td>
<td>1</td>
</tr>
<tr>
<td>32.4</td>
<td>1</td>
</tr>
<tr>
<td>34.0</td>
<td>99</td>
</tr>
<tr>
<td>68.0</td>
<td>99</td>
</tr>
</tbody>
</table>

2.3 Results and Discussion

2.3.1 RNA Isolation

The mirVana kit, which, through a series of centrifugation and extraction steps, collects all RNAs under 200 nucleotides, was used to isolate small RNAs from both *E. coli* and *S. oleracea*. An additional centrifugation step was performed to remove plant contaminants that interfere with the RNA extraction using the plant isolation aid, which
contains polyvinylpyrrolidone, a polymer that binds to polysaccharides and polyphenolics. The yield of RNA remained relatively constant for the organism from which it came. Isolating RNA from 1 gram of *E. coli* cells yielded between 600 and 780 µg of small RNA. Isolating RNA from 1 gram of *S. oleracea* leaves yielded between 100 and 160 µg of small RNA. The disparity between the numbers from *E. coli* to those for spinach is expected. *E. coli* cells are harvested at a point when cell division is extremely high. Amongst the cellular machinery for reproduction are tRNA and rRNA, which are abnormally elevated in abundance in the *E. coli* cells versus the dormant spinach leaf cells.

### 2.3.2 Polyacrylamide Gel Electrophoresis

The sequence for the fluorescent marker chosen could not be based on a *S. oleracea* microRNA sequence, because there are no available databases for such information. Instead, it was based on a sequence found in microRNAs of *Arabidopsis thaliana*, because it has been shown that the sequences of most flowering plants are likely to be conserved [62].

It is possible that the other RNAs present in the small RNA mixture could block the miRNAs and siRNAs from moving through the gel. The result of this would be that the fluorescently-labeled marker, which runs alone with no obstructions, would not correctly predict where the miRNAs and siRNAs elute on the gel. To test whether this is a valid concern, a gel was run which mixed the isolated RNAs with the fluorescently-labeled oligonucleotide. It is evident that the fluorescent marker moves to the same place in the gel, whether it is mixed with other RNAs or run by itself (Figure 2.4). Therefore, it
is not an issue that the movement of the smaller RNAs is impeded by larger RNAs present in the lane.

Figure 2.4 A gel showing that the fluorescent marker, and subsequently other smaller RNAs, moves the same distance on the gel whether in a mixture of RNAs or by itself

The addition of the fluorophore to the marker used admittedly would affect the way the oligonucleotide migrates through the gel. The movement of any molecule in gel electrophoresis is based on its charge and size. It is estimated that, because the molecular weight of the marker is near that of a nucleotide, the molecule may move more like a 21-mer than a 20-mer. However, since the marker has two negative charges versus the nucleotide’s one, it may migrate slightly less than a 21-mer would. Of course, more work would be necessary to confirm this supposition.

The denaturing polyacrylamide gels successfully separated the different components of small RNAs isolated from cells. Further, the RNA could be extracted from the gel using a previously described protocol. While it is difficult to provide an exact percent recovery, an estimation has been made by loading a known amount of tRNA (with known amounts of component nucleosides) onto the gel and extracting it. After the tRNA was extracted, it was then digested to nucleosides and the peak areas of
the four major nucleosides were compared against calibration curves from standard nucleoside HPLC runs (Figures 2.5 and 2.6). It is estimated that approximately 20 to 40% of the RNA can be recovered from a gel.

![HPLC Chromatogram](image)

**Figure 2.5** HPLC chromatogram depicting the separation of nucleosides from a digested tRNA molecule extracted from a PAGE gel

![Calibration Plot](image)

**Figure 2.6** A representative calibration plot used in estimating the percent recovery from gels

### 2.3.3 Analyzing the Gel Extract

In any given cell, it has been estimated that there are between a hundred and a thousand different siRNAs and miRNAs [63]. Assuming the lowest possible abundance of a modification, which would be that only one of the miRNAs or siRNAs contains a
methylation, this equates to approximately one modified nucleoside per every two thousand or twenty thousand non-modified nucleosides. Hence, effort is required to accumulate enough of these RNAs to be able to analyze them for modified nucleosides.

It is difficult to know exactly the mass of leaves necessary to detect one modified nucleoside. While it has been shown that miRNAs and siRNAs must be methylated to be active gene silencers, the proportion of methylated miRNAs and siRNAs to those not yet processed by *HEN1*, is presently unknown [29]. However, estimations may be made. In the supplemental information in the paper by B.Yu et al. concerning the detection of a methylated microRNA in plant cells, it was mentioned that in 60 mg of enriched small RNAs, there was approximately 5 ng of the methylated version of miR173 (the microRNA under study) [57]. Using the isolation methods described in our work, it would take approximately 100 grams of plant leaves to recover 60 mg of small RNAs. However, assuming a loss of approximately 2/3 of the sample during extraction from a polyacrylamide gel, one would instead need to start from 300 grams. Supposing only one of the miRNAs is methylated, one could assume a similar recovery of 5 ng of the methylated oligonucleotide. Estimating that the oligonucleotide weighs around 7000 Da, this equates to only $7 \times 10^{-13}$ moles of the miRNA. Assuming one modification per oligonucleotide, which was found to be the case in Yu’s study, this leaves only approximately 0.2 ng of modified nucleoside (assuming a molecular weight of 280 Da). Therefore, 0.2 ng per nucleoside would need to be at or above the limit of detection for the instrumentation used in experiments. In conclusion, given enough time, it certainly would be possible to perform such an analysis and obtain meaningful data.
Below is an HPLC-UV chromatogram (Figure 2.7) of the spinach RNA extracted from the lower band, where the microRNAs are thought to reside. The area of the peaks is significantly smaller than that from the chromatogram in Figure 2.5 produced from extracting tRNA. Further, no modified peaks can be seen. This makes evident how little RNA is present in such a sample.

![HPLC chromatogram showing the nucleosides recovered from the gel segment thought to contain microRNA](image)

**Figure 2.7** HPLC chromatogram showing the nucleosides recovered from the gel segment thought to contain microRNA

The large peak labeled with a question mark is a contaminant that enters the sample during the extraction protocol, as corroborated by performing control experiments. Phenol, a chemical used during the extraction of RNA from the gels, is 8% (w/w) soluble in water so it is a probable source of the contamination. Further, it is also a UV-absorber at the wavelength used in the study. However, this hypothesis has not yet been confirmed.

Since the Supelcosil stationary phase was not available in low ID columns, the Targa stationary phase was used to mimic the Supelcosil stationary phase, as both are endcapped. This column will be used for separations of nucleosides that are analyzed by the LTQ-MS.
2.4 Conclusion

It has been shown that small RNAs can be isolated from spinach leaves using a commercially available extraction kit. Further, these RNAs can be successfully separated using a denaturing polyacrylamide gel. RNA can be extracted from a gel, digested, and analyzed using HPLC-UV. A fluorescent label has been used to verify the location of the microRNAs.

The work that still needs to be performed is to accumulate enough of these microRNAs to analyze them for the presence of a modified nucleoside. This could be accomplished by performing mirVana extractions on many grams of spinach leaves, followed by separation by PAGE gel and extraction. Performing additional washing steps would help to reduce contamination.

To approximate how much time this would require, estimations may be made in regard to each step of the process. Performing the mirVana kit on one gram of plant leaves takes approximately two hours. Therefore, to perform the isolation procedure on 300 g of plant leaves could take up to 600 hours, which would isolate approximately 180 mg of small RNAs. However, since such a large starting amount is necessary, higher throughput methods of isolation should be investigated to decrease preparation time.

Next, extensive effort must be made to separate these small RNAs using gels. In a single preparative gel, one could load approximately 100 µg of small RNAs in the preparative lane. Hence, to run the entire 180 mg would take an unreasonable amount of time. Therefore, when evaluating the methods presented here, traditional gels seem to be a significant bottleneck in accumulating a sufficient amount of RNA. To make the sample preparation more amenable, one may consider alternative ways of separating
RNA molecules. Some possibilities are sucrose fractionation, preparative column gel electrophoresis, or size exclusion chromatography.

The modified nucleoside would be present at very low abundance amongst all the other unmodified nucleosides. This may seem like searching for a needle in a haystack. Fortunately, however, if one performs LC-MS to analyze these RNA hydrolyzate products, the modified nucleoside can be separated from the highly abundant major nucleosides and subsequently detected. Unfortunately, the limitation, which has proved to be quite startling in scope, is the incredible amount of sample preparation time. If a more efficient way of isolating these small RNAs than gels could be investigated, the preparation and successive analysis of modified nucleosides present in a heterogeneous mixture of miRNAs and siRNAs would be reasonable to accomplish.
Chapter 3: Comparing Direct ESI-MS to LC-MS for the Analysis of Modified Nucleosides

3.1 Introduction

Modified nucleosides are ubiquitous in many different types of RNA. In tRNA alone, there are 91 various nucleoside modifications, approximately 70 of which are unique to tRNA [2]. It is thought that modified nucleosides play a significant role in determining the structure and function of an RNA molecule. The exact role of each modified nucleoside is not yet known, although there have been many reasonable explanations for individual roles [9-12].

While mature tRNA molecules always contain modifications, there are other types of RNA, such as 5S rRNA, that may not contain a modification, depending on the organism from which they originate. For example, 5S rRNA in *E. coli* contains no modified nucleosides, whereas the 5S rRNA in *Sulfolobus solfataricus* and *Pyrodictium occultum* each contains two modified nucleosides [64-65]. Further, some miRNAs and siRNAs contain methylated nucleosides, such as those enzymatically processed by *HEN1*. Development of a screening method for nucleosides would be useful as a preliminary step in the study of nucleoside modifications.

Most mass spectrometry methods for determining posttranscriptional modifications of RNA at the nucleoside level are based on the use of a separation technique, primarily via online chromatography (e.g., LC-MS). In this work, a comparison was made between a standard LC-MS approach for nucleoside characterization to one based upon direct ESI-TOF-MS analysis of unseparated mixtures of nucleosides. These approaches were developed using tRNA Val and tRNA Tyr1.
### Table 3.1 The nucleosides found in tRNA\textsuperscript{Val}, the number of times they occur, and relative mole ratios

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>Abundance</th>
<th>Relative Mole Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytidine</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>Uridine</td>
<td>9</td>
<td>0.39</td>
</tr>
<tr>
<td>Guanosine</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>Adenosine</td>
<td>14</td>
<td>0.61</td>
</tr>
<tr>
<td>Pseudouridine</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>4-thiouridine</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Dihydrouridine</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>6-methyladenosine</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>7-methylguanosine</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Ribosylthymine (5-methyluridine)</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Uridine 5-oxyacetic acid</td>
<td>1</td>
<td>0.04</td>
</tr>
</tbody>
</table>

### Table 3.2 The nucleosides found in tRNA\textsuperscript{Tyr\textsuperscript{1}}, the number of times they occur, and relative mole ratios

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>Abundance</th>
<th>Relative Mole Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cytidine</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>Uridine</td>
<td>12</td>
<td>0.44</td>
</tr>
<tr>
<td>Guanosine</td>
<td>12</td>
<td>0.44</td>
</tr>
<tr>
<td>Adenosine</td>
<td>18</td>
<td>0.67</td>
</tr>
<tr>
<td>Pseudouridine</td>
<td>2</td>
<td>0.07</td>
</tr>
<tr>
<td>4-thiouridine</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>2-O-methylguanosine</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Queuosine</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>2-methylthio-\textsuperscript{N}\textsuperscript{6}-isopentenyladenosine</td>
<td>1</td>
<td>0.04</td>
</tr>
<tr>
<td>Ribosylthymine (5-methyluridine)</td>
<td>1</td>
<td>0.04</td>
</tr>
</tbody>
</table>
3.2 Experimental

3.2.1 Materials

\( E. coli \) tRNA\(^{\text{Tyr}}\) (catalog # R0258), \( E. coli \) tRNA\(^{\text{Val}}\) (catalog # R2645), ammonium bicarbonate (catalog # A6141), adenosine (catalog # A9251), cytidine (catalog # C4654), and venom phosphodiesterase I (catalog # P6877) were purchased from Sigma (St. Louis, MO, USA). Nuclease P1 is available from Sigma (catalog # N8630). Alkaline phosphatase (catalog # LS006124) was purchased through Worthington (Lakewood, NJ, USA). Methanol (catalog # MS1922) and acetonitrile (catalog # AS1122) were HPLC-grade, purchased from Tedia Company (Fairfield, OH, USA). Ammonium acetate (catalog # A639) and formic acid (catalog # A118\(^{\text{F}}\)) were purchased from Fisher Scientific (Hampton, NH, USA). Glacial acetic acid (catalog # 281000ACS) was purchased from Pharmco (Brookfield, CT, USA). Nanopure water (18M\(\Omega\)) was obtained from a Barnstead (Dubuque, IA, USA) System and was autoclaved prior to use.

3.2.2 Methods

3.2.2.1 Enzymatic Digestion

RNA was digested to its component nucleosides using the previously described protocol in Chapter 2 (§2.2.3.1).

3.2.2.2 HPLC-UV of Nucleoside Digests

The enzymatic hydrolyzate was separated using reversed-phase HPLC. The column used to verify the presence of nucleosides and a complete digestion was a 4.6 mm X 25 cm Supelcosil LC-18S column by the same method described in Chapter 2 (§2.2.3.2).
3.2.2.3 Direct Infusion Electrospray Mass Spectrometry

Samples were analyzed using a Perseptive Biosystems Mariner electrospray time of flight (ESI-TOF) mass spectrometer in positive-ion mode. Adenosine and cytidine standards were used to optimize instrumental parameters. Products of the enzymatic digest, dissolved in methanol, autoclaved water, and formic acid (50:49.9:0.1, v/v), were directly injected at 0.500 µL/min using a Harvard apparatus syringe pump.

Table 3.3 Instrumental parameters used during direct infusion mass spectrometry experiments

<table>
<thead>
<tr>
<th>Instrumental Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spray Tip Potential</td>
<td>1889.65 V</td>
</tr>
<tr>
<td>Nozzle Potential</td>
<td>73.00 V</td>
</tr>
<tr>
<td>Skimmer 1 Potential</td>
<td>7.50 V</td>
</tr>
<tr>
<td>Quadrupole DC Potential</td>
<td>5.49 V</td>
</tr>
<tr>
<td>Deflection Voltage</td>
<td>0.39 V</td>
</tr>
<tr>
<td>Einzel Lens Potential</td>
<td>-24.00 V</td>
</tr>
<tr>
<td>Quadrupole RF Voltage</td>
<td>689.94 V</td>
</tr>
<tr>
<td>Quadrupole Temperature</td>
<td>140.01 ºC</td>
</tr>
<tr>
<td>Nozzle Temperature</td>
<td>140.01 ºC</td>
</tr>
<tr>
<td>Push Pulse Potential</td>
<td>550.89 V</td>
</tr>
<tr>
<td>Pull Pulse Potential</td>
<td>234.89 V</td>
</tr>
<tr>
<td>Pull Bias Potential</td>
<td>2.52 V</td>
</tr>
<tr>
<td>Acceleration Potential</td>
<td>3999.94 V</td>
</tr>
<tr>
<td>Reflector Potential</td>
<td>1569.98 V</td>
</tr>
<tr>
<td>Detector Voltage</td>
<td>1975.21 V</td>
</tr>
</tbody>
</table>

3.2.2.4 Microbore and Capillary-LC Separation of Nucleoside Digests and LC-MS

A 1.0 mm x 30.0 cm LC-18S column, with 5 µm particle size and 120 Å pore size (catalog # 57920), was purchased from Supelco (Bellefonte, PA, USA). The gradient
Crain described for separating nucleosides was scaled down to accommodate the lower flow rate and smaller volume of a microbore column, using Equation 2-2. There were a few minor alterations, listed in Table 3.4.

**Table 3.4 Gradient used on microbore column**

<table>
<thead>
<tr>
<th>Time (minutes)</th>
<th>Percent A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>98</td>
</tr>
<tr>
<td>0.10</td>
<td>96</td>
</tr>
<tr>
<td>4.00</td>
<td>90</td>
</tr>
<tr>
<td>10.0</td>
<td>85</td>
</tr>
<tr>
<td>14.0</td>
<td>80</td>
</tr>
<tr>
<td>16.0</td>
<td>75</td>
</tr>
<tr>
<td>19.0</td>
<td>50</td>
</tr>
<tr>
<td>21.5</td>
<td>25</td>
</tr>
<tr>
<td>22.5</td>
<td>25</td>
</tr>
<tr>
<td>24.5</td>
<td>1</td>
</tr>
<tr>
<td>25.5</td>
<td>1</td>
</tr>
<tr>
<td>28.0</td>
<td>99</td>
</tr>
<tr>
<td>50.0</td>
<td>99</td>
</tr>
</tbody>
</table>

A graduated micro-splitter valve was used to split the flow at the pumps from 450 µL/min to 50 µL/min leading to the column. In addition, a back-pressure regulator was added to the tubing after the detector. The HPLC used was a Hitachi D-7000, with an L-7100 pump, an L-7400 UV detector, and a semi-micro flow cell.

A 0.15 mm x 10 cm TARGA C18 column, with 5 µm particle size and 120 Å pore size, was used to perform capillary LC on nucleoside digests. The flow rate used was 1.06 µL/ min, split based on the column’s measured CRP value of 266. The capillary LC system used was a Dionex/LC Packings Ultimate system, equipped with a LC Packings Famos autosampler (Sunnyvale, CA, USA). The gradient used on this column is identical to that mentioned in Chapter 2 (§2.2.2.3.3).
3.3 Results and Discussion

3.3.1 Enzymatic Digest

RNA samples were successfully digested according to a previously described protocol, depicted in Figure 3.1.

Figure 3.1 Schematic of enzymatic hydrolysis of RNA

3.3.2 HPLC-UV of Nucleoside Digests

Analytical scale chromatography was utilized to verify the success of a nucleoside digest (Figure 3.2). In the case of tRNA^{Val}, all unmodified and modified nucleosides showed peaks near predicted retention times, within ± 1 minute, with the exception of one nucleoside. 5,6-dihydrouridine does not possess a significant chromophore like the other nucleosides and therefore is very difficult to detect using UV spectrophotometry at 254 nm [66].
Figure 3.2 A representative chromatogram from the 4.6 mm x 25.0 cm LC-18S column, preceded by a 4.6mm guard column [57]. Sample was 46 µg of the nucleoside digestion product of tRNA$^{Val}$, corresponding to approximately 10 µg of each major nucleoside.

3.3.3 Direct Infusion Analysis of Modified Nucleosides

To gain further information about the modified nucleosides present, the nucleosides digested from tRNA$^{Tyr}$ were directly infused into a mass spectrometer. After the conditions for analyzing nucleosides on the ESI-TOF-MS were optimized, tRNA digestion mixtures were analyzed at various concentrations, all dissolved in 50% methanol, 49.9% water, and 0.1% formic acid. The best results were obtained at a concentration of approximately 8.5 µg/mL of the digestion, because at this concentration, five different modified nucleosides could be detected above the signal-to-noise ratio (Figure 3.3). Control experiments were also performed to verify the origin of other major peaks found in the mass spectrum.
Figure 3.3 Direct infusion ESI-MS of 8.5 µg/mL of tRNA\textsuperscript{Tyr\textsubscript{1}} digestion

Table 3.5 Origin of peaks within mass spectrum that are not attributable to nucleosides, confirmed by control experiments

<table>
<thead>
<tr>
<th>Peak</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>248</td>
<td>Digest</td>
</tr>
<tr>
<td>279</td>
<td>Buffer</td>
</tr>
<tr>
<td>291</td>
<td>Buffer</td>
</tr>
<tr>
<td>313</td>
<td>Buffer</td>
</tr>
<tr>
<td>335</td>
<td>Buffer</td>
</tr>
<tr>
<td>354</td>
<td>Buffer</td>
</tr>
<tr>
<td>363</td>
<td>Digest</td>
</tr>
<tr>
<td>368</td>
<td>Digest</td>
</tr>
<tr>
<td>391</td>
<td>Buffer</td>
</tr>
</tbody>
</table>

The known modifications in \textit{E. coli} tRNA\textsubscript{Tyr\textsubscript{1}} are pseudouridine (Ψ), 5-methyluridine (m\textsuperscript{5}U), 4-thiouridine (s\textsuperscript{4}U), 2’-O-methylguanosine (Gm), queuosine (Q) and N6-(3-Methyl-2-butenyl)-2-methylthioadenosine (ms\textsuperscript{2i6}A). The limit of detection for ms\textsuperscript{2i6}A is roughly ten times lower than the limits of detection for m\textsuperscript{5}U, s\textsuperscript{4}U, Gm and Q.
In all cases, these modified nucleosides can be detected even in the presence of the more concentrated major (i.e., unmodified) nucleosides.

Table 3.6 Limits of detection for various modified nucleosides from tRNA$^{\text{Tyr}}$

<table>
<thead>
<tr>
<th>Nucleoside</th>
<th>Symbol</th>
<th>Limit of Detection (ng)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-thiouridine</td>
<td>$s^4U$</td>
<td>0.40</td>
</tr>
<tr>
<td>2-O-methylguanosine</td>
<td>Gm</td>
<td>0.46</td>
</tr>
<tr>
<td>Queuosine</td>
<td>Q</td>
<td>0.64</td>
</tr>
<tr>
<td>2-methylthio-$N^6$-isopentenyladenosine</td>
<td>ms$^{2i6}$A</td>
<td>0.059</td>
</tr>
<tr>
<td>Ribosylthymine (5-methyluridine)</td>
<td>m$^3U$</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Further, the response of these nucleosides generated a linear response in all concentrations of the digestion that were run, except in the most concentrated mixture (42.5 µg/mL), when the detector had become saturated (Figures 3.4 and 3.5).

Figure 3.4 The linear response of the nucleosides with higher signal-to-noise values
Figure 3.5 The linear response of nucleosides with lower signal-to-noise values

Pseudouridine, the only mass-silent modified nucleoside, requires derivatization or HPLC separation for confirmation. This was the only modified nucleoside present in tRNA<sub>Tyr</sub> that could not be unambiguously identified. In general, any nucleoside modifications resulting in the same shift in molecular weight would not be distinguishable by this method. For example, a common modification to a nucleoside is methylation. However, by using direct infusion mass spectrometry with no separation, the three modified adenosines in Figure 3.6 would not be immediately distinguishable.

Figure 3.6 A. 1-methyladenosine, B. 2-methyladenosine, and C. 2'-O-methyladenosine. All have the molecular weight of 281.27 Da.
Differentiating between such nucleosides would be possible using LC-MS, because these three nucleosides, though identical in mass, have different retention times. The direct infusion portion of the experiment would serve as a screening technique. It shows whether there are any modifications to the four common nucleosides prior to performing more elaborate LC-MS experiments, which are necessary to fully characterize the modification.

3.3.4 LC-MS of Modified Nucleosides

Two different columns interfaced to the Thermo LTQ-MS were used to separate nucleoside digests. Below are chromatograms produced from each of the columns, depicting a separation of the same amount of the nucleoside test mixture. It is evident that the 0.15 mm i.d. column separated the nucleosides with similar resolution in a comparable amount of time, but resulted in significantly higher absorbances, due to the lower flow rate used. Therefore, the 0.15 mm i.d. column is the preferred column to interface to the mass spectrometer.

![Chromatogram from 1.0 mm column, separating components from Supelco’s nucleoside test mix. Each peak contains between 1-10 ng of a nucleoside.](image)

**Figure 3.7** Chromatogram from 1.0 mm column, separating components from Supelco’s nucleoside test mix. Each peak contains between 1-10 ng of a nucleoside.
Figure 3.8 Separation of nucleoside test mix using 0.15 mm column, containing 1-10 ng of each nucleoside.

LC-MS was performed by interfacing the LC Packings capillary LC system with the Thermo LTQ-MS. A mixture of four standard nucleosides (adenosine, cytidine, guanosine, and uridine) was injected in various concentrations. The limit of detection was found to be 0.2 ng per nucleoside. The limit of detection for LC-UV theoretically should be above that for the mass spectrometer, so this number provides insight regarding the minimal amount of sample necessary for dual detection by UV and MS. As stated previously, using LC-MS versus UV detection alone gives an increased dynamic range in analysis [61].

Figure 3.9 Chromatogram depicting the separation of 0.2 ng of each of the four major nucleosides
3.3.5 Comparison of Direct Infusion Mass Spectrometry to LC-MS

Direct infusion mass spectrometry allowed detection of all mass-shifting modified nucleosides in a digested sample of tRNA$^{\text{Tyr}}$. The modified nucleosides could be detected when the concentration of the tRNA from which they were digested was 8.5 µg/mL or greater. The molecular weight of tRNA$^{\text{Tyr}}$ is 27352 Da, so at a modified nucleoside molecular weight of 280 Da, the concentration in terms of a single modified nucleoside corresponds to 87 ng/mL. This concentration can be used to find a detection limit for nucleoside detection by taking into account the flow rate (0.500 µL/min) and the amount of time the experiment takes to run (10 minutes). Therefore, the limit of detection for finding a modified nucleoside in a tRNA molecule using direct infusion mass spectrometry is 0.4 ng.

Comparing this limit of detection to that found in capillary LC-UV experiments (0.2 ng), this shows that performing a direct infusion experiment prior to an LC-UV experiment uses only slightly more sample but could save a great deal of time, depending on the result of the experiment. One possible result is that the direct infusion experiment shows that no modifications are present. However, it could also pinpoint a modification that should be further investigated using LC-MS.

3.4 Conclusion

This chapter has described a method by which to screen a simple RNA sample for the presence of modified nucleosides. If a modified nucleoside were indeed present, it would then be necessary to perform LC-MS on the sample to fully characterize the
modification. If no modifications were shown in the direct infusion data, this would obviate the need for LC-MS experiments to search for a non-existent modification.

One possible problem could be that a modified nucleoside is present, but at extremely low abundance. Ion suppression effects by the other major nucleosides could prevent the minor nucleoside from being detected. In such a case, LC-MS would not be performed and the modification would go unnoticed. To prevent such a loss of information, one should consider the circumstances under which this method was developed. The modified nucleosides were detected amongst major nucleosides resulting from a tRNA digestion. tRNAs are relatively small RNA molecules, ranging from 70-90 nucleotides. A modified nucleoside would be more abundant, and thus easier to detect amongst the major nucleosides in a tRNA (1 out of 80) versus a more complex RNA sample, such as a ribosomal RNA, where there may be only one modification amongst thousands of unmodified nucleosides.

**Table 3.7** Modified nucleosides in *E. coli* 16S rRNA and mole abundances, relative to guanosine

<table>
<thead>
<tr>
<th>Modified Nucleoside</th>
<th>Relative Mole Ratio (to most abundant major nucleoside)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guanosine</td>
<td>1</td>
</tr>
<tr>
<td>Pseudouridine</td>
<td>0.002</td>
</tr>
<tr>
<td>7-methylguanosine</td>
<td>0.002</td>
</tr>
<tr>
<td>2-methylguanosine</td>
<td>0.006</td>
</tr>
<tr>
<td>5-methylcytidine</td>
<td>0.004</td>
</tr>
<tr>
<td>$N^4,2'-O$-dimethylcytidine</td>
<td>0.002</td>
</tr>
<tr>
<td>3-methyluridine</td>
<td>0.002</td>
</tr>
<tr>
<td>$N^6,N^2$-dimethyladenosine</td>
<td>0.004</td>
</tr>
</tbody>
</table>
Chapter 4: Conclusions and Future Perspectives

4.1 Summary

The objective of this work was to utilize mass spectrometry to identify the posttranscriptional modifications present in RNA. RNA is involved in protein synthesis and regulation of the expression of genes. Posttranscriptional modifications are frequently crucial to the structure and function of an RNA molecule; consequently, the characterization of modifications is necessary to fully understand the biochemical role of RNA.

To do so, work was presented concerning the development of a method to isolate and characterize relatively new types of RNA, miRNAs and siRNAs. Though the technique was successful at isolating miRNA and siRNA, it was not very practical, due to the extensive effort required to accumulate enough sample to analyze. However, the methods developed in this work may be used to guide future work regarding sample preparation for miRNAs and siRNAs and their subsequent analysis by LC-UV or LC-MS.

Further, a screening method for posttranscriptional modifications was developed. Posttranscriptional modifications were analyzed by first enzymatically digesting RNA to nucleosides. The digestion product, a mixture of nucleosides, was analyzed using positive-ion mode direct infusion mass spectrometry. When a peak resulting from a mass shift of a major nucleoside was observed, this signaled that a modified nucleoside was likely to be present. LC-MS experiments can be used to confirm the identity of any modifications.
4.2 Future Work

To achieve a successful analysis of the posttranscriptional modifications present in a heterogenous mixture of miRNAs and siRNAs, it would be beneficial to investigate higher throughput methods of isolation. Size exclusion chromatography or preparative gel columns are reasonable approaches. In this work, the analysis of nucleosides generated by enzymatic hydrolysis of RNA is currently limited to LC-UV. However, since LC-MS could provide the exact identity of modified peaks, work needs to be done to optimize the mass spectrometry. Because electrospray ionization is hampered by contaminants, it would be advantageous to include washing steps for the final RNA sample isolated.

Due to the limitations of direct infusion mass spectrometry on nucleoside digests, it is likely that modified nucleosides in a complex mixture of miRNAs and siRNAs could not be detected due to ion suppression from the more abundant major nucleosides. Direct infusion analysis could be used if a single miRNA or siRNA was isolated. This could be accomplished by using affinity purification with biotinylated beads that contain a sequence complementary to a specific miRNA or siRNA sequence [56]. Using the direct infusion method presented here would provide additional information to what was found by B.Yu, et. al, which was that a methylation was present but the location was not elucidated. Analyzing nucleosides from an enzymatic digest would show what type of nucleoside contains the methylation. By doing MS/MS or nozzle-skimmer dissociation, it could be confirmed that the methylation is on the sugar.

In 2002, *HEN1* was identified as a protein involved in the biosynthetic pathway of miRNAs and siRNAs in *A. thaliana* [27]. In 2005, the precise function of *HEN1* was
elucidated: it is a methyltransferase that acts on the 2’-OH of the 3’ terminal nucleotide of miRNAs and siRNAs [29]. In plants with nonfunctional copies of this protein, *hen1* mutants, it has been shown using gel mobility studies that small RNAs are not methylated, while they are methylated in *HEN1* wild type cells [67]. Performing such studies involves chemical treatment of RNA, which causes those nucleotides that are not modified at the sugar to have the terminal nucleotide cleaved, which causes a shift in gel mobility. However, if a modification is present on the sugar, no reaction happens and the mobility of the small RNA is unchanged [67].

In future studies, mass spectrometry could provide an alternative technique for studying the methylation status of the 3’ terminal nucleotide of miRNAs and siRNAs. Performing an enzymatic digest on the RNAs, followed by mass spectrometry analysis, should show numerous methylated nucleosides in the wild type cells, versus no methylated nucleosides in mutant cells. Doing so would provide a more direct measurement (a mass shift of 14 Da) than the slight change in gel movement seen in current methodology.

**Figure 4.1** Example of experimental methodology currently in use for detecting a methylated ribose sugar. In the lane labeled *hen1-1* –β, this shows the regular miRNA, before chemical treatment. In the lane labeled *hen1-1* +β, this corresponds to a miRNA that was sensitive to chemical reaction and shifted movement due to the lack of a 2’-OH methyl group. If the sugar is methylated, the result would look like the Ler lanes, with no shift in migration upon treatment [67].
References


[27] Park, W., Li, J., Song, R., Messing, J., Chen, X. “CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in *Arabidopsis thaliana*” *Current Biology* **12** 1484 (2002).


