Mass spectrometry-based rnomics: global surveys of ribonucleotide modifications as possible indicators of cell identity, epigenetic, metabolic and pathological state

Rebecca Erin Rose
University at Albany, State University of New York, rrose66@gmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Analytical Chemistry Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/1490

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive.
Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
MASS SPECTROMETRY-BASED RNOMICS:

GLOBAL SURVEYS OF RIBONUCLEOTIDE MODIFICATIONS AS POSSIBLE INDICATORS OF CELL IDENTITY, EPIGENETIC, METABOLIC, AND PATHOLOGICAL STATE

by

Rebecca Erin Rose

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy

College of Arts & Sciences
Department of Chemistry
2015
MASS SPECTROMETRY-BASED RNOMICS:

GLOBAL SURVEYS OF RIBONUCLEOTIDE MODIFICATIONS AS POSSIBLE INDICATORS OF CELL IDENTITY, EPIGENETIC, METABOLIC, AND PATHOLOGICAL STATE

by

Rebecca Erin Rose

COPYRIGHT 2015
**ABSTRACT**

Discovery of the regulatory roles of non-protein coding RNAs (ncRNAs), once considered “junk” or “transcriptional noise”, has prompted a reassessment of the significance of the multifaceted activities of RNA in cellular response, which can result in cell transformation and diseased states. This reassessment, however, can only be accomplished if techniques are developed to capture, characterize and quantify the more than 100 covalent post-transcriptional modifications (PTMs) that adorn natural RNAs. Undoubtedly, the creation of high-throughput platforms such as next-generation sequencing and RNA-seq have provided unparalleled accuracy and sensitivity, however, they rely on strand amplification procedures that are blind to the known variants. In contrast, mass spectrometry (MS) techniques operate directly on the genuine RNA sample, not a DNA copy, and can provide the identity and sequence position of all PTMs according to their mass and fragmentation characteristics. Often, these MS approaches are coupled to a front-end separation technique such as liquid chromatography which rely on the use of standards and are ridden with sample bias and carry-over issues.

For these reasons, this work focuses on the creation of a direct infusion MS platform to unambiguously characterize the global profiles of PTMs in complex cellular mixtures, which takes advantage of the complementary approaches of high-resolution MS and ion mobility spectrometry-MS. To date, most work on PTMs has focused on three major classes of RNAs (e.g., tRNA, rRNA and mRNA), but has failed to accurately capture overall PTM content. In contrast, our platform provided a unique way of assessing the global distribution and studying the effect of PTM on RNA function. Its utilization enabled the direct monitoring of distinct PTM profiles
associated with cell type and metabolism and was implemented to monitor the abundance of
PTMs as a function of environmental conditions capable of capturing the dynamic nature of the
epitranscriptome. As a whole, this work has provided the necessary technological innovation and
discussion of experimental strategies needed to address distinct biological roles that PTMs play
in different regulatory events governing disease. Ultimately, these approaches can be applied
towards many applications including, but not limited to, enhancing diagnostic techniques, drug
discovery and therapeutics.
PREFACE

Since the completion of the Human Genome Project, advances in analytical techniques capable of capturing specific RNAs within the cell have led to the discovery of many new functional roles for this essential biopolymer. Non-coding RNAs have become increasingly more interesting due to their ability to silence genomic DNA and mRNA, which makes them essential players in regulatory mechanisms. Despite the growing recognition of the activity of specific ncRNAs in pathways activated by stress response, which can often lead to various cellular pathologies, the analytical methods that have been implemented in lock-step to study these mechanisms fail to accurately capture the more than 100 covalent post-transcriptional modifications (PTMs) that may decorate these natural RNAs. This failure only hits upon the most primary level of information attributable to these particular species: their existence.

With the exception of 2-O’-methylation, adenosine-N6-methylation, cytosine-5-methylation, and pseudouridylation, all other PTMs elude established sequencing techniques that are based on hybridization, strand amplification and fluorescent labelling. Fortunately, this issue has not gone unnoticed, and multiple advancements in mass spectrometry (MS) techniques have been implemented. These techniques are extremely advantageous because they are able to accurately capture any PTM located on the native RNA strand. To date, most mass spectrometric approaches have been limited by the sensitivity of the instrumentation, and thus have required larger sample concentrations. Additionally, many approaches have coupled a front-end separation technique such as capillary electrophoresis or liquid chromatography. While these techniques have unparalleled sensitivity, they also rely on the use of standards and often
are burdened by carry-over issues and sample bias. As such, we have sought to overcome these intrinsic limitations by targeting the following aims:

a) developing MS-based approaches for PTM analysis at the entire transcriptome level;
b) enabling quantification of expression levels at the entire transcriptome level; and
c) exploring the significance of the epitranscriptome in stress response.

Ultimately, the ability to implement this streamlined mass spectrometric platform will allow for the characterization of entire transcriptomes based on PTM content. Given that this platform will be able to evaluate the PTM content in both qualitative and quantitative fashion, it will be expected to enable the differentiation of cells by type, metabolic state, stressed state, and cell pathology. Additionally, the investigation of PTM biogenesis has typically focused on individual pathways without exploring the possible communications between their expression controls. The ability to monitor simultaneously the up-/down-regulation of all observed PTMs will provide valuable insights into putative regulatory relationships. New evidence of participation in cellular sensing and response to environmental stress suggests that PTMs may represent essential cogs in signaling networks between the genome and the proteome, working perhaps in both directions. This hypothesis finds a rationale in the enzymatic genesis of modifications and their ability to affect essential binding interactions between RNA and both proteins and DNA. This project specifically sought the evidence necessary to prove/disprove this hypothesis by looking for correlations between PTM expression and corresponding proteome and interactome maps. Ultimately, the completed work provided a greater understanding of the functions of PTMs and their place in RNA biology.
DEDICATION

To Mom and Dad, without whom I would be lost.
For your continued unwavering love and commitment to fostering my individuality and growth.
   For helping me to see my potential amidst my self-doubts.
For giving me the confidence to take risks and follow my passions even in the face of adversity.
For instilling in me the courage to stand up for my convictions despite the masses who did not agree.
   For showing me how to start from little and build a remarkable life full of quality and achievements and things of which to be proud.
   For providing me with family, consistency, and stability.
For the values you have helped me cultivate: respect, passion, integrity, perseverance, confidence, honesty; and most importantly, love.

Cheers to you, Mom and Dad. This accomplishment is as much mine as it is yours.
ACKNOWLEDGMENTS

I would like to acknowledge, with extreme gratitude, my thanks to Dr. Daniele Fabris, who, at times, had more faith in me than I had in myself. Thank you for the many fruitful (and often sarcastic) conversations that continually and relentlessly pushed me toward the “bigger picture” that I could not always see. Finally, thank you for showing me that those who are unafraid to buck trends and think creatively are those who can live unencumbered by their own passions.
# TABLE OF CONTENTS

Abstract ........................................................................................................................................ iii
Preface ........................................................................................................................................ vi
Dedication ................................................................................................................................... vii
Acknowledgments ......................................................................................................................... viii
Table of Contents ......................................................................................................................... ix
List of Tables ................................................................................................................................ x
List of Schemes/Equations ........................................................................................................... xii
List of Figures .............................................................................................................................. xii
List of Abbreviations .................................................................................................................... xv

## Chapter 1. Introduction ............................................................................................................... 1

1.1 Discovery of RNA Post-transcriptional Modifications ......................................................... 1
1.2 Detection of Post-transcriptional Modifications ................................................................. 2
1.3 Structural Characterization of Ribonucleotide Variants by Mass Spectrometry .................. 6
1.4 Locating Post-transcriptional Modifications in Parent RNAs ............................................. 12
1.5 MS-based Characterization of Modified RNA Strands ....................................................... 17
1.6 RNA PTM Databases and Repositories ............................................................................. 23
1.7 Biological Functions of RNA: Significance of PTMs ......................................................... 24
1.8 Conclusions ......................................................................................................................... 29

## Chapter 2. MS-based Approaches for the Analysis of Post-transcriptional Modifications at the Entire Transcriptome Level ....................................................................................... 32

2.1 Introduction ......................................................................................................................... 32
2.2 Materials and Methods ....................................................................................................... 34
  2.2.1 Mononucleotide standards .............................................................................................. 34
  2.2.2 Cellular extracts .............................................................................................................. 35
  2.2.3 Molecular mass determination ....................................................................................... 36
  2.2.4 Tandem mass spectrometry .......................................................................................... 37
  2.2.5 Data analysis .................................................................................................................. 38
2.3 Results and Discussion ......................................................................................................... 41
  2.3.1 Identification of PTMs based on mass ........................................................................... 41
  2.3.2 Tandem MS approaches and isomeric PTMs ................................................................. 46
  2.3.3 IMS-MS techniques ..................................................................................................... 52
LIST OF TABLES

Table 2-1: Elemental compositions and monoisotopic mass of neutral 5’-phosphonucleosides .......................................................... 42

Table 2-2: Hit obtained by searching ESI data against non-redundant database .. 67

Table 3-1: Figures of merit obtained by analyzing isolated tRNA\textsuperscript{Phe} from \textit{S. cerevisiae} .............................................................................. 83

Table 3-2: Quantitative determination of ribonucleotides present in total RNA extract of \textit{S. cerevisiae} .............................................................................. 85

Table 3-3: Reproducibility of ESI-MS analysis of total RNA extract from \textit{S. cerevisiae} grown in YPD medium........................................ 87

Table 4-1: Hits provided by a total RNA extract from \textit{S. cerevisiae} grown in synthetic complete medium (SC) ........................................ 102

Table 4-2: Hits provided by a total RNA extract from \textit{E. coli} grown in synthetic complete medium (SC) ........................................ 103

Table 4-3: Quantitative determination of ribonucleotides present in total RNA extract of \textit{S. cerevisiae} grown under various stressors .......... 108

Table 4-4: Summary new PTMs expressed in WT, \textit{ssb1Δ, rit1Δ} mutant under either control or stress conditions ........................................ 113
LIST OF SCHEMES/EQUATIONS

Scheme 2-1: Proposed MS workflow .................................................................39

Scheme 2-2: Proposed experimental workflow ..................................................62

Scheme 4-1: Regulation of stress genes in fission and budding yeast ...............115

Scheme 4-2: Established and putative relationships between genes involved in response to hyperosmotic stress and heat shock .........................117

Equation 2-1: Regulation of stress genes in fission and budding yeast ............40
LIST OF FIGURES

Figure 1-1: Neutral structures of the four methyl-guanine isomers. ..................4
Figure 1-2: MS\textsuperscript{n} fragmentation acquired with an ion trap analyzer.........8
Figure 1-3: IMS-MS analysis of \textit{S. cerevisiae} ..........................................................11
Figure 1-4: Isolation and fragmentation of tRNA\textsubscript{Val} on Orbitrap analyzer ..........21
Figure 2-1: ESI-MS spectrum of standard nucleotide mixture in LTQ analyzer. .....44
Figure 2-2: MS\textsuperscript{n} of AMP and dGMP from standard nucleotide mixture ..........47
Figure 2-3: Negative ion mode MS/MS spectra of UMP and ΨMP ..................49
Figure 2-4: Ion mobility profiles obtained from UMP and ΨMP ..................54
Figure 2-5: Comprehensive representation of IMS-MS data obtained from the entire nucleotide mixture. ...............................................................56
Figure 2-6: Heat-map representation of time aligned parallel (TAP) dissociation products provided by the entire nucleotide mixture ..................58
Figure 2-7: IMS-MS profile of m/z 378 after isolation in the mass-selective quadrupole and separation in the ion mobility element. ..............60
Figure 2-8: Representative ESI-MS spectrum of total RNA digest obtained from \textit{S. cerevisiae} grown in YPD medium. ..................64
Figure 3-1: tRNA\textsuperscript{phe} from \textit{S. cerevisiae}. ..................................................................79
Figure 3-2: ESI-MS spectrum of digestion mixture obtained from \textit{S. cerevisiae} tRNA\textsuperscript{Phe} .................................................................81
Figure 4-1: IMS-MS heat-map and differential plot obtained from \textit{S. cerevisiae} grown in synthetic complete (SC) medium. ..................98
Figure 4-2: IMS-MS heat map and difference plot obtained from \textit{E. coli} grown in SC medium. .................................................................100
**Figure 4-3:** Representative heat maps obtained by anionic IMS-MS analysis of *S. cerevisiae* mutants grown in YPD medium. .................................................................106

**Figure 5-1:** Mass mapping of a tRNA\textsuperscript{Lys} sample submitted to limited RNase T1 digestion. ......................................................................................................................127
LIST OF ABBREVIATIONS

ac$^4$C  N4-acetylcytidine
ac$^4$Cm  N4-acetyl-2'-O-methylcytidine
ac$^6$A  N6-acetyladenosine
acp$^3$U  3-(3-amino-3-carboxypropyl)uridine
aix  absolute intensity
Am  2'-O-methyladenosine
Ar(p)  2'-O-ribosyladenosine (phosphate)
AvP  abundance versus proxy
C+  agmatidine
CARD  comparative analysis of RNA digests
cDNA  complementary DNA
CE  capillary electrophoresis
CE-MS  capillary electrophoresis-mass spectrometry
CESR  core environmental stress response
chm$^5$U  5-(carboxyhydroxymethyl)uridine
CID  collisional induced dissociation
Cm  2'-O-methylcytidine
cm$^5$U  5-carboxymethyluridine
CMC  N-cyclohexyl-N'-b(4-methylmorpholinium) ethylcarbodiimide p-tosylate
cmnm$^5$s$^2$U  5-carboxymethylaminomethyl-2-thiouridine
cmnm$^5$U  5-carboxymethylaminomethyluridine
cmnm$^5$UUm  5-carboxymethylaminomethyl- 2'-O-methyluridine
cmo$^5$U  uridine 5-oxyacetic acid
cnm$^5$U  5-cyanomethyl-uridine
CRAC  crosslinking and cDNA analysis
cri  corresponding to the respective absolute intensity
D  dihydrouridine
DNA  deoxyribonucleic acid
dNMPs  deoxyribonucleotide monophosphates
EDD  electron detachment dissociation
ESI  electrospray ionization
ESI-MS  electrospray ionization-mass spectrometry
f$^5$C  5-formylcytidine
f$^5$Cm  5-formyl-2'-O-methylcytidine
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>FID</td>
<td>flame ionization detector</td>
</tr>
<tr>
<td>FTICR</td>
<td>Fourier transform ion cyclotron resonance</td>
</tr>
<tr>
<td>FWHM</td>
<td>full width at half maximum</td>
</tr>
<tr>
<td>G+</td>
<td>archaeosine</td>
</tr>
<tr>
<td>g^6A</td>
<td>N6-glycinylcarbamoyladenosine</td>
</tr>
<tr>
<td>galQ</td>
<td>galactosyl-queuosine</td>
</tr>
<tr>
<td>geneΔ</td>
<td>deletion of gene</td>
</tr>
<tr>
<td>GC</td>
<td>gas chromatography</td>
</tr>
<tr>
<td>gcmnm^5s^2U</td>
<td>geranlated 5-carboxymethylaminomethyl-2-thiouridine</td>
</tr>
<tr>
<td>Gm</td>
<td>2'-O-methylguanosine</td>
</tr>
<tr>
<td>gmnm^5s^2U</td>
<td>geranlated 5-methylaminomethyl-2-thiouridine</td>
</tr>
<tr>
<td>Gr(p)</td>
<td>2'-O-ribosylguanosine (phosphate)</td>
</tr>
<tr>
<td>hm^5C</td>
<td>5-hydroxymethylcytidine</td>
</tr>
<tr>
<td>hn^6A</td>
<td>N6-hydroxynorvalylcarbamoyladenosine</td>
</tr>
<tr>
<td>ho^5U</td>
<td>5-hydroxyuridine</td>
</tr>
<tr>
<td>HOG</td>
<td>hyper-osmolarity glycerol pathway</td>
</tr>
<tr>
<td>I</td>
<td>inosine</td>
</tr>
<tr>
<td>i^6A</td>
<td>isopentenyladenosine</td>
</tr>
<tr>
<td>Im</td>
<td>2'-O-methylinosine</td>
</tr>
<tr>
<td>imG</td>
<td>wyosine</td>
</tr>
<tr>
<td>imG-14</td>
<td>4-demethylwyosine</td>
</tr>
<tr>
<td>imG2</td>
<td>isowyosine</td>
</tr>
<tr>
<td>IMS-MS</td>
<td>ion mobility spectrometry-mass spectrometry</td>
</tr>
<tr>
<td>inm^5s^2U</td>
<td>5-(isopentenylaminomethyl)-2-thiouridine</td>
</tr>
<tr>
<td>inm^5U</td>
<td>5-(isopentenylaminomethyl)uridine</td>
</tr>
<tr>
<td>inm^5Um</td>
<td>5-(isopentenylaminomethyl)-2'-O-methyluridine</td>
</tr>
<tr>
<td>io^6A</td>
<td>N6-(cis-hydroxyisopentenyl)adenosine</td>
</tr>
<tr>
<td>IRMPD</td>
<td>infrared multiphoton dissociation</td>
</tr>
<tr>
<td>k^3C</td>
<td>lysidine</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>LC-MS</td>
<td>liquid chromatography-mass spectrometry</td>
</tr>
<tr>
<td>IncRNAs</td>
<td>long noncoding RNAs</td>
</tr>
<tr>
<td>LOD</td>
<td>limit of detection</td>
</tr>
<tr>
<td>LTQ</td>
<td>linear trap quadrupole</td>
</tr>
<tr>
<td>m/z</td>
<td>mass-to-charge ratio</td>
</tr>
<tr>
<td>m^1A</td>
<td>1-methyladenosine</td>
</tr>
<tr>
<td>Symbol</td>
<td>Name</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>m(^1)acp(^3)Y</td>
<td>1-methyl-3-(3-amino-3-carboxypropyl) pseudouridine</td>
</tr>
<tr>
<td>m(^1)Am</td>
<td>1,2'-O-dimethyladenosine</td>
</tr>
<tr>
<td>m(^1)G</td>
<td>1-methylguanosine</td>
</tr>
<tr>
<td>m(^1)Gm</td>
<td>1,2'-O-dimethylguanosine</td>
</tr>
<tr>
<td>m(^1)I</td>
<td>1-methylinosine</td>
</tr>
<tr>
<td>m(^1)Im</td>
<td>1,2'-O-dimethylinosine</td>
</tr>
<tr>
<td>m(^1)Y</td>
<td>1-methylpseudouridine</td>
</tr>
<tr>
<td>m(^2)(^2)G</td>
<td>N2,N2,7-trimethylguanosine</td>
</tr>
<tr>
<td>m(^2)G</td>
<td>N2,N2-dimethylguanosine</td>
</tr>
<tr>
<td>m(^2)Gm</td>
<td>N2,N2,2'-O-trimethylguanosine</td>
</tr>
<tr>
<td>m(^2)Gm</td>
<td>N2,7-dimethylguanosine</td>
</tr>
<tr>
<td>m(^2)Gm</td>
<td>N2,7,2'-O-trimethylguanosine</td>
</tr>
<tr>
<td>m(^2)A</td>
<td>2-methyladenosine</td>
</tr>
<tr>
<td>m(^2)G</td>
<td>N2-methylguanosine</td>
</tr>
<tr>
<td>m(^2)Gm</td>
<td>N2,2'-O-dimethylguanosine</td>
</tr>
<tr>
<td>m(^3)C</td>
<td>3-methylcytidine</td>
</tr>
<tr>
<td>m(^3)U</td>
<td>3-methyluridine</td>
</tr>
<tr>
<td>m(^3)Um</td>
<td>3,2'-O-dimethyluridine</td>
</tr>
<tr>
<td>m(^3)Y</td>
<td>3-methylpseudouridine</td>
</tr>
<tr>
<td>m(^4)(^2)Cm</td>
<td>N4,N4,2'-O-trimethylcytidine</td>
</tr>
<tr>
<td>m(^4)C</td>
<td>N4-methylcytidine</td>
</tr>
<tr>
<td>m(^4)Cm</td>
<td>N4,2'-O-dimethylcytidine</td>
</tr>
<tr>
<td>m(^5)C</td>
<td>5-methylcytidine</td>
</tr>
<tr>
<td>m(^5)Cm</td>
<td>5,2'-O-dimethylcytidine</td>
</tr>
<tr>
<td>m(^5)D</td>
<td>5-methylidihydrouridine</td>
</tr>
<tr>
<td>m(^5)s(^2)U</td>
<td>5-methyl-2-thiouridine</td>
</tr>
<tr>
<td>m(^5)U</td>
<td>5-methyluridine</td>
</tr>
<tr>
<td>m(^5)Um</td>
<td>5,2'-O-dimethyluridine</td>
</tr>
<tr>
<td>m(^6)(^2)A</td>
<td>N6,N6-dimethyladenosine</td>
</tr>
<tr>
<td>m(^6)(^2)Am</td>
<td>N6,N6,2'-O-trimethyladenosine</td>
</tr>
<tr>
<td>m(^6)A</td>
<td>N6-methyladenosine</td>
</tr>
<tr>
<td>m(^6)Am</td>
<td>N6,2'-O-dimethyladenosine</td>
</tr>
<tr>
<td>m(^6)(^6)A</td>
<td>N6-methyl-N6-threonylcarbamoyladenosine</td>
</tr>
<tr>
<td>m(^7)G</td>
<td>7-methylguanosine</td>
</tr>
<tr>
<td>m(^8)A</td>
<td>8-methyladenosine</td>
</tr>
</tbody>
</table>

**MALDI** matrix-assisted laser desorption/ionization

**MALDI-MS** matrix-assisted laser desorption/ionization-mass spectrometry
manQ  mannosyl-queuosine
MAPK  mitogen-activated protein kinase
MaSTeR  mass-selective time-resolved dissociation
mchm<sup>5</sup>U  5-(carboxyhydroxymethyl)uridine methyl ester
mcm<sup>5</sup>s<sup>2</sup>U  5-methoxycarbonylmethyl-2-thiouridine
mcm<sup>5</sup>U  5-methoxycarbonylmethyluridine
mcm<sup>5</sup>Um  5-methoxycarbonylmethyl-2'-O-methyluridine
mcmo<sup>5</sup>U  uridine 5-oxyacetic acid methyl ester
MIKES  mass-analyzed ion kinetic energy spectrum
mimG  methylwyosine
mnm<sup>5</sup>s<sup>2</sup>U  5-methylaminomethyl-2-thiouridine
mnm<sup>5</sup>se<sup>2</sup>U  5-methylaminomethyl-2-selenouridine
mnm<sup>5</sup>U  5-methylaminomethyluridine
mo<sup>5</sup>U  5-methoxyuridine
MRM  multiple reaction monitoring
mRNA  messenger RNA
MS  mass spectrometry
MS/MS  tandem mass spectrometry
ms<sup>2</sup>hn<sup>6</sup>A  2-methylthio-N6-hydroxynorvalyl carbamoyladenosine
ms<sup>2</sup>i<sup>6</sup>A  2-methylthio-N6-isopentenyladenosine
ms<sup>2</sup>io<sup>6</sup>A  2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine
ms<sup>2</sup>m<sup>6</sup>A  : 2-methylthio-N6-methyladenosine
ms<sup>2</sup>t<sup>6</sup>A  2-methylthio-N6-threonyl carbamoyladenosine
MSn  multiple step tandem mass spectrometry
ncm<sup>5</sup>U  5-carbamoylmethyluridine
ncm<sup>5</sup>Um  5-carbamoylmethyl-2'-O-methyluridine
ncRNAs  non-protein coding RNAs
nm<sup>5</sup>s<sup>2</sup>U  5-aminomethyl-2-thiouridine
NMPs  ribonucleotide monophosphates
NMR  nuclear magnetic resonance
NTPs  triphosphonucleotides
o<sup>2</sup>yW  peroxywybutosine
OD600  optical density at 600 nm
OH<sup>4</sup>yW  hydroxywybutosine
OH<sup>4</sup>yW*  undermodified hydroxywybutosine
oQ  epoxyqueuosine
ORF  open reading frame
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>preQ0</td>
<td>7-cyano-7-deazaguanosine</td>
</tr>
<tr>
<td>preQ1</td>
<td>7-aminomethyl-7-deazaguanosine</td>
</tr>
<tr>
<td>PTM</td>
<td>post-transcriptional modification</td>
</tr>
<tr>
<td>PTMs</td>
<td>post-transcriptional modifications</td>
</tr>
<tr>
<td>Q</td>
<td>queuosine</td>
</tr>
<tr>
<td>Q-IMS-TOF</td>
<td>quadruple-ion mobility spectrometry-time-of-flight</td>
</tr>
<tr>
<td>Q-TOF</td>
<td>quadrupole time-of-flight</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>RSD%</td>
<td>relative standard deviations</td>
</tr>
<tr>
<td>RT</td>
<td>reverse transcriptase</td>
</tr>
<tr>
<td>s²C</td>
<td>2-thiocytidine</td>
</tr>
<tr>
<td>s²U</td>
<td>2-thiouridine</td>
</tr>
<tr>
<td>s²Um</td>
<td>2-thio-2'-O-methyluridine</td>
</tr>
<tr>
<td>s⁴U</td>
<td>4-thiouridine</td>
</tr>
<tr>
<td>SAPK</td>
<td>stress-activated protein kinase</td>
</tr>
<tr>
<td>SC</td>
<td>synthetic complete</td>
</tr>
<tr>
<td>snRNPs</td>
<td>small nuclear RNAs</td>
</tr>
<tr>
<td>t⁶A</td>
<td>N6-threonylcarbamoyladenosine</td>
</tr>
<tr>
<td>TAP</td>
<td>time aligned parallel dissociation</td>
</tr>
<tr>
<td>tₐ</td>
<td>arrival time</td>
</tr>
<tr>
<td>tm⁵s²U</td>
<td>5-taurinomethyl-2-thiouridine</td>
</tr>
<tr>
<td>tm⁵U</td>
<td>5-taurinomethyluridine</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TSQ</td>
<td>triple-stage quadrupole</td>
</tr>
<tr>
<td>Um</td>
<td>2'-O-methyluridine</td>
</tr>
<tr>
<td>WC</td>
<td>Watson-Crick</td>
</tr>
<tr>
<td>WT</td>
<td>wildtype S. cerevisiae</td>
</tr>
<tr>
<td>WT37</td>
<td>wildtype enduring heat shock</td>
</tr>
<tr>
<td>WTs</td>
<td>wildtype enduring salt stress</td>
</tr>
<tr>
<td>Y</td>
<td>pseudouridin</td>
</tr>
<tr>
<td>Ym</td>
<td>2'-O-methylpseudouridin</td>
</tr>
<tr>
<td>YPD</td>
<td>yeast extract, peptone, dextrose</td>
</tr>
<tr>
<td>yW</td>
<td>wybutosine</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 DISCOVERY OF RNA POST-TRANSCRIPTIONAL MODIFICATIONS

Since the first demonstrated report of a “fifth nucleotide” in 1957,\(^1\) which was later identified as 5-ribosyluracil (pseudouridine, \(\Psi\)),\(^2\) over 100 natural RNA variants have been discovered. These species originate from the activity of specialized enzymes that recognize target RNA strands and catalyze their post-transcriptional modification (PTM). A close examination of their chemical features allows for a general classification into well-defined groups.\(^3\) The first includes the above-mentioned pseudouridine that is formed by direct isomerization of uridine’s N-glycosidic bond into a C-glycosidic form. The second contains ribonucleotide variants produced by simple alterations of the nitrogenous bases, which may include methylation, deamination, reduction, thiolation, or alkylation. A separate class is generated by methylation of the 2’-hydroxyl group of ribose, whereas the final class consists of PTMs with multiple modifications or introduction of sizable modifying groups. At the molecular level, the different modifications are capable of influencing the folding of RNA higher-order structure and the stability of interactions with other nucleic acid or protein factors.\(^4,5\) Growing knowledge of their cellular origin shows that biogenetic pathways can be very complex and tightly regulated,\(^6,7\) hinting to the largely unrecognized value PTMs may have within the cell.
Chemical modification can affect the structure as well as the overall reactivity of the respective canonical precursor, which could be potentially used to support identification through direct chemical approaches.\textsuperscript{8} Over the years, the chemistry of nucleic acids has been investigated extensively to understand damage processes and their implication in carcinogenesis.\textsuperscript{9,10} The realization that base methylation could induce destabilization of the N-glycosidic bond, followed by base loss and strand cleavage, led to the development of effective sequencing strategies.\textsuperscript{11} In analogous fashion, the ability to target specific modifications to induce site-directed cleavage can be employed to identify the location of susceptible PTMs on parent strands.\textsuperscript{12} As described below, a variety of reagents have been developed to obtain specific labeling for sequencing purposes. However, the utilization of chemical approaches to perform direct structural elucidation has been very limited.

\subsection*{1.2 Detection of Post-transcriptional Modifications}

Identification and characterization of newly discovered PTMs relies on typical combinations of spectroscopic techniques, including nuclear magnetic resonance (NMR)\textsuperscript{4,13–15} and mass spectrometry (MS).\textsuperscript{16–19} Classic approaches involve isolation of the RNA of interest, enzymatic hydrolysis into individual nucleosides/nucleotides, followed by sample separation to obtain sufficient material for the following steps.\textsuperscript{20–22} At first inspection, the heterogeneity of PTM structures should be expected to facilitate their discrimination. Instead, separation and analysis can be severely hampered by the fact that PTMs stem from a limited number of precursor types and frequently display isomeric structures with identical functional groups and
chemical properties. This scenario is exemplified by the set of methyl-guanines shown in Figure 1-1, which differ solely by the position of the modifier onto a common ribonucleotide scaffold. The very similar physicochemical properties afforded by this group of isomers requires the utilization of high-resolution separation techniques to achieve isolation. At the same time, these characteristics increase the challenges faced by de novo synthetic operations aimed at producing standards necessary to validate the proposed structures and sustain the development of analytical approaches.\textsuperscript{23}

A wide variety of separation approaches have been developed to support initial elucidation, as well as detection of known PTMs. Paper\textsuperscript{24,25} and thin-layer\textsuperscript{26} chromatography have been proven to be very adept at separating RNA hydrolysates and enabling PTM detection by autoradiography.\textsuperscript{27} Given the propensity of PTMs under the proper conditions to exist in either the cationic or anionic form, ion exchange chromatography specifically, was implemented for the identification of these species.\textsuperscript{28-31} As such, mixtures of PTMs were extensively studied using these resins to judge the relative retention of the species in the mixtures. For some like guanine, adenine, and cytosine; the characteristic amino group became cationic in acidic solutions. For others like thymine and uracil, the lack of amino group caused weak or no retention on these types of exchangers.\textsuperscript{32,32,33} Despite these limitations, the presence of the acidic phosphate group on nucleotide species made this approach easily accomplishable by altering the pH of the resins and/or exchanger used.\textsuperscript{34,24} Thus, extensive binding studies were endeavored, and the resulting elution fractions from these procedures were monitored by ultraviolet absorption to identify the PTMs based on their retention. While this was the best approximation of cellular lysate content at the time, this technique lacked the ability to structurally confirm molecules and became
Figure 1-1. Neutral structures of the four methyl-guanine isomers (377.0762 u) isolated from *S. cerevisiae* total RNA digest which differ based on location of the modifying methyl group on either one of three positions on the nucleobase or on the 2'-O position of the ribose phosphate moiety. Common names taken from http://mods.rna.albany.edu/.
exponentially more difficult with the increasing number of PTMs present in the mixture as the likelihood of co-elution increased, making the identification of unknowns found in biological systems virtually impossible.

It has been demonstrated that capillary and polyacrylamide gel electrophoresis (CE and PAGE, respectively) can afford the resolution necessary to overcome the previously described limitations and allow for the discrimination of different variants. Each technique has made use of the highly chargeable nature of these compounds. In the case of electrophoresis, increased separation was seen in comparison to paper chromatography. This made possible the separation of previously unidentifiable PTMs. Additionally, both techniques proved to be much faster and utilized even smaller sample amounts than those necessary for filter paper analysis. In the case of electrophoresis specifically, detection was much more accurate due to the ability to predict the mobility of each PTM at any given pH and allowed for identification of unknowns.

Column separation techniques in the form of both gas and liquid chromatography (GC and LC, respectively) also have been developed to support PTM analysis. Reversed-phase LC techniques were employed by Gehrke and Kuo to successfully identify PTMs in a total RNA digest by separation with phosphate buffers and subsequent UV detection. Concurrently, Buck was able to use ammonium based buffers to accomplish the same goal and was able to recover the sample post analysis due to solvent volatility and evaporation. The ability to demonstrate that the separation of these analytes could be accomplished using such volatile solvents made the coupling of chromatography to a mass spectrometer seem like a natural transition, and the most advantageous approach for characterization individual PTMs. Alternatively, gas chromatography techniques were developed which exploited the easily volatile trimethylsilyl derivatives of the
PTMs. The limitation to this approach was encountered when trying to volatilize PTMs that were very polar in the nucleobase. Additionally, the utilization of rather non-specific readout methods, such as flame ionization detection (FID) for GC, or UV spectroscopy for LC, requires the application of synthetic standards to enable the determination of specific migration/retention characteristics, which can be used to support sample identification. The broad implementation of more specific readout methods, which can reveal structural features of eluting species, has progressively reduced the significance of migration/retention characteristics as unique identifying properties.

1.3 STRUCTURAL CHARACTERIZATION OF RIBONUCLEOTIDE VARIANTS BY MASS SPECTROMETRY

The ability to directly recognize the various PTMs according to their specific mass and fragmentation signatures sets MS-based methods apart from the other analytical approaches.\textsuperscript{21,44–46} The pioneering work by Crain and McCloskey established the basis for the ever expanding application of this platform to RNA analysis.\textsuperscript{16,21,47,48} Owing to favorable energetics that preserve the integrity of the labile N-glycosidic bond and to the ability of interfacing directly with separation methods, electrospray ionization (ESI) represents the technique of choice for PTM characterization.\textsuperscript{49–52} In contrast, the utilization of matrix-assisted laser desorption/ionization (MALDI)\textsuperscript{53–56} in ribonucleotide analysis is frequently discouraged by the concomitant presence of matrix signals in the low mass range. Positive ion mode is preferred for nucleosides, whereas the negatively charged phosphates present in nucleotides offer the opportunity of utilizing the opposite polarity. Accurate mass determination provides essential
information on the elemental composition, which may become the basis for initial identification through database searching.\textsuperscript{57} Tandem mass spectrometry (MS/MS) affords instead the unique fragmentation data necessary to recognize the specific structural features that corroborate PTM’s identity.\textsuperscript{58,59} Cleavage of the N-glycosidic bond with nucleobase loss tends to be the most favorable dissociation process in the gas phase,\textsuperscript{60} with the notable exception of pseudouridine that includes instead a C-glycosidic bond.\textsuperscript{61–63} The resulting fragments can immediately reveal whether the modifier may be present on the nucleobase or ribose moieties. Extensive studies have investigated the effects of modifications on the gas-phase stability of both ribo- and 2’-deoxyribonucleotides,\textsuperscript{64–68} which can be serve as structure diagnostics.\textsuperscript{62} In many cases, the ability to perform consecutive stages of ion selection and activation (MS\textsuperscript{n}) enables one to further fragment the nucleobase or ribose products to correctly identify the specific position of the modifier.\textsuperscript{57}

These types of approaches are particularly helpful in the analysis of isomeric species that share identical elemental composition and thus, molecular mass. These scenarios can be exemplified by the characterization of the species detected with mass 377.0762 u in a total RNA extract from \textit{S. cerevisiae}.\textsuperscript{57} Based on the accurate mass alone, this specie could possibly correspond to any of the methyl-guanine isomers displayed in \textbf{Figure 1-1}. Submitted to MS/MS analysis in negative ion mode, the precursor ion with mass-to-charge ratio (m/z) 376 produced fragments consistent with the presence of the modifying methyl group on both the ribose and nucleobase, which could be ascribed respectively to 2’-O-methyl-GMP (Gm) and any of the remaining methyl-guanines in the mixture (\textbf{Figure 1-2a}). The structure of the methylated nucleobase was further investigated by following its fate through successive MS\textsuperscript{n} steps in
Figure 1-2. MS^n fragmentation acquired with an ion trap analyzer. **a)** Anionic MS^2 spectrum of methyl-guanosine isomers isolated from *S. cerevisiae* total RNA digest, which was obtained by activating the m/z 376 species observed in Figure 1-1; **b)** Cationic MS^3 spectrum of obtained by activating m/z 378 → 166 →. The inset displays the MS^4 spectrum obtained by activating m/z 378 → 166 → 124 →. Solid arrows indicate possible methylated positions; dashed arrows suggest putative cleavages. Overall complexity of discerning multiple isobaric species can readily be appreciated by the need to implement multiple MS^n steps for conclusive identification of each specie; accomplished solely by trapping instruments. Taken with permission from Ref. 310.
positive ion mode, which confirmed the concomitant presence of 1-methyl-GMP (m\textsuperscript{1}G), N2-methyl-GMP (m2G), and 7-methyl-GMP (m\textsuperscript{7}G) in the S. cerevisiae's extract (Figure 1-2b).\textsuperscript{57} The observed fragmentation patterns matched those obtained from similar methyl-guanosine isomers on a Fourier transform ion cyclotron resonance (FTICR).\textsuperscript{69} These types of experiments can be effectively carried out on trapping analyzers, such as ion traps and FTICRs, but require additional in-source activation in popular beam instruments, such as triple-stage quadrupole (TSQ) and quadrupole-time of flight (Q-TOF) analyzers, which afford only limited ability to complete multiple selection/activation steps.

Similar strategies can be implemented by using the MS platform either in stand-alone fashion, or in combination with high-resolution separation techniques, such as liquid chromatography (LC-MS)\textsuperscript{13,40,43,45,47,48,52} and capillary electrophoresis (CE-MS).\textsuperscript{35,37,70} Hyphenated techniques are particularly helpful in the analysis of complex mixtures produced by hydrolysis of the initial RNA samples. In addition to differentiating the various species according to their distinctive physicochemical properties, the separation step has the beneficial effect of concentrating the analytes into narrow elution bands, which can boost the analytical sensitivity to low attomole levels. By itself, retention time has very limited value as a unique analytical property for sample identification. However, the ability to separate components of very similar structures is particularly significant for the correct identification of isomeric/isobaric species. The chromatographic approaches employed for LC-MS determinations are typically based on reversed-phase systems with and without ion pairing,\textsuperscript{40,43,71} in which all non-volatile additives have been replaced with MS-friendly counterparts (i.e., ammonium-based salts).\textsuperscript{47,72,73} Gradients consisting of water, acetonitrile, and low percentages of formic acid also have been employed to
accomplish the analysis of protonated nucleosides in positive ion mode.\textsuperscript{74} Although C18 stationary phases are the most widely used, C30 columns have been employed to improve the retention of hydrophilic nucleosides,\textsuperscript{75} whereas porous graphitic carbon can benefit the analysis of hydrophobic ones. Instrumentation of choice include TSQ and Q-TOF analyzers, with the former being preferred for quantification purposes.\textsuperscript{73} Furthermore, the former enable the implementation of linked scans, such as multiple-reaction monitoring (MRM) and similar, which limit the background contributions by detecting only gas-phase processes that are specific for the target PTM.\textsuperscript{62,76-77}

Ion mobility spectrometry-mass spectrometry (IMS-MS) is an emerging hyphenated technique with excellent potential for PTM analysis\textsuperscript{78-80} due to its ability to discriminate analytes according to both mass and ion mobility behavior. The latter is a function of size and conformation that together determine the extent by which ions interact with background gas while traveling in a moderate electric field. The process may significantly alter their time of arrival \((t_D)\) to the mass analyzer, which can be plotted against the corresponding \(m/z\) to produce 3D-plots or heat map diagrams. As shown in Figure 1-3a for the total RNA extract from \textit{S. cerevisiae},\textsuperscript{57} the relative abundance of detected ions is typically represented by a color gradient that constitutes the third dimension of the 3D-plot. These data show that the process can readily discriminate between isomeric/isobaric species, such as uridine and pseudouridine, which are recognizable as discrete spots on the observed heat maps. This capability also can be accessed by experiments in which isomeric/isobaric ions with identical \(m/z\) are first isolated in the mass-selective Q located at the front-end of the instrument. The different populations are then dispersed on the time domain in the ion mobility region, which are recorded as discrete signals
**Figure 1-3.** a) Anionic heat map obtained by IMS-MS analysis of the *S. cerevisiae* total RNA digest. Zoomed region shows assignment of PTMs based on their distinct mobility profiles. Color scale provides a qualitative measure of each PTM present in the digest. b) Ion mobility profile obtained from the precursor ion at m/z 324.04. c) Gaussian-fitted plots generated by deconvolution in PeakFit 4.1. Analyte assignment was confirmed by comparing the apparent drift time (t₀) of the deconvoluted signals with those obtained in the separate experiments utilizing isolate nucleotide standards. Taken with permission from Ref. 310 and 262.
in the ensuing mobilogram. Data obtained from a mixture containing both uridine and pseudouridine demonstrated that the conformations produced by their N- and C-glycosidic bonds were sufficiently different to enable their correct discrimination in the time domain (Figure 1-3b), discussed extensively in the following chapter.\textsuperscript{63} Widely available instruments with Q-IMS-TOF and analogous configurations afford the additional ability to corroborate the structure of time-resolved species by activating their gas-phase dissociation in different regions of the instrument. Preliminary work has shown that the observed fragmentation patterns conform with those observed on TSQ and Q-TOF platforms, which bodes well for the broader future application of this technique to PTM analysis.\textsuperscript{57,77,81–83}

1.4 Locating Post-transcriptional Modifications in Parent RNAs

Since the completion of the Human Genome Project, the rapid advances made by sequencing techniques have placed firmly within sight the prospect of performing the affordable determination of individual genomes.\textsuperscript{84–88} At the same time, these advances have greatly benefited the elucidation of the sequence position of PTMs in the respective parent RNAs, which is essential to the investigation of their biological functions.\textsuperscript{89,90} A common denominator of these sequencing-based strategies consists of an initial step in which the genuine strand bearing PTMs is replicated into a complementary DNA (cDNA). The enzymatic process requires the availability of complementary triphosphonucleotides (i.e., NTPs) that recognize template bases and sustain primer extension. The absence of specific NTPs capable of proper base-pairing with target PTMs may result in different scenarios. If the modification does not involve the Watson-Crick (WC) edge, then a canonical complementary nucleotide may be regularly incorporated in the
extending strand (i.e., readthrough), and the cDNA will bear no trace of the original PTM. If the modification involves the WC edge, then an incorrect canonical base may be added (i.e., misincorporation), and the cDNA will include a point mutation in the position originally occupied by the PTM. Finally, if the PTM is sufficiently bulky to interfere with base pairing or enzymatic activity, no base will be incorporated at all and strand extension will halt (i.e., elongation inhibition). These considerations prevent the direct application of sequencing techniques to elucidate the presence and location of the vast majority of known PTMs.\textsuperscript{89–91}

Different indirect strategies have been developed to circumvent and, in some cases, take advantage of these scenarios in the context of nucleic acid sequencing approaches. For example, the ability to stall, or at least delay, the strand-extension activity of reverse transcriptase (RT) has been successfully exploited to detect the presence of WC-edge modifications,\textsuperscript{89} such as those in m\textsuperscript{1}A,\textsuperscript{92} m\textsuperscript{3}U,\textsuperscript{92} m\textsuperscript{2}G,\textsuperscript{93,94} and m\textsuperscript{1}G\textsuperscript{95} (full names provided in the RNA Modification Database).\textsuperscript{96} The stalling effect is enhanced by performing the enzymatic reaction in the presences of low NTP levels.\textsuperscript{89,90,97} Elongation products ending at the modified sites can be subsequently analyzed by common dideoxy-termination or pyrosequencing approaches.\textsuperscript{98} Although this strategy is relatively straightforward, the presence of different PTMs on the same substrate may complicate the unambiguous assignment of the observed termination to any given one. In some cases, modifications that do not involve the WC edge, such as 2′-O-methylation, also have been proven capable of inhibiting RT.\textsuperscript{99,100} In others, further chemical derivatization of the target PTM is necessary to induce RT inhibition, as shown by treating target RNA with carbodiimide reagents, such as N-cyclohexyl-N′-β(4-methylmorpholinium) ethylcarbodiimide p-tosylate (CMC).\textsuperscript{101} Subsequent exposure to alkaline environment removes of all CMC groups except for those
bonded to pseudouridine, which can be then detected from its ability to stall strand elongation. Provided that the final product interferes with enzymatic activity, this general strategy can be potentially extended to additional PTMs with unique chemical reactivities that differentiate them from the remaining nucleotides in the strand. A common shortcoming of these approaches stems from the possible presence of stable secondary structure on the target substrate, which may induce non-specific stalling at unmodified sites.

The ability of RT to read through modified nucleotides has benefited approaches based on direct comparisons between sequence information obtained before and after PTM-specific treatments meant to achieve their elimination or further modification. In the case of 5-methylcytosine (m5C), which is known to modulate the accessibility of DNA in epigenetic processes, the modifying methyl group protects this structure from undergoing deamination to uridine mediated by bisulfite reaction. For this reason, its sequence position can be unambiguously located by comparing sequencing data acquired from the initial sample and a bisulfite-treated aliquot in which all canonical cytosines have been effectively converted to uridines. The absence of RT-stalling activity manifested by m5C enables the initial sequencing of the genuine strand, which is necessary to detect its failure to convert in the treated sample. Taking this strategy a step forward, the bisulfite-resistant 5-hydroxymethylcytosine (5hmC) can be recognized by treating the sample with TET proteins, which convert this PTM into bisulfite-susceptible 5-carboxylcytosine (5caC). In this way, the concerted application of TET and bisulfite can reveal the respective position of both 5mC and 5hmC on the RNA substrate of interest.

In principle, specific antibodies capable of recognizing target PTMs in the context of RNA
strands also could be utilized to achieve elongation inhibition at modified sites. However, the rather large footprint of typical antibody-RNA interactions would be expected to lead to poor nucleotide resolution. More frequently, the specificity of these types of interactions are employed to enrich samples generated by random shearing of RNA extracts prior to massive parallel sequencing analysis (i.e., RNA-seq),\textsuperscript{107,108} which can lead to the identification of conserved modification sites. For example, antibodies specific for 6-methyladenine (m\textsuperscript{6}A) have been used to immunoprecipitate RNA fragments in the \(\sim\)100-nucleotide range, which were subsequently submitted to RNA-seq. When the resulting reads were aligned, the detection of numerous common sequences clearly identified the position of m\textsuperscript{6}A sites.\textsuperscript{109,110} In this strategy, the accuracy of the analysis is predicated on the agreement between the large numbers of reads generated by massive parallel sequencing. In analogous fashion, the increasing availability of sequence information produced by these types of technologies has been driving the identification of mismatches with corresponding genomic sequences and hotspots of misincorporation errors, which could be ascribed to the presence of specific PTMs. For example, deamination of adenosine (A) produces inosine (I) that is regularly decoded as guanosine in the corresponding cDNA. Therefore, direct comparison between genomic and cDNA can immediately provide evidence of A to I editing in specific sites.\textsuperscript{111} In similar fashion, mismatches with original genomic sequences have been successfully utilized to identify m\textsuperscript{1}A, m\textsuperscript{2}G, and m\textsuperscript{2,2}G in tRNA and miRNA.\textsuperscript{112} This type of strategy can produce excellent results even in the absence of corresponding genomic data. Indeed, the incidence of incorporation errors in specific sites, which was monitored by aligning large numbers of reads obtained from small RNAs, was by itself sufficient to reveal the presence of modification, which was subsequently traced to specific PTMs by analyzing known tRNAs.\textsuperscript{113}
Any possible shortcomings associated with the rather indirect nature of these approaches are easily overshadowed by the fact that their findings are usually supported by overwhelming statistics obtained from the massive number of available reads.

In addition to strategies aimed at mapping the unknown distribution of PTMs in RNA samples of interest, different targeted approaches have been developed to monitor the actual incidence of modification in known sites. For example, a strategy based on deoxyribozymes was demonstrated, which relied on the ability of pseudouridine and 2’-O-methyl modifications to inhibit their typical RNA-cleaving activity.\textsuperscript{114} The recognition arms of the deoxyribozyme were designed to specifically base-pair with the sequences surrounding the conserved site, in such a way as to place the catalytic domain in the optimal position for strand hydrolysis. The actual presence of modification was revealed by a decreased production of cleaved fragments, which was detected by utilizing ancillary post-labelling techniques.\textsuperscript{115} In analogous strategies, the RNA-cleaving activity was provided by RNase H, whereas the specific positioning onto target site was accomplished by the annealing of appropriate oligonucleotide guides.\textsuperscript{116,117} Mirror approaches also have been explored in which strand cleavage was replaced with ligation performed by T4 DNA ligase.\textsuperscript{118,119} In this case, the target modified strand acted as a template for the ligation of antisense oligonucleotide pairs that were designed to specifically anneal next to one another, while leaving the junction region upon the modified site. The extent of modification was then recognized by the variation of ligation efficiency between oligonucleotides pairs that were affected by the presence of PTM versus those that were not. Suitable oligonucleotides were identified for at least seven different PTMs, including 2’-O-methylation.\textsuperscript{119}
1.5 MS-BASED CHARACTERIZATION OF MODIFIED RNA STRANDS

The same unique features that make modified units recognizable by MS analysis can be potentially employed to elucidate their sequence position in the context of the parent biopolymer. Over the years, a variety of approaches have been developed to accomplish nucleic acid sequencing, which rely on either gas-phase fragmentation during tandem MS experiments,\textsuperscript{120,121} or indirect methods requiring the application of ancillary chemical/biochemical techniques.\textsuperscript{122} Gas-phase activation induces cleavage of the phosphodiester linkage between consecutive nucleotides in the polymer chain, which leads to the formation of characteristic ion series.\textsuperscript{123,124} The mass spacing between contiguous signals in the series is characteristic of canonical nucleotides in unmodified samples, but can also reveal the presence of any chemical modification that introduced a change of mass.\textsuperscript{72} Indirect methods involve generating oligonucleotide ladders by chain termination during synthesis,\textsuperscript{125} or by the controlled chemical/enzymatic hydrolysis of consecutive phosphodiester linkages.\textsuperscript{126} The ladders are analyzed as mixtures of discrete products, in which the mass difference between contiguous signals identify canonical versus modified units. Direct gas-phase activation of intact samples, which is defined as top-down analysis, is capable of obtaining full-sequence coverage from unmodified oligodeoxynucleotides up to 50 nucleotides.\textsuperscript{127,128} Larger RNA samples are typically treated with endonucleases to obtain mixtures of smaller hydrolytic products, which can be readily mass mapped and sequenced according to typical bottom-up strategies.\textsuperscript{129,130} The availability of different enzymes that can specifically target well-defined ribonucleotides lends a great deal of flexibility to this type of analysis. With few variations, these approaches represent the foundations for the various MS-based strategies employed to investigate the sequence
position of PTMs.

Mass mapping involves determining the mass of hydrolytic mixtures generated by treatment of target RNA samples with specific endonucleases.\textsuperscript{131} This task can be readily accomplished by MALDI analysis, which may be preceded\textsuperscript{40,43,132,133} or not by appropriate chromatographic steps.\textsuperscript{130,134–136} For example, the entire tRNA pool extracted from \textit{E. coli} was digested by using RNase T1 to obtain a complex mixture of hydrolysis products, which was analyzed directly by MALDI-MS.\textsuperscript{131} The observed masses were compared with those calculated from the known sequences and modifications of all tRNAs expressed by this microorganism. The detection of specific signature products allowed for the correct identification of approximately 90\% of the tRNA isoacceptors present in the \textit{E. coli} lysate. Although digestion products can be isolated by chromatographic procedures before separate analysis by direct infusion ESI-MS,\textsuperscript{48} complex mixtures are more conveniently tackled by integrated LC-MS approaches.\textsuperscript{75,137,138} For this purpose, separation methods based on reversed-phase chromatography have been developed in which the highly polar character of oligonucleotide analytes is modulated by ion pairing agents, while full compatibility with subsequent ESI operations is maintained by including 1,1,1,3,3,3-hexafluoro-2-propanol or similar additives.\textsuperscript{139} LC-MS analysis has been proven capable of supporting the mass determination of products reaching the size of intact tRNA molecules (i.e, 77 nucleotides, ~24kDa).\textsuperscript{140}

In many instances, the presence of PTMs in specific hydrolytic fragments can be deduced before their sequencing is completed by comparing the observed masses with those expected from known substrates, thus boosting the value of mapping operations. In order to increase the sequence coverage afforded by these experiments, different endonucleases can be used in
parallel to obtain complementary overlapping maps. Nucleotide-specific enzymes such as RNase A and T1, which target respectively C/U and G,\textsuperscript{141,142} tend to produce small hydrolysis products that reflect the relatively high frequency of these bases in natural RNA. For this reason, the resulting mixtures may contain numerous oligonucleotides with similar or identical base composition, which are difficult to assign unambiguously. Different solutions have been proposed to address the mass degeneracy associated with endonuclease products, which may involve performing limited hydrolysis (e.g., shorter incubation time, lower reaction temperature, or smaller enzyme to substrate ratio), or utilizing structure-specific enzymes that provide larger products. This concept has been explored by testing the application of Mung bean and V1 nucleases, which attack respectively single- or double-stranded regions of folded structure,\textsuperscript{143–145} and thus can provide complementary maps containing larger, more readily assignable products.\textsuperscript{146}

The gas-phase fragmentation processes employed to sequence digestion products or intact RNA molecules in bottom-up or top-down approaches, respectively, can count on different activation methods. Alternative to classic collisional processes (i.e., CID and CAD),\textsuperscript{147} infrared multiphoton dissociation (IRMPD)\textsuperscript{148–151} and electron detachment dissociation (EDD)\textsuperscript{152,153} have been explored in an effort to increase the size of accessible analytes. The fact that base-pairing and stacking interactions may be preserved during MS analysis\textsuperscript{154,155} has been suggested as one of the possible factors that limit the efficiency of fragmentation displayed by large folded RNA. This possibility has been addressed by introducing an additional activation step to induce dissociation of these types of interactions –with melting of compact higher-order structures– prior to actual backbone fragmentation.\textsuperscript{156,157} Manipulating the pH and composition of the initial
solutions may provide another avenue for achieving at least partial denaturation, as well as increased charging that may facilitate dissociation. Taking advantage of these principles, an entire tRNA spanning 76 nucleotides and containing a full complement of PTMs was sequenced by collisional activation on an FTICR instrument. The data in Figure 1-4 shows that the fragmentation processes covered the entire length of the strand with no significant effects associated with the presence of modifications.

Sequencing tasks can be effectively completed off-line by both ESI and MALDI, or directly on-line by LC-MS/MS. In order to streamline the operations and facilitate the automation of data collection, exclusion lists based on corresponding DNA sequences have been utilized to restrict LC-MS/MS determinations only to digestion products with high probability of containing PTMs. Although the majority of known PTMs display characteristic mass increases over the corresponding canonical ribonucleotides, isomeric/isobaric species may not be directly recognizable in the context of sequencing experiments. In the case of pseudouridine, the different efficiency of fragmentation manifested by C- versus N-glycosidic bonds is detectable by tandem MS and, thus, may enable correct identification in the context of sequencing experiments. More frequently, however, strategies based on chemical derivatization are preferred, which rely on the production of either acrilonitrile adducts, or the above-mentioned carbodiimide derivatives. These types of strategies are potentially applicable to any type of PTM with distinctive chemical reactivity that may be leveraged to highlight its presence in RNA strands.

While chemical derivatization can be effectively employed in conjunction with a variety of analytical platforms, numerous strategies based on stable-isotope labeling are made uniquely possible by MS analysis. Common approaches tend to rely either on the direct comparison of
**Figure 1-4.** a) Representative full ESI spectrum of tRNA$\text{Val}^{\text{Val}}$ (2µM) containing H$_2$O/CH$_3$OH (1:1) with piperidine (10 mM) and quinuclidine (10 mM); b) isolation of ions between $m/z$ 700 and 920, corresponding to species with net negative charges of 27-35; c) mass spectrum acquired after activation by EDD. Resulting spectrum shows $d$ and $w$ ions from the backbone cleavage at 60/75 sites on tRNA$\text{Val}^{\text{Val}}$. The inset shows the isotopic resolution achieved by this approach; d) Fragment-ion map acquired after CAD and EDD of tRNA$\text{Val}^{\text{Val}}$. Production of $c$ and $y$ ions from CAD (not shown here) were coupled with those obtained from EDD for complete coverage. PTM content had no effect on identification of such large RNAs. Taken with permission from Ref. 153.
labeled and unlabeled samples analyzed in separate experiments, or the simultaneous detection of labeled and unlabeled species in the same sample. For example, $^{15}$N and $^{13}$C isotopes were included in *E. coli* and *S. cerevisiae* cultures to achieve their metabolic introduction in cellular RNA. The nucleoside mixtures obtained by enzymatic digestion of their RNA extracts were analyzed by LC-MS and LC-MS/MS. The results were compared with those obtained from samples grown in normal media, which allowed the detection of PTMs characteristic of these organisms, as well as several previously undescribed modifications. The possibility of introducing an $^{18}$O label by enzymatic hydrolysis in H$_2^{18}$O environment enables the quantitative determination of RNA strands in complex mixtures. When samples treated in H$_2^{18}$O and regular water are mixed in equal volumes, signals detected 2-Da apart can immediately reveal the ratio of concentrations possessed by the target hydrolytic product in the “heavy” and “light” samples. This type of application can provide valuable information on the variation of RNA expression levels in different samples. In analogous fashion, the incorporation of $^{18}$O labels at the 3’-end of hydrolysis products can support the comparative analysis of RNA digests (CARD) for the possible identification of sequence-specific modifications. The mixtures obtained from “heavy” and “light” samples also are examined for the presence of doublet signals 2-Da apart, which are produced by species sharing the same sequence and modification. In contrast, possible singlets identify species that differ in their sequence/modification makeup, which can be immediately submitted to LC-MS/MS characterization. This type of application will be expected to greatly benefit studies aimed at combining genotyping with PTM analysis.
1.6 RNA PTM DATABASES AND REPOSITORIES

All of these advancements in instrumentation and scientific applications designed at isolating and characterizing PTMs located on the major subsets of RNAs have resulted in over 140 PTMs being identified. Ultimately, these discoveries created a need for publicly available repositories. As such, the first repository was created in 1999, by Pamela Crain, Jef Rozenski and James McCloskey. This publicly available database, which is still actively updated, provides a comprehensive listing of the naturally occurring PTMs. Each PTM, in particular, has its own page which includes the common name and symbol, the subtype of RNA to from which it came, and the phylogenic origin. In addition, publications related to the isolation and characterization of that specific PTM have provided a way by which to investigate each compound. Additionally, links to other publications related to the biological function (if any has been determined) is made accessible.

This available PTM data opened the eyes of many to the potential of the vast roles RNAs could be playing in the structure and function of RNA within the cell. The result was an enormous increase in the exploration of these roles within the cell and, ultimately led to advancement of the field of transcriptomics. As many new biological processes and gene correlations were being made, and questions about the roles in which RNA played were being answered, a need for larger more extensive repositories became imminent. In 2006, the Genesilico Modomics Database was created with the goal of making these new findings publicly available. Unlike the RNA Modifications Database, Modomics focuses on the biology of RNA modifications. Modomics includes comprehensive information about PTMs, their respective location in RNA sequences, pathways for their biosynthesis and the enzymes responsible for their catalysis. In addition to all
of this new biological information, this database merged the information provided by proteomics studies, included structural data, and provided easy access to any and all publications related to a particular PTM. To date, approximately 140 of these ribonucleotide variants, along with numerous metabolic pathways, are catalogued in either the RNA Modifications and MODOMICS databases. Although some PTMs are well known for stabilizing RNA structure and mediating molecular recognition, the function of most of them has yet to be determined.

1.7 BIOLOGICAL FUNCTIONS OF RNA: SIGNIFICANCE OF PTMs

Since their first discovery, most advances in PTM analysis have focused on their identity and location within the cell. However, these findings leave huge gaps in the knowledge about PTM biological function. Dating back to the 1980s, it became evident that some RNA molecules exhibited catalytic capabilities. As such, many groups took on the daunting role of determining which enzymes catalyze specific PTMs. However, limitations in the ability to detect these enzymes were a direct result of inadequate substrates to test enzyme activities in vitro and the low abundance of enzymes within the cell. These limitations often caused groups to abandon biological characterization of PTMs altogether. Luckily, the advancements in RNA technologies and chemical methods aimed at locating PTMs within the cell, revitalized the efforts. Also, the availability of genomic sequences and the ability to produce purified recombinant enzymes from any organism undoubtedly opened the door to understanding biological significance.

In 1986, Benne et al. demonstrated that the presence and absence of several uridine
modifications occurred in mRNA in some protozoa. This transcription event was originally observed in the coding region of mRNAs but further research suggested that similar events took place on rRNA and tRNA as well. The ability of RNA to insert or delete in this instance was seen to be similar to the RNA machinery involved in the conversion of bases, such as the conversion of C to U in mRNA. Ultimately, characterization of the different methods in which RNA was capable of being modified by enzymatic processes spurred more thought about what these modified RNAs actually do within the cell.

Studies aimed at mRNA functions employed the use of methylation inhibitors in order to stop the occurrence of base modification. As such, it was determined that mRNA methylations affect the efficiency of pre-mRNA splicing and transport. These modifications, located in the cap structure and in positions within the pre-mRNA, occur after transcription. Additionally, the PTMs found in mRNA seem to be more conserved and limited in nature in comparison to those located on tRNA and rRNA residues. The base methylations found in the mRNA cap has been found to increase translation in *Xenopus* oocytes, along with increases in cells that are directly transfected with mRNAs. The methylated species have been shown to enhance the binding of the mRNA to the 40S ribosomal subunit. In comparison, the absence of the cap hinders this event. These cap methylations have been shown to be regulators of translation, capable of decreasing the susceptibility of mRNA molecules to nuclease degradation, and be involved in splicing and transport. The discovery of these diverse roles resulted in the discovery of several biogenetic enzymes by Moss et al.

In the nucleolus, rRNAs undergo an extensive maturation process involving small nuclear RNAs (snRNPs). These snRNPs are capable as acting as guides for the 2’O-methylation and
pseudouridine formation within rRNAs. However, the functional role of these pseudouridines is still mostly unknown. In contrast, transfer RNAs have been shown to be the most heavily modified RNA species, to date.\textsuperscript{45,168} It is hypothesized that the extensive modification of tRNA produces residues that prevent misfolding events during maturation processes in the cell. On tRNAs, PTMs that are located in the first wobble position of the anticodon loop participate in decoding events that are regulated by codon-anticodon interactions.\textsuperscript{13,191} Studies targeting the wobble position have proven that this modification must be conserved in order for certain biological functions to persist. Often, the lack of the wobble modification causes translational defects that can be linked to certain mitochondrial diseases. Given the ability to easily isolate this class of RNA, tRNA modifications are some of the most well understood. They have been implicated in controlling the subcellular localization of RNA molecules and play extensive roles in decoding. Additionally the presence and exact position of the PTM on the tRNA molecules is crucial as it has been known to affect mutations found on mitochondrial DNA, which have been known to result in diseases such as diabetes. It is important to note that while these modifications have been shown to be important and sometimes necessary for cellular health due to certain interactions, the overall biological importance of each PTM still remains relatively unknown.

More recently, several classes of RNAs have been implicated in cellular adaptation to heightened stress states. In order to maintain cellular viability, tRNAs, mRNAs and long non-coding RNAs undergo chemical modification in order to alter their activity, ensure proper localization and maintain stability. Specifically, modifications such as N\textsuperscript{6}-methyladenosine (m\textsuperscript{6}A) have been determined to be located in rRNA, snRNA and mRNA of cells infected with heightened stress. This modification also has been identified in several yeasts, plants and animals and tends to be more
abundant in eukaryotes. Studies focused on m⁶A regulation have unveiled its pathway and have shown that m⁶A is regulated by a multicomponent complex in mammals. Defects in this complex have been shown to cause downstream effects in meiosis; altering growth patterns, or causing developmental arrest and cell division defects. In total, these studies prove that PTMs such as m⁶A are physiologically relevant and essential for cellular adaptation and survival. While this information proves useful in studying pathways related to these specific biological processes, m⁶A is only one of over 100 known natural variants to RNA. Ultimately, until new strategies can be employed which are able to accurately capture the entirety of PTM content, what is actually happening within the transcriptome will remain unknown. We believe that the use of mass spectrometric strategies established herein will ultimately enable targeted investigation of many different important biological roles of PTMs.

Stress response is the process by which a cellular system maintains viability in reaction to a variety of external conditions.¹⁹²,¹⁹³ These external stimuli result in internal responses which are responsible for maintaining cell growth and overall cell survival. The adaptation of these cells in response to stressors such as heat, osmotic changes, and growth nutrients has resulted in known fluctuations of enzyme activity and metabolism.¹⁹⁴,¹⁹⁵ Historically, it has been found that these cells persist due to rapid regulatory changes which secure cellular protection.¹⁹⁶ Specifically, over the past few decades, extensive work has been focused on understanding these adaptation characteristics by the exploration of lethal doses of stressors, which has resulted in the discovery of mitogen-activated protein kinase (MAPK) pathways found to be activated during heat stress and osmostress states within many Eukaryotic systems.¹⁹⁷,¹⁹⁸,¹⁹⁹ These pathways are initiated by sensors on the cell surface that recognize the external changes.²⁰⁰ After sensing
stress, a signal is converted by GTPase nodes to MAPK phosphorylation cascades. These cascades generally involve a MAPK kinase kinase, a MAPK kinase and a MAPK, which are activated to target various downstream processes often involving transcriptional regulation. These MAPK pathways are highly conserved, and as such, have been extensively monitored in *S. cerevisiae* in order to deduce important mechanisms governing cellular survival in the presence of stress.

 Particularly, the biological focus of this project is the hyper-osmolarity glycerol (HOG) pathway. This pathway contains the MAPK Hog1 and MAPK Mpk1 which responds to osmostress and heat stress, respectively. The HOG pathway is known to regulate ion channels, glycerol export, protein machinery and cell cycle progression, which are vital to maintaining homeostasis in the presence of external stimuli. In this pathway, the transmembrane protein, Sho1, is connected to protein kinases and is essential for HOG signaling as it is responsible for activating specific MAPKs including Mpk1. These MAPK pathways result in the activation of DNA-binding transcriptional activators such as Tec1 which is found to be over-expressed in the absence of Hog1. Additionally, Hog1 has been described as a stress-activated protein kinase (SAPK) that is involved in multiple phases of the cell cycle and has the ability to reprogram gene expression in the presence of stress. This is performed by the enlistment of Hog1 to chromatin to recruit RNA polymerase II. Hog1 has the ability to remodel chromatin structure and plays a key role in the biogenesis of mRNA. During osmostress, Hog1 is shown to induce a set of ~200 long noncoding RNAs (lncRNAs), confirming that Hog1 plays a key role in transcriptional regulation. Moreover, these lncRNAs are capable of enacting changes in chromatin structure and nucleosome occupancy, which have lasting effects on cellular memory. Together, all this information contributes to the hypothesis that RNA is a key
player in mechanisms of biological regulation within the cell. Considering that little is known about the location and function of PTMs, this particular yeast system could serve as a well-developed model by which to examine the role played by PTMs.

1.8 CONCLUSIONS

The awareness of the pervasive regulatory functions enacted by RNA has grown to the point where the significance of its post-transcriptional modifications cannot be discounted anymore. The greater focus being placed on the investigation of RNA PTMs has increased the demand for effective analytical approaches capable of supporting not only the identification of modification sites, but also the determination of expression levels as a function of experimental variables. The concepts addressed in this chapter clearly show that no individual technology can fulfill all the needs of this growing research field. Approaches based on massive parallel sequencing enable the transcriptome-level analysis of a handful of individual variants through the concerted utilization of PTM-specific chemical derivatization, enzymatic activity, and antibody recognition. However, no such methods are available for the vast majority of known PTMs. Approaches based on MS detection can immediately identify ribonucleotide variants with characteristic mass and fragmentation signatures, which represent the majority of known PTMs, and tackle mass-silent modifications by using a variety of ancillary strategies. However, transcriptome-level applications are hampered by the relatively low throughput afforded by this platform. Sequencing-based approaches enjoy the high sensitivity afforded by sample amplification steps, but may provide only indirect evidence of the presence of PTMs. MS
techniques utilize genuine RNA to enable direct detection, but sample requirements may hurt the overall sensitivity. Quantitative analyses are difficult by using the former, whereas the latter enable direct comparisons between expression levels in different samples. The strengths and weaknesses of these platforms are very well matched and will be expected to invite increasing concerted applications to pursue specific questions in RNA biology.

In recent years, the development of MS-based technologies for RNA analysis has proceeded in the shadow of the astonishing advances made by the sequencing field. During this time, their evolution has followed in virtual lockstep the constant progress of the much more popular technologies employed in proteomics research. For this reason, virtually any MS-based strategy employed today in protein analysis can be potentially repurposed for the characterization and investigation of RNA. It is just a matter of time before the broader RNA community will realize the great opportunities afforded by this analytical platform, and will incorporate it with increasing regularity in the available toolkit.

To this end, the mass spectrometric platform developed herein to measure global PTM profiles will be employed to assess *S. cerevisiae* in the presence of various environmental stressors and/or in the presence or absence of key gene players associated with MAPK or SAPK pathways. Publicly available interactome databases will be used to deduce key gene plays in stress response pathways that may alter PTM expression amidst heightened stress. Ultimately, this analytical approach will enable the advancement of strategies aimed at addressing the biological significance of PTMs. In this way, the dearth of knowledge left by previous techniques will be supplemented to more fully characterize the transcriptome. It is hopeful to think that with the advancement of new techniques and strategies, the scientific community will be able to
assess any role of PTMs across various organisms and cellular states. Specifically, unveiling the
roles of PTMs could lead to designing better and cheaper diagnostic platforms, drug targets and
ultimately, therapeutic avenues.
Chapter 2

MS-based approaches for the analysis of post-transcriptional modifications at the entire transcriptome level

2.1 INTRODUCTION

The ability to address biological questions regarding the significance of RNA in cellular processes is greatly hampered by a lack of effective approaches for monitoring their presence within a particular biological system. As was indicated in the previous chapter, strand amplification techniques fail to accurately capture the entirety of PTM content. Alternative approaches aimed at identification of PTMs have included liquid chromatography or capillary electrophoresis coupled to mass spectrometry which has led to the discovery and characterization of all known nucleosides over the past 50 years.\textsuperscript{16,47,120} Their typical implementation relies on analytical standards to enable the matching of retention times of known standards to unknowns exhibiting identical chemical properties.\textsuperscript{13,40,43,47,48,225} Additionally, these standards provide a measurement of response based on their concentration, which enables the investigator to quantitatively assess the content of the sample. Other challenges consist of carry-over issues and sample bias. During chromatographic analysis, samples are introduced and pass through multiple parts of the HPLC system in order to reach the detector. The retention of such a sample on any or all of these parts (syringe, injector, valves, tubing and column) is referred to as carry-over and can have lasting effects on the quality of the
analysis. As such, signals from previous injections may be mistaken for the current chromatographic run and introduce bias in analysis. Often, this is represented while addressing quantitation because the carry-over can result in a falsely amplified analyte signal, making the concentration of the analyte seem higher. To offset these issues, careful consideration has to be given to the proper instrumentation setup, sample injection concentrations, and mobile and stationary phases. Testing of these parameters relies on the monitoring of the chromatographic signal obtained through blank runs.

Previous studies also have utilized tandem mass spectrometry approaches to obtain readily observable fragmentation patterns or the use of $^{32}$P-post labeling, immunochemical and fluorescence-based methods.$^{226-228}$ These methods typically are performed on an isolated RNA strand which is labelled through the use of isotopically labelled solvents during digestion procedures. Once the labelled digestion products are formed, they are subjected to mass analysis. In total, these issues take an enormous amount of time and energy to tackle, and their complexity increases almost exponentially with complex mixtures of samples and the advancement of better instrumentation with increased sensitivity. For these reasons, we aimed at investigating possible analytical avenues for characterizing nucleotide mixtures using direct infusion electrospray ionization without the assistance of a front-end separation technique. Together, these methods have provided the foundation with which we are currently able to validate our particular avenues of study.

This chapter illustrates the challenges of investigating nucleotide mixtures by direct infusion electrospray ionization (ESI) combined with either Fourier transform mass spectrometry or ion mobility spectrometry-mass spectrometry (IMS-MS). It demonstrates that these
techniques can unambiguously identify analytes with very similar or identical elemental compositions by examining both a standard mixture and a complex cellular extract. The high resolving power and sub-ppm accuracy afforded by the linear trap quadrupole (LTQ)-Orbitrap instruments can provide the means to resolve these complex analyte mixtures and enable the determination of their elemental compositions. This platform also is capable of multi-step tandem (MS\textsuperscript{n}) mass spectrometry that provide the additional layers of structural information needed to confirm any putative hits and thus, unambiguously identifying them as nucleotide variants. Additionally, IMS-MS is able to probe the three-dimensional structure of each ion by determining its drift time as it travels through a low pressure region, where it may encounter a continuous or oscillating electric field. Using the intrinsic relationship between drift time and the probability of the given ion to interact with background gas, the structure and conformation of each isomeric and chiral specie are able to be differentiated. Together, these techniques have been investigated for their individual merits in enabling the complete characterization of a standard mixture of nucleotides and those nucleotide variants found on naturally occurring RNAs in whole cell lysates.

2.2 MATERIALS AND METHODS

2.2.1 Mononucleotide standards.

Individual nucleosides 5'-monophosphate (nucleotides) employed in the study were purchased from Sigma-Aldrich (St Louis, MO) and used without further purification. For this reason, stock solutions contained characteristic unidentified contaminants detected in all ESI spectra of solutions containing such samples. Pseudouridine monophosphate (ΨMP) was instead
produced as a custom synthesis by Dharmacon Research Inc. (Lafayette, CO). Appropriate mixtures of these analytes can aptly mimic typical samples obtained by enzymatic digestion of DNA and RNA by the conventional nucleases employed in nucleic acid analysis. Sample concentrations were determined by UV absorbance at 260 nm by using extinction coefficients reported previously. Immediately before analysis, stock solutions were typically diluted to a final concentration of 9.0 µM of total NMPs in 100 mM ammonium acetate and 10% isopropanol.

2.2.2 Cellular extracts.

*Saccharomyces cerevisiae* strain BY4741 was grown in either yeast extract, peptone, dextrose (YPD) or synthetic complete (SC) media. Cell suspensions were streaked onto YPD agar plates and incubated at 30°C overnight. Five individual colonies were typically selected from each plate and placed into individual tubes containing 6 mL of either YPD or SC medium. All growth tubes were incubated at 30°C with 200 rpm gyration. Optical density at 600 nm (OD$_{600}$) was monitored on a ThermoFisher Scientific (Waltham, MA) Nanodrop 2000c spectrophotometer until a value greater than 0.3 units was reached. Each liquid culture was diluted to a final 0.3 OD$_{600}$. A 3-mL aliquot of each culture was centrifuged at 6000 g for 5 min. to obtain pellets that contained approximately the same number of cells. *Escherichia coli* K-12 strain MG1655 was grown in synthetic complete (SC) medium according to established procedures. Harvesting was carried out in analogous way.

Each pellet was disrupted by using Denaturation Solution (Life Technologies, Green Island, NY). When required by the standard-additions protocol, accurately known aliquots of *S. cerevisiae* tRNA$^{\text{Phe}}$ (Sigma-Aldrich, St. Louis, MO) were introduced at this point into the lysate to
serve as an internal standard. Total RNA was extracted by using the ToTALLY RNA Extraction Kit (Life Technologies, Green Island, NY), which is based on a typical phenol/chloroform procedure. The RNA was precipitated by using cold isopropanol and then treated with DNase 1 (New England Biolabs, Ipswich, MA) in 1X DNase buffer to remove any remaining DNA. The recovered RNA was subsequently desalted by ethanol precipitation overnight and reconstituted in 50 µL of RNase-free water (Sigma-Aldrich, St. Louis, MO). The concentration of intact total RNA from each sample was measured by UV absorbance at 260 nm. Nuclease P1 and phosphodiesterase 1 from snake venom (Sigma-Aldrich, St. Louis, MO) were employed to complete the digestion of RNA into individual mononucleotides, as previously described. Immediately before analysis, final samples were diluted 1:10 in 150 mM ammonium acetate and 10% isopropanol.

2.2.3 Molecular mass determination.

Samples were analyzed by direct infusion ESI on either a Thermo Fisher Scientific (West Palm Beach, CA) LTQ-Orbitrap Velos mass spectrometer, or a Waters (Milford, MA) Synapt G2 HDMS IMS mass spectrometer. All analyses were performed in nanoflow ESI mode by using quartz emitters produced in house by a Sutter Instruments Co. (Novato, CA) P2000 laser pipette puller. Up to 5µL samples were typically loaded onto each emitter by using a gel-loader pipette tip. A stainless steel wire was inserted in the back-end of the emitter to supply an ionizing voltage that ranged between 0.6 and 1.4 kV. Source temperature and desolvation conditions were adjusted by closely monitoring the incidence of ammonium adducts and water clusters. For high-resolution determinations, the LTQ-Orbitrap instrument was calibrated by using an anion mixture that contained sodium dodecyl-sulfate, sodium taurocholate, and Ultramark. These
standards enabled the calibration of the instrument over a range of m/z 50-2000 with a ~200 ppb mass accuracy. Samples were collected for 5 min. runs from m/z 300-700 at 5 times a piece. Raw data was then subjected to database searches to find putative PTMs.

Apparent drift time (t_D) was determined by allowing ions to move through the travelling wave (Tri-WAVE) element of the IMS mass spectrometer, which were then transferred for mass analysis into the time-of-flight (TOF) stage operated in single reflectron mode. The instrument was calibrated by using a 2 mg/mL solution of cesium iodide in 50:50 water/methanol, which afforded a ~9 ppm mass accuracy. For comprehensive mixture analysis, the Tri-WAVE region was held at a pressure of approximately 4.40 mbar (uncalibrated gauge reading) by a 90 mL/min flow of N_2 and 180 mL/min of He. It was operated with an approximately 650 m/s IMS wave velocity, a 40 V wave height, a 109 m/s transfer wave velocity, and a 2.0 V transfer wave height. Analysis of mass-selected isobars was performed by raising the cell pressure to ~4.60 mbar (uncalibrated gauge reading) with a flow of 140 mL/min N_2 and 180 mL/min He. At the same time, IMS wave velocity was raised to ~700 m/s, transfer wave velocity to ~600 m/s, and transfer wave height to 4.0 V. The observed mobility profiles were compared with theoretical Gaussian distributions obtained from the curve-fitting algorithm present in the PeakFit 4.1 package. In our hands, PeakFit deconvolution allowed for the discrimination of contiguous profiles with a resolving power of ~40, which exceeded the value of ~25 quoted for these types of experiments by the instrument’s manufacturer.

2.2.4 Tandem mass spectrometry.

Tandem mass spectrometry (MS/MS) was accomplished by isolating the precursor ion of
interest in the linear trap quadrupole (LTQ) element, which was then collided with N₂ to activate fragmentation. The ensuing products were mass analyzed either in the LTQ or the Orbitrap region of the instrument. Where applicable, multiple activation steps (MSⁿ) were employed by isolating the first generation ion and subsequent fragments prior to activation. The fragmentation of mass-selected ions was activated by using typical 25eV collision voltage. Ensuing products were mass analyzed either in the LTQ or the Orbitrap region of the instrument. Consecutive reaction monitoring (CRM) experiments were performed by dialing the selected precursor → fragment transitions in the instrument data system. Series of diagnostic CRM experiments were performed in systematic fashion by inputting lists of precursor → fragment transitions specific for the different modifications, which were completed by the instrument with no further user intervention (Scheme 2-1).

2.2.5 Data analysis.

High-resolution and fragmentation data obtained on the LTQ-Orbitrap instrument were processed by Xcalibur 2.1 software (ThermoFisher Scientific, Waltham, MA). Mass calculations and predictions of elemental composition were performed by using the Molecular Weight Calculator software made available by the Pacific Northwest National Laboratory. A data reduction step was implemented prior to database searching to simplify the operations and minimize the incidence of false positives. Instead of relying on a predefined intensity threshold to discriminate signal from noise, the experimental masses to be employed in the searches were selected according to a deconvolution algorithm included in the Xcalibur 2.1 software. This algorithm requires the detection of full-fledged ¹²C and ¹³C signals to correctly assign the charge...
**Scheme 2-1.** Proposed MS workflow.
state of observed species. If the $^{13}$C peak of a low-abundance component was not recognized from the background (and a plausible charge was not assigned), then the mass of the corresponding $^{12}$C was filtered out regardless of whether its intensity afforded an acceptable signal-to-noise ratio. The resulting mass list was then searched against the METLIN database (http://metlin.scripps.edu/index.php) and a non-redundant registry obtained in house by combining the entries present in the RNA Modifications (http://mods.rna.albany.edu/) and MODOMICS databases.\textsuperscript{240} Matching between experimental data and database information was carried out by using software developed in house.

IMS-MS data were displayed in the form of heat-map plots with arrival time ($t_D$) and mass to charge ratio ($m/z$) placed on the x- and y-axis, respectively, by using OriginPro 9.1 (Origin Lab, North Hampton, MA). A color gradient provided in each plot was used to indicate the signal intensity expressed in arbitrary ion counts. For data subtraction analysis, appropriate scaling factors were utilized to align the intensity scales of the selected plots. Such factors were calculated to match the combined intensities of the four canonical ribonucleotides (i.e., $\sum_1^4 cr_i$, with $cr_i$ corresponding to the respective absolute intensity in arbitrary counts) observed in each plot. Taking advantage of this proxy, de facto internal reference, it also was possible to express the abundance of each species in relation to that of the canonical ribonucleotides according to:

$$AvP_x = \frac{a_{ix}}{\sum_1^4 cr_i} \cdot 100$$  \hspace{1cm} \text{Equation 2-1}$$

in which $AvP_x$ is the abundance versus proxy of a certain species obtained from its absolute intensity ($a_{ix}$) normalized to the combination of the abundances of the canonical ribonucleotides ($\sum_1^4 cr_i$). Home-built software was employed to process the experimental data, calculate $AvP$
values and, when necessary, apply appropriate scaling factor for data alignment. The same application also was employed to perform point-by-point subtraction of aligned data. The software itself is the object of a manuscript in preparation. The results were visualized in heat-map format by utilizing OriginPro 9.1 (Origin Lab, North Hampton, MA).

2.3 RESULTS AND DISCUSSION

2.3.1 Identification of PTMs based on mass.

Investigation of the capabilities of using high-resolution mass spectrometry to unambiguously identify PTMs based on mass was initiated by exploring the resolving power—the ability to distinguish two peaks of slightly different mass-to-charge values—of the LTQ-Orbitrap instrument; a hybrid instrument containing both a linear trap and an Orbitrap mass analyzer. A mixture of the DNA and RNA canonical bases were prepared and combined in equimolar amounts. Table 2-1 provides the complete list and elemental compositions of the mixture studied. Upon examination, it can be seen that several of the species are isobaric species, and as such, they exhibit the same elemental composition but have different structures. This is a common problem encountered when looking at complex mixtures, therefore, it was paramount that the analytical approach utilized here be capable of identifying all species despite their overlap in elemental composition.

First approximation of the content of the mixture was recorded in the ion trap analyzer of the LTQ-Orbitrap using negative ion mode detection (Figure 2-1a). In this mode, the analytes can be detected as singly charged ions produced by the deprotonation of the respective
<table>
<thead>
<tr>
<th>Name</th>
<th>Elemental composition</th>
<th>Monoisotopic mass (u)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guanosine monophosphate (GMP)</td>
<td>C₁₀H₁₄N₅O₇P</td>
<td>363.05800</td>
</tr>
<tr>
<td>Adenosine monophosphate (AMP)</td>
<td>C₁₀H₁₄N₅O₇P</td>
<td>347.06308</td>
</tr>
<tr>
<td>Cytidine monophosphate (CMP)</td>
<td>C₉H₁₄N₃O₈P</td>
<td>323.05185</td>
</tr>
<tr>
<td>Uridine monophosphate (UMP)</td>
<td>C₉H₁₃N₂O₉P</td>
<td>324.03587</td>
</tr>
<tr>
<td>Pseudouridinemonophosphate (ΨMP)</td>
<td>C₉H₁₃N₂O₉P</td>
<td>324.03587</td>
</tr>
<tr>
<td>Deoxyguanosine monophosphate (dGMP)</td>
<td>C₁₀H₁₄N₅O₇P</td>
<td>347.06308</td>
</tr>
<tr>
<td>Deoxyadenosine monophosphate (dAMP)</td>
<td>C₁₀H₁₄N₅O₆P</td>
<td>331.06817</td>
</tr>
<tr>
<td>Deoxycytidine monophosphate (dCMP)</td>
<td>C₉H₁₄N₃O₇P</td>
<td>307.05693</td>
</tr>
<tr>
<td>Deoxythymidine monophosphate (dTMP)</td>
<td>C₁₀H₁₅N₂O₈P</td>
<td>322.05660</td>
</tr>
</tbody>
</table>

**Table 2-1** Elemental compositions and monoisotopic mass of neutral 5’-phospho-nucleosides calculated by using PNNL’s Molecular Weight Calculator. Taken with permission from Ref. 262.
phosphate group. The majority of the signals observed were correctly assigned to individual nucleotides and respective ammonium, sodium or potassium adducts based on their theoretical calculated masses. As electrospray ionization is a gentle, less energetic, ionization technique, no products corresponding to loss of a nucleobases were observed. Due to the labile nature of the N-glycosidic bond, more energetic ionization techniques often result in the aforementioned loss.

It was expected from their chemical composition that some canonical DNA and RNA species would show overlapping signals in the analyzer. This was readily observable when looking at AMP and dGMP (m/z 346.08) as well as UMP and YMP (m/z 323.08). A way to address this issue is by monitoring the $^{12}\text{C}$ and $^{13}\text{C}$ signals for each species present. However, inspection of the signals obtained from the ion trap analyzer only showed a single Dalton spacing, preventing us from observing these isotopic distributions; a major limitation of this particular analyzer. The result of this analysis were $^{12}\text{C}$ and $^{13}\text{C}$ signals that were blurred. Consequently, in the mass range of 320-325, only the $^{12}\text{C}$ signals of dTMP, CMP and U/YMP were readily assigned, whereas the $^{13}\text{C}$ peaks were not (Figure 2-1b).

In order to overcome the drawbacks of the ion trap analyzer, we set out to explore the entire range of capabilities afforded by this hybrid instrument by investigating the merits of the Orbitrap mass analyzer in the analysis of the same nucleotide mixture. The full scan obtained in Orbitrap mode revealed that its high resolution enabled a clearer picture of both the $^{12}\text{C}$ and $^{13}\text{C}$ signals (Figure 2-1c). In particular, the signal for the $^{12}\text{C}$ of [U/ΨMP-H]$^-$ was now resolved from the $^{13}\text{C}$ signal of [CMP-H]$^-$ (Figure 2-1c insets). These results showed that the ~150k full width at half maximum (FWHM) resolving power achieved by the Orbitrap analyzer significantly outweighed the ~2.3k FWHM capability of the ion trap analyzer. For this reason, the Orbitrap mode was used
Figure 2-1  a) ESI-MS spectrum of the nucleotide mixture recorded in the LTQ analyzer; b) detail of the same spectrum showing isotopic distribution overlap between species that differ by a single mass unit; c) detail of the same region of an analogous ESI-MS spectrum recorded in the orbitrap analyzer. The enlarged insets demonstrate the ability of orbitrap experiments to resolve the $^{13}$C and $^{12}$C signals of contiguous nucleotides. Take with permission from Ref. 262.
for all subsequent mass determinations that, despite the resolving power afforded by this analyzer, still faced some intrinsic difficulties. For instance, our data clearly showed that the significant $^{13}$C contribution of an overlapping neighbor could still potentially overshadow the weaker $^{12}$C signal and skew the mass assignments for some of the lower intensity species present in more complex mixtures.

For any type of mass analysis, it is of essential importance that the selected instrumental technique be capable of providing accurate results. Thus, the accuracy of this particular technique was assessed by first comparing the experimental mass obtained from the Orbitrap mass analyzer with the chemical formulae generated by the PNNL Molecular Weight Calculator. This fact can be exemplified by the species observed at m/z 322.04406. When this mass was searched for possible matching formulae by using the PNNL Molecular Weight Calculator, only two possible formula combinations were obtained: $C_8H_7N_1O_3P$ and $C_9H_{13}N_3O_8P$. Given the implausible number of nitrogen atoms present in the former, it was determined that the latter was indeed more likely and matched the expected formula for deprotonated CMP. When comparing the theoretical masses of the deprotonated species to the experimental masses obtained, ~200 ppb accuracy was obtained on average. The values obtained here are not surprising given the vast improvements in the field of high resolution mass spectrometers. Even still, these improvements only show enough ability to differentiate simple mixtures of isobaric species such as UMP/ΨMP and AMP/dGMP where the respective $^{12}$C and $^{13}$C signals can be easily assigned. The complexity of this issue increases exponentially with the increase in complexity of the sample being analyzed due to the likelihood of overlapping signals. Additionally, whole cell extracts may contain multiple species (upwards of six) appearing at one
mass-to-charge ratio. For this reason, other avenues should be investigated in addition to isotope analysis, such as tandem mass spectrometry, to resolve these more complex sample mixtures.

2.3.2 Tandem MS approaches and isomeric PTMs.

In light of the fact that mass alone may not be sufficient to unambiguously assign the identity of nucleotides, we explored additional approaches that focused on tandem mass spectrometry to obtain specific structure information. Extensive work on the application of MS/MS approaches have been accomplished by pioneers like Crain, McLuckey and Limbach in order to accurately determine the placement of the various modifying groups on the canonical bases.\textsuperscript{16,22,96,122–124,245–247} The earliest approach for elucidating structure of nucleic acids employed mass-analyzed ion kinetic energy spectrum (MIKES) followed by collision induced dissociation (CID) of both DNA and RNA. Use of this method showed the ability to unambiguously discern 1-, 2- and N\textsuperscript{6}-methyladenines from a DNA lysate.\textsuperscript{248} Consequently, this same approach was applied to a tRNA sample isolated from \textit{E. coli}. Since then, CID has been used to identify modified bases in unpurified human and rat urine.\textsuperscript{249} Various techniques involving the coupling of a front-end separation before MS analysis also have made vast advancements leading to the identification of >100 natural variants of the canonical ribonucleotides. As such, tandem approaches seemed like the most advantageous avenue for the complete characterization of the canonical DNA and RNA bases and their natural variants.

To this end, the hypothesis that diagnostic fragments may be directly traceable to their respective precursor was investigated by first examining the fragmentation exhibited by activating [AMP-H]\textsuperscript{+} alone (\textbf{Figure 2-2a}). The resulting spectrum shows the cleavage of the 5\textsuperscript{'-}
Figure 2-2 a) Negative ion mode MS/MS spectrum of deprotonated AMP obtained from a separate sample; MS$^3$ spectrum obtained by isolation and activation the 5'-phosphoribose fragment produced in the previous dissociation step (i.e., m/z 346.08 $\rightarrow$ 211.02 $\rightarrow$); c) negative ion mode MS/MS spectrum of coexisting AMP/dGMP in the nucleotide mixture; d) positive ion mode MS/MS spectrum of protonated AMP/dGMP from the nucleotide mixture. Taken with permission from Ref. 262.
phosphoribose moiety, which was subsequently subjected to MS³ analysis in which ring cleavages and the loss of water could be observed (Figure 2-2b). Similarly, activating the species observed at 346.08 m/z in the nucleotide mixture (Figure 2-2c), showed fragments belonging to both deprotonated AMP and dGMP. Consistent with the presence of coexisting populations, the process of base elimination resulted in the release of the 5’-phosphoribose moiety from the former and that of the equivalent 5’-phospho-2’-deoxyribose from the latter (Figure 2-2c). The assignment of these fragments was confirmed in separate MS³ experiments, as described earlier. Additionally, the mixed precursor yielded products corresponding to loss of water from the 5’-phosphoribose moiety (i.e., [C₅H₆O₆P]⁺, observed separately for [AMP-H]⁻) and the opening of the five-member ring (i.e., [C₃H₄O₅P]⁺, observed separately for both [AMP-H]⁻ and [dGMP-H]⁻). In positive ion mode, the mixed precursor provided signals corresponding to protonated adenine and guanine (Figure 2-2d), consistent with the gas-phase behavior of the individual nucleotides. Also in this case, each first-generation fragment was isolated and activated in MS³ analyses that confirmed the respective purine structures.

In spite of their identical elemental composition, the different placement of the oxygen on the pentose or purine system ensured that the same dissociation process would lead to unique diagnostic fragments for AMP and dGMP. In isomeric species, however, base loss may not be sufficient to discriminate more subtle structural variations, unless such features are conducive to distinctive fragmentation pathways, as in the case of UMP and ΨMP. Analyzed individually, the former displayed the typical 5’-phosphoribose fragment from cleavage of the N-glycosidic bond, in addition to a species produced by formal elimination of CONH from the pyrimidine ring (Figure 2-3a).²⁵⁰,²⁵¹ In contrast, no fragments associated with base loss were observed for
Figure 2-3. Negative ion mode MS/MS spectra obtained from a) an individual UMP sample; b) individual ΨMP; c) coexisting UMP/ΨMP in the nucleotide mixture. Taken with permission from Ref. 262.
deprotonated ΨMP (Figure 2-3b). Instead, an intense signal was detected for the nucleoside moiety left by cleavage of the phosphate group (i.e., [C₉H₉N₂O₅]), which was not observed for any other species under the same experimental conditions. The absence of base loss is consistent with the fact that, unique among the nucleotides in this study, the ring of ΨMP is connected to the pentose by a C-glycosidic bond that provides greater stability than that of typical N-glycosidic bonds. For this reason, when the mixed precursor at 324.04 m/z was activated, the resulting spectrum contained diagnostic features produced exclusively by either component, such as the 5’-phosphoribose fragment ascribable to UMP and the phosphate-loss product traceable only to ΨMP (Figure 2-3c). In this case, mutually exclusive fragmentation pathways provided the key for differentiating isomeric populations coexisting in the same precursor ion.

While extensive work has previously shown that there are known stabilizing properties attributed to the 2’-deoxy- versus ribonucleotides under selected experimental conditions, this stabilization effect did not seem to contribute significantly to the results afforded by these particular experiments. Indeed, the stabilizing effects did not produce any difference in the patterns observed for the mononucleotide species. Instead, both the dNMPs and NMPs exhibited similar behaviors lending to the idea that the differences observed in previous work between gas-phase cleavage patterns of DNA and RNA are a product of their relative overall stabilities, which is defined by the contributions of the 3’-phosphate and other functional groups present on the specific oligonucleotide. Additionally, since analysis was mainly performed in negative ion mode, no signals were observed for the nucleobase moieties themselves. This is not surprising given the inability of such species to hold a negative charge due to the electron-rich nature of nitrogenous bases. However, when fragmentation experiments were performed in positive ion
mode, the cleavage of the N-glycosidic bond resulted in the nucleobase moieties that were observed as BH$_2^+$ species. The only nucleobases that veered from this particular pattern were [dTMP+H]$^+$ and [U/ΨMP+H]$^+$. This particular deviation could be attributed to the relatively basic character of the nucleobases which can be substantiated by their respective proton affinities. This character could help to explain why the ability of these bases are observable in their protonated forms in the gas phase. Even the absence of protonated T and U in the respective spectra could be ascribed to their low rank in the relative scale of proton affinities (i.e. G > C > A >> T ≈ U).$^{253-255}$ Acid-base properties alone cannot explain the results obtained from the deprotonated analytes. Indeed, it has been noted that loss of neutral base (i.e., BH) is the predominant dissociation channel for 5′-phospho-2′-deoxynucleosides,$^{256,257}$ whereas 3′-phospho-nucleosides, dinucleotides, and larger oligonucleotides display a greater propensity to eliminate the deprotonated form B$^-$.$^{251,258,259}$ These observations further emphasize the importance of the structural context in determining the mechanism of base loss and, by extension, the exact form in which the lost fragment may be detected.

In some instances, certain PTMs can display multiple variants with the same chemical formula. Therefore, employing tandem analysis to identify unique fragments becomes increasingly more difficult. Often, the ability to accomplish the analysis can be limited by low analyte concentrations or a lack of discernable unique fragments. These difficulties were exemplified by the analysis of the species detected at m/z 376.0684 in a more complex sample in Figure 1-2a, which could potentially match different methyl-G isomers. In anionic mode, characteristic phosphoribose fragments produced by base loss were readily observed upon collisional activation of the deprotonated precursor in the LTQ-Orbitrap analyzer. The fact that
such products were detected in both methylated and unmethylated form was consistent with the presence of alternative isomers with the methyl group on the ribose (i.e., 2’-O-methyl-GMP (mG) only possible match), or the purine system (i.e., 1-methyl-GMP (m1G), N2-methyl-GMP (m2G), or 7-methyl-GMP (m7G)). In positive ion mode, activation of the protonated species detected at m/z 378.0840 led to complementary products corresponding to free and methylated nucleobase. The first-generation fragment obtained in positive ion mode was subsequently submitted to both MS3 (i.e., m/z 378 → 166 →, Figure 1-2b) and MS4 analysis (i.e., m/z 378 → 166 → 124 →, Figure 1-2b inset) to identify the position of the methyl group onto the purine system. Ultimately, the observed fragments were consistent with N1-, N2- and N7-methylation proving that tandem analysis is indeed capable of capturing unambiguous differentiation of even complex isobaric mixtures.

2.3.3 IMS-MS techniques.

Gas-phase activation can enable the elucidation of structural information by the identity of fundamental units, functional groups, and bonds, which can be predicted directly from the masses of observed fragments. In contrast, IMS-MS experiments can access structural information in the form of the placement of structures and functional groups in the three dimensions, which define the collisional cross-section provided by drift time (tD) determinations. Therefore, the use of IMS-MS was explored as an alternative way to discriminate between analytes in the mononucleotide mixture that shared the same elemental composition but differing structures. In the case of AMP/dGMP which showed the overlapping unresolvable signal in both the ion trap and the Orbitrap analyzer, the mixed precursor ion at 346.17 m/z was isolated
in the mass-selective quadrupole and then injected in the travelling wave (Tri-WAVE) element of the IMS-MS instrument. The corresponding signal was recorded as a mobility profile, which was plotted as relative intensity vs. apparent t₀, and showed two maxima separated by a minimum, consistent with the presence of at least two main components with different propensity to undergo low-energy interactions with background gas (Figure 2-4a). When curve-fitting was employed to resolve the individual contributions, the presence of two distinct populations could be more clearly delineated (Figure 2-4b). The apparent t₀ values afforded by such populations matched those determined individually for deprotonated AMP and dGMP (i.e., 6.08 and 5.74 ms, respectively), thus confirming the signal assignments. In a similar fashion, the mobility profile recorded for the UMP/ΨMP mixed precursor displayed a major signal with a smaller, partially resolved peak (Figure 2-4c). The curve-fitting algorithm helped distinguish two components with different t₀ values (Figure 2-4d) that matched those determined from individual samples (i.e., 5.24 and 5.04 ms for deprotonated UMP and ΨMP, respectively). Therefore, these experiments demonstrated the ability of IMS-MS to detect the effects of significant structural differences, such as the distinctive placement of a discrete functional group in AMP and dGMP, but also those of more subtle variations, such as the different position of the bond between nucleobase and pentose in UMP/ΨMP.

In general, the mobility of each ion can be evaluated individually after isolation in the mass-selective quadrupole of the instrument. Alternatively, this type of information can be obtained in parallel for all ions in the sample by dispensing with individual isolation steps. In this approach, all ions are transmitted through the quadrupole without mass selection, are allowed to disperse during the travel across the Tri-WAVE, and are finally analyzed in the time-of-flight
Figure 2-4. Ion mobility profiles obtained from the precursor ions at m/z 346.17 a) and 324.04 c). Panel b) and d) are the respective curve-fitted plots generated by PeakFit 4.1 (see Experimental). Analyte identity was assigned by comparing the apparent drift time (t₀) of the deconvoluted signals with those obtained in separate experiments from isolated nucleotides. Taken with permission from Ref. 262.
analyzer. The recorded m/z is then plotted against the corresponding t_D in two-dimensional maps in which the relative intensity is expressed by color-coding. These types of heat-maps allow for each ion to be uniquely identified by its respective mass and mobility characteristics. Depending on the intensity threshold selected for peak detection, the maps can provide very comprehensive representations of the complexity of the samples under investigation. In the case of the nucleotide mixture, the recorded signals corresponded not only to the analytes of interest, but also to their salt adducts and unidentified background contaminants (Figure 2-5). As highlighted in the enlarged region, the signal for each nucleotide could be unambiguously recognized from its characteristic coordinates. Indeed, the map shows that an identical elemental composition placed AMP and dGMP on the same m/z coordinate, but their structural differences translated into broad dispersion on the t_D dimension. Owing to the more modest structural variations between UMP and ΨMP, the couple was not nearly as resolved as AMP/dGMP in the t_D dimension, consistent with the results obtained from mass-selected precursor ions (Figure 2-4c).

It is important to note that, if necessary, ion mobility profiles in the format shown in Figure 2-4a and c could be readily extracted from the original data to enable curve-fitting analysis and facilitate the recognition of partially resolved species.

This type of approach can take mixture characterization a step further by allowing the parallel activation of all ions after they are dispersed in the time dimension by the Tri-WAVE and before they are mass analyzed. The fragments generated at this stage maintain the t_D value of the respective precursor ion, but are now spread in the m/z dimension of the heat-map (see for example Figure 2-6). The data afforded by time aligned parallel (TAP)^260 dissociation can be analyzed by extracting longitudinal sections of the map, which correspond to actual mass spectra.
**Figure 2-5.** Comprehensive representation of IMS-MS data obtained from the entire nucleotide mixture. In the heat-map, the m/z of each ion is plotted against its apparent t₀, whereas its relative intensity is expressed by progressively lighter colors for the more abundant species. Red dots mark the maxima corresponding to the various nucleotides. The enlarged inset highlights the separation between mixture components, which was achieved in the t₀ and m/z dimension of the plot. A variety of salt adducts and unidentified background contaminants were also detected. No signal corresponding to typical nucleotide fragments was recognized in experiments completed without ion activation. Taken with permission from Ref. 262.
(i.e., relative intensity vs. m/z) displaying all precursor ions with the same $t_0$ and their respective fragments. Given that ions tend to exhibit relatively broad $t_0$ profiles (see for example Figure 2-4a and c), the extracted TAP spectra are likely to contain the products of multiple precursors. For example, the longitudinal section taken at 3.6 ms (i.e., solid vertical line in Figure 2-6) included the maximum of deprotonated AMP, but impinged also onto the broad profiles of dAMP and dCMP, as well as those of unidentified background species recorded also in the absence of activation (Figure 2-6). The corresponding spectrum displayed both deprotonated AMP and dAMP, in addition to familiar fragments observed in their separate MS/MS spectra, such as the 5’-phosphoribose and 5’-phospho-2’-deoxyribose moieties and phosphate fragments (Figure 2-6 inset). As discussed earlier for MS/MS spectra of mixed precursor ions (e.g., Figure 2-2c, d and 2-3c), with which TAP dissociation share many similarities, the interpretation of these data can greatly benefit from prior knowledge of the fragmentation patterns afforded by individual components.

Given the complexity of biological samples, exploration of an alternative to TAP was used a as characterization strategy for these more difficult cases. This strategy, called mass-selective time-resolved (MaSTeR) dissociation, mimicked the more traditional tandem MS spectrometry by activating the fragmentation of only those species that were recognized as potential nucleic acid constituents. As illustrated for the methyl-G species described in the previous section, this experiment was completed by isolating the precursor ion at m/z 378 in the mass-selective quadruple, by allowing the various isomers to disperse on the time domain in the ion mobility element, and by then activating their gas-phase dissociation before final mass analysis. In the MaSTeR dissociation experiment, the data obtained at different intervals displayed
Figure 2-6. Heat-map representation of time aligned parallel (TAP) dissociation products provided by the entire nucleotide mixture. The data were collected by using a transfer voltage of 25 V (see Experimental). The inset represents the mass spectrum extracted at 3.6 ms of the plot (white solid line). The spectrum contains quasi-molecular ions that shared the same \( t_D \) across the Tri-WAVE, as well as their fragment ions produced in the transfer region before the TOF analyzer. Such fragments were not observed in analogous experiments performed. Asterisks mark unidentified background contaminants. Taken with permission from Ref. 262.
fragmentation patterns that were characteristic of the various methyl-G isomers (Figure 2-6) and matched those observed in LTQ-Orbitrap experiments (Figure 1-2). As discussed above, dissociation of the N-glycosidic bond differentiated isomers with the methyl group located either in the phosphoribose or nucleobase moiety, whereas further fragmentation of the latter identified the position of methylation sites on the purine system. A close examination of these data revealed that characteristic fragments, such as the methylated/unmethylated purine moieties, could be detected with very distinctive intensities in different sections of the time domain (compare for example panel a) and b) of Figure 2-6), thus suggesting possible overlap between contiguous isomers. The possible ambiguity was readily resolved by extracting the mobility profiles of diagnostic fragments unique for each isomer, which allowed for a clear differentiation of the individual species (Figure 2-7). Further, Gaussian-fitting the mobility profile afforded by m/z 378 revealed that the m7G isomer was the most abundant isomer in the sample and contributed the lion’s share of the spectral overlap observed in the time domain (Figure 2-7a and b). A manuscript in preparation is dedicated to discuss in greater detail the benefits and perils this type of approach on all PTMs found in S. cerevisiae samples.

It should be noted that our strategy did not count on specific tD values to achieve positive identification of the various species, but relied instead on corroborating fragmentation information. The time domain was employed here to achieve separation between isomers and to enable the observation of their specific fragmentation patterns. In our hands, tD determinations showed an average ±0.006 msec over repeated analyses on different days, which could potentially support the utilization of tD as a unique identifying characteristic. However, in light of the number of experimental variables that may affect such quantity, it will be important
Figure 2-7. a) IMS-MS profile of m/z 378 after isolation in the mass-selective quadrupole and separation in the ion mobility element. Reconstructed ion chromatograms (RICs) unique for the different methyl-G isomers: b) m/z 151, m₁G; c) m/z 110, m₁G; d) m/z 68, Gm; e) m/z 151, m²G. The traces in panel a) represent the contribution of the different species revealed by Gaussian fitting (see Materials and Methods). Taken with permission from Ref.310.
to verify the reproducibility of this type of experiment across different instruments before \( t_D \) values can be employed directly for identification purposes. Further exploration of this potential has been the object of multiple studies which are now being drafted into a manuscript.

### 2.3.4 Direct infusion analysis of cellular nucleotide mixtures.

Employed in concert, these techniques offer an excellent platform for the analysis of a ribonucleotide mixture obtained from the mononucleotide standard. Additionally, these techniques were capable of discerning isobars from complex mixtures (i.e., cellular extracts); as seen from the methyl-G isomers. Further implementation of these approaches were investigated on a total RNA sample extracted from \( S. \) cerevisiae. As is typical to any analytical process, a sensible but deceptively challenging way to reduce sample losses and analyte bias consists of reducing the number of sample-handling steps included in the workflow. With the goal of minimizing these potential problems and increasing the analysis performance, we explored the utilization of a classic phenol-chloroform extraction to simultaneously lyse cells and expedite the isolation of nucleic acid components.\(^{261}\) The following step consisted of exonuclease digestion into separate mononucleotides.\(^{232}\) **Scheme 2-2** shows the typical analytical workflow employed in our studies. Given that each of the more than 100 RNA variants have masses that exceed those of the canonical bases, these approaches should be streamlined for the identification of mixtures of PTMs within the cellular extract. The advantage of using specific nucleases come from the ability of the analyst to predict the resulting content of the mixture. As is the case with nuclease P1 and phosphodiesterase, complete hydrolysis of the RNA strands to the mononucleotide level is readily achieved in under 5 hours. This procedure overcomes the limitations seen from complex
Scheme 2-2. Proposed experimental workflow (see Material and Methods for details). Taken with permission from Ref. 310.
ESI spectra of larger oligoribonucleotides. However, while this approach is advantageous for detecting any PTMs present within the transcriptome, it is not able to pinpoint the location of the PTM on a particular parent strand. As RNA function is known to be linked to its structure, the location of these PTMs will be of grave importance to understanding the biology associated with a particular strand. Regardless, the ability to monitor even the most subtle changes in expression level of PTMs can already unearth a much greater amount of information about PTM function than was previously attainable from other techniques.

The direct infusion of the mixture explored herein was immediately appreciated by examining the representative data obtained from a digest of total RNA from a sample of *S. cerevisiae* grown in yeast extract, peptone, dextrose (YPD) medium, which are shown in Figure 2-8. A broad distribution of signals with very different intensities covered the entire range between m/z 300 and 700—the region in which PTMs are typically observed. Abundant signals were easily identified as the deprotonated molecular ions of canonical ribonucleotides (i.e., \([\text{NMP-H}]^-\), where N indicates any nucleoside).²⁶² Their experimental masses exhibited an average of ~100 ppb deviation from values calculated from the corresponding elemental compositions, which matched the typical accuracy afforded by these types of LTQ-Orbitrap determinations.²⁶² Owing to the absence of high-resolution separation steps, the final samples were anticipated to contain not only the desired mononucleotide analytes, but also unrelated cellular components carried through the entire workflow (Scheme 2-1). Many of such signals were identified with common metabolites (vide infra).

The complexity of whole-cell extracts was also tackled by utilizing IMS-MS, which enables the differentiation of ions according to their size and/or conformation as described above in
Figure 2-8. Representative ESI-MS spectrum of total RNA digest obtained from *S. cerevisiae* grown in YPD medium. The enlargement shows the region containing the majority of the PTMs. Signals marked with * are hits from our custom modifications registry; † METLIN hits; o species detected also in the blank. Taken with permission from Ref. 310.
Section 2.3.3. In this instance it can be hypothesized that ions of greater mass or larger structure will exhibit increased low-energy collisions in comparison to lower mass species. Ultimately, these collisions will result in the smaller or more compact ions reaching the detector prior to the larger or more extended ions. In the case of PTMs in which there are several isobaric species, something as simple as the placement of a methyl group on a nucleobase could potentially correlate to vast differences in travel time through the low pressure region of the instrument. The corresponding data obtained from this type of experiment are displayed in the form of heat-maps or 3D plots, in which the different dimensions consist of arrival time (t_D), mass to charge ratio (m/z), and signal intensity. As such, a representative heat-map obtained from the aforementioned yeast extract is shown in Figure 1-3. The canonical ribonucleotides also were immediately recognizable on the basis of their characteristic m/z and t_D values. This aspect is exemplified in the enlarged region of the map. Given that this sample was derived directly from a whole-cell extract with no RNA fractionation, the corresponding heat-map provided an immediate and comprehensive representation of all the PTMs present in the cells. This comprehensive representation is analogous to the abundance of peaks in the lower intensity range enlarged in Figure 2-8. Consistent with the idea that smaller ions would travel at a faster rate, it can be directly seen that the ions with lower m/z values are in fact those exhibiting the shorter drift times. This technique, therefore, provided a means by which to assess the unique structural differences exhibited by each of these species present in the whole cell lysate.

Tackling identity of each of the numerous signals observed during LTQ-Orbitrap or IMS-MS experiments were confronted by employing the aid of database searches. The vast majority of the detected signals were readily assigned upon initial searching. However, a more stringent
approach was investigated by the application of data reduction steps which followed a conservative approach that eschewed the application of a pre-determined threshold to eliminate background noise on the basis of signal intensity, but relied instead on the detection of recognizable isotopic envelopes to differentiate signal from noise. This task employed a deconvolution algorithm included in the instrument’s data system, which was designed to infer the charge state of any given signal from the respective isotopic distribution. Those signals lacking a discernable charge state were therefore eliminated as a potential hit. When the ESI-MS data shown in Figure 2-8 were processed, this stringent filtering operation returned only 1,206 of the 14,639 entries contained in the initial mass list, which were subsequently employed for database searching. The minimization of data prior to database searching provided increased assurance that the list of obtained hits contained a lower incidence of false positives (see Section 2.2.5). After all redundant entries were eliminated, we ensured that the mass of each PTM appeared in both the nucleoside and nucleotide form to allow for the recognition of all possible products present in the nuclease digests. For each entry, the mass of the deprotonated and protonated species (i.e., [M-H]- and [M+H]+) were calculated to enable proper matching of data obtained in either polarity. The final custom registry included 254 searchable entries.

A total of 268 database hits were obtained when the reduced experimental data were searched against METLIN, whereas 41 were found in the custom registry (Table 2-2). The incidence of the higher number of hits obtained when searching against the METLIN database can be attributed to potential carry-over through extraction steps of metabolites with similar chemical characteristics to PTMs. As such, special attention was focused on pipetting only portions of each desired layer to ensure that other metabolites designed to be partitioned based
<table>
<thead>
<tr>
<th>Exp. mass (u)</th>
<th>Monoisotopic (u)</th>
<th>Hit$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>323.0519</td>
<td>323.05185</td>
<td>C$^+$</td>
</tr>
<tr>
<td>324.0359</td>
<td>324.03587</td>
<td>Y$^+$, U$^+$</td>
</tr>
<tr>
<td>326.0515</td>
<td>326.05152</td>
<td>D$^+$</td>
</tr>
<tr>
<td>337.0675</td>
<td>337.06750</td>
<td>m$^3$C, m$^5$C, Cm$^+$, m$^4$C</td>
</tr>
<tr>
<td>338.0515</td>
<td>338.05152</td>
<td>m$^3$Y$^*$, Um$^+$, m$^5$U, m$^3$Y, Ym$^+$, m$^3$U</td>
</tr>
<tr>
<td>347.0631</td>
<td>347.06308</td>
<td>A$^+$</td>
</tr>
<tr>
<td>348.0471</td>
<td>348.04710</td>
<td>I$^+$</td>
</tr>
<tr>
<td>361.0787</td>
<td>361.07873</td>
<td>m$^3$A, m$^2$A$^<em>$, m$^6$A, m$^8$A$^</em>$, Am$^+$</td>
</tr>
<tr>
<td>363.0580</td>
<td>363.05800</td>
<td>G$^+$</td>
</tr>
<tr>
<td>365.0623</td>
<td>365.06242</td>
<td>ac$^4$C$^<em>$, f$^6$Cm$^</em>$</td>
</tr>
<tr>
<td>375.0942</td>
<td>375.09438</td>
<td>m$^6$Am, m$^1$Am, m$^6$A$^*$</td>
</tr>
<tr>
<td>377.0762</td>
<td>377.07365</td>
<td>m$^1$G$^+$, m$^2$G$^+$, m$^7$G$^+$, Gm$^+$</td>
</tr>
<tr>
<td>379.0762</td>
<td>379.07807</td>
<td>ac$^4$Cm$^*$</td>
</tr>
<tr>
<td>381.0572</td>
<td>381.05733</td>
<td>ncm$^5$U</td>
</tr>
<tr>
<td>391.0893</td>
<td>391.08923</td>
<td>m$^1$Gm$^<em>$, m$^2$G, m$^3$Gm, preQ1, m$^2$G$^</em>$</td>
</tr>
<tr>
<td>427.0429</td>
<td>427.04505</td>
<td>cmnm$^5$s$^2$U$^*$</td>
</tr>
<tr>
<td>492.1005</td>
<td>492.10059</td>
<td>f$^6$A$^+$</td>
</tr>
<tr>
<td>588.1580</td>
<td>588.15811</td>
<td>yW$^+$</td>
</tr>
</tbody>
</table>

Table 2-2. Hits obtained by searching the ESI-MS data in Fig. 2-8 against the non-redundant database generated in house (see Results and Discussion). The experimental mass of the neutral species is expressed in mass units (u). Monoisotopic mass was calculated from the respective elemental composition. Taken with permission from Ref.310.

$^1$ Full names available at [http://mods.rna.albany.edu/](http://mods.rna.albany.edu/)

$^*$ Assignments corroborated by tandem mass spectrometry (i.e., MS$^n$ and CRM determinations).

* Modifications previously unreported in S. cerevisiae.
on their vast differences in chemical properties were not carried through to subsequent extraction steps.

The observed experimental masses matched very closely those found in the databases. The majority of hits provided an average deviation between experimental and calculated masses that fell within the accuracy assessed from canonical ribonucleotides, whereas weaker signals displayed slightly higher deviations. The majority of such hits corresponded to modifications typically observed in \textit{S. cerevisiae}’s ribosomal-RNA (rRNA) or transfer-RNA (tRNA), but 13 of them had not been previously reported for this organism (marked with asterisk in Table 2-2). These unreported PTMs may be attributable to the fact that these experiments did not target specific RNA fractions, but rather the whole transcriptome. Interestingly, with only a few exceptions, the majority of the hits afforded by the custom registry were also found in METLIN. This observation provided an indication of the excellent but not absolute overlap between the databases employed in the study. In particular, the greater breadth afforded by METLIN enabled the putative assignment of other notable but unrelated species of cellular origin mentioned above are marked with a red tick in Figure 2-8, such as UDP-L-arabinose and UDP-D-xylose. Their detection in the sample mixture –an unintended outcome of the broad nature of phenol-chloroform extraction– confirmed the potential for carryovers anticipated for the proposed workflow. Despite the presence of these species, the characterization of low-abundance PTMs did not appear to be hampered.

The m/z values obtained from the IMS-MS determination, which are plotted on the y-axis of the heat-map in Figure 1-3, were submitted to the same data treatment described above, and then used to search against the custom registry. This operation yielded the same database hits
produced by the ESI-MS data. In this case, the experimental values afforded by the canonical ribonucleotides displayed an average ~13 ppm deviation from the theoretical values provided by their elemental composition, which is consistent with the typical accuracy achieved with this type of instrumentation. In analogy with the ESI-MS data displayed in Figure 2-8, carry-over species contributed significantly to the complexity of the observed heat-map, but did not have any adverse effect on PTM detection.

Although the ability to achieve positive corroboration hinged in large part on the unique structural information obtained by gas-phase dissociation techniques, an important contributing factor was identified also in the conservative strategy employed for the initial data reduction. Filtering signals that did not possess recognizable isotopic patterns increased the efficiency of database searches and the effectiveness of subsequent analyses. While it cannot be excluded that this criterion might have caused the occasional rejection of potentially valid information, conservative data reduction minimized the incidence of false positives by removing questionable signals from the initial mass lists. In this way, subsequent analyses targeted only the species that had a legitimate probability of yielding viable fragmentation data for assignment confirmation.

2.4 CONCLUSIONS

With the rare exceptions of 2-O′-methylation, 5′-cytosine, adenosine-N6-methylation, and pseudouridylation, the high-throughput sequencing approaches (e.g., RNA-seq and similar next-generation techniques) that are the pillars of genomics research are incapable of detecting PTMs, owing to the fact that analysis takes place on DNA copies, rather than genuine RNA
samples bearing the PTMs. The lack of sufficient data on PTM expression and distribution has significantly hampered the elucidation of their biological functions. MS-based approaches have the potential to fill this gap by enabling PTM recognition and quantification on the basis of their unique mass and fragmentation signatures. These types of approaches have traditionally relied on liquid chromatography and capillary electrophoresis to reduce chemical background and provide separation before analysis. This chapter has provided the necessary proof of principle for their possible implementation without any high-resolution separation, which will greatly simplify their incorporation in large scale, high-throughput applications. The results reported here demonstrate that this strategy is capable of providing comprehensive surveys of ribonucleotide modifications at the full-transcriptome level.

Direct infusion analysis with either high-resolution MS or IMS-MS detection enabled the positive identification of PTMs in both the mononucleotide standard and the complex *S. cerevisiae* cellular extract based on their individual molecular masses, unique fragmentation patterns, and characteristic conformational features. Eliminating typical front-end chromatographic steps streamlined the operations without affecting detection sensitivity and characterization capabilities. Combining lysis and nucleic acid extraction in a single step led to minimal carryover of cellular components, which did not have any appreciable consequence on the ability to detect modified ribonucleotides. Additionally, the proposed workflow provided comprehensive PTM information by using as little as ~800 µg of wet cell pellet or ~69 µL of culture at 0.3 OD$_{600}$ (corresponding to ~1.2 × 10$^7$ cells). The approaches developed here relied on database searching and gas-phase activation techniques to positively identify the observed PTMs. For this reason, they are not necessarily geared toward the identification of new PTMs.
that are absent from the available databases. Indeed, it is likely that many of the observed signals that did not return hits in our experiments may correspond to yet undiscovered PTMs. Their presence in total cell extracts should not be considered a surprise in light of the almost exclusive emphasis placed by earlier studies on tRNA/rRNA analysis.\textsuperscript{16,21,47,73} This chapter has clearly demonstrated that the LTQ-Orbitrap and IMS-MS platforms, in concert or individually, are capable of providing the information necessary to support full-fledged structural characterization.
Chapter 3

Standard-free quantitation of post-transcriptional modifications in whole cell lysates

3.1 INTRODUCTION

The comprehensive identification of the entire complement of PTMs present in a certain cell can provide valuable insights into its genetic and metabolic state.\textsuperscript{44,77,100,112,263} This type of information, however, would be incomplete without the associated quantitative data necessary to explore the correlation between expression levels and specific experimental variables. Regardless the selected platform, LTQ-Orbitrap or IMS-MS, the analytical workflow must be capable of providing valid information on the abundance of the detected PTMs in order to accurately appreciate their up- or down-regulation. Only in this way, can these approaches be used to address important biological hypotheses regarding the role of PTMs in the cell’s lifecycle and/or amidst stressed or diseased states.

Classic quantitative approaches aimed at handling more complex mixtures of PTMs from biological samples typically require the availability of target analyte in neat form to generate a calibration curve through serial dilutions, or to perform incremental additions to the original sample according to the standard additions method.\textsuperscript{264–267} Unfortunately, the large number of natural PTMs and the endemic shortage of appropriate standards represent major obstacles to the broad implementation of these quantitative strategies, especially at the whole-transcriptome
level. In most cases, securing PTMs in neat form involves multistep synthetic procedures that are both lengthy and low-yield. These considerations provided strong motivations to our exploration of alternative quantitative strategies that did not require synthetic standards. This chapter describes two novel strategies for the determination of absolute and relative abundances of PTMs in whole-cell RNA extracts. The former employs purified tRNA samples as intrinsic sources of PTMs in standard-addition determinations, whereas the latter relies on the canonical ribonucleotides, the abundances of which are under direct genetic control, as intrinsic internal standards. The utilization of heterologous tRNAs was investigated by adding accurately known amounts to the cell pellets prior to cell lyses. In this way, the heterologous tRNA was processed through the entire workflow together with the actual sample and released all its PTMs at once during the RNase digestion step (Scheme 2-2). Samples containing incremental amounts of tRNA were subsequently used to generate the signal/concentration curves necessary to determine unknown concentrations of select PTMs in the whole cell lysate. The utilization of canonical ribonucleotides as intrinsic standards was realized by combining their respective signal intensities into a proxy “base peak” that provided the basis for expressing the abundance of individual PTMs in percentage terms. This strategy proved to be very well suited for comparative studies in which the ability to monitor over-/under-expression of PTMs was more important than determining their actual absolute abundances.
3.2 MATERIAL AND METHODS

3.2.1 Preparation of cellular extracts.

*Saccharomyces cerevisiae* strain BY4741 was grown in yeast extract, peptone, dextrose (YPD) and synthetic complete (SC) media. Cell suspensions were streaked onto YPD agar plates and incubated at 30°C overnight. Five individual colonies were selected from each plate and placed into individual tubes containing 6 mL of either YPD or SC medium. All growth tubes were incubated at 30°C with 200 rpm gyration. Optical density at 600 nm (OD$_{600}$) was monitored on a ThermoFisher Scientific (Waltham, MA) Nanodrop 2000c spectrophotometer until a value greater than 0.3 units was achieved. Each liquid culture was diluted to a final 0.3 OD$_{600}$. A 3-mL aliquot of each culture was centrifuged at 6000 g for 5 min. to obtain pellets that contained approximately the same number of cells. *Escherichia coli* K-12 strain MG1655 was grown in synthetic complete (SC) medium according to established procedures.$^{230,268}$ Harvesting was carried out in analogous way.

Each pellet was disrupted by using Denaturation Solution (Life Technologies, Green Island, NY). When required by the standard-additions protocol, accurately known aliquots of *S. cerevisiae* tRNA$^{\text{phe}}$ (Sigma-Aldrich, St. Louis, MO) were introduced at this point into the lysate to serve as internal standard. Total RNA was extracted by using the ToTALLY RNA Extraction Kit (Life Technologies, Green Island, NY), which is based on a typical phenol/chloroform procedure. The RNA was precipitated by using cold isopropanol and then treated with DNase 1 (New England Biolabs, Ipswich, MA) in 1X DNase buffer to remove any remaining DNA. The recovered RNA was subsequently desalted by ethanol precipitation overnight and reconstituted in 50 µL of RNase-free water (Sigma-Aldrich, St. Louis, MO). The concentration of intact total RNA from each sample
was measured by UV absorbance at 260 nm. Nuclease P1 and phosphodiesterase 1 from snake venom (Sigma-Aldrich, St. Louis, MO) were employed to complete the digestion of RNA into individual mononucleotides, as previously described.\textsuperscript{232} Immediately before analysis, final samples were diluted 1:10 in 150 mM ammonium acetate and 10% isopropanol. The MS analyses and data treatment were performed as described in Sections 2.2.3, 2.2.4 and 2.2.5.

### 3.3 RESULTS AND DISCUSSION

#### 3.3.1 Relationship between nucleotide concentration and signal intensity.

The tenet of quantitative determinations is the proportionality between analyte concentration and signal intensity. In the case of RNA units, the ionization efficiency is critical to this proportionality. Over the years, this issue have been extensively investigated. Therefore, in order to assess the quantitative capabilities of the technique, we initially examined the signals afforded by the various nucleotides present in the standard equimolar mixture of NMPs (discussed extensively in the previous chapter). We observed that they produced markedly different intensities despite the fact that they were present in the same amount (Figure 2-1a). As a first approximation, their relative intensities exhibited trends that matched more closely their gas-phase proton affinities (i.e., AMP > GMP > CMP > UMP > TMP\textsuperscript{255,269}), than the reverse order of solution pK\textsubscript{a} (i.e., TMP > UMP > CMP > AMP > GMP\textsuperscript{270}). However, the same relative scale was observed when the analysis was performed in positive ion mode to detect the corresponding protonated species, which raised doubts about whether there might be a direct relationship between signal response and acid-base properties. The detection of positively charged ions from very acidic analytes, under conditions that should favor the formation of deprotonated rather
than protonated species, represents a classic example of “wrong-way-round” electrospray.\textsuperscript{271,272} This phenomenon has been explained with the prominent role played by analyte hydrophobicity in determining the composition of electrosprayed droplets and, consequently, overall signal response.\textsuperscript{273,274} In the case of nucleic acid components, an equimolar mixture of 5\textsuperscript{'-}triphospho-2\textsuperscript{'-}deoxynucleosides (dNTPs) was reported to provide a dTTP > dATP > dCTP > dGTP scale of relative intensities, which matched the order of hydrophobicities exhibited by the respective nucleobases.\textsuperscript{275} In our analyses, a mixture of 5\textsuperscript{'-}monophospho-2\textsuperscript{'-}deoxynucleosides (dNMPs) displayed a dGMP>dUMP>dCMP>dTMP scale of relative intensities, which confirmed the data in Figure 2-1a without the possible complications introduced by the dGMP/AMP isotopic overlap. The same results were obtained when the initial solvent, a 100 mM solution of ammonium acetate with 10\% volume of 2-propanol (see Section 2.2.1), was replaced with the significantly more hydrophobic system employed in the earlier study (i.e., a 20\% aqueous solution with 60\% acetonitrile and 20\% 2-propanol),\textsuperscript{275} thus suggesting that the discrepancy may be ascribable to more subtle differences between 5\textsuperscript{'-}triphospho- and 5\textsuperscript{'-}monophospho-nucleosides, which will require further investigation.

Additionally, the limit of detection of the nucleotide mixture observed on the Orbitrap, discussed in Chapter 2, was quantified by performing serial dilutions of the initial sample. In this case, a 200 pM solution of total nucleotide mixture displayed an approximately 5:1 signal-to-noise ratio (S/N) for deprotonated dTMP, the weakest of the signals in Figure 2-1a, which was employed here as a benchmark. As a reference, it is important to note that signal averaging was completed by accumulating 16 Orbitrap acquisitions over a total span of 30 s. Considering that static nano-ESI involves typical flow rates in the order of \textasciitilde20nL/min,\textsuperscript{276} it could be estimated that
this determination required the consumption of approximately $2 \times 10^{-18}$ mol (attomol) of material. When analysis was performed in positive ion mode, an approximately 30% decrease of the overall ion current was noted across the board. The sensitivity for dTMP in the mixture was assessed also by consecutive reaction monitoring (CRM).\textsuperscript{277} The determination followed the cleavage of the N-glycosidic bond upon gas-phase activation, which results in nucleobase elimination and production of a 5’phospho-2’-deoxyribose fragment (vide infra). This approach allowed the detection of target analyte in a ~5 nM solution of total nucleotide mixture, with an overall sample consumption of $75 \times 10^{-18}$ mol.

### 3.3.2 Absolute assessment of PTM levels.

When the proportionality between analyte concentration and signal intensity cannot be determined a priori, different strategies can be applied, which can be grouped into two distinctive categories: one that relies on building appropriate calibration curves, the other involves the addition of standard to the initial sample. Both strategies require the availability of neat compound to be weighed and used as standard, which is not always possible for all the more than 100 known RNA variants. In classic standard-addition strategies, the standard consists of the actual analyte in neat form, which can be weighed (or otherwise aliquoted) and added in accurately known amounts to the sample of interest. The signals recorded before and after addition are then employed to generate curves that are extrapolated to obtain the unknown amount in the initial sample.\textsuperscript{278} In the absence of neat-form standards for all the PTMs that may be present in a whole RNA extract, we decided to explore the utilization of various tRNAs as convenient sources. The proposed strategy involved adding known amounts of a tRNA standard
to the initial sample and then completing all the steps included in the process workflow (Scheme 2-2).\textsuperscript{279,280} In this way, accurately known amounts of PTM standards are released directly in situ during the RNase digestion step, thus supporting the subsequent mathematical treatment.

The outcome of this strategy hinges on two essential conditions. The first is that the standard tRNA must undergo with minimal losses and no bias the preliminary steps of cell lysis, nucleic acid extraction, and DNA elimination (Scheme 2-2). The second is that RNase digestion must be complete to support the assumption that the amount of each PTM released in solution corresponded to that of the initial amount of tRNA standard. We investigated these conditions by utilizing \textit{S. cerevisiae} tRNA\textsuperscript{Phe} as a known source of selected PTMs (Figure 3-1). For instance, we processed in parallel a known aliquot of tRNA\textsuperscript{Phe}, a weighed amount of \textit{S. cerevisiae}'s pellet, and a mixture of the two. After each step, the recovery was assessed by monitoring the concentration of RNA by UV absorption. In all cases, an average \textasciitilde 25\% recovery was readily observed through the workflow procedures preceding RNase digestion. The fact that each sample showed comparable decreases of RNA concentration demonstrated that the experimental procedures induced predictable losses across the board and did not introduce any detectable bias against the added tRNA\textsuperscript{Phe}. The second condition was verified by monitoring the products of tRNA\textsuperscript{Phe} digestion by ESI-MS analysis. The spectrum provided in Figure 3-2 displayed signals for the entire complement of PTMs represented in this type of tRNA, with no trace of undigested substrate, thus indicating that the covalent modifications did not hamper nuclease activity.

This analysis was repeated on samples that contained decreasing concentrations of tRNA\textsuperscript{Phe} to evaluate the limit of detection of the proposed workflow. The results revealed that,
Figure 3-1. tRNA^{phe} from *S. cerevisiae*. 
on average, the various PTMs required a sample consumption in the low amol range (e.g., ~50 amol) to produce a 3:1 or better signal-to-noise ratio (Table 3-1). In addition to putative detection limits, this exercise enabled us to assess the various ionization energies of the different PTMs present. As such, individual signal/concentration curves were obtained and provided the signal response for each ribonucleotide in tRNA\textsuperscript{Phe} (Table 3-1). Given that each PTM had measurable unique responses in the gas phase, these response curves were implemented to facilitate accurate calculations of the absolute concentration of each PTM within the \textit{S. cerevisiae} whole cell extract.

In this particular case, tRNA\textsuperscript{Phe} found endogenous in \textit{S. cerevisiae} was added in known incremental amounts to an \textit{S. cerevisiae} extract containing the entire transcriptome and together, underwent the entire workflow. UV determinations were performed and confirmed the ~25% recovery observed for the isolated standard. The excellent match between recoveries observed in the absence/presence of cell material indicated that lysis debris did not significantly interfere with phenol-chloroform extraction and subsequent workflow operations. The data obtained from the standard-addition series were used to generate the curves necessary to complete the quantitative determination of the 41 PTMs found in the \textit{S. cerevisiae} investigated in Chapter 2 (Table 2-2). The absence of accurate estimates of cellular volumes precluded a correct translation of extract concentrations into actual cellular concentrations. For this reason, the total amounts of PTMs in the sample were more conveniently expressed in terms of mol per gram of wet pellet (mol/g, Table 3-2), which were based on the weight of initial cell material employed in the determination. For conversion purposes, we estimated that 1 g of wet pellet corresponds to ~86 µL of a culture suspension with 0.3 OD\textsubscript{600}. The results clearly displayed the
Figure 3-2. ESI-MS spectrum of digestion mixture obtained from S. cerevisiae tRNA\textsuperscript{Phe}. The enlargement shows the region containing the majority of the PTMs. Signals marked with * are hits from our custom modifications registry; o species detected also in the blank. No undigested tRNA\textsuperscript{Phe} was observed in the high m/z range. Taken with permission from Ref. 310.
typical gulf between abundant canonic ribonucleotides and low-abundance modifications representing the bulk of the observed analytes, which showcased the excellent dynamic range afforded by this approach. In the context of the detection limits obtained from isolated tRNA\(^{\text{Phe}}\) (Table 3-1), the observed values indicated that valid determinations could be comfortably accomplished for even the least abundant modifications (i.e., \(\text{ac}^4\text{Cm}\) and \(\text{cmnm}^5\text{s}^2\text{U}\)) with as little as 800 \(\mu\text{g}\) of wet pellet (\(\sim 69\ \mu\text{L}\) of a culture suspension with 0.3 \(\text{OD}_{600}\)). It should be noted that, although this determination covered only a subset of the entire complement of cellular PTMs – those represented in tRNA\(^{\text{Phe}}\) – the utilization of different tRNAs or other controlled sources of natural PTMs could extend the coverage to virtually any type of modified ribonucleotides, thus making this strategy viable for a wide range of possible applications.

### 3.3.3 Relative assessment of PTM levels

The proposed strategy can circumvent but not eliminate the hurdles associated with the significant lack of suitable standards for rigorous quantitative determinations. In many cases, however, obtaining the absolute amount of a given PTM is not as crucial as monitoring its relative abundance versus others in the sample. In proteomics studies, for example, the ability to appreciate mutual variations of post-translational modifications – reliable indicators of up- and down regulation – is at least as valuable as the ability to determine their absolute levels. For this reason, we explored the possibility of utilizing the four canonical ribonucleotides, whose overall amounts and distributions are dependent on the cell’s genetic makeup, as a convenient internal reference to observe relative variations within a given organism. More specifically, we combined their signal intensities to establish a multicomponent reference, which could fit the MS definition
<table>
<thead>
<tr>
<th>Name</th>
<th>Exp. mass (u)</th>
<th>Equivalent per mole</th>
<th>Detection limit (mol)</th>
<th>Response (m, q)</th>
<th>Theoretical AvP (%)</th>
<th>Exp. AvP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>323.0519</td>
<td>15</td>
<td>3.44 \times 10^{-17}</td>
<td>1.61 \times 10^{11}, 5.01 \times 10^{10}</td>
<td>24.7</td>
<td>24.6</td>
</tr>
<tr>
<td>U/Ψ</td>
<td>324.0279</td>
<td>14</td>
<td>2.68 \times 10^{-17}</td>
<td>1.53 \times 10^{11}, 2.03 \times 10^{11}</td>
<td>21.9</td>
<td>21.6</td>
</tr>
<tr>
<td>D</td>
<td>326.0514</td>
<td>2</td>
<td>1.80 \times 10^{-17}</td>
<td>1.82 \times 10^{11}, 3.01 \times 10^{11}</td>
<td>3.73</td>
<td>3.75</td>
</tr>
<tr>
<td>m^3C, m^3C, Cm</td>
<td>337.0674</td>
<td>3</td>
<td>5.50 \times 10^{-17}</td>
<td>1.41 \times 10^{11}, 8.00 \times 10^{11}</td>
<td>4.33</td>
<td>4.34</td>
</tr>
<tr>
<td>A</td>
<td>347.0630</td>
<td>17</td>
<td>9.58 \times 10^{-17}</td>
<td>1.31 \times 10^{11}, 9.20 \times 10^{10}</td>
<td>22.8</td>
<td>22.1</td>
</tr>
<tr>
<td>mA</td>
<td>361.0787</td>
<td>1</td>
<td>9.29 \times 10^{-17}</td>
<td>3.52 \times 10^{10}, 3.03 \times 10^{11}</td>
<td>3.60 \times 10^{-1}</td>
<td>3.75 \times 10^{-1}</td>
</tr>
<tr>
<td>G</td>
<td>363.0579</td>
<td>18</td>
<td>1.55 \times 10^{-17}</td>
<td>1.71 \times 10^{11}, 3.78 \times 10^{11}</td>
<td>31.5</td>
<td>31.7</td>
</tr>
<tr>
<td>m^2G, Gm</td>
<td>377.0718</td>
<td>3</td>
<td>5.29 \times 10^{-18}</td>
<td>1.24 \times 10^{11}, 5.88 \times 10^{11}</td>
<td>3.81</td>
<td>3.61</td>
</tr>
<tr>
<td>m^2G</td>
<td>391.0892</td>
<td>1</td>
<td>2.44 \times 10^{-17}</td>
<td>1.71 \times 10^{11}, 2.07 \times 10^{11}</td>
<td>1.75</td>
<td>1.60</td>
</tr>
<tr>
<td>yW</td>
<td>588.1571</td>
<td>1</td>
<td>4.14 \times 10^{-17}</td>
<td>1.10 \times 10^{11}, 1.71 \times 10^{11}</td>
<td>1.13</td>
<td>1.16</td>
</tr>
</tbody>
</table>

**Table 3-1.** Figures of merit obtained by analyzing isolated tRNA^{Phe} from *S. cerevisiae*. The abbreviation for each ribonucleotide is provided together with the corresponding monoisotopic mass in Da and the number of equivalents present in each mole of initial tRNA^{Phe}. Taken with permission from Ref. 310.

1 The limit of detection (LOD) was obtained by calculating the moles of each ribonucleotide, which provided at least a 3:1 signal to noise ratio. The results are the average of five repeat determinations. The calculation accounted for an average of ~0.093 μL sample consumption during ESI-MS analysis and a ~25% sample recovery for the entire work flow (see Experimental).

2 The response for each ribonucleotide was calculated by averaging the signals of five repeat determinations for each decreasing concentration. Each signal average was plotted against the respective concentration to obtain signal/concentration curves with the indicated slopes (m in counts/M) and intercepts (q in counts).

3 For each ribonucleotide, the value of abundance versus proxy (AvP) was calculated by dividing the respective intensity by the sum of the intensities of the four canonic ribonucleotides (see Experimental). The experimental AvP was calculated directly from the ESI-MS data. In contrast, the theoretical value was obtained from the intensity that would be expected from the analysis of exactly 1 M of tRNA^{Phe}, which was calculated by substituting the equivalents per mole of each species into the respective response curve. The excellent match between theoretical and experimental values provides support for the utilization of AvP to monitor fluctuations of PTM expression (see Results and Discussion).
of a proxy base-peak, for quantifying the various PTMs in terms of abundance versus proxy (i.e., AvP, see Materials and Methods). As shown in Table 3-2, this figure was readily attainable for all PTMs in the extract, regardless of their representation in a putative standard, thus providing a self-consistent and comprehensive measure of their relative abundances in the sample. Careful attention was taken when analyzing extracted samples as to ensure that the upper maximum of detection was not reached by the canonical bases. If the canonical bases had reached this upper limit, the trapping instrument would have shown a steady saturated signal for each base and skew the overall AvPs calculated. Thus, experiments were performed to determine the upper limit and stay well within a reasonable range of detection for this particular instrument. Ultimately, we evaluated the effectiveness of this approach by comparing experimental AvP values obtained from isolated tRNA\textsuperscript{Phe} with putative figures calculated for a fixed concentration by using the respective signal/concentration curves (Table 3-1). The excellent match between corresponding values provided the justification for a broader application of this treatment to monitor the relative variations of PTM levels in actual cell material.

### 3.3.4 Reproducibility of PTM analysis.

In order to completely validate this approach as an effective tool for monitoring the global profiles of PTMs in whole cell lysates, the reproducibility of these analyses needed to be assessed. Therefore, separate aliquots of the same *S. cerevisiae* pellet were submitted in parallel to the entire workflow to monitor the technical reproducibility. The analysis of the five individual samples consistently produced the same 41 hits seen in Table 2-2. Their relative abundances were expressed in AvP units to enable direct comparisons of their distributions (Table 3-3).
### Table 3-2
Quantitative determination of ribonucleotides present in total RNA extract of *S. cerevisiae*. This standard-additions determination used tRNA^Phe^ purified from *S. cerevisiae* to achieve in situ release of PTM standards (see Materials and Methods). Name abbreviation and neutral experimental mass in mass units (u) are provided for each ribonucleotide. Taken with permission from Ref. 310.

1. Full names available at [http://mods.rna.albany.edu/](http://mods.rna.albany.edu/)

2. The concentration of each PTM in the extract was calculated from the respective curve afforded by the standard-additions determination. This figure accounts also for the ~25% recovery estimated from the standard tRNA^Phe^ added to each sample. NA indicates PTMs that could not be determined due to their absence in the standard tRNA^Phe^.

3. The amount of each PTM per gram of wet pellet was calculated from the respective extract concentration by taking into account the initial weight of intact *S. cerevisiae* material.

4. For each PTM, the value of abundance versus proxy (AvP) was calculated from the respective signal intensity as percentage of the sum of the intensities of the four canonic ribonucleotides (see Materials and Methods). Each value was the average of five repeat analyses. This figure represents a relative measure of the abundance of each PTM in the sample, which can be always calculated across the board in the absence of PTM standards (see Results and Discussion).

<table>
<thead>
<tr>
<th>Hit</th>
<th>Exp. mass (u)</th>
<th>Conc. (M)</th>
<th>Amount (mol/g)</th>
<th>Exp. AvP</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>323.0519</td>
<td>2.60 x10^-5</td>
<td>1.25 x10^-7</td>
<td>28.7</td>
</tr>
<tr>
<td>U/P</td>
<td>324.0359</td>
<td>2.26 x10^-5</td>
<td>1.09 x10^-7</td>
<td>22.0</td>
</tr>
<tr>
<td>D</td>
<td>326.0515</td>
<td>2.40 x10^-6</td>
<td>1.16 x10^-8</td>
<td>2.95</td>
</tr>
<tr>
<td>m^4C, m^5C, Cm, m^4C</td>
<td>337.0675</td>
<td>2.48 x10^-7</td>
<td>1.19 x10^-9</td>
<td>2.45 x10^-1</td>
</tr>
<tr>
<td>m^3Y, Um, m^5U, m^3Y, Ym, m^3U</td>
<td>338.0515</td>
<td>NA</td>
<td>NA</td>
<td>1.03</td>
</tr>
<tr>
<td>A</td>
<td>347.0631</td>
<td>2.41 x10^-5</td>
<td>1.16 x10^-7</td>
<td>21.6</td>
</tr>
<tr>
<td>I</td>
<td>348.0471</td>
<td>NA</td>
<td>NA</td>
<td>1.44 x10^-4</td>
</tr>
<tr>
<td>m^1A, m^2A, m^6A, Am</td>
<td>361.0787</td>
<td>4.40 x10^-7</td>
<td>4.40 x10^-9</td>
<td>1.02 x10^-1</td>
</tr>
<tr>
<td>G</td>
<td>363.0580</td>
<td>2.42 x10^-5</td>
<td>1.17 x10^-7</td>
<td>27.7</td>
</tr>
<tr>
<td>ac^4C, f^3Cm</td>
<td>365.0623</td>
<td>NA</td>
<td>NA</td>
<td>4.75 x10^-4</td>
</tr>
<tr>
<td>m6Am, m1Am, m62A</td>
<td>375.0942</td>
<td>NA</td>
<td>NA</td>
<td>5.01 x10^-3</td>
</tr>
<tr>
<td>m^2G, m^2G, m^3G, Gm</td>
<td>377.0762</td>
<td>1.32 x10^-7</td>
<td>6.34 x10^-10</td>
<td>1.02</td>
</tr>
<tr>
<td>ac^4Cm</td>
<td>379.0762</td>
<td>NA</td>
<td>NA</td>
<td>6.85 x10^-4</td>
</tr>
<tr>
<td>ncm5U</td>
<td>381.0572</td>
<td>NA</td>
<td>NA</td>
<td>3.37 x10^-2</td>
</tr>
<tr>
<td>m^1Gm, m^2G, m^3Gm, preQ1, m^2G</td>
<td>391.0893</td>
<td>5.64 x10^-7</td>
<td>2.72 x10^-9</td>
<td>6.26 x10^-1</td>
</tr>
<tr>
<td>cmn^3s^2U</td>
<td>427.0429</td>
<td>NA</td>
<td>NA</td>
<td>1.02 x10^-2</td>
</tr>
<tr>
<td>t^6A</td>
<td>492.1005</td>
<td>NA</td>
<td>NA</td>
<td>1.34 x10^-1</td>
</tr>
<tr>
<td>γW</td>
<td>588.1580</td>
<td>2.50 x10^-7</td>
<td>1.20 x10^-9</td>
<td>1.84 x10^-1</td>
</tr>
</tbody>
</table>
Overall, the values displayed an average of ±4.4% relative standard deviations (RSD%) for all the PTMs, which offered a measure of the reproducibility of these determinations. Not surprisingly, the species at the higher end of the AvP scale displayed better reproducibility (i.e., smaller RSD% values) than those at the lower end, owing to the greater susceptibility of the latter to possible fluctuations of experimental conditions throughout the workflow. For comparison purposes, the reproducibility of the ESI-MS analysis itself was evaluated separately by repeating the determination of the same digestion mixture for a total of five times. The results provided an average RSD% of ±1.6% calculated from all the PTMs in the sample, thus suggesting that workup operations, such as extraction/lysis, digestion, etc., contributed to the majority of the overall ±4.4% uncertainty intrinsic in these determinations. It should be noted that in general the observed reproducibility benefited significantly from the utilization of relative rather than absolute notations. Indeed, any experimental inconsistency affecting detection is typically expected to influence analyte and reference in the same direction, when both undergo simultaneously the same procedure. When abundances are expressed in relation to the reference, these effects tend to cancel out, thus minimizing the impact of analytical fluctuations. This explains the observation that RSD% obtained directly from ion counts (absolute notation) were distinctively larger than those calculated from the corresponding AvPs (relative notation, Table 3-3).

In order to assess the sample-to-sample variability (i.e., biological reproducibility) against the observed technical reproducibility, we performed parallel analyses of individual samples grown in separate cultures under otherwise identical conditions (see Materials and Methods). These experiments also produced consistently the same database hits obtained from the
### Table 3-3. Reproducibility of ESI-MS analysis of total RNA extract from *S. cerevisiae* grown in YPD medium. Taken with permission from Ref. 310.

1 Full names available at [http://mods.rna.albany.edu/](http://mods.rna.albany.edu/)

2 Assessed by applying the proposed workflow to five separate aliquots of the same *S. cerevisiae* pellet. For each PTM, abundance versus proxy (AvP) was calculated from the respective signal intensity as percentage of the sum of the intensities of the four canonical ribonucleotides (see *Materials and Methods*). Average and relative standard deviation (RSD%) are reported.

3 Assessed from five different samples of *S. cerevisiae* grown under identical conditions in separate YPD cultures.

<table>
<thead>
<tr>
<th>Hit 1</th>
<th>Exp. mass (u)</th>
<th>Ave. AvP</th>
<th>AvP RSD%</th>
<th>Ave. AvP</th>
<th>AvP RSD%</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>323.0519</td>
<td>28.7 ±5.0</td>
<td>29.1 ±4.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y, U</td>
<td>324.0359</td>
<td>22.0 ±4.1</td>
<td>22.1 ±9.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>326.0515</td>
<td>2.95 ±3.7</td>
<td>1.18 ±14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m(^3)C, m(^3)C, Cm, m(^3)C</td>
<td>337.0675</td>
<td>2.45 x10(^{-1}) ±5.5</td>
<td>7.04 x10(^{-1}) ±7.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m(^3)Y, Um, m(^3)U, m(^3)Y, Ym, m(^3)U</td>
<td>338.0515</td>
<td>1.03 ±3.6</td>
<td>7.28 x10(^{-1}) ±5.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>347.0631</td>
<td>21.6 ±4.7</td>
<td>21.3 ±6.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>348.0471</td>
<td>1.44 x10(^{-1}) ±3.8</td>
<td>7.31 x10(^{-1}) ±6.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m(^3)A, m(^2)A, m(^6)A, m(^5)A, Am</td>
<td>361.0787</td>
<td>1.02 x10(^{-1}) ±6.9</td>
<td>4.53 x10(^{-1}) ±11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>363.0580</td>
<td>27.7 ±2.6</td>
<td>27.4 ±2.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ac(^4)C, f(^6)Cm</td>
<td>365.0623</td>
<td>4.75 x10(^{-1}) ±2.2</td>
<td>4.87 x10(^{-1}) ±2.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m(^6)Am, m(^3)Am, m(^5)A</td>
<td>375.0942</td>
<td>5.01 x10(^{-3}) ±9.9</td>
<td>5.15 x10(^{-2}) ±11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m(^1)G, m(^2)G, m(^3)G, Gm</td>
<td>377.0762</td>
<td>1.02 ±2.5</td>
<td>7.52 x10(^{-1}) ±4.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ac(^4)Cm</td>
<td>379.0762</td>
<td>6.85 x10(^{-3}) ±4.0</td>
<td>7.34 x10(^{-3}) ±5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ncnn(^5)U</td>
<td>381.0572</td>
<td>3.37 x10(^{-2}) ±6.3</td>
<td>2.76 x10(^{-2}) ±13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m(^1)Gm, m(^2)G, m(^3)Gm, preQ1, m(^2)G</td>
<td>391.0893</td>
<td>6.26 x10(^{-1}) ±4.9</td>
<td>2.25 x10(^{-1}) ±9.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cmnn(^5)s(^4)U</td>
<td>427.0429</td>
<td>1.02 x10(^{-1}) ±1.2</td>
<td>4.62 x10(^{-3}) ±6.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(^4)A</td>
<td>492.1005</td>
<td>1.34 x10(^{-1}) ±6.5</td>
<td>5.89 x10(^{-2}) ±8.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yW</td>
<td>588.1580</td>
<td>1.84 x10(^{-1}) ±1.3</td>
<td>1.86 x10(^{-2}) ±9.7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ave. ±4.4% Ave. ±7.8%
technical repeats, however their relative abundances displayed an average RSD% of ±7.8 (Table 3-3). At least at first sight, sample-to-sample fluctuations are typically ascribable to variations of total RNA in each sample. However, great effort was placed into growing the cultures in parallel under identical conditions, harvesting them at the same growth phase, and diluting the culture before aliquoting to approximate the same number of cells per sample. In addition, a closer look at the results revealed that the PTMs manifested widely different fluctuation levels from one another (e.g., compare ±14% for D with ±2.9% for ac4C/f5Cm). Any variation of overall RNA content would be expected to affect all PTMs in the same direction, leading to comparable swings. Therefore, these considerations ruled out possible variations of total RNA as a source of uncertainty and suggested the influence of uncontrolled experimental variables that will warrant further investigation. When evaluating the uncertainty intrinsic in these determinations, it is important to recognize that the observed biological reproducibility of ±7.8% included also the ±4.4% contribution of the underlying technical reproducibility present in every determination. Taken together, these figures provided a measure of the typical range within which the incidence of PTMs may vary sample-to-sample under strictly controlled conditions, which is necessary in order to recognize with confidence whether a certain variation is significant and may be unambiguously attributed to actual biological factors rather than mere sample variability.

3.3.5 Addressing the significance of quantitative determinations.

In order to fully comprehend the magnitude by which any fluctuation in a given data set is relevant, rigorous statistical testing must be applied to support the findings. Given that the deviation associated with the average AvP measurements taken for each PTM was consistently
averaged as ±7.8%, the deviation that swung outside of this measurement was deemed as a significant change due to a perturbation in the experimental protocol in the initial findings. While this strategy provided a reasonable assessment of variation between the aforementioned samples, this logic could be deemed as rudimentary. Given that the assessment of PTMs to monitor the entire transcriptome is a relatively new concept, the inception of the statistical testing needed to support its findings is also primitive. However, extensive proteomics and genomics studies have laid the ground work for the most appropriate ways in which to handle large data sets given their advancements in analytical techniques yielding extreme amounts of data.281,282

In the case of these “omics” platforms, analyses are large-scale and are aimed at discerning the structure and function of individual components within the context of a cellular system. This type of approach is no different than the approach used herein to look at the PTMs present in the transcriptome of any cell investigated. In any case, data sets collected across various samples undoubtedly need to be acquired under the same experimental conditions.283 Additionally, in order to find the underlying fluctuations associated with any biological effect, the data sets need to be normalized to one another. Normalization is capable of removing aberrant signals that can result from intrinsic experimental variation which can be caused from sample processing and/or differences in experimental runs. In the case of the PTMs investigated here, the AvP values were calculated to provide an overall normalized value for each PTM. It was assumed that the canonical content present in the cell would remain relatively unchanged despite the modification content observed. This hypothesis was realized in the case of the tRNAphe described in Chapter 1. However, this one tRNA does not account for the multitudes of
PTMs that decorate the transcriptome. While the values for theoretical AvPs matched closely those values calculated experimentally for this analyses, this may not always be the case. For instance, the intrinsic ionization energy associated with a single methylated specie is sure to be different than a hypermodified RNA like Ar(p). Therefore, this study ultimately will need to investigate the gas phase properties of all PTMs to conclusively comment on the gas phase responses of each.

Additionally, it is important to discuss the variation of each sample in its context. Instead of only assigning significance based on the value of ±7.8% RSD, it is more appropriate to correlate this value to a p-value for each PTM investigated. The p-value takes into account the mean difference, the sample size and the variance. As was mentioned before, using the RSD only took into account the variance, not the entirety of the population. P-values give the investigator a measure of the probability of obtaining a significant result if no real difference in samples actually existed. Typically, the cut-off value for the probability is 5%. This means that there is a 5% chance that the result is a false positive. Therefore, if a p-value is calculated that is less than 0.05, then the result investigated is deemed significant. Future studies will indeed implement the Student’s t-test in order to assign whether or not two data sets are actually significantly different from each other based on p-value outcome. Thus, credence will be added to assigning significance based on the ±7.8% RSD variation reported.
3.4 CONCLUSIONS

The utilization of isolated tRNA standard provided an excellent avenue for accomplishing accurate quantification in the absence of pure stocks of ribonucleotide variants. Reinterpreting a classic standard-additions strategy, purified tRNA from commercial sources was added to total RNA extracts immediately before ribonuclease digestion, which enabled the in situ release of accurately known amounts of specific PTMs. In this way, proper signal-concentration curves were obtained in parallel for all the PTMs in the standard, thus enabling their multiplexed determination in the total ribonucleotide mixture. Further, we evaluated also the possibility of utilizing the endogenous canonic ribonucleotides as a proxy internal reference. This approach allowed us to determine the relative abundance of all PTMs with no addition of individual standards. The fact that the results matched the quantitative data from standard-additions determinations provided validation and enabled us to use AvPs to accurately monitor changes of expression levels across multiple samples. The results demonstrated also that the typical technical reproducibility (i.e., sample to sample of the same culture) exceeded significantly the biological one (i.e., culture to culture), thus substantiating the robustness of the proposed workflow. In addition, the observed deviations served to draw well-defined boundaries for deciding whether any fluctuation might be simply ascribable to experimental inconsistencies, or assumed legitimate biological significance.

The heat-maps afforded by IMS-MS analysis clearly substantiated the possibility of visualizing in a very direct and compact format the full complement of PTMs produced by a cell, which will be expected to promote large scale comparative studies of complete epitranscriptomes. The unique features identified by dispersing the signals on the t\(_{\delta}\) and m/z
dimensions can lead to an immediate appreciation of qualitative variations between the types of
PTMs in different samples. The ability to complete direct data subtraction offers the opportunity
to detect more subtle variations of expression levels manifested by common PTMs. The
possibility to observe concomitant variations of all modifications in comprehensive and self-
consistent fashion will enable the investigation of their functional relationships at the system
biology level. In particular, it will be possible to investigate the complex network of metabolic
interactions that may lead to up- or down regulation of specific PTMs as a function of
environmental conditions, cellular states, or cell types. Taking advantage of these capabilities,
exploration into whether RNA modification profiles may faithfully reflect the difference between
healthy and diseased cells, or between distinctive epigenetic states could be explored. The results
will provide new insights into the biological significance of ribonucleotide modifications and their
roles in the cell lifecycle. However, the next chapter will focus on the application of this MS
platform described extensively in Chapter 2 and 3 to address stress response in S. cerevisiae.
Chapter 4

Addressing essential biological questions by global PTM profiling

4.1 INTRODUCTION

The viability of a cellular system revolves around its ability to recognize and respond to any changes in environmental conditions. This is accomplished by cellular activation of compensatory responses intended to maintain growth and ensure survival in response to the external stimuli. Cellular adaptation in response to stressors, such as sudden variations of temperature, osmotic conditions, and availability of nutrients, produces significant variations of enzyme activity and metabolism, which are orchestrated by well calibrated regulatory controls. Owing to its multifaceted activities in protein synthesis and newly discovered functions in gene regulation, RNA is uniquely positioned among cellular components to act as a communication node between metabolic and regulatory mechanisms of stress response. Extensive work has aimed at addressing the biological significance of PTMs within various cellular systems. These variants of RNA are generated by specific biogenetic enzymes and possess the ability to stabilize single base pairs and alternative hydrogen bonding patterns, which contribute to the vast diversity of RNA structure. At the same time, the observation that unmodified tRNAs were not loaded efficiently with the respective amino acyl groups, whereas specific modifications prevented incorrect loading altogether, revealed that PTMs play essential roles in protein-RNA recognition. Analogous effects were observed in RNA-RNA recognition for
hypermodified nucleotides in the tRNA anticodon loop, which were shown to induce translational recoding by affecting the accuracy of codon-anticodon interactions. Additionally, an increased incidence of modification upon treatment with hydrogen peroxide indicated that modification of anticodon loop and translational recoding were essential components of a tRNA-based mechanism of response to oxidative stress.

Owing to the almost exclusive emphasis on tRNA and rRNA as abundant sources of PTMs, little else is known about their biological functions and distribution in other classes of RNAs. However, their enzymatic biogenesis supports the hypothesis that PTMs may constitute integral elements of general feedback mechanisms for modulating the activity of regulatory RNAs. This principle is realized in the activity of mitogen-activated protein kinase (MAPK) pathways induced in many eukaryotic systems by heat and osmotic stress. In S. cerevisiae, activation of the hyper-osmolarity glycerol (HOG) MAPK pathway produces variations of the activity of ion channels, glycerol export, general protein machinery, and cell cycle progression. In this pathway, the transmembrane protein, Sho1, is connected to protein kinases and is essential for HOG signaling as it is responsible for activating specific MAPKs including Mpk1. These MAPK pathways result in the activation of DNA-binding transcriptional activators such as Tec1 which is found to be over-expressed in the absence of Hog1. Additionally, Hog1 has been described as a stress-activated protein kinase (SAPK) that is involved in multiple phases of the cell cycle and has the ability to reprogram gene expression in the presence of stress. This is performed by the enlistment of Hog1 to chromatin to recruit RNA polymerase II. Hog1 has the ability to remodel chromatin structure and plays a key role in the biogenesis of mRNA. During osmostress, Hog1 is shown to induce a set of ~200 long noncoding RNAs
(IncRNAs), confirming that Hog1 plays a key role in transcriptional regulation.\textsuperscript{194,222–224} Moreover, these IncRNAs are capable of participating changes in chromatin structure and nucleosome occupancy which have lasting effects on cellular memory.\textsuperscript{214,224} Given the increasing evidence that transcriptional regulation plays an overwhelming role in this well-defined pathway, the investigation into the use of the aforementioned MS platform was explored in this chapter to see what role, if any, PTMs played in such well-studied regulatory events.

Specifically, this chapter is aimed at the exploration of entire epitranscriptomes by employing our MS global profiling approach to investigate the possible existence of a broad, general link between PTM expression and stress response, which goes beyond the known mechanism of tRNA-based translational recoding. Preliminary experiments evaluated whether different types of stresses, such as heat shock and hyperosmotic stress, produced distinctive PTM profiles that could be ascribed to different response mechanisms. We subsequently focused on the HOG pathway because of the wealth of available information on its role in the hyperosmotic stress response,\textsuperscript{296,297} including the well-characterized induction and stabilization of mRNAs and IncRNAs by comparing PTM profiles obtained from wildtype \textit{S. cerevisiae} (i.e., WT) in the absence and presence of osmostress to identify possible PTM variations that correlated with the known induction of stress-response coding and noncoding genes.\textsuperscript{299,300–303} Then, we performed the profiling of a mutant strain in which the \textit{HOG1} gene was deleted (i.e., \textit{hog1}\textsuperscript{Δ}) to discriminate \textit{hog1}-dependent PTMs from those that were induced by osmostress without falling under direct control of \textit{HOG1}. Additionally, we explored the implementation an interactome-walk strategy in which the genes connecting \textit{HOG1} with PTM biogenetic enzymes in putative interactomics maps were sequentially deleted to retrace the signaling cascade and demonstrate a direct relationship
between PTM expression and this stress response pathway.

4.2 MATERIAL AND METHODS

4.2.1 Preparation of cellular extracts.

Saccharomyces cerevisiae strain BY4741, as well as hog1Δ::kanMX, ssb1Δ::kanMXi, rit1Δ::kanMX, and hog1Δ::kanMX derivatives obtained according to established procedures, were purchased directly from Open Biosystems. Yeast strains were grown in yeast extract, peptone, dextrose (YPD) medium. Each strain was streaked onto YPD agar and incubated at 30°C. For all studies, an individual colony was selected from a plate and placed into an individual tube containing 20 mL of YPD. The culture was incubated at either 30°C or 37°C with 200 rpm gyration. Optical density at 600 nm (OD_{600}) was monitored on a ThermoFisher Scientific (Waltham, MA) Nanodrop 2000c spectrophotometer until a value greater than 0.3 units was achieved. Liquid cultures left untreated were diluted to a final 0.3 OD_{600} and centrifuged at 6000 g for 5 min. to obtain pellets that contained approximately the same number of cells. In order to assess the possible effects of osmostress on RNA modification profiles, S. cerevisiae was grown to mid-log phase and induced with 0.4 M NaCl. Liquid cultures were then incubated for an additional 15 min. as these conditions have been found to specifically overexpress ~343 coding genes and ~173 lncRNAs. For RNA preparation and MS analysis procedures, please refer to Sections 2.2.3, 2.2.4 and 2.2.5.
4.3 RESULTS AND DISCUSSION

4.3.1 Epitranscriptomic profiling.

We challenged the performance of our approach by pursuing the study of specific biological samples. At the same time, we wanted to investigate how this approach could be effectively utilized to tackle specific questions. Therefore, our methods were applied to analyze *S. cerevisiae* grown in YPD media and harvested at 0.3 OD$_{600}$. First explored were the merits of using the heat maps generated by IMS-MS in order to assess global profiles of PTMs. In this type of plot, the independent variables describing molecular mass and ion mobility behavior are dispersed onto orthogonal dimensions. Their intersection is unique for each analyte and enables their accurate differentiation. For this reason, the heat-map was able to provide a comprehensive view of the distribution of all species in the sample. As shown in Figure 1-3, the full complement of cellular PTMs and carryovers from the original lysate was observed in a single experiment that placed the abundances on a common scale.

Given the visual nature of these types of plots, their application to inspect only the differences between two samples seemed imminent. As such, work became focused on using the available resources to create a platform in which data sets could be normalized to one another in order to enable a point-by-point subtraction of the datasets in an A minus B format. This type of algorithm was employed on the original sample of *S. cerevisiae* grown in YPD versus *S. cerevisiae* grown in synthetic complete. Synthetic complete is a media which only contains nitrogen base, dextrose and the essential amino acids. Given the simplicity of this media, the cell must act to use the provided primary nutrients to create the biomolecules essential for replication. SC is a minimal media whereas YPD is a rich media due to its incorporation of yeast
Figure 4-1. a) IMS-MS heat-map obtained from *S. cerevisiae* grown in synthetic complete (SC) medium. b) Differential plot obtained by subtracting the plot in panel a) from the one provided by *S. cerevisiae* grown YPD medium (Fig. 1-3). Taken with permission from Ref. 310.
extract which already contains these required biomolecules. Growth of *S. cerevisiae* in YPD is known to increase doubling time due to its excess of nutrients. As such, we expected these blatant metabolic changes to result in vastly different PTM profiles.

Direct inspection of the map obtained from the sample grown in SC (Figure 4-1a) revealed numerous features in common with the ones obtained in the YPD sample (Figure 1-3). This observation could be attributed to the intrinsic essential PTM baseline observed for this microorganism. These findings were readily confirmed by performing database searching, which led to the positive identification of 49 hits corroborated by gas-phase activation experiments (Table 4-1). Of this total, 41 hits matched PTMs observed in the YPD analysis (Table 3-3), whereas the remaining 8 were unique for this sample. The vast majority of the discrepancies between the SC and YPD samples corresponded instead to the respective carryover components, which were a direct consequence of the widely different compositions of these growth media and their putative metabolic effects.

Visual comparison of these two plots yielded many noticeable differences, therefore, implementation of a point-by-point subtraction as a means to observe only these changes seemed like the next logical approach for these types of data representations. Analysis was accomplished by expressing the intensity scales in AvP units to enable axes alignment and subsequent point-by-point subtraction. The result of this type of manipulation to the data sets can be seen in Figure 4-1b. The resulting differential plot highlighted the subtle changes experienced by low-abundance species, such as ac^4^C and f^5^Cm (Figure 4-1b). Overall, 31 of the 41 common PTMs were found to be more abundant in the SC than in the YPD sample, whereas
Figure 4.2: a) IMS-MS heat map obtained from *E. coli* grown in SC medium. b) Difference plot obtained by subtracting the plot in panel a) from the one provided by *S. cerevisiae* grown in the same medium (Fig. 4.1a). Taken with permission from Ref. 310.
the remaining 10 were less abundant. A closer look at the relative deviations featured in the differential plot showed that several of them exceeded the average RSD% of ±7.8% that expresses the biological reproducibility (Table 4-1). A more accurate assessment of the individual variations was obtained by comparing such deviations with the corresponding individual uncertainties provided in Table 4-1. This analysis indicated that the differences between SC and YPD profiles were confidently ascribable to the effects of the distinct growth media on *S. cerevisiae* metabolism, which represented the controlled variable between these datasets.

Environmental conditions and metabolic states can influence the expression of PTMs through the different metabolic pathways responsible for their biogenesis. At the same time, the enzymatic infrastructure that constitutes such pathways is coded by the genome of the organism under consideration. Therefore, global PTM profiles reflect the intersection of the very specific genetic and metabolic makeups of the respective cells. We explored the ability of the proposed approach to tackle this source of diversity by analyzing different microorganisms and comparing their PTM profiles. To this end, *E. coli* cultures were grown in the same SC medium utilized for *S. cerevisiae*, in such a way as to eliminate the type of available nutrients as an environmental variable. As expected, the recorded heat-maps (Figure 4-2a) differed significantly from those afforded by the corresponding *S. cerevisiae* sample (Figure 4-1a). The plot obtained by subtracting the former from the latter served to accurately assess such differences and to guide subsequent analysis (Figure 4-2b). The enlargement helps illustrate the type of variations afforded by low-abundance modifications. Overall, the *E. coli* sample provided a total of 30 hits, of which only 23 were in common with *S. cerevisiae* (Table 4-2). The common hits displayed relative deviations ranging from $1.68 \times 10^{-2} \%$ to 173%, many of which exceeded the RSD% values
<table>
<thead>
<tr>
<th>Hit(^1)</th>
<th>Exp. mass (u)</th>
<th>Exp. AvP(^2)</th>
<th>Rel. dev. (%)(^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>323.0519</td>
<td>21.5</td>
<td>+28.8</td>
</tr>
<tr>
<td>Y, U</td>
<td>324.0359</td>
<td>25.7</td>
<td>-15.4</td>
</tr>
<tr>
<td>D</td>
<td>326.0515</td>
<td>2.62</td>
<td>+11.9</td>
</tr>
<tr>
<td>m(^3)C, m(^3)Cm, m(^3)Cm</td>
<td>337.0675</td>
<td>1.13</td>
<td>-129</td>
</tr>
<tr>
<td>m(^3)Y, m(^3)U, m(^3)Y, Ym, m(^3)U</td>
<td>338.0515</td>
<td>1.24</td>
<td>-18.4</td>
</tr>
<tr>
<td>ho(^3)U</td>
<td>340.0308</td>
<td>9.94 x10(^{-2})</td>
<td>NA</td>
</tr>
<tr>
<td>A</td>
<td>347.0631</td>
<td>23.7</td>
<td>-9.20</td>
</tr>
<tr>
<td>I</td>
<td>348.0471</td>
<td>1.42 x10(^{-1})</td>
<td>+1.61</td>
</tr>
<tr>
<td>m(^1)A, m(^3)A, m(^5)A, Am</td>
<td>361.0787</td>
<td>4.78 x10(^{-1})</td>
<td>-130</td>
</tr>
<tr>
<td>m(^1)I, m(^1)Im</td>
<td>362.0628</td>
<td>2.52 x10(^{-2})</td>
<td>NA</td>
</tr>
<tr>
<td>G</td>
<td>363.0580</td>
<td>29.1</td>
<td>-5.13</td>
</tr>
<tr>
<td>ac(^4)C, f(^3)Cm</td>
<td>365.0624</td>
<td>5.72 x10(^{-1})</td>
<td>-18.6</td>
</tr>
<tr>
<td>m(^6)Am, m(^1)Am, m(^5)A</td>
<td>375.0944</td>
<td>3.53 x10(^{-2})</td>
<td>-150</td>
</tr>
<tr>
<td>m(^2)G, m(^2)G, Gm</td>
<td>377.0737</td>
<td>1.32</td>
<td>-25.3</td>
</tr>
<tr>
<td>ac(^4)Cm</td>
<td>379.0781</td>
<td>1.75 x10(^{-2})</td>
<td>-87.4</td>
</tr>
<tr>
<td>ncm(^3)U</td>
<td>381.0573</td>
<td>9.02 x10(^{-2})</td>
<td>-91.2</td>
</tr>
<tr>
<td>m(^1)Gm, m(^2)G, m(^3)Gm, preQ1, m(^2)G</td>
<td>391.0892</td>
<td>4.85 x10(^{-1})</td>
<td>+25.4</td>
</tr>
<tr>
<td>ncm(^3)Um</td>
<td>395.0730</td>
<td>2.33 x10(^{-2})</td>
<td>NA</td>
</tr>
<tr>
<td>mcm(^3)U</td>
<td>396.0570</td>
<td>6.24 x10(^{-2})</td>
<td>NA</td>
</tr>
<tr>
<td>mcm(^3)s(^3)U</td>
<td>412.0342</td>
<td>6.37 x10(^{-2})</td>
<td>NA</td>
</tr>
<tr>
<td>i(^6)A</td>
<td>415.1257</td>
<td>1.54 x10(^{-1})</td>
<td>NA</td>
</tr>
<tr>
<td>t(^6)A</td>
<td>492.1006</td>
<td>1.49 x10(^{-1})</td>
<td>-10.7</td>
</tr>
<tr>
<td>Ar(p)</td>
<td>559.0717</td>
<td>9.71 x10(^{-3})</td>
<td>NA</td>
</tr>
<tr>
<td>yW</td>
<td>588.15811</td>
<td>4.20 x10(^{-2})</td>
<td>+126</td>
</tr>
</tbody>
</table>

**Table 4-1.** Hits provided by a total RNA extract from *S. cerevisiae* grown in synthetic complete medium (SC). Taken with permission from Ref. 310.

\(^1\) Full names available at [http://mods.rna.albany.edu/](http://mods.rna.albany.edu/)

\(^2\) Abundance versus proxy (AvP) calculated from the respective signal intensity as percentage of the sum of the intensities of the four canonic ribonucleotides (see *Materials and Methods*). Each value was the average of five repeat analyses.

\(^3\) Relative deviation between AvPs obtained from *S. cerevisiae* grown in YPD and SC under otherwise identical conditions. NA indicates deviations that could not be calculated due to the absence of the corresponding species in the YPD samples.
<table>
<thead>
<tr>
<th>PTM</th>
<th>Exp. mass (u)</th>
<th>Average AvP (%)</th>
<th>Deviation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>323.0519</td>
<td>21.4</td>
<td>+2.40 x10^{-1}</td>
</tr>
<tr>
<td>Y, U</td>
<td>324.0357</td>
<td>22.1</td>
<td>-15.0</td>
</tr>
<tr>
<td>D</td>
<td>326.0510</td>
<td>2.27 x10^{-1}</td>
<td>+1.68 x10^{2}</td>
</tr>
<tr>
<td>m^3C, m^5C, Cm, m^6C</td>
<td>337.0673</td>
<td>8.08 x10^{-2}</td>
<td>173</td>
</tr>
<tr>
<td>m^3Y, Um, m^5U, m^3Y, Ym, m^3U</td>
<td>338.0513</td>
<td>2.44 x10^{-1}</td>
<td>134</td>
</tr>
<tr>
<td>A</td>
<td>347.0628</td>
<td>25.7</td>
<td>-7.95</td>
</tr>
<tr>
<td>I</td>
<td>348.0468</td>
<td>3.54 x10^{-3}</td>
<td>190</td>
</tr>
<tr>
<td>m^5Cm, m^4Cm, m^4C</td>
<td>351.0830</td>
<td>1.31 x10^{-2}</td>
<td>NA</td>
</tr>
<tr>
<td>mo^5U</td>
<td>354.0463</td>
<td>3.53 x10^{-3}</td>
<td>NA</td>
</tr>
<tr>
<td>G</td>
<td>363.0575</td>
<td>30.8</td>
<td>-5.61</td>
</tr>
<tr>
<td>ac^4C, f^6Cm</td>
<td>365.0617</td>
<td>4.95 x10^{-1}</td>
<td>14.3</td>
</tr>
<tr>
<td>mnm^5U</td>
<td>367.0776</td>
<td>2.86 x10^{-3}</td>
<td>NA</td>
</tr>
<tr>
<td>m^1G, m^2G, m^7G, Gm</td>
<td>377.0720</td>
<td>2.26 x10^{-1}</td>
<td>141</td>
</tr>
<tr>
<td>cmo^5U, chm^5U</td>
<td>398.0362</td>
<td>1.58 x10^{-1}</td>
<td>NA</td>
</tr>
</tbody>
</table>

**Table 4-2.** Hits provided by a total RNA extract from *E. coli* grown in synthetic complete medium (SC). Taken with permission from Ref. 310.
obtained from biological repeats of either organism. This observation provided excellent indications that these deviations were statistically significant, consistent with the considerable evolutionary distance between *E. coli* (a prokaryote) and *S. cerevisiae* (a eukaryote) in the phylogenetic tree. Therefore, the outcome of these preliminary experiments bodes well for the utilization of transcriptome-wide PTM profiles as effective tools for taxonomic determinations.

4.3.2 Global profiles of stress.

The cellular response orchestrated by an organism under stress conditions involves the concurrent and coordinated activation of numerous metabolic pathways. These events include, but are not limited to, specific enzymatic activities, the disruption of structures, and perturbations of gradients within the cell. In total, these events are compensatory and are aimed at restoring and maintaining the homeostasis of a particular organism. In this context, the investigation of an individual PTM can provide only a limited view any potential role it plays in the overall response mechanism. In contrast, the ability to monitor multiple PTMs at the same time can provide valuable insights into their synergistic activities and coordination at the system level; similar to approaches established in metabolomics studies. This type of information can be obtained through analytical strategies capable of multiplexing PTM detection and accurately assessing their mutual expression levels.

In this direction, the developed MS strategy investigated herein was used to obtain complete PTM profiles capable of capturing these biological fluctuations. The complexity of these types of mixtures can be readily resolved by high-resolution MS analysis, followed by tandem MS for structure verification. Alternatively, IMS-MS analysis can accomplish the task by
discriminating analytes according to both mass and ion mobility behavior. The latter is determined by the probability of ions to interact with background gas while traveling in a moderate electric field, which is a function of size and conformation (see Chapter 2). Application of these techniques to wild-type *S. cerevisiae* grown in YPD at 30°C (WT) resulted in the heat map plot shown in Figure 4-3a. The representative heat map shows the intensity of each detected ion by the color gradient provided. In comparison upon preliminary inspection, wild-type enduring heat shock by growth at 37°C (WT\textsubscript{37}) shows radical overexpression of the detected ions. It is true from these representative samples that the use of heat maps can provide comprehensive visual representations of all species in the samples, which enabled immediate comparisons between strains grown under these different environmental conditions (e.g., Figure 3a and b). Additionally, mutant strains of the *S. cerevisiae* in which a gene was knocked out, i.e. *rit1Δ* (Figure 4c), also proved to have a different trend than either the WT or the WT\textsubscript{37} (Figure 4a, b and c).

In order to inspect the content of these plots, the mass information provided on the y-axis was searched against the in house databases. The initial database “hits” included numerous isomeric/isobaric species that shared the same elemental composition and, thus, could not be readily discriminated on the basis of mass alone. However, as demonstrated earlier for the uridine/pseudouridine couple\textsuperscript{262} and a series of methyl-guanine products (see Chapter 2),\textsuperscript{310} the different structures associated with the isomeric species translated into distinctive ion mobility behaviors that were readily differentiated on the t\textsubscript{D} dimension (see for example Figure 4-3a inset). Efforts to appropriately calibrate the t\textsubscript{D} axis in different hardware platforms are currently underway to enable the utilization of characteristic t\textsubscript{D} values as unique identifiers for
**Figure 4-3.** Representative heat maps obtained by anionic IMS-MS analysis of *S. cerevisiae* grown in YPD medium at either **A)** 30°C (WT, control) or **B)** 37°C (WT<sub>37</sub>, heat shock); and **C)** rit1Δ strain grown at 30°C (rit1Δ). Panel **D)** shows the expanded regions containing the Ar(p) modification, which enable one to readily compare its expression levels in the respective samples. The intensity axis of the heat maps were scaled to the same level to enable direct comparisons by using the color gradient provided on the right. The expanded region in panel **A)** shows the separation achieved on the time domain for the U/Ψ isobars.
individual PTMs. Both MaSTeR and MS\textsuperscript{n} techniques described in Chapter 2 were applied here to confirm the identity of all putative hits returned after the database searching. The PTMs corroborated by these analyses are shown in Table 4-3a.

The experimental strategies summarized above were employed to obtain the global profiles of samples that had been subjected to different types of stresses. For example, the heat map in Figure 4-3b was obtained from a sample of wildtype S. cerevisiae BY4741 that was grown in YPD medium at 37°C (WT\textsubscript{37}, see Materials and Methods), under conditions known to induce a typical heat shock response.\textsuperscript{315} Visual comparison between these data and those provided by WT control (Figure 4-3a) revealed remarkable differences, which were confirmed by completing the identification of the PTMs in the sample (Table 4-3b). The baseline represented by the profile of WT control consisted of a total of 41 distinct PTMs (Table 4-3a). In contrast, the WT\textsubscript{37} sample displayed 50 PTMs, out of which 32 were present also in WT control and the remaining 18 were newly expressed PTMs unique for this type of sample (marked in red in Table 4-3b). The observed increase of PTM content was consistent with previously published reports that heat shock activation caused broad overexpression of RNA components within the cell.\textsuperscript{316,317} Among the 32 PTMs in common between the two types of samples, 4 were detected with comparable relative abundances in both, whereas 14 and 14 were respectively under- and over-expressed in WT\textsubscript{37} (marked respectively with darker and lighter shades of blue in Table 4-3b). It must be emphasized that at least five biological repeats were analyzed for each type of sample to assess the uncertainty intrinsic in the quantitative determination (see Materials and Methods). Only PTMs with a p-value less than 0.05 were marked as being under-/over-expressed. Values greater than 0.05 matched the typical biological reproducibility obtained from yeast samples submitted to this
<table>
<thead>
<tr>
<th>PTM(^1)</th>
<th>Mass(^2)</th>
<th>Class(^3)</th>
<th>A)(^4)</th>
<th>B)(^4)</th>
<th>C)(^4)</th>
<th>D)(^4)</th>
<th>E)(^4)</th>
<th>F)(^4)</th>
<th>G)(^4)</th>
<th>H)(^4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ac(^\circ)C, F(^\circ)Cm</td>
<td>365.024</td>
<td>r, t</td>
<td>6.34 E-1</td>
<td>5.14 E-1</td>
<td>5.26 E-1</td>
<td>6.60 E-1</td>
<td>6.03 E-1</td>
<td>3.82 E-1</td>
<td>3.13 E-1</td>
<td>2.67 E-1</td>
</tr>
<tr>
<td>ac(^\circ)Cm</td>
<td>379.078</td>
<td>r, t</td>
<td>6.63 E-1</td>
<td>1.34 E-2</td>
<td>1.45 E-2</td>
<td>6.38 E-3</td>
<td>1.31 E-2</td>
<td>1.16 E-2</td>
<td>1.69 E-3</td>
<td>1.78 E-3</td>
</tr>
<tr>
<td>ac(^\circ)p(^\circ)Um, m(^\circ)ap(^\circ)Y(^4)</td>
<td>439.099</td>
<td>r</td>
<td>4.87 E-1</td>
<td>1.87 E-3</td>
<td>1.87 E-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ar(p)</td>
<td>559.072</td>
<td>t</td>
<td>5.60 E-3</td>
<td>5.03 E-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cm(^\circ)U(^*)</td>
<td>381.034</td>
<td>r</td>
<td>1.46 E-3</td>
<td>5.68 E-3</td>
<td>5.68 E-3</td>
<td>3.49 E-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cmm(^\circ)s(^*)U</td>
<td>427.045</td>
<td>?</td>
<td>1.02 E-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>326.052</td>
<td>r, sn, t</td>
<td>2.95</td>
<td>0.86</td>
<td>1.82</td>
<td>2.32</td>
<td>2.91</td>
<td>2.06</td>
<td>1.98</td>
<td>2.46</td>
</tr>
<tr>
<td>pu(^*)</td>
<td>352.031</td>
<td>?</td>
<td>1.02 E-2</td>
<td>1.32 E-2</td>
<td>1.53 E-2</td>
<td>1.69 E-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h(^m)A(^<em>), m(^t)(^4)A(^</em>)</td>
<td>506.116</td>
<td>t</td>
<td>5.39 E-3</td>
<td>4.49 E-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ho(^U)U(^*)</td>
<td>340.031</td>
<td>t</td>
<td>1.43 E-3</td>
<td>1.25 E-3</td>
<td>9.72 E-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>348.047</td>
<td>m, t</td>
<td>1.44 E-1</td>
<td>9.87 E-2</td>
<td>1.40 E-1</td>
<td>1.21 E-1</td>
<td>1.26 E-1</td>
<td>1.53 E-1</td>
<td>2.09 E-1</td>
<td>1.04 E-1</td>
</tr>
<tr>
<td>jA</td>
<td>415.126</td>
<td>t</td>
<td>3.93 E-2</td>
<td>1.34 E-1</td>
<td>8.34 E-3</td>
<td>1.99 E-2</td>
<td>2.55 E-1</td>
<td>4.15 E-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>m(^1)A, m(^2)A, m(^4)A, m(^5)A*, Am*</td>
<td>361.079</td>
<td>m, r, sn, sno, t</td>
<td>1.02 E-1</td>
<td>5.08 E-1</td>
<td>2.95 E-1</td>
<td>4.67 E-1</td>
<td>3.35 E-1</td>
<td>5.67 E-1</td>
<td>3.29 E-1</td>
<td></td>
</tr>
<tr>
<td>m(^1)G, m(^2)G, m(^5)G, Gm</td>
<td>377.074</td>
<td>m, r, sn, sno, t</td>
<td>1.02</td>
<td>9.96 E-1</td>
<td>1.52</td>
<td>1.11</td>
<td>1.19</td>
<td>1.56</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>m(^1)Gm*, m(^2)G*, m(^5)G*, preQ1*, m(^6)G</td>
<td>391.089</td>
<td>m, r, t</td>
<td>6.26 E-1</td>
<td>3.20 E-1</td>
<td>4.56 E-1</td>
<td>2.78 E-1</td>
<td>3.54 E-1</td>
<td>3.16 E-1</td>
<td>6.04 E-1</td>
<td>1.98 E-1</td>
</tr>
<tr>
<td>m(^1)l, lm*</td>
<td>362.063</td>
<td>r, t</td>
<td>2.53 E-2</td>
<td>2.83 E-2</td>
<td>1.08 E-2</td>
<td>2.44 E-2</td>
<td>1.52 E-2</td>
<td>3.15 E-2</td>
<td>3.64 E-3</td>
<td></td>
</tr>
<tr>
<td>m(^2)C, m(^3)C, Cm, m(^4)C*</td>
<td>338.052</td>
<td>m, r, sn, sno, pre-t</td>
<td>2.45 E-1</td>
<td>8.91 E-1</td>
<td>1.17</td>
<td>8.19 E-1</td>
<td>1.04</td>
<td>9.95 E-1</td>
<td>1.34</td>
<td>8.16 E-1</td>
</tr>
<tr>
<td>m(^2)Y*, Um, m(^4)U, m(^5)Y, Ym*, m(^6)U</td>
<td>338.052</td>
<td>m, r, sn, sno, t</td>
<td>1.03</td>
<td>7.11 E-1</td>
<td>6.89 E-1</td>
<td>1.11</td>
<td>8.24 E-1</td>
<td>1.28</td>
<td>9.31 E-1</td>
<td></td>
</tr>
<tr>
<td>m(^1)Cm*, m(^2)Cm*, m(^3)Cm*, m(^4)Cm*</td>
<td>350.076</td>
<td>m, r, t</td>
<td>3.60 E-3</td>
<td>2.21 E-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m(^3)Am*, m(^4)Am*, m(^5)Am*, m(^6)Am*</td>
<td>375.094</td>
<td>m, r, sn, sno, t</td>
<td>5.01 E-3</td>
<td>3.60 E-2</td>
<td>6.05 E-2</td>
<td>2.24 E-3</td>
<td>5.30 E-3</td>
<td>5.21 E-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mcm(^\circ)U</td>
<td>412.034</td>
<td>t</td>
<td>1.34 E-2</td>
<td>7.86 E-2</td>
<td>7.96 E-3</td>
<td>4.50 E-2</td>
<td>9.60 E-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mm(^\circ)U</td>
<td>396.057</td>
<td>t</td>
<td>7.83 E-2</td>
<td>5.69 E-2</td>
<td>7.83 E-2</td>
<td>1.57 E-2</td>
<td>9.47 E-2</td>
<td>6.85 E-2</td>
<td>6.67 E-2</td>
<td></td>
</tr>
<tr>
<td>mm(^\circ)U*</td>
<td>381.057</td>
<td>t</td>
<td>6.85 E-3</td>
<td>3.36 E-2</td>
<td>4.29 E-2</td>
<td>3.74 E-2</td>
<td>5.56 E-2</td>
<td>5.22 E-2</td>
<td>3.16 E-2</td>
<td></td>
</tr>
<tr>
<td>mm(^\circ)Um</td>
<td>395.073</td>
<td>t</td>
<td>6.25 E-3</td>
<td>7.10 E-3</td>
<td>1.81 E-3</td>
<td>6.50 E-3</td>
<td>1.02 E-2</td>
<td>4.32 E-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mm(^s)(^s)U</td>
<td>369.040</td>
<td>t</td>
<td>1.54 E-3</td>
<td>1.30 E-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t(^*)A</td>
<td>492.101</td>
<td>t</td>
<td>1.34 E-1</td>
<td>1.46 E-1</td>
<td>2.11 E-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X*</td>
<td>364.042</td>
<td>?</td>
<td>1.21 E-3</td>
<td>1.86 E-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y, U</td>
<td>324.036</td>
<td>r, sn, sno, t</td>
<td>22.0</td>
<td>24.5</td>
<td>26.6</td>
<td>23.6</td>
<td>27.6</td>
<td>23.6</td>
<td>23.7</td>
<td>26.5</td>
</tr>
<tr>
<td>yW</td>
<td>588.158</td>
<td>r, t</td>
<td>1.84 E-1</td>
<td>6.75 E-3</td>
<td>3.16 E-2</td>
<td>6.57 E-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3. Quantitative determination of ribonucleotides present in total RNA extract of S. cerevisiae grown at A) 30°C (WT, control); B) 37°C (WT\(_{37}\), heat shock); C) 30°C after treatment with 0.4 M NaCl (WT\(_{37}\), osmoston); D) 37°C after treatment with 0.4 M NaCl (WT\(_{37}\), hog1\(^\Delta\) grown at E) 30°C (hog1\(^\Delta\)); F) 37°C (hog1\(^\Delta\)_37); G) 30°C after treatment with 0.4 M NaCl (hog1\(^\Delta\)); and H) 37°C after treatment with 0.4 M NaCl (see Materials and Methods).  

1 Full names available at [http://mods.rna.albany.edu/](http://mods.rna.albany.edu/)
2 Neutral experimental mass expressed in mass units (u).
3 Putative classes of RNAs include: messenger RNA (m); ribosomal RNA (r); transfer RNA (t); pre-transfer RNA (pre-t); small nuclear RNA (sn); small nucleolar RNA (sno); unknown class (?).
4 For each PTM, a value of abundance versus proxy (AvP) was calculated from the respective signal intensity as percentage of the sum of the intensities of the four canonic ribonucleotides. This figure of merit represents a relative measure of the abundance of each PTM in the sample, which was introduced and validated in reference 34. Each reported value was the average of five repeat analyses.
5 Red color marks newly expressed PTMs absent in the WT sample; blue marks PTMs with expression levels within ±7.8% of those of WT; darker/lighter blue indicates PTMs with levels that deviated by more than ±7.8% from those of WT.
6 Modifications previously unreported for this microorganism.
type of analysis and, thus, were considered as normal experimental fluctuations.

In analogous fashion, *S. cerevisiae* strain BY4741 was grown in YPD medium at 30°C and then exposed for 15 min. to an environment containing 0.4 M NaCl (WT<sub>S</sub>, see Materials and Methods), which induces a characteristic osmostress response. Also in this case, the respective heat map (not shown) provided an outcome that was significantly different from that of WT control and, remarkably, from that of the WT<sub>37</sub> sample, as well. The PTM profile revealed by completing positive identification (Table 4-3c) showed that WT<sub>S</sub> contained a total of 38 PTMs, with 8 newly expressed (red) and 30 in common with WT control (blue), out of which 7 and 22 were respectively under- and over-expressed (shades of blue). In addition, 7 of the 8 newly expressed PTMs (absent in WT control) were also present in WT<sub>37</sub>, whereas the remaining one was unique for this sample.

These experiments clearly indicated that the cellular responses to these different types of stresses involved profound changes in the production of specific RNA PTMs. The changes were not limited to variations of the expression levels of common PTMs, but comprised also the activation of new pathways that were previously dormant in the absence of stressors. According to the information included in the RNA Modification<sup>239</sup> and MODOMICS<sup>240</sup> databases, 11 of the 18 new PTMs activated by heat shock and 3 of 8 associated with osmostress have been previously described in eukaryotic, but not necessarily yeast tRNA. In this regard, however, it is important to note that the data available on the distribution of PTMs in different organisms and classes of RNAs is still rather fragmentary. In most cases, the provenance of a certain PTM cited only the initial study that reported its first detection/characterization, rather than the results of comprehensive surveys. The vast majority of studies have traditionally focused on tRNA and
rRNA, not only because of their abundant PTM content, but also for the availability of convenient isolation protocols. Further, it is not clear whether any distribution information present in the common databases was obtained from samples harvested under stress conditions. Therefore, the provenance information included in Table 4-3 should be considered with discretion. These considerations aside, it was very significant that none of the new PTMs detected under stress corresponded to those implicated in the mechanism of oxidative stress response involving tRNA-based translational recoding. This observation suggested that a sizeable portion of the newly activated pathways were likely to rely on very different mechanisms.

A closer examination of the new PTMs expressed in WT revealed that 7 were in common with the WT sample, but 1 and 11 were unique for either the former or the latter, respectively, thus pointing to the existence of distinct response pathways involving PTM biogenesis. It is well established that thermic stress activates a response based on the prominent expression of heat shock proteins (HSP), which exert their protective effects by limiting protein denaturation. However, it has been shown that the HOG pathway can be also activated at higher temperatures, thus raising the possibility that the respective signaling pathways may be closely intertwined. If proven correct, this hypothesis would provide an excellent explanation for the 7 common PTMs that were induced by both heat shock and osmostress in our experiments.

In addition to the induction of new PTMs absent in WT control, stress conditions affected the expression levels of existing PTMs in different ways. Indeed, the relative abundance of 14 and 22 existing PTMs were increased above the normal experimental fluctuations (i.e., p-value <0.05) in the WT and WT samples, respectively. In contrast, 14 and 7 existing PTMs fell below the fluctuations level in the same samples, whereas 4 and 1 ceased to be detectable altogether under
the standard analytical conditions. Also in this case, the various PTMs could be readily categorized in distinctive groups that were either stress-specific or not. Of the 20 downregulated/vacated modifications in WT37, 14 displayed the same behavior in WTs. At the same time, of the 32 species upregulated/unique in WT37, 20 displayed the same behavior in WTs. In some cases, however, the variation directions did not match in both samples, as shown for $t^6A$ that was upregulated under osmostress as compared to the control, but disappeared under heat shock, or the methyl-A isomers that had the exact opposite fate (Table 4-3). The disappearance of specific PTMs under stress conditions was particularly intriguing and suggested the possibility that they may not be essential to the cellular response and that their functions may be effectively taken over by other stress-specific PTMs. The picture emerging from these initial experiments clearly ruled out the notion that PTM biogenesis might be controlled by a single regulatory pathway, but suggested instead the presence multiple individual controls.

4.3.3 Identification of biogenetic enzymes.

In order to unambiguously identify the stress-induced PTMs that were under HOG control, we analyzed a hog1Δ strain, which lacks the stress-activated protein kinase Hog1; the master regulator of gene expression reprogramming in response to osmostress. The hog1Δ mutant was first grown at 30°C to identify its baseline profile in the absence of stress, which comprised the induction of 4 new PTMs absent in WT control (Table 4-3e). The mutant strain was subsequently grown at 37°C (i.e., hog1Δ37) and treated with 0.4 M NaCl at both 30°C (hog1Δs) and 37°C (i.e., hog1Δ37,s). The samples displayed complex profiles with up/down variations of the majority of
the PTMs observed in WT under corresponding environmental conditions, as well as production of distinctive sets of new ones absent in WT (marked respectively with different shades of blue, or solid red in Table 4-3f-h).

The newly activated PTMs observed exclusively under stress conditions were examined according to a process of elimination to identify their putative regulatory relationships (summarized in Table 4-3). In particular, the mutation-induced elimination of any of the newly expressed PTMs (i.e., absent in WT control) was interpreted as evidence of HOG control. The reasoning behind the elimination process could be illustrated here for the newly expressed ho5U modification. This species was absent in WT control, but present in both WT37, WT5, as well as the double-stressed WT37,5 sample (see Materials and Methods), thus cementing its status as a stress-induced PTM. At the same time, this species was prominently absent in all of the hog1Δ samples regardless of conditions, which indicated beyond doubt that its expression was linked to the HOG pathway. In contrast, the i6A and ncm5Um modifications observed in both WT37 and WT5 were not eliminated by the mutation, thus suggesting that their stress-induced biogenesis was independent from the HOG pathway.

It is interesting to note that while the production of these PTMs was triggered by both heat shock and osmostress, others displayed prominent specificity for either type of stress. For example, X was detected only in the WT5 sample (and, consistent with this finding, in the double-stressed WT37,5) to earn the status of osmostress-induced and hog1-dependent. Its reproducibility and abundance in the analyzed samples dissipated any doubt about its significance. At the same time, 10 putative PTMs were found to be heat shock-induced and hog1-dependent, but only one was heat shock-induced and hog1-independent (Table 4-4). Finally, f5U
Table 4-4. Summary new PTMs expressed in WT, ssb1Δ, rit1Δ mutant under either control or stress conditions.

1 Full names available at [http://mods.rna.albany.edu/](http://mods.rna.albany.edu/)
2 Putative classes of RNAs include: ribosomal RNA (r); transfer RNA (t); unknown class (?).
3 WT, hog1Δ, ssb1Δ, and rit1Δ strains were grown at 30°C under normal conditions (no subscript); 37°C (37 subscript); 30°C after treatment with 0.4 M NaCl (S subscript); and 37°C after treatment with 0.4 M NaCl (S,37 subscript) (see Materials and Methods).
4 Red color marks newly expressed PTMs absent in the WT sample.
*Modifications previously unreported for this microorganism.
was observed only when the \textit{hog1A} mutant was subjected to higher temperature, thus suggesting the possible activation of an alternative mechanism of heat response, which may compensate for the elimination of the Hog1 kinase.

4.3.4 \textbf{Integrating PTM biogenesis with regulatory networks.}

The results provided by these experiments clearly showed that the effects of heat and osmotic stress on PTM biogenesis were not mediated by a single regulatory pathway, but resulted instead from the coordinated activities of a more complex regulatory network. In \textit{S. cerevisiae} (budding yeast), extensive investigation of gene expression under stress has led to the identification of genes that are consistently up- or down-regulated in response to specific stimuli\textsuperscript{320}. These genes have been defined as specific environmental stress response genes (SESR, \textbf{Scheme 4-1}).\textsuperscript{321} In \textit{S. pombe} (fission yeast), instead, common sets of genes defined as core environmental stress response genes (CESR) are involved in nearly all types of stresses.\textsuperscript{321} In this case, a central hub integrates the stimuli provided by various stressors and coordinates a general cellular response. The detection of PTMs that were activated by both heat and osmotic stress suggests that a similar organization may be present in the stress-induced PTM biogenesis observed in \textit{S. cerevisiae}.

In order to test this possibility, we devised a strategy that utilized gene deletion strains to retrace putative regulatory pathways. This strategy required the prerequisite formulation of valid hypotheses on the network’s composition and architecture to guide the selection of the genes to be eliminated. For example, \textbf{Scheme 4-2a} depicts the organization of genes involved in the typical cellular response to heat shock and osmostress in \textit{S. cerevisiae}, as deduced from a wealth
Scheme 4-1. Regulation of stress genes in fission and budding yeast. In the former, different stresses share a common core of environmental stress response genes (CESR), whereas stress-specific environmental response genes (SESR) predominate in the latter (17, 58).
of available information. In contrast, **Scheme 4-2b** was drawn by using the Cytoscape server to examine databases of genetic and protein–protein interactions, as well as transcriptional and post-transcriptional regulation networks. This specific subsection of the extensive stress-response interactome delineates a putative network connecting essential components of the established stress-response pathways, such as HOG and various heat shock genes, to the biogenetic enzymes responsible for the formation of target PTMs, which are listed in the MODOMICS database. It is interesting to note that, while the organization of the established response genes (**Scheme 4-2a**) fully conformed to the prevailing view of parallel, largely independent pathways for different stimuli, the downstream region covering the putative regulation of PTM expression (**Scheme 4-2b**) assumed a much more complex architecture characterized by numerous crossovers. The interactomic map clearly suggested that genes such **HOG1** and **FUS3** could provide high-level integration between the heat shock and osmostress signaling pathways, whereas **SSA1**, **SSB1**, **SSB2**, **SSE1**, and **HSP82** could provide additional low-level integration and fine tuning of PTM production. This set of genes codes for proteins with broad chaperone activities, which are known to prevent aggregation, promote folding of proteins to their native states, and solubilize and refold aggregated proteins. Often, these heat shock chaperones work closely with co-chaperones which are involved in signal transduction, cell cycle regulation, cellular differentiation and cell death. Under stress, eukaryotic systems rely heavily upon Hsp70 chaperones such as Hsp82 to stabilize death-inducing mutations, such as those propagated by oncogenes.

Based on this map, we have initiated the systematic investigation of selected gene deletion strains to substantiate the putative signaling pathways responsible for PTM regulation.
**Scheme 4-2.**  
a) Established relationships between genes involved in response to hyperosmotic stress and heat shock.  
b) Putative interactomic network connecting established stress-response genes with PTM biogenetic enzymes. Arrows indicate proven functional relationships. Vertical lines indicate possible functional links. Horizontal lines indicate possible co-expression. Red and blue boxes highlight known biogenetic enzymes and respective PTMs. Light blue arrows and boxes identify the $HOG1 \rightarrow SSB1 \rightarrow RIT1$ pathway corroborated by gene deletion.
The PTMs of most interest were the 11 induced in either heat shock or osmostress, which were found to be $hog1$-dependent (Table 4-4). The results obtained from the $ssb1\Delta$ mutant showed that 11 of the 19 PTMs found unique in stressed states were $ssb1$-dependent, indicating that this pathway may indeed be acting as an integration hub responsible for the downstream initiation of multiple enzymatic processes involved in PTM biogenesis. The fact that the majority of these PTMs are controlled by genes located downstream in these established stress-response pathways appears to rule out their possible participation in sensing mechanisms, which are typically located upstream. Additionally, the two PTMs which appeared in WT, but not in any specific stress states, were also found to be $ssb1$-dependent (Table 4-4). Further, a total of 8 PTMs were found to be expressed even in the absence of $SSB1$ indicating that $SSB1$ cannot be the only hub responsible for signaling PTM biogenesis.

The next step in retracing the proposed map consisted of analyzing strains in which the specific biogenetic enzymes were deleted. The case of the $RIT1$ gene, which appears to be connected directly to $SSB1$ in Scheme 4-2b, could serve here to exemplify possible outcomes. The Rit1 enzyme is known to catalyze the formation of Ar(p) by the addition of a 2′-O-ribosyl phosphate group to adenine. A typical substrate is represented by adenine 64 of the initiator tRNA$\text{Met}$(i), which prevents its utilization during the elongation step of protein synthesis in yeast. Our data indicated that this PTM was uniquely expressed in the WT strain during heat shock and was both $hog1$- and $ssb1$-dependent. When the $rit1\Delta$ mutant was examined (see for example the heat map obtained from $rit1\Delta$ under normal growth conditions, Figure 4-3c), no Ar(p) could be detected with or without stress, thus confirming that $RIT1$ is indeed essential for the production of this PTM. Further, in the context offered by the mutants in the study, this
outcome corroborated the putative $HOG1 \rightarrow SSB1 \rightarrow RIT1$ pathway as the signaling control for Ar(p) (marked in light blue in Scheme 4-2b). Surprisingly, close examination of $rit1\Delta'$'s global profile revealed the absence of 9 PTMs that were both $hog1$- and $ssb1$-dependent (Table 4-4), which potentially places them under the same $HOG1 \rightarrow SSB1 \rightarrow RIT1$ regulatory pathway. In addition, three other PTMs that were not $hog1$- or $ssb1$-dependent were found to be $rit1$-dependent. These observations suggest that $RIT1$ may itself act as an additional hub in the control network. The fact that the majority of $rit1$-dependent PTMs do not involve adenine modification rules out the possibility that Ar(p) itself might serve as their direct biosynthetic precursor. An alternative hypothesis could be that activation of $RIT1$ may result in the co-expression of other biogenetic enzymes yet to be identified. In this direction, it must be noted that, of the 12 stress-induced PTMs that are $rit1$-dependent, Ar(p) is the only one with a known biogenetic enzyme. Considering the large number of genes without assigned function in the yeast genome, the functional link with $RIT1$ activation may provide the basis for further studies aimed at the identification of the missing biogenetic enzymes.

4.4 CONCLUSIONS

Monitoring the epitranscriptomic profile of $S. \ cerevisiae$ as a function of different experimental variables has unambiguously demonstrated that PTM biogenesis is profoundly influenced by environmental conditions. The fact that the production of numerous new PTMs was readily activated by increasing temperature or salt concentration provides very strong evidence for the broad implication of PTMs in stress response. This finding is consistent with the role of selected tRNA modifications that are known to participate in translational recoding
triggered by oxidative stress \cite{291,292}. However, the majority of the activated modifications observed in our study are not known to engage in these types of mechanisms. The direct comparison of PTM profiles has clearly revealed the presence of stress-specific modifications, as well as non-specific ones, which could be part of a more general cellular response. These observations support the presence of alternative mechanisms involving PTM activity, which are likely to include the broad participation of diverse classes of RNAs.

With the goal of investigating the relationships between established stress-response signaling and PTM activation, we devised a strategy that involves evaluating the simultaneous effects of gene deletion on the expression of all PTMs in the cell. This system-biology approach was inspired by the wealth of genomic and functional data available for \textit{S. cerevisiae}. In what could be described as an interactome-walk, genes connecting the putative biogenetic enzymes with known response genes were systematically eliminated to retrace a putative regulatory network inferred from available information. Preliminary results have allowed us to clearly differentiate PTMs that fall under HOG control from those that do not. The HOG pathway consists of a cascade of MAP kinases, which is preeminently responsible for the response to osmostress and has more tenuous communication with parallel heat shock controls \cite{322-328}. Our results support the presence of numerous integration hubs between these stress-specific pathways, which coordinate the expression of selected PTMs. This type of regulatory architecture could help explain cross-tolerance effects, in which preconditioning to a specific type of stress can increase the tolerance to a different one \cite{335}. These observations should be considered also in the context of the effects of \textit{HOG1} on chromatin remodeling \cite{336,337} and activation of DNA-binding transcriptional activators \cite{338-342}. The facts that \textit{HOG1} increases the stability of a well-defined
group of osmostress mRNAs and induces specific lncRNAs under osmostress leave ample space for the possible existence of feedback mechanisms based on PTM activity. In this direction, we could speculate that chromatin remodeling by stress-induced lncRNAs may affect the expression of genes involved in the general cellular response, as well as PTM regulatory pathways. In turn, the latter could integrate signals from other (likely protein-based) pathways and modify lncRNAs to reduced/enhance their chromatin remodeling activity. Therefore, determining the distribution of PTMs in these classes of RNAs and monitoring their levels under stress will provide the basis for exploring these possible mechanisms.

After establishing a direct link between stress and selected PTMs, the next challenge will be the identification of the actual roles played by such PTMs in cellular response. This important task is severely hampered by the fragmentary knowledge of their distribution outside tRNA and rRNA, and the absence of annotation on the extensive genomic/transcriptomic data available to the public. Towards this goal, we are currently developing technologies capable of determining the class of RNA containing a specific PTM, the identity of the parent species, and the sequence position of the modification. This information will help interrogate the available transcriptomic data, which will lead to testable hypotheses by suggesting possible correlations with known functions associated with the parent RNA. This information will also guide the implementation of strategies based on the mutation of nucleotides susceptible to modification, or silencing of the RNA bearing the modification, aimed at testing their effects on the metabolic and epigenetic state of the phenotype. The results will provide valuable insights into the functions of PTMs in stress response and, more broadly, will be expected to drive a reevaluation of their overall biological significance.
5.1 INTRODUCTION

The focus of PTM analysis has historically revolved around identifying new PTMs within well-defined classes of RNA, which comprised almost exclusively the typical abundant sources (i.e., tRNA, rRNA, and mRNA), rather than studying their distribution and functions at the entire transcriptome level. Therefore, these investigations have resulted in extensive information about the identity, structure, and biogenesis of most PTMs. However, the discovery of different functions performed by ncRNAs, such as gene and mRNA silencing, and the discovery of translational recoding mechanisms mediated by modified tRNA suggests that PTMs play essential roles in the regulation of the multifaceted activities of their parent RNAs. This idea has prompted a general reassessment of the significance of RNA classes and the biological significance of PTMs. As shown in Chapter 4, the MS platform developed herein is able to capture details about specific pathways activated by external stimuli. Indeed, exposure to a high-salt environment represented a simple mechanism for triggering on demand specific PTMs. Additionally, the link between the HOG pathway and RNA biology was significantly made when osmostress was shown to activate Hog1-dependent stabilization of mRNA and induction of ~200 long noncoding RNAs (lncRNAs).224
Consistent with this observation, the data presented in the previous chapter indicated that a novel set of PTMs is induced by osmostress in Hog1-dependent fashion. These results provide strong motivations for the development of novel strategies to investigate the putative effects of osmostress PTMs on the structure, stability, and interactions of their parent ncRNAs. In this direction, this chapter advances possible strategies that could be employed to address these, as well as many other biological questions. As a model system that may exemplify the broader implication of PTMs in IncRNA activity, we decided to explore a set of Hog1-dependent antisense transcripts of the *CDC28* gene, which may be linked to the PTMs observed in Chapter 4. Their transcription from the *CDC28* 3′-UTR promotes *CDC28* gene looping, induces *CDC28* transcription and promotes recovery from osmostress-induced cell cycle arrest. In addition, these osmostress IncRNAs may facilitate the induction of stress-response mRNAs, which occurs while global transcription is repressed. The osmostress PTMs could affect these activities by interfering with mechanisms involving protein recruiting, RNA recognition, and even catalysis. Further, the fact that many PTMs possess “writers” as well as “erasers” implies possible mechanisms by which protein-based pathways could participate in regulating the IncRNA-mediated processes. The well-defined regulation of osmostress IncRNAs provides an excellent handle to investigate the function of Hog1-induced PTMs on IncRNAs, and thus to take a major step forward in elucidating the role of PTMs as modulators of RNA function.

Many aspects of osmoadaptation are conserved across eukaryotes and constitute a valid paradigm for other types of stress responses. Therefore, everything that will be learned from these models will constitute a valuable framework for supporting future targeted investigations of systems with direct disease implications, which could be used to discern cellular pathology.
Ultimately, the elucidation of the universe of transcriptome functions will open the doors for the discovery of new drug targets and the development of new diagnostic avenues and drug therapies, which will aid in the advancement of the human race.

5.2 CHARACTERIZATION OF MODIFIED RNAs BY COMBINING AFFINITY CAPTURE WITH MASS MAPPING AND SEQUENCING TECHNIQUES.

The PTM profiling workflow discussed in the previous chapters involves the complete digestion of total RNA into mononucleotide units. This modus operandi precludes the ability to identify the parent RNA that bore the PTMs and to identify their sequence position. This type of information is absolutely essential to support the functional investigation of PTMs by mutating the modification sites and observing the effects on the phenotype. In order to take the next step, we must identify ways for accessing this type of information. A possible strategy involves the utilization of antisense affinity capture to assist the classification and isolation of target RNAs. In recent years, extensive data have proven the ability of affinity capture technologies to target specific RNAs within the cell. The heart of these approaches consist of the specific recognition and annealing of target sequences with corresponding oligonucleotide probes which are immobilized on solid beads. In hybridization-based technologies, the interaction typically serves to anchor the target to a solid phase for separation and strand amplification purposes. In some cases, binding may induce changes in the environment of specific fluorescent labels, which enables detection even in the absence of actual separation. We plan instead to apply affinity capture to cell lysates and total RNA extracts to isolate the material necessary to perform subsequent MS mapping and sequencing. In the case of lncRNAs, antisense captures could be employed in a step-wise fashion to first achieve the separation of entire pools of lncRNAs for
classification purposes, and then to isolate individual RNA species of interest for targeted PTM analysis.

Preliminary work ongoing in our laboratory utilized sets of magnetic beads derivatized with antisense oligonucleotide probes complementary to specific osmostress induced IncRNAs. After salt induction, isolated total RNA containing the osmo-induced IncRNAs was placed in a tube with the designed probes. After several wash steps to rid the captured material of any unbound RNAs, global profiling was performed on the targeted immobilized strands, which displayed heavily modified content. Coupling our PTM profiling with affinity capture could create a platform for the comprehensive analysis of RNA PTMs in biological samples, which would be capable of providing different levels of information, as needed. Indeed, the above example described the isolation of individual RNAs. However, different applications could be conceived in which the expression levels of an entire class of RNA can be monitored without further differentiation. In this case, class-specific beads could be prepared by utilizing multiple probes complementary to a large number of sequences characteristic of the class. Companies like Roche can supply libraries containing up to 2.1 million oligonucleotides complementary to user-defined sequences (see for example the SeqCap system used to enrich for human IncRNAs). Derivatizing a set of beads with this library would afford a very effective tool for capturing a representative sample of the desired class. Class-specific sets could be generated for any known type of RNA, based on accurate genomic information, thus providing a general classification tool viable for many other applications beyond PTM analysis. Alternative strategies could target classes characterized by a common sequence, such as mRNAs and IncRNAs that display prominent poly-A tails. In the absence of a common sequence, the desired class could be isolated according to
size by either electrophoresis or size exclusion, followed by ligase-assisted introduction of a
specific unique linker.\textsuperscript{347}

As required by the need to design appropriate antisense probes, the affinity capture
procedures will be expected to lead to the isolation of very specific RNAs with known (or at least
partially known) sequences. However, adding the next layer of information –the location of
modified sites– will require performing the sequencing of captured strands. Toward this goal, we
plan to replace exonucleases with appropriate endonucleases to obtain the hydrolytic
oligonucleotide products necessary to support mass mapping and sequencing. In these
techniques, intact RNA strands of interest are hydrolyzed to smaller fragments by site-specific
endonucleases, i.e. RNase T1 which cleaves at every guanosine. These digested products can
then be subjected to sequencing by direct infusion followed by multiple tandem mass
spectrometry steps. The gas-phase activation of oligonucleotide ions produces characteristic
series of fragments, which reveal the sequence of the precursor ion, including the position of
possible covalent modifications.\textsuperscript{120,123,348} The utilization of specific endonucleases capable of
cleaving larger RNA strands into smaller oligonucleotides enables one to cover the entire
sequence of the original RNA strand. We initially investigated a tRNA model of known sequence
to assess the figures of merit afforded by the MS mapping and sequencing approaches. Figure
\textbf{5-1} shows the ESI-MS spectrum of a sample mixture obtained by digesting \textit{E. coli} tRNA\textsubscript{Lys} with
RNase T1\textsuperscript{146}. The x15 enlargement in \textbf{Figure 5-1a} illustrates the isotopic resolution (i.e., 73,000 in
this case) achieved by the LTQ-Orbitrap instrument employed for the analysis. Given that these
types of analyses must be completed on genuine samples, non-stoichiometric modification can
be immediately detected by observing signals spaced by the incremental mass characteristic of
Figure 5-1. a) Mass mapping of a tRNA\textsuperscript{Lys} sample submitted to limited RNase T1 digestion. b) Tandem MS spectrum of the m/z 1239.9 precursor ion from the digest mixture. The x15 enlargement in panel a) shows two contiguous isotopic envelopes that were fully resolved under the selected experimental conditions. The x20 enlargement in panel b) enables one to appreciate the limited abundance of the m/z 1239.9 precursor ion employed in the tandem MS experiment. The observed fragments matched the A32:C46 region of tRNA\textsuperscript{Lys}. 
the PTM of interest. For example, the two isotopic envelopes in the x15 enlargement are spaced by exactly 16 Da (calculated according to their 4- charge state), which could potentially correspond to species with different base compositions, or to the same species before/after hydroxylation. If this were the case, the intensities of the envelopes would immediately reveal the ratio between modified and unmodified species, which could thus be employed to express the incidence of modification.\textsuperscript{167,169,170}

In order to obtain sequence information, each hydrolytic product is typically submitted to tandem MS that involves isolating the ion of interest in the mass spectrometer, followed by gas-phase activation to induce the fragmentation of successive phosphodiester linkages.\textsuperscript{123,348} An example of this concept is the tandem MS spectrum of a 17-nt product from the tRNA digest, which was isolated as [M-4H]\textsuperscript{4-} with m/z 1239.9 (Figure 5-1b). These representative data show characteristic ion series that matched the sequence of the A32:G46 section of tRNA\textsuperscript{Lys}. The fragmentation processes covered the entire length of the 17-nt precursor and provided full-sequence coverage. The spacing between successive members of the ion series confirmed the presence of mnm\textsuperscript{5}s\textsuperscript{2}U and t\textsuperscript{6}A modifications in position 35 and 38 of the original tRNA\textsuperscript{Lys} sample. It should be noted that tandem MS has been demonstrated for strands spanning up to 100 nt on a Fourier transform ion cyclotron resonance (FTICR) mass spectrometer,\textsuperscript{349} but oligonucleotides up to ~20-25 nt are typically considered routine. Full-size tRNA\textsuperscript{Val} spanning 76 nt was recently sequenced on a similar instrument without prior digestion step, in what is commonly called top-down analysis.\textsuperscript{222} We envision that limiting the reactivity of the endonucleases employed in the digestion step will produce hydrolytic fragments that are closer to this desirable size. This adjustment will make tandem MS analysis more efficient and will facilitate the sequencing of
progressively larger RNA targets.\textsuperscript{153,159} This could be achieved, for example, by using additives like piperidine and imidazole to increase charge state distributions of precursor ions and produce higher fragmentation yields due to increased intramolecular Coulomb repulsion. These analyses result in the detection of discernable RNA ions that can be readily fragmented and identified based on unique cleavage patterns.\textsuperscript{159}

The example in \textbf{Figure 5-1b} purposely focused on one of the least abundant species in the endonuclease digest to probe the sensitivity of the technique. In spite of its relatively low abundance, the precursor at \textit{m/z} 1239.9 displayed excellent signal-to-noise ratio and its fragmentation provided full-sequence coverage. This tandem MS experiment consumed only \textasciitilde{}135 fmol, which is well within the sample capacity afforded by antisense capture methods (i.e., 300 nmol per gram of beads, or per well). This figure of merit bodes well for the future combination of mass mapping and sequencing techniques with affinity capture to identify the sequence position of selected RNA targets. We envision that, with few variations, this combination will support the characterization of any individual RNA bearing the PTM of interest. These capabilities will allow us to pursue the functional investigation of the osmostress lncRNAs described above.

Moving forward, we are also investigating possible ways for integrating the capture/characterization workflow with the leading techniques of RNA analysis. We envision that the captured material could be submitted in parallel to RNA-seq analysis to validate the results of the MS experiments. In these types of techniques, a population of RNA is converted to cDNA fragments with adaptors that are attached to one end or both ends of the fragments. Each one of the fragments is then sequenced in parallel to create multiple reads, thus increasing the
chances of achieving full sequence coverage in shorter amounts of time. The availability of sequence information from RNA-seq analysis would corroborate the identity of target species and would provide a framework to facilitate the MS mapping and sequencing operations. For example, this information could be used to generate theoretical maps to be compared with the experimental data provided by MS mapping experiments. Any mass discrepancy could be tentatively attributed to the presence of PTMs. In this way, specific hydrolytic fragments could be identified as candidates for tandem MS analysis, thus absolving one from the need to re-sequence the entire RNA strand of interest. Further, the availability of sequence information will facilitate data interpretation in those cases in which full-sequence coverage was not achieved. As an added value, it is very possible that examination of the RNA-seq reads may lead to the identification of high-frequency stops that are typical signs of modification. In this case, direct comparison with the MS sequencing data would help assign the observed hotspots to specific modifications, thus expanding the applicability of massive parallel sequencing techniques to the analysis of additional PTMs.

With the goal of mirroring the high-throughput nature of the RNA-seq techniques, the entire capture/analysis workflow could be performed by using material immobilized on a 96-microtiter well plate. This type of platform is one of the industry standards for high-throughput robotics operations and matches typical microarray formats employed in “omics” technologies, which would make it ideal for studying large amounts of samples, such as the entire class of osmostress lncRNAs. The utilization of dedicated microfluidic robotics to take advantage of the microtiter well format would greatly facilitate PTM detection and quantification, classification and identification of parent RNA, and sequence localization. Dedicate microfluidics would
minimize sample losses during handling/transfer operations, which would greatly improve the overall sensitivity of the analysis. This type of implementation could increase the expected throughput, and prompt the possible interfacing with other techniques. For example, microarray technologies have been used to build massive arrays with oligonucleotide probes for virtually any target sequence, up to entire genomes, which can be utilized for many different purposes. Adopting a similar format would enable the utilization of our MS technology as a complementary readout to perform PTM analysis in parallel (or subsequent) to any of such applications. At the end, we believe that combining the capture and sequencing approaches with the MS platform would provide an excellent toolset for mining the epitranscriptome and elucidating the significance of PTMs in the RNA World.

5.3 ELUCIDATION OF PTM EFFECTS ON PARENT IncRNAs

The Hog1-dependence of X and ho5U described in the previous chapters clearly implicates their modifying enzymes in osmoadaptation. At the same time, the detection of these PTMs in salt-induced IncRNAs implicates them in the activities performed by these RNAs in osmostress response. For these reasons, it is important that we determine the effect of such PTMs on parent RNAs. This can be accomplished by confirming the identity of specific RNAs targeted by modifying enzymes. For example, in the case of PTMs for which there is a known biogenetic enzyme, the RNA substrate can be investigated fairly easily. Previous work in our lab has already successfully shown the ability of the platform to monitor PTMs that are specific to certain biogenetic enzymes. Additionally, this work has shown that knockouts of these enzymes have a direct effect on PTM content. Preliminary studies on specific RNAs isolated from total extracts has also shown

131
direct correlation with particular biogenetic enzymes. In order to continue these types of studies, specific strains from GE Dharmacon’s Yeast TAP-fusion ORF Collection will be employed to achieve specific biogenetic enzyme overexpression and purification. Each of the tagged proteins will be employed to establish specific binding interactions with any cognate RNA that may be present in the cell lysates, or in total RNA extracts from control samples and/or salt-treated cultures. The specific interaction will offer the ability to isolate desired RNA-protein complexes containing the TAP-tagged protein. Additionally, it will allow us an additional way to confirm the RNP interaction with the targeted IncRNA of interest. Indeed, following this step, samples will be affinity purified and the co-purifying RNA will be isolated from the complex and subjected to sequencing to confirm its identity.\textsuperscript{351}

An alternative approach to this method will involve the crosslinking of the tagged proteins to their target RNAs followed by purification and sequencing. The stability of crosslinked conjugates will enable the digestion of the unbound regions and the preparation of cDNA from the protected RNA for sequencing purposes. This procedure represents the basis for the crosslinking and cDNA analysis (CRAC) technique used to reveal not only the identity of bound RNAs, but also the sequence position of the protein-RNA contacts corresponding to the modification sites.\textsuperscript{352,353} This method works by immobilizing tagged RNA-protein complexes on solid supports for purification. Subsequently, complexes are treated with protease and purified further by affinity chromatography techniques. Ultimately, this enables the identification of the binding sites of putative RNAs with specific proteins.\textsuperscript{354}

This analysis could be accomplished on at least two biological replicates of co-purified/crosslinked RNA and total RNA from the same strain to RNA-seq following depletion of
ribosomal RNA using commercially available kits. These kits will be used to generate cDNA amplified by PCR with PCR primers to allow multiplexing and sequencing. The RNA-seq data will then be aligned with the yeast genome by using a publicly available sequence alignment programs. We will then apply rigorous statistical algorithms will be applied to assign significance of the RNA enrichment in the co-purified/crosslinked sample relative to total RNA and to the variation in the two biological replicates. Modification of co-purifying/crosslinked RNAs of interest, particularly ncRNAs with known functions, will be verified in parallel by using the MS profiling approach discussed extensively in the previous chapters. Ultimately, these comparisons will unveil the broadest possible view of RNA substrates that are associated with PTM expression.

Using the aforementioned techniques will enable us to investigate these RNA-protein interactions in the midst of different stressed states. In the instance of the candidate PTMs, X and ho$^5$U, the consequences of their absence on the stability and function of specific RNAs could be explored. Once we identify targeted RNAs, the level of these specific RNAs will be assessed in the presence and absence of the modifying enzyme related to the targeted PTMs by Northern blotting using strand-specific riboprobes. Genes that encode the specific class of PTM-bearing RNAs will also be considered, even if the above pull-downs did not detect them, with particular focus on RNAs that have a known and quantifiable function. For targets in which the sequence position of modification sites could be identified, introduction of a nucleotide substitution at the modification site in the gene that encodes the target RNA will be accomplished by site mutagenesis followed by gene replacement in yeast. Consequently, we will determine the effect of substituting the modified nucleotide on target RNA abundance.

In the case of target RNAs with known functions, we will compare the effects of depleting
the modifying enzyme and introducing a base substitution in the modification site of its target
RNA. In the context of osmostress, an antisense IncRNA is present in the \textit{CDC28} gene, whose
Hog1-induced expression mediates the production of the Cdc28 protein known to restore
progression through the cell cycle after high salt exposure.\textsuperscript{224} If the \textit{CDC28} antisense IncRNA were
identified as a target of X- or ho\textsubscript{5}U-modification, then depletion of the corresponding modifying
enzyme or introduction of a modification site substitution in the IncRNA, followed by
measurement of Cdc28 protein levels by Western blot analysis will be done. Finally, we will
explore the plausible role for the modification in cell cycle progression by using flow cytometry
determinations. We will implement similar approaches for the investigation of any target RNA
with known activities.

Knowledge of the biogenetic enzymes and their specific RNA targets is essential to the
elucidation of the significance of PTMs in biology. The mutation/deletion experiments that we
are employing will eliminate putative target RNAs and is expected to provide significant
information on the function of these modifications in stress response and, in the process, offer
valuable insights into their broader roles in the cell lifecycle. Yeast and osmostress have been
selected here as convenient general models for elucidating the interplay between PTMs and
parent RNAs. However, the functional studies accomplished will help establish general strategies
for probing such interplay and understanding its effects on essential cellular processes.
Undoubtedly, the combined approaches hit at the heart of the central dogma of biology by
enabling investigators to more fully understand expression of transcriptome components as they
relate to the genome and the proteome.
5.4 CONCLUSIONS

The overarching goal of this project was to develop a MS-based platform to globally survey the more than 100 PTMs that are found on naturally occurring RNAs. The fruition of this goal was achieved by the concerted application of high-resolution mass spectrometry and ion mobility spectrometry-mass spectrometry. Rigorous studies demonstrated the ability of monitoring the reproducibility, relative responses and quantitative values for each PTM in these complex mixtures. We have been able to monitor PTM content specific to various microorganisms, metabolic states, heightened stress states, and captured RNA subsets. Additionally, we have been able to confirm PTM relationships with biogenetic enzymes by matching these data with the information provided by publicly available interactomic networks and databases. Investigations into stress response pathways have unveiled important stress-induced PTMs, like ho\textsuperscript{5}U and X. Additionally, the hypothesis that PTMs may play major roles in these types of pathways has been investigated by monitoring PTM expression in the presence/absence of certain genes.

The strategies proposed in this chapter will provide the next layer of information necessary to determine the biological significance of each PTM. The preliminary data provided by the initial investigations have unveiled evidence of a direct and generalized involvement of PTMs in activities by parent RNAs. This realization is not surprising given that signs of PTM involvement are already present in the translational recoding performed by tRNA PTMs in response to oxidative stress. However, this preliminary data show that PTM involvement is much more extensive than generally thought. Therefore, the proposed activities will probe for this widespread participation of PTMs in ncRNA activities and will aim at identifying their
mechanisms. The ability to obtain both qualitative and quantitative information will drive the functional investigation of PTMs, which will take full advantage of the identification of their specific biogenetic enzymes and target RNAs. For this reason, the knowledge produced by these future studies will help establish a general framework for understanding the activities of PTMs in any type of RNA process. Overall, the stepwise approach (classification followed by individual identification) will enable the systematic analysis of all possible classes of RNAs, which may not be necessarily involved in stress response. At the same time, the universal nature of mass analysis will generate a wealth of precious information on the distribution of all known PTMs. Therefore, in addition to providing specific information on the osmostress PTMs considered here, this project has provided the necessary foundation by which to initiate the painstaking process of annotating the transcriptome, which will require marrying the unlimited supply of sequence information from genomic techniques with newly acquired knowledge of PTM identity and distribution. The annotation of the entire transcriptome will result in the unveiling of its multitudes of functions. Ultimately, this will lead the way for the discovery of new diagnostic avenues, drug targets, and drug therapies which will undoubtedly advance the medical field and result in the betterment of the human race.
REFERENCES


239. http://mods.rna.albany.edu/


