Characterization of two unique pathways for wyosine biosynthesis in Kinetoplastids

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Paul Joseph Sample, B.S.
Graduate Program in Microbiology

The Ohio State University
2014

Dissertation Committee:
Dr. Juan D. Alfonzo, Advisor
Dr. Jane Jackman
Dr. Michael Ibba
Dr. Birgit Alber
Abstract

Strategies for reading frame maintenance were likely prerequisites to establish the early genetic code and must have inevitably involved tRNA post-transcriptional modifications. Given their well-documented effects on translational fidelity, wyosine/hydroxywybutosine found in tRNA\textsuperscript{Phe} of Archaea and Eukarya probably played a central role as one of the drivers of translational accuracy. In its most common form in eukaryotes, hydroxywybutosine is the product of five highly conserved enzymes that use a methylated guanosine (m\textsuperscript{1}G\textsubscript{37}) as a precursor, but its function has been relegated to cytosolic translation. Here we present genetic, molecular and mass spectrometry data demonstrating the first example of a wyosine pathway in mitochondria, a situation thus far unique to the kinetoplastid lineage. We also show that under conditions of low glucose, when oxidative phosphorylation is most needed, mitochondrial wyosine and hydroxywybutosine become important for mitochondrial function, demonstrating the biological significance of these findings. Although two forms of wyosine/hydroxywybutosine-modified tRNA\textsuperscript{Phe} exist in \textit{Trypanosoma brucei} mitochondrion, the organellar pathway has features in common with that of Archaea. Based on molecular phylogeny arguments, we suggest that mitochondrial wyosine biosynthesis represents an ancestral pathway dating back
to the last common ancestor with the Archaea. These findings are discussed in the context of the extensive RNA editing in the trypanosomatid mitochondrion, whereby editing, in generating potentially “slippery” U-rich sequences, provided the selective pressure to maintain mitochondrial wyosine. (Chapter 2)

The second focus of this thesis discusses the development of a computer program, named RoboOligo, designed to automatically determine the sequence of modified RNA oligomers via assignment of peaks generated by tandem mass spectrometry. Using this program, a user can resolve the sequence of a modified RNA oligomer of up to 14 nucleotides, de novo – that is, in the absence of any prior sequence information. In total, 73 of 77 independently verified oligomers were correctly sequenced by the automated de novo sequencing algorithm. (Chapter 3)

The final topic discusses work on identifying proteins that associate with the *Trypanosoma brucei* proteins ADAT2 and ADAT3 (adenosine deaminase acting on tRNA). We have previously shown that cytosolic ADAT2/3 heterodimer is responsible for adenosine to inosine (A to I) editing of the first anticodon residue $A_{34}$ of a subset of *T. brucei* tRNAs and thereby generates the complete anticodon set by allowing the edited tRNAs to decode multiple codons. However, we hypothesize that ADAT2 and ADAT3 may play a ‘secondary’ role in the nucleus facilitated by associating with other nuclear proteins. Through the use of affinity-purification and mass spectrometry, we investigated this possibility by
analyzing proteins that co-eluted with an epitope-tagged ADAT3 and present possible candidates for future investigation. (Chapter 4)
Dedication

I would like to dedicate this document to my mother, stepfather, and sister for their advice, support, and love.
Acknowledgments

I would like to thank my fellow laboratory members for their professional support and friendship: Dr. Zdenek Paris for his experimental expertise, knowledge, and good humor; Dr. Kirk Gaston for his essential contributions in the development of RoboOligo and lessons in mass spectrometry; Dr. Jessica Spears for her help in establishing myself within the laboratory; Ian Fleming for his extensive experimental knowledge, fruitful discussions, and laboratory camaraderie; Scott Hinger for his assistance in immunofluorescence and general good humor; and Alan Kessler, Katie Anderson, and Raphael Soares who will undoubtedly continue the success of the laboratory.

I appreciate the generosity of Dr. Valerie de Crecy-Lagard, Dr. John J. Perona, and Dr. Uttam T. RajBhandary for their contributions of isolated tRNA used to test RoboOligo, and Dr. Patrick Limbach for his guidance.

I extend my gratitude to Katie and the Guagenti family for their support and kindness over the years.

Finally, I would like to thank Dr. Juan Alfonzo and Dr. Mary Anne Rubio, not only for their contributions in my development as a scientist, but also for being friends and confidants. I will remember my time in the Alfonzo Laboratory fondly.

vi
Vita

2003 ........................................ Springfield Catholic Central

2009 .............................................. B.S. Microbiology, Ohio State University

2009 to present ............................... Graduate Teaching and Research

Associate, The Ohio State University

Publications


Fields of Study

Major Field: Microbiology
# Table of Contents

Abstract ........................................................................................................................................... ii
Dedication ........................................................................................................................................ v
Acknowledgments ................................ ............................................ vi
Vita ................................................................................................................................................ vii
Publications ................................................................................................................................... vii
Fields of Study ........................................................................................................................... viii
Table of Contents ......................................................................................................................... ix
List of Tables ................................................................................................................................. xii
List of Figures .............................................................................................................................. xiii

## Chapter 1: Introduction

1.1 Transfer RNA ......................................................................................................................... 1
   1.1.1 tRNA Transcription ........................................................................................................ 6
   1.1.2 tRNA 5′ Maturation ....................................................................................................... 7
   1.1.3 tRNA 3′ Maturation ....................................................................................................... 9
   1.1.4 tRNA Introns .................................................................................................................. 11
   1.1.5 tRNA transport .............................................................................................................. 13
   1.1.6 tRNA Nuclear Export ................................................................................................... 14
   1.1.7 tRNA Mitochondrial Import ......................................................................................... 15

1.2 RNA Editing .......................................................................................................................... 18
   1.2.1 tRNA Editing ................................................................................................................ 19

1.3 tRNA Modifications .............................................................................................................. 24
   1.3.1 tRNA Modifications and Disease ............................................................................. 24
   1.3.2 Modified Nucleotides and tRNA Structure and Stability ........................................... 26
   1.3.3 tRNA Modifications and Identity ............................................................................. 28
   1.3.4 tRNA Modifications and Decoding ............................................................................ 33
   1.3.5 tRNA Modifications and Reading Frame Maintenance ............................................ 35
   1.3.6 Wyosine-derivatives Limit Ribosomal Frame Shifting .............................................. 38

1.4 Trypanosome Biology ........................................................................................................... 42
   1.4.1 Trypanosome Mitochondrion ....................................................................................... 43
   1.4.2 U-insertion/deletion Editing Within the Mitochondrion .............................................. 44

## Chapter 2: A New Biosynthetic Pathway for Mitochondria Unveils the Origin of Wybutosine in Eukaryotes

i x
# Table of Contents

2.1 Introduction........................................................................................................... 50
2.2 Results .................................................................................................................. 56
  2.2.1 Paralogous Wyosine Biosynthesis Genes in *T. brucei*............................... 56
  2.2.2 Mitochondrial Localization of the Wyosine Modification Enzymes .. 60
  2.2.3 Two versions of tRNA\(^\text{Phe}\) co-exist in mitochondria ....................... 65
  2.2.4 RNAi of TbTYW1L and TbTYW1S and status of OHyW and imG... 69
  2.2.5 The evolution of TYW1 .............................................................................. 76
2.3 Discussion ............................................................................................................. 78

Chapter 3: RoboOligo: Software for Mass Spectrometry Data to Support Manual and De Novo Sequencing of Post-transcriptionally Modified Nucleic Acids........... 82
  3.1 Introduction........................................................................................................... 82
  3.2 Results ............................................................................................................... 87
    3.2.1 Automated *De Novo* Sequencing......................................................... 87
    3.2.2 The Automated De Novo Sequencing User Interface ......................... 94
    3.2.3 Accuracy of automated *de novo* sequencing ..................................... 97
  3.3 Manual Sequencing......................................................................................... 104
  3.4 Variable Sequencing....................................................................................... 108
  3.5 Discussion ....................................................................................................... 111

Chapter 4: ADAT2 and ADAT3 – Roles Beyond A to I tRNA Editing?.......... 113
  4.1 Introduction........................................................................................................ 113
  4.2 Results .............................................................................................................. 117
    4.2.1 Affinity purification of epitope-tagged ADAT2 and ADAT3.......... 117
    4.2.2 MS Analysis of ADAT3-ProC Affinity Purification Elution ....... 119
  4.3 Conclusions..................................................................................................... 123

Chapter 5: Concluding Remarks and Future Directions......................................... 129
  5.1 Automated *de novo* sequencing of modified RNA oligomers........ 129
  5.2 A novel eukaryotic pathway for wyosine biosynthesis............................. 130
  5.3 Concluding Remarks...................................................................................... 134

Appendix A: Supplementary Figures .................................................................... 135

Appendix B: Supplementary Tables ........................................................................ 143

Appendix C: Materials and Methods....................................................................... 149
  C.1 DNA Miniprep ............................................................................................... 150
  C.2 Column Purification of tRNA ................................................................. 151
  C.3 RNA Preparation......................................................................................... 153
  C.4 Protein-C Purification.................................................................................. 155
  C.5 Trypanosome Mitotracker Immunofluorescence Protocol .................. 157
  C.6 Reactive Oxygen Species Assay............................................................... 159
  C.7 Reverse Transcription – Polymerase Chain Reaction (RT-PCR)......... 160
C.8 Preparing tRNA for Nucleoside Analysis via HPLC .......................... 162
C.9 Nucleoside Analysis via HPLC................................................................. 163
C.10 Silver Stain of Protein Gels................................................................. 165
C.11 Trypanosoma brucei Cyto and Mito Fraction Isolation Protocol.......... 167
C.12 Mitochondria Gradient Purification Protocol (Alfonzo Lab protocol).... 170
C.13 Mitochondrial Membrane Potential Assay........................................ 171

References .................................................................................................... 172
List of Tables

Table 1 Analysis of phenylalanine codon usage. .................................................. 48

Table 2 Subcellular localization prediction ............................................................. 60

Table 3 Summary of *E. coli* tRNA$^{\text{Asp}}_{\text{GUC}}$ and tRNA$^{\text{Gln}}_{\text{UUG}}$. sequencing .......... 101

Table 4 MS analysis of ADAT3 Affinity Purification ............................................ 125

Table 5 Protein accession numbers........................................................................ 144

Table 6 Automated sequencing results.................................................................... 145
List of Figures

**Figure 1.1** The standard genetic code ................................................................. 2

**Figure 1.2** The four main ribonucleic acids .......................................................... 4

**Figure 1.3** Transfer RNA secondary structure and numbering scheme ............... 5

**Figure 1.4** Crystal structure of native *Saccharomyces cerevisiae* tRNA\(^{\text{Phe}}\) ........ 6

**Figure 1.5** tRNA processing in trypanosomes .......................................................... 8

**Figure 1.6** Mitochondrial tRNA import in *T. brucei* ............................................. 17

**Figure 1.7** RNA deamination editing ..................................................................... 21

**Figure 1.8** Structures of G, m\(^1\)G, and imG derivatives ........................................ 40

**Figure 1.9** Structure of nucleotides G\(_{34}\) – A\(_{38}\) of *S. cerevisiae* tRNA\(^{\text{Phe}}\) ... 41

**Figure 1.10** Uridine-insertion/deletion in the trypanosome mitochondria ............... 46

**Figure 1.11** The edited mRNA of the mitochondrial ND7 .................................... 47

**Figure 2.1** Biosynthetic pathways for wyosine/wybutosine and derivatives ............ 54

**Figure 2.2** *Trypanosoma brucei* wybutosine-synthesizing homologs .................. 58

**Figure 2.3** Intracellular localization of TbTYW1L and TbTYW1S ......................... 62

**Figure 2.4** Intracellular localization of TYW2, TYW3A, TYW3B, and TYW4/5 ... 63

**Figure 2.5** Nucleoside analysis of wyosine derivatives in mitochondria .............. 66

**Figure 2.6** Identification of wyosine in the trypanosome mitochondrion .............. 68

**Figure 2.7** Wyosine status after TYW1L and TYW1S RNAi ................................. 71
Figure 2.8 TbTyW1S and TbTyW1L RNAi growth in the low glucose media ..... 74
Figure 2.9 RNAi of TbTYW1L and TbTYW1S affects mitochondrial function .... 75
Figure 2.10 TYW1 phylogenetic tree ................................................................. 77
Figure 3.1 Typical CID RNA oligomer fragmentation products ..................... 83
Figure 3.2 The logical design of the de novo sequencing algorithm ................. 90
Figure 3.3 Automated de novo sequencing interface ....................................... 96
Figure 3.4 Effects of increasing nucleotide pool size on sequencing efficiency . 97
Figure 3.5 Analysis of E. coli tRNA$^{\text{Asp}}_{\text{GUC}}$ and E. coli tRNA$^{\text{Gln}}_{\text{UUG}}$ ................. 100
Figure 3.6 Sequencing efficiency in relation to oligomer length ..................... 104
Figure 3.7 The RoboOligo primary interface ....................................................... 107
Figure 3.8 Variable sequencing interface .......................................................... 110
Figure 4.1 The three tRNA$^{\text{Thr}}$ isoacceptors ............................................... 114
Figure 4.2 Affinity purification of ADAT3-Pro C................................................ 119
Figure 4.3 SDS-PAGE ADAT3-ProC affinity purification............................... 120
Figure 5.1 Biosynthetic pathways of OHyW and imG in trypanosomes .......... 131
Figure A.1 TYW1 alignment.............................................................................. 136
Figure A.2 TYW2 alignment............................................................................. 138
Figure A.3 TYW3 alignment............................................................................. 139
Figure A.4 TYW4/5 alignment......................................................................... 140
Figure A.5 MS analysis of OHyWpA and imGpA from mitochondrial tRNA ..... 142
Chapter 1: Introduction

1.1 Transfer RNA

The process of producing proteins based on genes, the functional units of a genome, follows a specific order of events known as the 'Central Dogma of Molecular Biology' and can be stated simply as: DNA is transcribed to mRNA, mRNA is translated to protein. That is, genes in DNA are rewritten as an intermediary molecule, messenger ribonucleic acid (mRNA), which is then decoded and translated into proteins. The individual ‘words’ of DNA that are translated to peptides are spelled out as nucleotide triplets, forming a sequence called a codon. Each codon either encodes for one of twenty-two amino acids or is read as a signal that tells the translation machinery to stop and release the finished protein product (Figure 1.1). For each codon there exists a corresponding tRNA with an anticodon composed of three nucleotides that base pair with the codon. If the codon and anticodon match, then the ribosome catalyzes a peptidyl transfer reaction of the amino acid found in the acceptor stem of the tRNA to the growing polypeptide chain.
Figure 1.1 The standard genetic code.

For this document, we are primarily concerned with transfer ribonucleic acids (tRNA) — the molecules that decode messenger RNA during translation. Following transcription, tRNAs are composed of the four main unmodified ribonucleotides: cytidine (C), uridine (U), guanosine (G), and adenosine (A)
Shortly after transcription, the nascent tRNA molecule undergoes a series of processing events (discussed later in this chapter) that produces a mature tRNA that is approximately 75 nucleotides long. The secondary structure of tRNAs resembles the shape of a cloverleaf, beginning with the 5’ and 3’ ends at the top to form the acceptor stem, two central stem loops to either side forming the arms of the cloverleaf — the D and TΨC arms, respectively — and a third stem loop distal to the acceptor stem called the anticodon stem loop (Figure 1.3). The first crystal structure of a tRNA was that of the *Saccharomyces cerevisiae* tRNA$_{\text{Phe}}$, which revealed the now canonical tRNA L-shape. The crystal structure of *S. cerevisiae* tRNA$_{\text{Phe}}$ has since been updated with a greater resolution (Figure 1.4) (Shi and Moore, 2000).

Newly transcribed tRNAs, with an initial nucleotide composition of A, U, C and G, are incapable of participating in translation. Instead, in one of the processes of tRNA maturation, new chemical groups are added at specific nucleotides. These are called RNA modifications and effectively change the chemistry of tRNAs in much the same way that the different chemical groups of amino acids can affect the structure and function of proteins. tRNA modifications are critical for the overall function of tRNAs within the cell and are specifically important for maintaining tRNA structure, stability, and ensuring accuracy during translation (Phizicky and Hopper, 2010). As inferred by their evolutionary conservation, tRNA modifications were essential to the early establishment of life and continue to serve critical roles in cellular biology.
Figure 1.2 The four main ribonucleic acids. Structural depiction of A. adenosine (A), B. guanosine (G), C. cytidine (C), and D. guanosine (G). Atomic numbering scheme of purines: A and G; and pyrimidines: C and U. The numbering scheme of the ribose carbons differs with the addition apostrophe (‘), and are the same for the ribose moieties of G, C, and U.
Figure 1.3 Transfer RNA secondary structure and numbering scheme.
In congruence with other eukaryotes, trypanosome tRNA genes are transcribed by RNA polymerase III (RNA Pol III). Sequence elements “A” and “B” are intragenic for tRNA genes and serve as promoters (Schramm and Hernandez, 2002), not only to tRNA, but also for some small RNAs such as the...
U6 snRNA and 5S rRNA. The transcription factors (TF) that control RNA pol-III are understudied, but the trypanosome genome contains a potential homolog of the eukaryotic TFIIIB, while the highly conserved eukaryotic TFIIIA and TFIIIC have not been identified (Bindereif, 2012). Newly transcribed pre-tRNA have extra sequences at the 5’ (leader) and 3’ (trailer) ends that must be removed. The order of leader and trailer removal depends on the tRNA species, but the most prevalent maturation sequence begins with the 5’ end.

1.1.2 tRNA 5’ Maturation

Removal of the 5’ leader is completed by RNase P – commonly a ribonucleoprotein endonuclease that is highly conserved in each domain of life and is one of few ribozymes capable of multiple turnover (Altman and Kirsebom, 1999; Guerrier-Takada et al., 1983). Bacterial RNase P is predominantly composed of RNA with as few as one associated protein. Archaeal RNase P utilizes at least five proteins and the trend of less RNA and more protein involvement is continued with eukaryotic RNase P, where humans have at least ten associated proteins. In all of these cases, the RNA component acts as the catalyst (Jarrous, 2002). Curiously, some mitochondria utilize a different mechanism for 5’ leader removal that involves a protein-based enzyme that entirely lacks an RNA component (Holzmann et al., 2008). This class of RNA-less RNase P are now known as ‘proteinaceous RNase P, or PRORP (Taschner et al., 2012).
Figure 1.5 tRNA processing in trypanosomes. RNA polymerase III transcribes tRNA from the nuclear genome. The 5' leader sequence is removed by the proteinaceous RNase P 1 (PRORP1). Presumably, the tRNase Z homolog will then cleave the 3' trailer sequence. If the tRNA contains an intron (in trypanosomes, only tRNA$^{\text{Tyr}}$ has an intron), then it will be removed by a tRNA intron endonuclease and a tRNA ligase will rejoin the two halves. The tRNA is then ready for CCA addition at the 3' end by CCA nucleotidyltransferase – a required sequence for aminoacylation. During this maturation process, various tRNA editing and modifications steps are also taking place. The nearly mature tRNA is then exported to the cytoplasm by Xpo-t where additional editing and modifications will occur.

Many eukaryotic genomes, especially those within plant and protist families, are devoid of any identifiable RNase P RNA. Instead, growing evidence indicates that these organisms utilize PRORP homologs to remove 5' leader sequences of pre-tRNAs. Trypanosomes also lack the RNA component of RNase P, but encode two PRORP homologs, each single polypeptides: PRORP1 and
PRORP2. The *Trypanosoma brucei* PRORP1 localizes to the nucleus, cleaves 5’ leader sequences *in vitro* without any additional protein or RNA co-factors and, when expressed ectopically in yeast, can substitute for the yeast nuclear RNase P (Taschner et al., 2012). PRORP2 localizes to the mitochondrion and also cleaves 5’ leader sequences *in vitro*; however, in trypanosomes, tRNA 5’ end processing occurs in the nucleus and all mitochondrial tRNAs are imported from the cytoplasm (tRNA mitochondrial import is discussed in Chapter 1.1.7). This implies that the mitochondrial PRORP2 may have an alternative role from tRNA 5’ leader removal. In any case, all *T. brucei* pre-tRNAs are 5’ matured by the strictly proteinaceous PRORP1 and potentially PRORP2 (Taschner et al., 2012).

(Figure 1.5)

1.1.3 tRNA 3’ Maturation

Processing of the 3’ end is accomplished in two steps. The first involves the removal of the 3' trailer by tRNase Z. In yeast and many other eukaryotes, two tRNase Z genes are common, which interact with different substrates (Takaku et al., 2004) and are predicted to localize to different cellular compartments. tRNase ZS, found in all three domains of life, localizes to the nucleus in eukaryotes, while the strictly eukaryotic tRNase ZL contains N-terminal sequences that predict its localization to mitochondria (Dubrovsky et al., 2004) or chloroplasts (Vogel et al., 2005). The trypanosome tRNase Z has not been extensively studied, but a BLAST search using the two *Saccharomyces cerevisiae* tRNase Z paralogs as queries revealed that only one homolog is
identifiable within the trypanosome genome. The lack of a mitochondrial tRNase Z makes sense considering that all trypanosome mitochondrial tRNAs are encoded in the nucleus and are imported into the mitochondrial as fully mature molecules (Bindereif, 2012). (tRNA mitochondrial import is discussed in Chapter 1.1.7)

Alongside of tRNase Z, the RNA pol III transcript-binding protein La is an important factor in 3’ tRNA cleavage in yeast. The T. brucei La also binds to pol III transcripts; however, RNAi of La in T. brucei exhibits a lethal phenotype (Arhin et al., 2005; Foldynová-Trantírková et al., 2005), whereas deletion of La in S. cerevisiae is not lethal (Yoo and Wolin, 1997). In yeast, exonucleases compensate for the lack of La-tRNase Z-based 3’ trimming, allowing production of fully matured tRNAs (Yoo and Wolin, 1997). In trypanosomes, tRNA Tyr is the only tRNA with an intron, which must be removed before the tRNA can participate in translation. RNAi of the trypanosome La had no effect on 3’ trailer removal from tRNA Tyr, but did result in a four-fold decrease in intron removal, suggesting La has an additional role in tRNA splicing (Foldynová-Trantírková et al., 2005).

Following 3’ trailer removal, the sequence CCA must be added to the 3’ end. This sequence is universally conserved for all eukaryotic, bacterial, and archaeal tRNAs and is required for aminoacylation. This requires the enzyme CCA nucleotidyltransferase of which there are two classes: the archaeal class I and bacterial/eukaryotic class II; both add CCA in a manner independent of a
nucleic acid template. Instead, in a fascinating example of structural acrobatics, the active site only allows for C74 and C75 addition via CTP and then changes conformation to catalyze the addition of A76, using ATP as substrate (Hou, 2010). The importance of CCA at the 3’ terminus of tRNAs cannot be understated, for even in organisms where CCA is genomically encoded, and thus contain the sequence after 3’ trailer removal, the CCA-adding enzyme exists and serves as a way to repair tRNAs with damaged 3’ ends. This theme of quality control is extended by studies that show that tRNAs with damaged backbones are not as efficiently processed by the nucleotidyltransferase, limiting the potentially negative effects of aberrant tRNAs on translation and translation products (Dupasquier et al., 2008).

CCA nucleotidyltransferases have not been explored within trypanosomes. However, a homolog of the eukaryotic class II enzyme is present within the genome and all trypanosome tRNA genes lack encoded CCA 3’ ends, supporting the reasonable assumption that it is expressed and functions similarly to better studied examples in other organisms.

### 1.1.4 tRNA Introns

Along with the extra sequence in the 5’ leader and 3’ trailer, some tRNAs contain introns that must be removed. tRNAs bearing introns are found within bacteria, eukaryotes, and archaea and are most commonly found one nucleotide 3’ of the anticodon, but vary in length. Removal of the intron and the rejoining of the two tRNA halves, or splicing, is required for the use of intron-containing tRNA
in translation and the means by which cells splice introns depends on their lineage. Bacteria utilize self-splicing introns that do not require protein cofactors (Reinhold-Hurek and Shub, 1992). Archaea and Eukarya have protein-based tRNA splicing machineries; the endonuclease – composed of four subunits in yeast – measures and precisely cleaves the intron (Trotta et al., 1997). This produces a 5' tRNA half with a cyclic phosphate at its 3' termini and a 3' tRNA half with a hydroxyl group at its 5' termini that must be rejoined by an endonuclease. In eukaryotes, a cyclic phosphodiesterase rearranges the cyclic phosphate to the 2' hydroxyl group. A kinase phosphorylates the free 5' hydroxyl of the 3' tRNA half, which is then used by the tRNA ligase to create a phosphodiester bond to the 3' hydroxyl of the 5' tRNA half. A 2'-phosphotransferase then removes the 2' phosphate that was the result of the cyclic phosphate rearrangement. Bacteria and archaea utilize mechanistically different tRNA ligation pathways (Popow et al., 2012). Work on the trypanosome endonuclease is discussed in the following paragraph, but trypanosome tRNA ligase is yet to be identified (Bindereif, 2012).

All sequenced eukaryotic and archaeal tRNA^{Tyr} contain an intron 3' adjacent to the anticodon (Chan and Lowe, 2009). In fact, the only tRNA encoded within the T. brucei genome that contains an intron is that of tRNA^{Tyr} and, at 11 nucleotides, is one of the shortest ever discovered (Bindereif, 2012). In yeast, the tRNA^{Tyr} intron is required for modification of the uridine within the anticodon GUA to pseudouridine (Johnson et al., 1983), but in T. brucei the
relationship between the intron and RNA editing has taken a dramatic turn. The *T. brucei* tRNA\textsuperscript{Tyr} gene exists as a single copy with the intron sequence of, 5' -AGUAUGCGGGU- 3'. We have shown that tRNA\textsuperscript{Tyr} with the genomic intron is not cleaved by the *T. brucei* endonuclease until the intron undergoes a series of noncanonical edits of up to three positions within the intron, to produce the edited species: 5’ -AGU\textsubscript{U}UGC\textsubscript{A}GGU- 3’ and 5’ -A\textsubscript{A}U\textsubscript{U}UGC\textsubscript{A}GGU- 3’, representing 64% and 36% of sequenced pre-tRNAs (pre-tRNAs with the genomic sequence were not detected) (Rubio et al., 2013). The mechanism behind these unusual editing events remains to be determined, but it is clear that editing is required for tRNA\textsuperscript{Tyr} intron splicing and, thus, cell viability.

1.1.5 tRNA transport

tRNAs are transcribed and partly processed in the nucleus, but translation occurs in the cytoplasm, making tRNA export from the nucleus an essential process. Additionally, despite the need for tRNAs in mitochondrial protein synthesis, many mitochondrial genomes, such as those from *Toxoplasma*, *Tetrahymena*, yeast, mammals, and plants, lack at least one required tRNA species. Instead, these missing tRNA genes are encoded in the nucleus and must be imported into mitochondria after transcription, processing, and export from the nucleus (Simpson et al., 1989). What is unusual, however, is the complete absence of all tRNA genes from the trypanosome mitochondrial genome, which means that all mitochondrial tRNAs are imported from the cytoplasm (Schneider et al., 1994). These facts imply that there must be
mechanisms for the organized transport of tRNAs around cellular compartments. The next two sections will discuss what is generally known about tRNA transport in eukaryotes and what is known about these mechanisms in kinetoplastids, specifically.

1.1.6 tRNA Nuclear Export

The means by which tRNA are exported from the nucleus have been extensively studied, primarily in yeast and vertebrates. Before the export of tRNAs into the cytoplasm, the cell has evolved mechanisms to prevent under-processed or misfolded tRNAs from entering the cytoplasmic tRNA repertoire. The tRNA nuclear export machinery plays a significant role in this quality control process. The protein Xpo-t, a member of the Ran-binding β-importin family, and the GTPase Ran together facilitate the export of tRNA from the nucleus to the cytoplasm through nuclear pore complexes (Phizicky and Hopper, 2010). Xpo-t, when bound to Ran-GTP, recognizes and binds only those tRNAs that have been fully processed at the 3’ and 5’ ends – both the 3’ trailer and 5’ leader removed and after CCA addition to the 3’ end. Xpo-t recognizes the canonical tRNA L-shape structure and binds to the TΨC loop and D loop (Cook et al., 2009); this limits the recognition of tRNAs with unusual structure as export substrates by Xpo-t, preventing them from participating in cytoplasmic translation (Phizicky and Hopper, 2010). Another tRNA export quality control mechanism lies in the aminoacylation state of nuclear tRNAs. Once believed to be a strictly cytoplasmic event, aminoacylation of 19 of 20 (one representative tRNA for each amino acid)
*S. cerevisiae* tRNAs occurs in the nucleus before export as well as in the cytoplasm (Steiner-Mosonyi and Mangroo, 2004). In experimental scenarios where nuclear aminoacylation is prevented, export of such tRNAs is significantly reduced, suggesting that the export machinery monitors the aminoacylation status of tRNAs and those suitable for aminoacylation may also be suitable for translation and, thus, exported from the nucleus (Lund and Dahlberg, 1998; Steiner-Mosonyi and Mangroo, 2004).

Once bound to tRNA, the entire complex of Ran-GTP-Xpo-t-tRNA then travels through a nuclear pore and releases the tRNA after the hydrolysis of GTP to GDP. Ran-GDP and Xpo-t then return to the nucleus to continue the process of tRNA export (Phizicky and Hopper, 2010). Unfortunately, nuclear export of tRNA in kinetoplastids has not received much research attention, but homologs of Xpo-t-related export machinery are present within the *T. brucei* genome, suggesting that they may employ a similar strategy.

### 1.1.7 tRNA Mitochondrial Import

Once tRNAs reach the cytoplasm, they are free to undergo additional modification / processing steps, enter the cytoplasmic translation pathway, or be imported into the mitochondrial. In *Leishmania tropica*, the RNA import complex (RIC), a protein-based multimeric complex composed of 11 proteins that assemble at the mitochondrial inner membrane, is essential for transporting cytoplasmic tRNAs into the mitochondrion (Bhattacharyya et al., 2003; Mukherjee et al., 2007). However, this works has been called into question by a number of...
investigators. *T. brucei* mitochondrial import – the organism with which most of the work shown in this document pertains to – is perhaps better defined, currently, by what it is not. For example, the RNA interference of the *T. brucei* Rieske protein, which is essential for the function of the *L. tropica* RNA import complex, does not affect mitochondrial tRNA import (Paris et al., 2009). Furthermore, the voltage dependent anion channel (VDAC) in plants is essential for tRNA import, but in *T. brucei* down regulation of the VDAC homolog has no effect on mitochondrial tRNA import (Pusnik et al., 2009). *(Figure 1.6)*

Experiments conducted with isolated mitochondria from *L. tarentolae* have demonstrated that mitochondria are capable of *in vitro* import of tRNA in the absence of cytosolic factors (Rubio et al., 2000). Incubation of isolated *L. tarentolae* mitochondria with proteinase K inhibited tRNA import, suggesting that import is protein-mediated. In *L. tarentolae* (Rubio et al., 2000) and *T. brucei* (Paris et al., 2009) a mitochondrial membrane potential is not required for tRNA import, but the close relative *Leishmania tropica* (Mukherjee et al., 2007) does require a membrane potential. Again, this underscores the diversity of these systems between kinetoplastid species.

These *in vitro* import studies indicate that cytoplasmic factors play a limited role in tRNA import, however, there is one example that shows that this is not entirely the case. It appears that the eukaryotic cytoplasmic translation elongation factor (eEF1α) serves as a specificity determinant for the import of tRNA<sup>Met</sup>, tRNA<sup>Ile</sup>, and tRNA<sup>Lys</sup> and does so by detecting a specific base pair in
the TΨC-stem in these tRNAs. In fact, the introduction of this base pair in the strictly cytoplasmic tRNA^{Sec} caused it localize to the mitochondrion (Bouzaidi-Tiali et al., 2007).

**Figure 1.6** Mitochondrial tRNA import in *T. brucei*. No tRNA genes are encoded by the mitochondrial genome; they are instead encoded in the nuclear genome and must be imported into the mitochondrion. The cytoplasmic translation elongation factor, EF1α, recognizes a specific base pair in the T loop of a subset of tRNAs and facilitates their import – the mechanism has not been characterized. In plants, voltage-dependent anion channels (VDAC) are required for tRNA mitochondrial import, however, down-regulation of a *T. brucei* VDAC homolog had no effect on mitochondrial import. Similarly, the Rieske protein (RIC6) is required for tRNA import in the closely related *L. tropica*, but is not required in *T. brucei*. In vitro, tRNA has been shown to enter isolated *T. brucei* mitochondria without any soluble cytoplasmic materials.
1.2 RNA Editing

RNA editing is the post transcriptional change in the information content of an RNA molecule from its genomic DNA origin. There are two ways in which RNA editing can occur: single-nucleotide conversion, where the nitrogenous base is chemically changed such that it becomes another base; or through the incorporation or removal of nucleotides from an original RNA precursor. One of the best studied cases of single-nucleotide conversion RNA editing is that of the apolipoprotein B editing complex 1 (APOBEC1) in vertebrates. When expressed, APOBEC1 deaminates a specific cytidine of the apolipoprotein B mRNA to uridine (C to U editing). This has the effect of introducing an early stop codon, which truncates the translated product (Navaratnam et al., 1993). A similar protein, APOBEC3, is involved in inhibiting the proliferation of retroviruses in primates, including the human APOBEC3G which is a factor that restricts HIV proliferation (Sheehy et al., 2002). The mechanism of the APOBEC3G protein is the widespread use of cytidine deamination, where as much as 3% of the retroviral RNA genome is mutated (Mangeat et al., 2003).

The first example of RNA editing via post-transcriptional nucleotide incorporation into an RNA polymer was discovered in kinetoplastids (Benne et al., 1986). Within the mitochondrial genome of these organisms exist protein coding sequences that, when initially transcribed, are untranslatable. This is due to missing sequence elements, often within the coding region, that are corrected by the site-specific insertion of uridines (U-insertion) and, in the reverse process,
the site-specific deletion of genomically encoded uridines (U-deletion) from the precursor mRNA. The end result of this RNA editing process is an mRNA with a restored open reading frame that will translate into a functional protein. U-insertion / U-deletion editing in kinetoplastids is an important theme for this document and is covered extensively in Chapter 1.4.2. The slime mold, *Physarum polycephalum*, was the second organism found to edit mitochondrial mRNA by nucleotide insertion (Mahendran et al., 1991). Similar to kinetoplastids, these editing events restore mRNA open reading frames, but differ in the mechanism — editing is co-transcriptional in *P. polycephalum* — and in the types and quantities of nucleotides inserted — 90% of nucleotide insertion edits in *P. polycephalum* are single cytidines, while the rest are either single uridines or the dinucleotides CU, GU, AA, UU, GC, or UA (Bundschuh et al., 2011).

Both types of editing are used to affect the structure or function of tRNAs. The following section covers tRNA editing in general with kinetoplastid-specific information where applicable.

1.2.1 tRNA Editing

Our laboratory studies one of the most common tRNA editing events in all kingdoms of life: the adenosine to inosine (A to I) conversion at the wobble base (Crick, 1966) of certain tRNAs. In this reaction, an amine at N1 of adenosine is replaced by a carbonyl group to produce inosine (*Figure 1.7*), which has increased base-pair flexibility and can interact with uracil, adenosine, and cytosine, allowing tRNAs with inosine at the wobble position (*I*₃₄) to have
expanded decoding capacity. This, in turn, decreases the number of required
tRNA isoacceptors needed for translation, as a single tRNA can decode multiple
codons (Curran, 1995). In bacteria, the homodimer adenosine deaminase acting
on tRNA (ADATa), utilizing two bound Zn$^{2+}$ for catalysis, deaminates A$_{34}$ to I$_{34}$ of
a subset of tRNAs (Elias and Huang, 2005). We have shown that the enzyme
responsible for this activity in *T. brucei* is the heterodimer ADAT2/ADAT3
(ADAT2/3). The *T. brucei* ADAT2/3 is similar to the bacterial ADATa homodimer
in that it coordinates two Zn$^{2+}$, but differs in that the Zn$^{2+}$ jointly coordinated by
both ADAT2 and ADAT3 catalyzes A$_{34}$ to I$_{34}$ deamination, while the other Zn$^{2+}$,
coordinated solely by ADAT2, is proposed to play a structural role (Spears et al.,
2011).

In addition to the A$_{34}$ to I$_{34}$ editing *T. brucei* tRNA$^{Thr}_{AGU}$, a cytidine to
uridine (C to U) edit occurs two nucleotides 5’ of I$_{34}$ at position 32, as well as at
C$_{32}$ of the other two threonyl-tRNAs. This editing event takes place in the nucleus
before export into the cytoplasm and is the first example of C to U editing outside
of an organelle (Rubio et al., 2006). Curiously, RNAi of ADAT2 caused a
decrease in C$_{32}$ to U$_{32}$ editing. This observation, when coupled with data showing
that C$_{32}$ to U$_{32}$ editing occurs in the nucleus, suggests that ADAT2 may play a
direct role in the catalysis of this C to U conversion, perhaps by associating with
a yet to be identified cofactor (Gaston et al., 2007). This topic is explored further
in Chapter 4.
C to U editing works by a similar deamination reaction as A to I editing, and was first discovered in one tRNA in the mitochondria of marsupials (Janke, 1993). In this case, a mitochondrial tRNA gene has sequence features resembling tRNA\textsuperscript{Asp}, but lacks the typical aspartyl GUC anticodon. Instead, the anticodon for this tRNA gene is GCC — the anticodon for two out of four
mitochondrial glycine codons. Marsupial mitochondria correct this discrepancy by deaminating C\textsubscript{35} to U\textsubscript{35} in 50% of these tRNAs, producing two tRNA populations from a single tRNA gene: tRNA\textsubscript{Asp, GUC} and tRNA\textsubscript{Gly, GCC}. Furthermore, both tRNAs are properly charged by their cognate aminoacyl tRNA synthetases (Börner et al., 1996) and, in combination with the other mitochondrial tRNAs, complete the anticodon repertoire required for decoding all mitochondrial codons.

The trypanosome mitochondrion exhibits a similar C to U editing event, but this one changes the decoding ability of a mitochondrial tRNA. In the trypanosome cytoplasm, the codon UGA is used as a stop codon, while in the mitochondrion UGA codes for tryptophan. The anticodon for tRNA\textsuperscript{Trp, CCA}, is capable for decoding the canonical UGG tryptophan codon in the cytoplasm but, in its genomically encoded state, would be unable to decode the mitochondrial tryptophan codon UGA. As discussed in Chapter 1.1.7, all tRNA genes are encoded in the nucleus and must be imported from the cytoplasm into the mitochondrion, eliciting the question of how tRNA\textsuperscript{Trp} decodes the mitochondrial tryptophan codon UGA. Remarkably, C\textsubscript{34} of the imported tRNA\textsuperscript{Trp, CCA} undergoes C to U deamination editing in the mitochondrion, creating the new anticodon UCA, which allows the tRNA to decode the mitochondria-specific tryptophan codon through canonical Watson-Crick base pairs with UGA. The ability of U and G to form a ‘wobble’ base pair also allows for the edited tRNA\textsuperscript{Trp} to decode the other, more common tryptophan codon UGG (Alfonzo et al., 1999). The enzyme responsible for this editing event remains to be identified.
As mentioned previously, editing can also involve the incorporation or removal of nucleotides from a tRNA. In *Acanthamoeba castellanii*, 16 out of 19 tRNAs encoded in the mitochondrial genome contain mismatches within the last three base pairs of the acceptor stem (1:72, 2:71, and/or 3:70) that would disrupt the standard helical structure of this region (Lonergan and Gray, 1993). These mismatches are corrected post-transcriptionally by an exonuclease that recognizes and cleaves the mismatched nucleotides from the 5’ end of the tRNA and is followed by the activity of a 3’ to 5’ nucleotidyltransferase that incorporates nucleotides that restore Watson-Crick base pairing with the 3’ end that also serves as the template (Gray, 2003).

Another 3’ to 5’ nucleotide polymerase catalyzes the addition of a single guanosine to the 5’ end of tRNA\textsuperscript{His}, creating the nucleotide G\textsubscript{1}. This RNA editing process occurs in eukaryotic tRNAs and is used as a means for the histidyl tRNA synthetase to recognize tRNA\textsuperscript{His} and proceed to charge it with histidine (Jackman et al., 2012) (The topic of tRNA identity is discussed in Chapter 1.3.3). The enzyme that facilitates G\textsubscript{1} addition, tRNA\textsuperscript{His} guanylyltransferase (THG1), was initially identified in *S. cerevisiae* and, unlike the activity of the *A. castellanii* 3’ – 5’ nucleotidyltransferase, the enzyme that completes G\textsubscript{1} addition does so in a template-independent manner (Gu et al., 2003). Interestingly, trypanosomes do not encode a THG1 homolog and, at the time of publication, the status of the 5’ end of tRNA\textsuperscript{His} (whether or not it contains G\textsubscript{1}) has not yet been investigated.
1.3 tRNA Modifications

The post transcriptional modification of tRNA is of paramount importance, as is evident in the amount of cellular resources invested in these processes. Nearly 100 modified nucleotides have been found on tRNAs (Jühling et al., 2009) with a median of 8 modified nucleotides per tRNA molecule (Phizicky and Alfonzo, 2010). Seventy-five different modifications decorate specific positions of tRNAs providing a wealth of chemical properties that would be unachievable if only the four main nucleotides were used. Information on the precise order of nucleotide modification events is often limited to specific modifications within specific organisms and these events are likely to vary significantly from system to system. Furthermore, research on the spatiotemporal order of modification events has been limited, but it is known via localization experiments that tRNAs are modified in the nucleus, cytoplasm, mitochondria, and chloroplasts. For these reasons, the following sections will break away from the chronological order of tRNA processing in favor of providing a broad background of the diverse roles of the modified nucleotides found within tRNA and how they relate to the biology of the cell.

1.3.1 tRNA Modifications and Disease

Numerous human diseases have been attributed to deficiencies in tRNA modifications. For example, five mutations in the mitochondrial gene encoding tRNA\textsubscript{Leu\_UUA} were each identified as the causative agent of MELAS (Mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes). In healthy
individuals, $U_{34}$ of $\text{tRNA}^{\text{Leu}}_{\text{UA}}$ is modified to 5-taurinomethyl ($\tau m^5U$) and is required for efficient wobble decoding of the leucine codon UUG (Kirino et al., 2005). Similarly, the mitochondrial disorder MERRF (mitochondrial epilepsy with ragged red fibers) is caused by point mutations in mitochondrially-encoded tRNA$^{\text{Lys}}$(UUR); for 80% of people afflicted with MERFF, a single A-to-G transition at position 8344 (Becker et al., 2013) prevents $U_{34}$ from being modified to 5-taurinomethyl-2-thiouridine ($\tau m^5s^2U$), again affecting wobble base pairing during translation. The deletion of MTU1, the gene responsible for the thiolation at C-2 of the same nucleotide, however, does not affect mitochondrial translation (Sasarman et al., 2011). This example highlights the diversity in the essentiality of tRNA modifications, where some are required for viability while the loss of others, perhaps only in isolation, have no effect on cellular function within tested experimental contexts.

Modified ribonucleosides have also been proposed as potential markers for disease. As animal cells breaks down rRNA and tRNA to free nucleosides, the modified ribonucleosides are concentrated in the kidneys and excreted as waste in urine. These nucleosides are not recycled, likely due to the lack of enzymes capable of removing modifications and thus preventing their salvage and future use as unmodified A, U, C, and G. Patients with certain types of cancer, for example, show increased levels of various modified nucleosides (Xu et al., 1999). This is probably due to an overall increase in translation, which also means an increase in tRNA and rRNA (Marvel et al., 1994). A method has been
recently developed for the precise quantitation of modified nucleotide using liquid chromatography – tandem mass spectrometry (LC-MS/MS) (Chan et al., 2010). A proof-of-principle demonstration in *Saccharomyces cerevisiae* revealed the dynamic extent of specific tRNA modifications in response to four different types of cellular stress. For example, the amount of $m^5C$, $Cm$, $m_2G$, and $t^6A$ increased significantly after exposure to hydrogen peroxide. This led to the investigation of $m^5C$ at position 34 of tRNA$^{\text{Leu}}_{\text{CCA}}$ (the anticodon wobble base) and how it selectively increases the translation of genes enriched in the TTG codon. In fact the ribosomal protein Rpl22a (100% of Leu codons are TTG) is expressed higher, at the point of translation, after oxidative stress exposure compared to the expression of the homolog Rpl22b (34% of Leu codons are TTG). Knockout of Rpl22a causes an increase in susceptibility to hydrogen peroxide exposure, thus, the increase in the amount of $m^5C_{34}$ of tRNA$^{\text{Leu}}_{\text{CCA}}$ causes in increase in expression of genes involved in responding to oxidative stress (Chan et al., 2012). This method of quantifying the dynamics of tRNA modifications could lead to new avenues of disease research and has already been used to identify tRNA modification biosynthetic pathways (Chan et al., 2010).

1.3.2 Modified Nucleotides and tRNA Structure and Stability

The central region of tRNA is considered to be the structural core. It is the interactions of nucleotides within it that determine the overall shape of the folded tRNA and modified nucleotides contribute significantly to this structure. One of the simplest and most commonly used modifications is the addition of methyl
groups, either on the ribose moiety or at a position on the base that disrupts the
nucleotides ability to form a Watson-Crick base pair (Motorin and Helm, 2010).
One such modification is that of 1-methyladenosine (m1A) at position 9 of the
human mitochondrial lysine tRNA. The absence of the methylation disrupts the
ability of the tRNA to fold in the canonical tRNA L-shape (Helm et al., 1999) and
instead causes it to form an extended hairpin that is incapable of being
aminoacylated by the lysyl-tRNA synthetase (Sissler et al., 2004). The
methylation of A9 occurs early in the tRNA maturation process (Helm and Attardi,
2004), which is logical when considering that the tRNA maturation machinery,
such as RNase P, recognize the canonical tRNA L shape (Mondragón, 2013).

Three of the most conserved tRNA modifications occur in the D-loop and
TΨC-loop: dihydouridine at position 16 and 17 in the D-loop and ribothymidine
and pseudouridine in the TΨC-loop. The hydrogen bonding between G18-Ψ55
and G19-C56 of the two loops are considered to be important for proper tRNA
folding and stability, but studies in yeast have shown that deletion of the gene
responsible for Ψ55, pus4, is not lethal – a surprising result given that all native
tRNAs contain a pseudouridine at position 55 (Grosshans, 2001). The role of
pseudouridine in tRNA structural stability at other positions is clearer. The human
tRNA_{Lys}^{UUU} contains Ψ at position 39, three nucleotides away from the anticodon,
that base pairs with the adjacent A31 – the last base pair before the anticodon
stem loop. At a physiological pH, tRNA_{Lys}^{UUU} with Ψ39 instead of U39 displayed an
increase in the melting temperature (Tm) of the tRNA by 4.7 °C (Durant et al.,
The increase in thermostability of hairpins containing pseudouridine is not due to changes in base pairing interactions, but is instead due to heightened base stacking with neighboring nucleotides, especially if those nucleotides are purines (Davis, 1995).

A modification found in the extremophile *Thermus thermophilus* tRNAs, 2-thiolation of T₅₄ (s²T₅₄) within the T-loop, increases the thermostability of tRNAs by 3 degrees Celsius and the amount of the s²T₅₄, relative to unmodified T₅₄, increases as the environmental temperature of *T. thermophiles* is raised. The added thiol group to T₅₄ stabilizes the C3'-endo conformation of its ribose moiety and allows for a reverse Hoogsteen interaction between it and A₅₈ (Yokoyama et al., 1987). Interestingly, experiments comparing *E. coli* tRNA¹⁰⁰ME₅₄(T₅₄), tRNA¹⁰⁰ME₅₄(T₅₄), and *T. thermophilus* tRNA¹⁰⁰ME₅₄(s²T₅₄) show a step-wise increase in thermostability with each modification: U₅₄ to T₅₄ increases the Tₘ by 6 °C and T₅₄ to s²T₅₄ increases the Tₘ by 7 °C (Davanloo et al., 1979). This property, and the fact that the mesophile *E. coli* utilizes T₅₄ while the thermophile *T. thermophilus* utilizes s²T₅₄, suggests that there exists an ideal balance between the two extremes of tRNA thermostability and that it can be controlled by ribonucleoside modification.

1.3.3 tRNA Modifications and Identity

The correct charging of a tRNA with its intended amino acid is as equally important to translation fidelity as is the proper codon-anticodon interaction between mRNA and tRNA. Aminoacyl tRNA synthetases (aaRS) are a class of
enzymes that serve the purpose of recognizing their substrate tRNA and supplying it with its cognate amino acid. The aaRS first binds its amino acid and a single ATP molecule and then uses the energy of hydrolyzing the ATP to covalently attach the amino acid to AMP. Hydrolysis of AMP catalyzes the addition of the amino acid to either the 2’ hydroxyl or the 3’ hydroxyl, depending on the class of the aaRS. Class I transfers it to the 2’ hydroxyl, which later moves the amino acid to the 3’ hydroxyl by transesterification and Class II transfers it directly to the 3’ hydroxyl. The product of an aaRS is a charged tRNA, which can then be used during translation.

The recognition of a tRNA by its cognate aminoacyl tRNA synthetase is dependent on ‘identity elements’ – structural features that must be conserved between tRNA isoacceptors that allow an aaRS to differentiate between cognate and non-cognate tRNAs. These elements can be read positively as ‘determinants’, the presence of which promotes proper aminoacylation and negatively as ‘anti-determinants’, where the presence of an element discourages mischarging. The identity elements of all 20 of the \textit{E. coli} aminoacylation systems, nearly all of the \textit{S. cerevisiae} identity elements, and fewer elements in other organisms have been determined and, when combined, highlight common themes (Giegé et al., 1998).

The majority of identity elements involve the four unmodified ribonucleotides A, U, C, and G. As might be expected, bases in the anticodon are often vital for ensuring proper aaRS recognition and represent some of the
The most conserved elements shared between Archaea, Bacteria, and Eukarya. For example, in order for tRNA^{Asp}_{GUC} to be charged, G_{34}, U_{35}, C_{36}, and C_{38} are required for all three domains of life. The same is true for tRNA^{Phe}, tRNA^{Gly}, and tRNA^{Thr}; however, these tRNAs have identity elements elsewhere that do differ from domain to domain and species to species. In *E. coli*, of the 16 tRNAs with determinants in the anticodon, all utilize position 35 as an identity element. Similarly, *14 S. cerevesiae* tRNAs have identity elements in their anticodon region (Giegé et al., 1998).

Identity elements in the acceptor stem are also widely used. Perhaps the most extensively studied determinant in this region is that of G_{-1} of tRNA^{His}. In all domains, tRNA^{His} contains an extra nucleotide, a guanosine, at its 5' terminus. Bacteria encode the extra base in tRNA^{His} gene, which is retained after 5' end maturation due to tRNA^{His}-specific aberrant activity of RNase P (Himeno et al., 1989; Orellana et al., 1986; Rosen and Musier-Forsyth, 2004). As discussed in Chapter 1.2.1, eukaryotes post-transcriptionally add a guanosine to the mature 5' end by tRNA^{His} guanylyltransferase (THG1) (Gu et al., 2003; Nameki et al., 1995) and archaea utilize either strategy (Heinemann et al., 2012). In any case, the uniqueness of G_{-1} of tRNA^{His} allows it to be easily differentiated by His RS.

The nucleotide immediately 5' of CCA, position 73, is another commonly used identity element in the acceptor stem. This nucleotide, referred to as the ‘discriminator base’, was hypothesized to be a primary participant in cognate aaRS identification by subdividing tRNAs into groups that represent the chemical
nature of the amino acids to which they belong — amino acids that are chemically similar will have the same discriminator base (Crothers et al., 1972).

For the most part, this hypothesis is supported by numerous studies, but the importance of \( N_{73} \) in cognate aaRS recognition varies from tRNA to tRNA (Hou, 1997).

There are fewer instances where a modified base is used as an identity element, which is surprising given the chemical uniqueness that modifications can provide. This observation may be a case of underrepresentation as the majority of tRNA and aaRS recognition studies have been conducted with *in vitro* transcripts and thus lack all tRNA modifications, perhaps glossing over the importance of modifications *in vivo*. At any rate, the theme of using the anticodon nucleotides as identity elements is continued in the relationship of modifications and tRNA recognition by cognate aaRS.

For example, position 34 in the anticodon of the *E. coli* tRNA\(^{\text{Ile}}\) is modified from cytidine (C\(_{34}\)) to lysidine (L\(_{34}\)), which involves the replacement of the cytidine carbonyl group for the amino acid lysine. The isoleucine aaRS (IleRS) is incapable of charging tRNA\(^{\text{Ile}}\)CAU in the absence of L\(_{34}\); instead, the methionine aaRS (MetRS) recognizes the position 34 unmodified tRNA and mistakenly charges it. IleRS can charge both isoleucine codons, GUA and LUA, which indicates that the nature of the nucleotide at position 34 is not used as a positive determinant for aminoacylation by IleRS, but rather as anti-determinants to
prevent the misacylation of both tRNA\textsuperscript{Ile} isoacceptors by MetRS (Muramatsu et al., 1988).

This theme of using modified nucleotides as anti-determinants to prevent misacylation of tRNAs is continued with \textit{E. coli} tRNA\textsuperscript{Asp}. \textit{In vitro}, unmodified transcripts of tRNA\textsuperscript{Asp} are robustly charged by the arginyl-tRNA synthetase (ArgRS), whereas fully-modified native tRNA\textsuperscript{Asp} shows very little mischarging by ArgRS. The introduction of a single methylation at G\textsubscript{37} of tRNA\textsuperscript{Asp} to form 1-methylguanosine (m\textsuperscript{1}G\textsubscript{37}) showed a 400-fold decrease in mischarging by ArgRS and retained efficient charging by the cognate aspartyl-tRNA synthetase (AspRS). This single modification restored the aminoacylation specificity between cognate and non-cognate aaRSs to the same levels as the fully-modified tRNA\textsuperscript{Asp} and demonstrates the importance of m\textsuperscript{1}G\textsubscript{37} of tRNA\textsuperscript{Asp} as an anti-determinant in the misacylation by ArgRS (Putz, 1994). This same strategy may also be conserved in eukaryotes, as similar results were observed with \textit{S. cerevisiae} tRNA\textsuperscript{Asp} (Perret et al., 1990).

In eukaryotes, the canonical start codon for translation, AUG, is decoded by the initiator methionyl-tRNA (tRNA\textsuperscript{iMet}), which begins polypeptide synthesis with a methionine. Another eukaryotic methionyl-tRNA decodes AUG codons that occur within protein coding sequences and are called elongator methionyl-tRNA (tRNA\textsuperscript{Met}) (Smith and Marcker, 1970). tRNA\textsuperscript{iMet}, alone, specifically targets start AUG codons which it identifies in the ribosomal P site. Elongator tRNA\textsuperscript{Met}, on the other hand, is specific for decoding AUG codons in the A site of the ribosome.
and can only do this as a complex with the eukaryotic elongation factor-1α (eEF-1α), which leads to the question: how does eEF-1α recognize tRNA\textsuperscript{Met} and discriminate against tRNA\textsuperscript{iMet}? For the yeast initiator methionyl-tRNA, the adenosine at position 64, found within the TΨC stem, is phosphorylated at the 2'\textsuperscript{-}O-ribose (Ar(p)). It was shown that when the gene responsible for the modification, RIT1, is down-regulated, eEF-1α is capable of binding tRNA\textsuperscript{iMet}. In fact, this Δrit1 mutant is viable with the additional deletion of the elongator tRNA\textsuperscript{Met}, showing that tRNA\textsuperscript{iMet} could be used for translation initiation as well as elongation, provided simply by the lack of the phosphorylated ribose of A\textsubscript{64} (Aström and Byström, 1994)

The trypanosome mitochondria-imported tRNA\textsubscript{Trp}\textsubscript{CCA} undergoes C\textsubscript{34} to U\textsubscript{34} deamination, allowing the tRNA to decode the mitochondria-specific tryptophan codon UGA, as well as the typical UGG tryptophan codon (see Chapter 1.1.8). These two tRNA\textsubscript{Trp} isoacceptors are independently charged by two separate tryptophanyl-tRNA synthetases; TrpRS1 charges the cytosolic tRNA\textsubscript{Trp}CCA and TrpRS2 charges the mitochondrial tRNA\textsubscript{Trp}UCA. In addition to the U\textsubscript{34} editing, tRNA\textsubscript{Trp}UCA contains a mitochondria-specific thiol modification at U\textsubscript{33} (s\textsuperscript{2}U\textsubscript{33}). These two unique features act as negative determinants and prevent charging of tRNA\textsubscript{Trp}UCA by the cytosolic TrpRS1 (Charrière et al., 2006).

1.3.4 tRNA Modifications and Decoding

The anticodon loop is a seven nucleotide-long single-stranded structure with two nucleotides 5' and two nucleotides 3' of the three nucleotides that
compose the anticodon. As ribosomes process mRNA during translation, tRNAs are rapidly sampled in the ribosomal A-site where tRNA anticodon meets the mRNA codon and, based on how they interact, will determine whether or not the tRNA-associated peptide will be joined to the nascent polypeptide. The selective process of choosing cognate from noncognate tRNAs is acted out by the 30S ribosomal subunit, which is capable of analyzing the codon-anticodon duplex that may or may not form. When tRNA meets its cognate codon, the highly conserved bases A1493 and A1492 of the ribosomal 16S RNA are structurally flipped from an internal loop structure to form interactions with the first and second base pairs of the codon-anticodon helix. The third base position of the codon is monitored by G530 and can accommodate non-Watson-Crick base pairs and thus allows for wobble base pairing (Crick, 1966; Ogle et al., 2001). Correct codon-anticodon pairing leads to peptidyl transfer and subsequent steps of translation.

Modifications that occur in the anticodon loop generally serve to either affect anticodon decoding (by changing specificity: broadening recognition or reducing recognition) or affect structural properties of the anticodon, usually by constraining its ability for movement (Ogle et al., 2001). For example, U34 of tRNAs is often thiolated to form s^2U34. This simple modification increases the tRNAs specificity for codons ending with A rather than G by increasing the stability of an s^2U-A base pair by 6 fold and destabilizing an s^2U-G base pair by 2-fold (Testa et al., 1999). Furthermore, in E. coli tRNA_Lys, s^2U34 is additionally modified to 5-methylaminomethyl-2-thiouridine (mcm^5s^2U34) which increases
base stacking between nucleotides 34 and 35 and promotes the proper anticodon structure for interacting with its cognate codon (Durant et al., 2005).

In the previous chapter, the role of the lysidine modification of tRNA$^{\text{Ile}}_{\text{CAU}}$ in *E. coli* IleRS recognition and charging was discussed. This modification is also, perhaps, one of the best examples of a modification that affects tRNA decoding. In bacteria and archaea, the isoleucine codon of AUA is decoded by an isoleucyl-tRNA with the anticodon CAU that is actually the canonical anticodon for methionyl-tRNA which decodes the Met codon AUG. Thus, the genomically encoded tRNA$^{\text{Ile}}_{\text{CAU}}$ anticodon sequence cannot decode AUA, due to the mismatch between C34 of the anticodon and the first A of the codon. To circumvent this problem, the oxygen at C2 is substituted for the amino acid lysine to form the modified base lysidine ($k^2$C34). Remarkably, the lysidine forms a stable base pair specifically with the Watson-Crick face of adenosine, effectively changing the identity of C$_{34}$ of isoleucyl-tRNA such that it can decode the isoleucine codon AUA (Grosjean and Björk, 2004; Muramatsu et al., 1988).

1.3.5 tRNA Modifications and Reading Frame Maintenance

To produce proteins that accurately reflect their genomic sequences, the ribosome must process mRNA codon-by-codon without slippage into an alternative reading frame – otherwise slippage could cause aberrant peptide synthesis that would likely lead to a nonfunctional protein. There exist mRNA sequences that are particularly prone to ribosomal frameshifting, termed ‘slippery sequences’. When such sequences are encountered during translation the
ribosome has an increased chance of slipping into an alternative reading frame, from which point the resulting amino acid sequence will differ from the amino acid sequence of the original reading frame. In fact, these sequences cause ribosomal frameshifting so predictably that numerous viruses use them to encode two proteins with overlapping mRNA sequences — frameshifting at a defined site can, for example, bypass an early stop codon to allow the synthesis of additional peptides to produce a longer protein with different properties (Brierley and Dos Ramos, 2006; Brierley et al., 1987; Jacks et al., 1988). Slippery sequences used by viruses often have a pseudoknot structure downstream that can dramatically increase the propensity to frameshift. For instance, the HIV-1 sequence 5’ – UUUUUUA – 3’, which alone promotes frameshifting, showed a five- to tenfold increase in frameshifting when the infectious bronchitis virus pseudoknot was placed downstream of it (Brierley et al., 1992). Although often utilized by viruses, ribosomal frameshifting is not limited to viruses, instead it is an intrinsic property of translation.

Immediately 3’ of the tRNA anticodon lies nucleotide 37 - commonly a purine that is modified to become more hydrophobic. This causes increased base stacking of the anticodon, especially when the neighboring nucleosides are also purines, and stabilizes the interaction between the first base pairs of the codon-anticodon helix. Generally, modifications at 37 limit the flexibility of the anticodon loop and contribute to making the anticodons of each tRNA structurally uniform, despite having different sequences (Ledoux and Uhlenbeck, 2008). This property
allows the ribosomal A-site to maintain an accommodating structure for incoming tRNAs without making tRNA-specific structural adaptations, which is proposed to aid in the rapidity of translation (Agris, 1996, 2008)).

Early investigations in *Salmonella typhimurium* showed that ribosomal frameshifting increases when the gene responsible for the RNA modification 1-methylguanosine at position 37 of several tRNAs (m$^1$G$_{37}$) was deleted (Björk et al., 1989). The authors proposed that G$_{37}$ could base pair with a C or U that is 5’ adjacent to the intended codon and if this were to occur the ribosome could slip backwards to produce a -1 frameshift. Methylation at N1 of guanosine, however, disallows Watson-Crick base pairing with cytidine and uridines, which would negate the ability of m$^1$G$_{37}$ to form a base pair and thus prevent frameshifting. Due to its simplicity, this is a logical and attractive model but biochemical evidence required to support it remains elusive. m$^1$G$_{37}$ is found on eukaryotic, bacterial, and archaeal tRNAs and is considered to be a ‘primordial’ modification, suggesting that strategies for reading frame maintenance were established in the last common ancestor to all life on Earth (Björk et al., 2001).

The point at which frameshifting occurs is another area of active research. Urbonavicius *et al.* posited that hypomodified tRNA induce frameshifting by either causing the ribosome to pause as it enters the A site or after translocation to the P site, which then leads to slippage into an alternative reading frame. In a clever experiment conducted in *E. coli*, it was shown that tRNA$^{Leu}$ decoding CUC and tRNA$^{Arg}$ decoding CGG without m$^1$G$_{37}$ are 30% and 290% more likely to
frameshift after translocation into the P site (Urbonavicius et al., 2001). However, the same tRNA\textsuperscript{Leu} lacking m\textsuperscript{1}G\textsubscript{37} shows similar levels of increased frameshifting as the tRNA enters the A site while decoding CUU, CUA, and CUG. At the risk of obfuscating things further, the lack of some modifications at position 37, such as ms\textsuperscript{2}io\textsuperscript{6}A\textsubscript{37} of tRNA\textsuperscript{Tyr}, produces frameshifting as the hypomodified tRNA enters the A site, as well as when it translocates to the P-site (Urbonavicius et al., 2001). This may indicate that the means by which these modifications contribute to limiting frameshifting is not uniform for all contexts, but rather that they contribute to reading frame maintenance in a sequence dependent manner.

Although the biophysical mechanics of frameshifting are unclear, the data overwhelmingly shows that modified nucleotides at position 37 of tRNAs are important for reading frame maintenance.

### 1.3.6 Wyosine and Wyosine-derivatives Limit Ribosomal Frame Shifting

The codons UUU and UUC are decoded by the single tRNA\textsuperscript{Phe} isoacceptor (anticodon: GAA). In nearly all Eukarya, the nucleotide at the position immediately 3' of the tRNA\textsuperscript{Phe} anticodon (position 37) is wybutosine, a highly conserved modification with several naturally occurring derivatives (Figure 1.8). The role of wybutosine was assessed in *Xenopus* oocytes as researchers attempted to translate a reporter gene with the known slippery sequence 5'-AAAUUUU-3' in cells injected with either tRNA\textsuperscript{Phe} +wybutosine or tRNA\textsuperscript{Phe} -wybutosine. Frameshifting occurred nearly 3-fold more at the slippery sequence when wybutosine was not present at position 37 of tRNA\textsuperscript{Phe} (Carlson et al., 2001).
Similarly to $m^1G$, the wyosine family of modified bases are predicted to increase base stacking of neighboring purines within the anticodon of $tRNA_{Phe}$ (Figure 1.9).

Wybutosine ($yW$), the wyosine-derivative found in *S. cerevisiae* and other eukaryotes, is synthesized by five enzymes: tRNA methyltransferase 5 (TRM5) and tRNA wybutosine synthesizing protein 1 – 4 (TYW1 – TYW4) (Jühling et al., 2009; Noma et al., 2006). The effectiveness of wybutosine in limiting frameshifting was compared to two biosynthetic intermediates, $m^1G$ (the product of TRM5) and imG-14 (the product of TYW1). The results show that as $G_{37}$ of $tRNA_{Phe}$ becomes increasingly modified the frequency of frameshifting decreases. That is to say, $m^1G_{37}$ limits frameshifting less than imG-14$_{37}$, which limits frameshifting less than $yW$ (Waas et al., 2007). The authors suggest that this data may reflect the selective pressure that caused the step-wise evolution of each wybutosine synthesizing enzyme.

The nature of wyosine and related derivatives within trypanosomes is a central theme to this document. Chapter 3 discusses the biology and biosynthetic pathways of these modified bases in greater detail.
Figure 1.8 Guanosine, 1-methylguanosine, and the relevant wyosine (imG) derivatives. A. guanosine (G) B. 1-methylguanosine (m\(^1\)G) C. 4-demethylwyosine (imG-14) D. wyosine (imG) E. iso-wyosine (imG2) F. 7-methylwyosine (mimG) G. 7-aminocarboxypropyl-demethylwyosine (yW-86) H. 7-aminocarboxypropyl-wyosine (yW-72) I. wybutosine (yW) J. hydroxywybutosine (OH\(\text{y}\)W)
Figure 1.9 Structural arrangement of nucleotides $G_{34} - A_{38}$ of \textit{S. cerevisiae} tRNA$^{\text{Phe}}$. Modified purines at position 37, such as $yW_{37}$, limit intra-loop hydrogen bonding and promote base stacking. These properties help stabilize the interaction between the first base of the codon-anticodon duplex.
1.4 Trypanosome Biology

Trypanosomatids are an order of single-celled flagellated protists that include a number of medically important organisms. African trypanosomiasis (African Sleeping Sickness), caused by infection with *Trypanosoma brucei*, is an affliction shared by 500,000 people, with 30 million people currently at risk in sub-Saharan Africa. In Central and South America approximately 350 million people are at risk while 11 to 18 million are infected by *Trypanosoma cruzi* — the cause of Chagas disease (Despommier, 2005). From the same order, species of the genus *Leishmania* cause Leishmaniasis worldwide, but most prevalently in parts of Central and South America, as well as the Middle East, Africa, and Asia. An estimated twelve million people are currently infected and 1.5 million new cases of cutaneous Leishmaniasis and 500 thousand new cases of visceral Leishmaniasis occur each year. Left untreated, these parasites can lead to death after invasion of the central nervous system by *Trypanosoma* or invasion of susceptible organs, such as the liver or spleen, by *Leishmania*.

The mechanism of infection differs between *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania*, but all rely on the transfer from an insect vector to a mammalian host. Shortly after inoculation, *Leishmania* are phagocytized by host macrophages where they avoid phagolysosomal digestion, whereas *Trypanosoma brucei* thrives in the host’s bloodstream and *Trypanosoma cruzi* replicates within various host cells. Once inside the mammalian host, they change from an insect specific stage to that of mammalian
specific stage that is particularly well suited for the changes in the metabolic conditions of their new environment and in immune system avoidance.

The current treatment options for these diseases are not ideal. One of the more commonly used drugs in the treatment of African trypanosomiasis, Melarsoprol, causes encephalopathy and fails to cure patients as high as 10% of the time. Two nitroheterocyclic drugs, Nifurtimox and Benznidazole, are used in the treatment of Chagas disease but cause severe side effects and have little efficacy in treating the chronic form of the disease. The situation in the treatment of leishmaniasis is at a similar point; pentamidine successfully treats 90% of people cutaneous leishmaniasis (Fat et al., 2002) and miltefosine cures 95% of those with visceral leishmaniasis (Jha et al., 1999) — both with minimal side effects. While significant scientific and medical progress has alleviated the suffering of millions of those affected or at risk, the issues outlined here and the ever growing cases of drug-resistant trypanosomes raise the need for new therapies that exploit nuances of trypanosome biochemistry.

1.4.1 Trypanosome Mitochondrion

Trypanosomes have a single, large mitochondrion that is reticulated about the entire cell. The mitochondrial genome, the kinetoplast DNA (kDNA), is unusual in that it is composed of ring-like DNA structures that are concatenated to form a vast network (Shapiro, 1993). The larger DNA rings are called maxicircles and primarily contain protein coding sequences for genes involved in respiration. The second smaller but much more numerous DNA structures are
called minicircles and contain guide RNAs (gRNAs), which are described in the next section about U-insertion mRNA editing (Simpson, 1987). As stated previously, the trypanosome mitochondrial genome has a markedly reduced set of genes; in *T. brucei* there exists the 12S rRNA, 9S rRNA, ribosomal protein S12 (RPS12), and only 17 other protein coding sequences that are mostly involved in respiration. This implies that all other components required for mitochondrial function are encoded in the nuclear genome and must be imported into the mitochondrion. This encompasses the machinery required for translation, including most ribosomal protein subunits and all tRNAs (Bindereif, 2012; Simpson et al., 1989).

### 1.4.2 U-insertion/deletion Editing Within the Mitochondrion

Of the 18 protein coding genes in *T. brucei*, 12 do not match the sequences of their mature, translatable mRNA (Simpson et al., 2000). This is due to post-transcriptional editing of these mRNAs that occurs by a process in which uridines are site-specifically introduced or removed from mitochondrial transcripts (Benne et al., 1986). This process is facilitated by the guide RNAs encoded on the minicircles and, less frequently, the maxicircles. gRNAs are short RNA transcripts that have complimentary sequence to a region 3’ of an editing region, which allows for the mRNA and gRNA to hybridize and form a duplex with one another. This interaction ‘guides’ the editing complex to this position. In the case of uridine deletion and endonuclease cleaves the mRNA, a 3’ uridine-specific exonuclease removes a defined number of uridines, and an RNA ligase rejoins
the two tRNA halves (Seiwert and Stuart, 1994; Seiwert et al., 1996). Uridine-insertion works in a similar way where an endonuclease cleaves the mRNA, a defined number of uridines are inserted by a terminal-U-transferase (determined by a complimentary sequence in the gRNA), and an RNA ligase rejoins the mRNA halves (Kable et al., 1996, 1997; Sollner-Webb, 1996) (Figure 1.10).

The extent of post-transcriptional U-insertion / deletion varies between genes. The NADH dehydrogenase subunit 7 (ND7) mRNA undergoes the largest number of both types of editing events with 553 U insertions and 89 U deletions (Figure 1.11). Other mRNAs are not far behind: A6 mRNA, 447-28 and ND9 mRNA, 345-20, U insertions – U deletions, respectively) (Alfonzo et al., 1997). A consequence of this large scale uridine insertion is the production of long stretches of consecutive uridines and, in turn, an abundance of phenylalanine codons – 92.5% of which are the UUU codon, rather than UUC. (Table 1). In fact, greater than 20% of codons in 6 of the edited mRNAs (ND9, CO3, CR3, MURF2, CR4, and ND3) code for phenylalanine in the T. brucei mitochondrion.
Figure 1.10 The mechanism of uridine-insertion/deletion mRNA editing in the trypanosome mitochondria. A protein coding gene on a maxicircle is transcribed, producing Pre-mRNA. A The editosome machinery guided by guide RNA (gRNA) site-specifically inserts a set number of uridines. B U-deletion also uses gRNA to site-specifically remove a set number of uridines from Pre-mRNA. It is only after the editing process is finished that the mRNA sequence is made translatable. REN - RNA editing endonuclease, RET - RNA editing TUTase, REX - RNA editing exonuclease, REL - RNA editing ligase.
Figure 1.11 The edited mRNA of the mitochondrial NADH dehydrogenase subunit 7 (ND7). Black uppercase letters are encoded in maxicircle DNA. Red lowercase u’s are added post transcriptionally via uridine-insertion editing. Not shown: Maxicircle encoded uridines that are post transcriptionally removed by U-deletion editing. Start codon: italics and underlined AuG. Stop codon: bold and underlined uAG.
Table 1 Analysis of phenylalanine codon usage

<table>
<thead>
<tr>
<th>Gene</th>
<th>Codons</th>
<th>%Phe</th>
<th>UUU</th>
<th>UUC</th>
<th>%UUU</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND8</td>
<td>145</td>
<td>17%</td>
<td>24</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>ND9</td>
<td>117</td>
<td>22%</td>
<td>26</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>ND7</td>
<td>387</td>
<td>12%</td>
<td>44</td>
<td>2</td>
<td>96%</td>
</tr>
<tr>
<td>CO3</td>
<td>289</td>
<td>22%</td>
<td>64</td>
<td>1</td>
<td>98%</td>
</tr>
<tr>
<td>Cyb</td>
<td>371</td>
<td>14%</td>
<td>50</td>
<td>2</td>
<td>96%</td>
</tr>
<tr>
<td>A6</td>
<td>232</td>
<td>24%</td>
<td>56</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>MURF1</td>
<td>444</td>
<td>19%</td>
<td>68</td>
<td>16</td>
<td>81%</td>
</tr>
<tr>
<td>CR3</td>
<td>71</td>
<td>30%</td>
<td>21</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>ND1</td>
<td>320</td>
<td>16%</td>
<td>41</td>
<td>9</td>
<td>82%</td>
</tr>
<tr>
<td>CO2</td>
<td>211</td>
<td>9%</td>
<td>20</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>MURF2</td>
<td>357</td>
<td>25%</td>
<td>82</td>
<td>6</td>
<td>93%</td>
</tr>
<tr>
<td>CO1</td>
<td>539</td>
<td>14%</td>
<td>67</td>
<td>11</td>
<td>86%</td>
</tr>
<tr>
<td>CR4</td>
<td>147</td>
<td>31%</td>
<td>46</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>ND4</td>
<td>438</td>
<td>13%</td>
<td>49</td>
<td>7</td>
<td>88%</td>
</tr>
<tr>
<td>ND3</td>
<td>151</td>
<td>21%</td>
<td>31</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>RPS12</td>
<td>83</td>
<td>12%</td>
<td>10</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>ND5</td>
<td>591</td>
<td>20%</td>
<td>105</td>
<td>11</td>
<td>91%</td>
</tr>
<tr>
<td>Total</td>
<td>4893</td>
<td>18%</td>
<td>804</td>
<td>65</td>
<td>18%</td>
</tr>
</tbody>
</table>

A All 17 protein coding sequences in the Trypanosoma brucei mitochondrial genome. Red gene names indicate those that undergo RNA editing.
Table 1 continued

<table>
<thead>
<tr>
<th>Gene</th>
<th>Codons</th>
<th>%Phe</th>
<th>UUU</th>
<th>UUC</th>
<th>%UUU</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYW1Large</td>
<td>836</td>
<td>4%</td>
<td>15</td>
<td>19</td>
<td>44%</td>
</tr>
<tr>
<td>RISP</td>
<td>298</td>
<td>4%</td>
<td>6</td>
<td>7</td>
<td>46%</td>
</tr>
<tr>
<td>Enolase</td>
<td>430</td>
<td>3%</td>
<td>4</td>
<td>8</td>
<td>33%</td>
</tr>
<tr>
<td>Isd11</td>
<td>100</td>
<td>6%</td>
<td>4</td>
<td>2</td>
<td>67%</td>
</tr>
<tr>
<td>Total</td>
<td>1664</td>
<td>4%</td>
<td>29</td>
<td>36</td>
<td>48%</td>
</tr>
</tbody>
</table>

B Phenylalanine codon usage in *T. brucei* nuclear encoded genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Codons</th>
<th>%Phe</th>
<th>UUU</th>
<th>UUC</th>
<th>%UUU</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND1</td>
<td>319</td>
<td>5%</td>
<td>4</td>
<td>12</td>
<td>25%</td>
</tr>
<tr>
<td>ND3</td>
<td>115</td>
<td>7%</td>
<td>2</td>
<td>6</td>
<td>25%</td>
</tr>
<tr>
<td>CO1</td>
<td>514</td>
<td>8%</td>
<td>12</td>
<td>29</td>
<td>33%</td>
</tr>
<tr>
<td>Total</td>
<td>7%</td>
<td>18%</td>
<td>18</td>
<td>47</td>
<td>28%</td>
</tr>
</tbody>
</table>

C Phenylalanine codon usage in *H. sapiens* mitochondrial genes.
Chapter 2: A New Biosynthetic Pathway for Mitochondria Unveils the Origin of Wybutosine in Eukaryotes

2.1 Introduction

Crucial to early life was the establishment of efficient mechanisms that enhance translational accuracy during protein synthesis. Thus given its role in codon recognition, tRNA must have played a central function in determining the genetic code. One of the defining features of all tRNAs is the presence of numerous posttranscriptional chemical modifications (Jackman and Alfonzo, 2013). Because some modifications are common to all tRNAs in all domains of life, it has been suggested that such “primordial” nucleosides were essential in ensuring reading-frame maintenance early in the evolution of translational systems (Björk et al., 2001; Grosjean, 2009; Jackman and Alfonzo, 2013; Phizicky and Hopper, 2010). Although many positions in a tRNA can affect translational accuracy, position 37 of the anticodon loop plays an important, if not a critical, role in reading-frame maintenance (Urbonavicius et al., 2001). This universally modified position may harbor simple base methylations (such as 1-methylguanosine, m\(^1\)G), which in the context of a specific tRNA anticodon and
depending on the inherent stability of anticodon-codon interactions, may be sufficient to prevent translational errors. Other tRNAs may require more complex chemical groups such as 6-isopentenyladenosine ($t^6A$) and 6-threonyladenosine ($t^6A$) (Caillet and Droogmans, 1988; Deutsch et al., 2012).

By far, the most chemically intricate frameshift-preventing modification at position 37 involves the nucleoside wyosine (imG) and its derivatives, including wybutosine (yW) and hydroxywybutosine (OHyW) (4). Wyosine derivatives are exclusively found in Archaea and Eukarya in the single tRNA$^{Phe}_{GAA}$, which is responsible for decoding the two synonymous phenylalanine codons, UUU and UUC. This unique evolutionary conservation may be related to the observed propensity of ribosomes to shift reading frames during translation in response to “slippery” sequences, most prominently those rich in uridines. By inference, genes encoding phenylalanine-rich proteins should contain an abundance of potentially slippery sequences. Evidence suggests that wybutosine promotes stable codon-anticodon pairing by base stacking interactions with anticodon loop nucleotides, such that the wobble-base, G$_{34}$ of tRNA$^{Phe}$, may more favorably form a G:U base pair with the U at the third codon position (Atkins and Björk, 2009). Therefore, the potential frameshifting problem caused by U-rich slippery sequences is partly solved by the unique presence of wyosine and its derivatives, such as wybutosine in tRNA$^{Phe}_{GAA}$.

Despite the omnipresence of wyosine and its derivates in archaeal and eukaryal tRNA$^{Phe}$, the pathway varies greatly between these organisms. The
eukaryotic pathway for wybutosine (yW) involves at least four sequential reactions that use 1-methyguanosine (m$^1$G$_{37}$), as a precursor (Björk et al., 2001). The m$^1$G$_{37}$ modification is itself the product of a phylogenetically widespread methylation found in all domains of life. Formation of m$^1$G$_{37}$ is followed by the incorporation of the C-1 and C-2 carbons of pyruvate by the radical-SAM enzyme TYW1 (tRNA yW-synthesizing protein) to form an additional ring on the methylated purine (Perche-Letuvee et al., 2012; Young and Bandarian, 2011). The resulting product constitutes the minimal core of the modification, with 4-demethylwyosine (imG-14) (Young and Bandarian, 2011) as its simplest form.

What follows this core tricyclic component is a series of reactions that vary among different organisms. For example, in most eukaryotes the end product of the reaction is the nucleoside yW, the product of 3 additional enzymes in yeast (TYW2 – TYW4) (Figure 2.1A) (Noma et al., 2006). In some organisms, including humans, yW can be further modified by a fifth enzyme, TYW5, to hydroxywybutosine (OHyW) (Kato et al., 2011; Noma et al., 2010).

In Archaea, the situation varies even more, with most species having at least one type of wyosine derivative ranging from 4-demethylwyosine (imG-14, the product of TAW1) to isowyosine (imG2), wyosine (imG), methylwyosine (mimG), 7-carboxypropyl-demethylwyosine (yW-86) and/or 7-amino-carboxypropylwyosine (yW-72) (de Crécy-Lagard et al., 2010). This diversity is created by TAW1, TAW2 and TAW3, which can act in a strictly sequential pathway or in a combinatorial manner. Among these, most striking is
the conserved pathway shared by Crenarchaeota (Figure 2.1B), which lacks TAW2 and involves only TAW1 and TAW3, yielding imG as the end product (de Crécy-Lagard et al., 2010).

Wyosine and derivatives are thus far absent in Bacteria and instead bacterial tRNA$^{\text{Phe}}$ has an encoded A$_{37}$, which then can be modified to i$^6$A (or ms$^2$i$^6$A) as the key modification ensuring reading frame maintenance; given their shared ancestry, the same is expected of mitochondria (Agris, 2008; Jühling et al., 2009; Urbonavicius et al., 2001). However, in protists such as *T. brucei*, there is a complete lack of tRNA genes in the mitochondrial genome. These organisms are therefore forced to import all of their tRNAs from the cytosol for organellar translation; their strategy for reading frame maintenance during translation then becomes of particular interest (Simpson et al., 1989).
Figure 2.1 Biosynthetic pathways for wyosine/wybutosine and derivatives. All enzymatic steps require S-adenosyl-L-methionine (SAM), except the third ring closure by TYW1 and TAW1. A The eukaryotic hydroxywybutosine (OHyW) and B the proposed biosynthetic pathway of Archaeal (specifically-Creanarchaeota) wyosine (imG). TAW1 and TAW3 are the archaeal homologs of eukaryotic TYW1 and TYW3. Initial methylation at N1 of guanosine (G₃₇) to form 1-methylguanosine (m1G₃₇) via eukaryotic or archaeal TRM5 (tRNA methyltransferase) homologs provides the substrate for TYW1 uses radical S-adenosyl-L-methionine (SAM) chemistry to transfer 2 carbons from pyruvate to form the pathway-defining tricyclic imidazopurine ring (imG-14). Eukaryotic TYW1 contains a flavodoxin-1-binding domain, which is essential for activity. Archaeal TAW1 lacks this domain and it is not clear what protein provides the FMN function. Also shown are homologs of eukaryotic TYW4 and TYW5; these have never been found in archaea. In addition most euryarchaeota contain a TYW2 homolog and incorporate the α-amino-α-carboxypropyl (acp) sidechain to form yW-86 derivatives. All steps other than hydroxylation of yW-72 via TYW5 require SAM.
In the present study, we report two distinct pathways for wyosine biosynthesis and its derivatives in *T. brucei*: one cytosolic and the other mitochondrial. The coexistence of these two pathways is made possible by the trypanosomatid-specific presence of two TYW1 and two TYW3 paralogs yielding different wyosine derivatives as end products: hydroxywybutosine, found in the cytosol and mitochondrion, and wyosine, strictly confined to the mitochondrion. This represents the first example of an organellar wyosine pathway. Both pathways are important for growth under conditions of low glucose, highlighting their role in mitochondrial translation. Surprisingly, the mitochondrial TYW1 has a close evolutionary affinity to the equivalent archaeal enzyme as supported by molecular phylogeny arguments presented here. We suggest that the *T. brucei* mitochondrial enzyme represents the ancestral state in wyosine biosynthesis (still shared with Archaea). The cytosolic yW pathway then reflects a more recent acquisition of a key FMN-domain. Our findings suggest that the maintenance of the ancestral wyosine pathway in the mitochondrion is possibly the result of the widespread mechanism of U-insertion/deletion RNA editing of kinetoplastids; although necessary for mRNA processing, editing also provides a wealth of potentially slippery mRNA sequences.
2.2 Results

2.2.1 Unique Occurrence of Paralogous Wyosine Biosynthesis Genes in *T. brucei*

Wybutosine has been extensively studied in *Saccharomyces cerevisiae*, but little is known about its biosynthesis in protists. To identify *T. brucei* homologs of the yW enzymes, we performed database searches using *S. cerevisiae* yW protein sequences as BLAST queries for the kinetoplastid genome databases (Tri-Tryp) (See Supplementary Table 5 for protein accession numbers). Four potential reading frames with significant similarities with the *S. cerevisiae* sequences were identified (TbTYW1–TbTYW4) (Appendix Figure A.1). In addition, a sequence extension in TbTYW4 showed similarity to the TYW5 protein found in some eukaryotes but is absent in *S. cerevisiae* (Appendix Figure A.4). This extension contains a highly conserved Jumonji C (JmjC) domain important for the formation of hydroxywybutosine (OHyW) as the end product of the pathway (Noma et al., 2010). Fusion of the last two enzymes in the pathway is reminiscent of *Aspergillus oryzae* and thus strongly suggests the presence of OHyW in *T. brucei* tRNA^{Phe} (Sample and Alfonzo, unpublished results).

These searches also revealed that two of the genes (*tyw1* and *tyw3*) were duplicated in the kinetoplastid lineage (which includes the genus *Trypanosoma*). The larger of the two putative genes encodes an 835 amino acid (aa) predicted protein we termed TbTYW1L. This protein bears all the conserved features of
this family of enzymes including: a flavodoxin-1 domain at the N-terminus needed for FMN binding, two 4Fe-4S iron-sulfur cluster domains, a radical SAM domain, and a recognizable ‘wyosine base formation motif’ (Figure 2.2 and Appendix Figure A.1) (Punta et al., 2012). The second gene encodes a smaller, 395 aa protein termed here TbTYW1S, which despite containing all the key above-mentioned catalytic domains, lacks the flavodoxin-1 domain, and more closely resembles the equivalent archaeal enzyme, TAW1 (Figure 2.2 and Appendix Figure A.1B) (de Crécy-Lagard et al., 2010; Noma et al., 2006).
Figure 2.2 Graphical representation of the *Trypanosoma brucei* wybutosine-synthesizing homologs. The two TYW1 paralogs differ in respect to one protein domain - flavin mononucleotide (FMN) binding domain (or flavodoxin-1) is present at the N terminus of TbTYW1L and absent in TbTYW1S. Both TBTYW1 paralogs contain two iron-sulfur (4Fe-4S) clusters. The two TYW3 paralogs have been named TbTYW3A and TbTYW3B. TbTYW4 and TbTYW5 are fused in trypanosomes, while in mammals they exist as separate proteins. Protein domains were identified by Pfam. Met-10+: methionine-10. LCM: Leucine carboxyl Methyltransferase. Kelch: Galactose oxidase central domain. JmjC: jumonji C.

A similar situation occurs with the third enzyme in the pathway TYW3, with two open reading frames of different sequences encoding potential TbTYW3 homologs (Figure 2.2 and Appendix Figure A.3). In yeast TYW3 catalyzes the methylation at N4 of the yW intermediate yW-86, the product of TYW2, to produce yW-72 (Noma et al., 2006). The archaeal homolog, TAW3, performs the same reaction in species that encode TAW1 and TAW2. In addition, TAW3 may
directly methylate imG-14 (the product of TAW1) to form wyosine (imG) as the end product in species that do not encode TAW2 (TYW2 homolog). This last pathway, consisting of a TAW1 (lacking a flavodoxin-1 domain) and TAW3, is predominant in Crenarchaeota (Figure 2.1B), but the analogous pathway has so far been described in only one eukaryote, Candida utilis (de Crécy-Lagard et al., 2010; Kasai et al., 1976; Takemura and Kasai, 1974).

Based on these observations, we explored the possibility that the duplicated genes may point at different versions of wyosine derivatives in T. brucei, perhaps in different cellular compartments. Indeed, the program Target-P predicted TbTYW1L to localize to the cytosol and TbTYW1S to the mitochondrion (Table 2). The cytosolic localization of TbTYW1L is not unusual given that in all eukaryotes studied so far yW is a nucleo-cytosolic modification. However, the potential localization of TbTYW1S to the single mitochondrion is surprising and would represent the first example of a pathway for wyosine and/or derivatives in an organelle. Likewise, the two TYW3 paralogs are also predicted to localize to different cellular compartments, TbTYW3A to the cytosol and TbTYW3B to the mitochondrion.
<table>
<thead>
<tr>
<th>TargetP Prediction</th>
<th>Experimentally Determined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mTP</td>
</tr>
<tr>
<td>Tyw1L</td>
<td>0.061</td>
</tr>
<tr>
<td>Tyw1S</td>
<td>0.664</td>
</tr>
<tr>
<td>Tyw2</td>
<td>0.509</td>
</tr>
<tr>
<td>Tyw3A</td>
<td>0.048</td>
</tr>
<tr>
<td>Tyw3B</td>
<td>0.755</td>
</tr>
<tr>
<td>Tyw4</td>
<td>0.621</td>
</tr>
</tbody>
</table>

Table 2: Subcellular localization prediction as analyzed by TargetP. Amino acid score relating to the strength of any present mitochondrial targeting sequence. SP: ‘secretory pathway’ - strength of any present secretory pathway sequence. Other: Strength of localization prediction some place other than the mitochondrion or secretory pathway. Loc: Final localization prediction based on the scores of ‘mTP’, ‘SP’, and ‘Other’. S: secretory pathway. M: mitochondrion. ‘-‘ : other. RC (reliability class): Confidence of localization prediction based on the difference between the highest and second highest localization scores. 1 is the highest and 5 is the lowest. TPlen: number of amino acids residues between the N-terminus and a predicted presequence cleavage site. ‘-‘ indicates that a cleavage site was not predicted.

2.2.2 Mitochondrial Localization of the Wyosine Modification Enzymes

To further test the potentially novel localization of the paralogous genes listed above, transgenic cell lines expressing epitope-tagged versions of each putative gene in the pathway were generated. Western blot analysis with subcellular fractions (total, cytosolic, and mitochondrial) from *T. brucei*
expressing C-terminus V5-epitope-tagged TbTYW1L revealed its cytosolic localization (Figure 2.3A). In these experiments, antibodies against compartment-specific markers (enolase and lsd11 for cytosol and mitochondrion, respectively) were used as localization controls and also as controls for fraction purity (Figure 2.3A). These results were further confirmed by immunofluorescence analysis using the same anti-V5 antibody (Figure 2.3B). Furthermore, in S. cerevisiae, TYW1 localizes to the endoplasmic reticulum (ER) (Noma et al., 2006); the same is true for TbTYW1L-V5, which co-localizes with BiP (an ER marker), thus confirming the ER localization prediction made using the Target-P program. Similar experiments were performed with V5-epitope-tagged TbTYW1S. Western blot analysis revealed a signal for TbTYW1S in the mitochondrial fraction from transgenic cells (Figure 2.3A). Immunofluorescence (as above) also confirmed the mitochondrial localization of TbTYW1S-V5, based on its co-localization with the mitochondrion-specific MitoTracker Red dye (Figure 2.3C).

Additional Western blot analysis and immunofluorescence microscopy were performed with the remaining putative members of the pathway. These results show that TbTYW3A is cytosolic, while TbTYW3B (the other duplicated gene product) showed mitochondrial localization. In addition, TbTYW4/5 localizes to the cytosol and TbTYW2 localizes to both cytosol and mitochondrion (Figure 2.4 and Table 2). Taken together, these experiments support the existence of two potential pathways for wyosine biosynthesis in T. brucei: a typically
eukaryotic pathway occurring in the cytosol involving TbTYW1L, 2, 3A and 4/5 and a novel mitochondrial pathway potentially involving TbTYW1S, TbTYW2 and TbTYW3B.

**Figure 2.3** Intracellular localization of TbTYW1L and TbTYW1S. **A** Total (T), mitochondrial (M), and cytosolic (C) protein fractions were analyzed via Western blot with V5 epitope-specific antibodies to determine the intracellular localization of V5-epitope-tagged TbTYW1L and TbTYW1S. Similar blots were performed with antibodies specific for *T. brucei* enolase (a cytosolic marker) and Isd11 (a mitochondrial marker). These serve as controls for fraction purity, ruling out significant cross-contamination between the fractions. **B** and **C** Immunofluorescence-localization experiments with the same V5-antibodies (α-V5) (green fluorescence) and cells expressing epitope-tagged proteins as indicated. DAPI (blue) was used to show the location of the nucleus (N) and kinetoplast (K). Nomarski refers to a phase-contrast image of the same cells. MitoTracker dye (red) was used as a mitochondrial marker. Yellow fluorescence indicates co-localization of either protein with the MitoTracker dye in merged images (α-V5&MitoTracker). “Merge All” refers to a composite of all the fluorescence images for each set.
Figure 2.4 Subcellular localization of TYW2, TYW3A, TYW3B, and TYW4/5. A Western blot analysis of epitope-tagged constructs from total, cytoplasmic, and mitochondrial protein fractions. Enolase is used as a cytoplasmic marker and as a means to determine mitochondrial fraction purity. Isd11 is both a mitochondrial marker and control for cytoplasmic fraction purity. TYW2-Myc localizes to both the cytoplasmic and mitochondrial fractions. TYW3A-HA is localized to the cytoplasm, while TYW3B-FLAG is strictly found in the mitochondrial. TYW4-Myc is strictly cytoplasmic. B Immunofluorescence of the same epitope-tagged proteins. Green fluorescence indicates presence of the tagged constructs. MitoTracker Red localizes to the mitochondrion and fluoresces red. When merged, colocalization is identified by the color yellow.
Figure 2.4

A

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>C</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enolase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isd11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYW2-Myc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYW3A-HA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYW3B-FLAG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TYW4-Myc</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

- **TYW2-Myc**
  - Nomarski
  - DAPI
  - α-Myc
  - MitoTracker
  - α-Myc & MitoTracker
  - Merge All

- **TYW3A-HA**
  - Nomarski
  - DAPI
  - α-HA
  - MitoTracker
  - α-HA & MitoTracker
  - Merge All

- **TYW3B-FLAG**
  - Nomarski
  - DAPI
  - α-FLAG
  - MitoTracker
  - α-FLAG & MitoTracker
  - Merge All

- **TYW4-Myc**
  - Nomarski
  - DAPI
  - α-Myc
  - MitoTracker
  - α-Myc & MitoTracker
  - Merge All

C

- **TYW1-V5**
  - Nomarski
  - DAPI
  - α-V5
  - α-BIP
  - α-V5 & α-BIP
  - Merge All
2.2.3 Two differentially modified versions of tRNA$^{\text{Phe}}$ co-exist in mitochondria

To further establish the chemical nature of the modification(s) at G$_{37}$ of tRNA$^{\text{Phe}}_{\text{GAA}}$, total cytosolic and mitochondrial tRNA fractions were isolated, digested to nucleosides and analyzed by liquid chromatography tandem mass spectrometry (LC/MS/MS). In the cytosolic tRNA fraction we observed a peak with a mass/charge ($m/z$) ratio of 854, which elutes from the reversed phase chromatography column with a retention time of 42 min (average: 42.1 min, standard deviation: +/- 0.15 min.) (Figure 2.5A). The observed $m/z$ value of 854 is consistent with that of a hydroxywybutosine adenosine dinucleoside with a bridging phosphate (OHyWpA). The appearance of such dinucleoside monophosphate is due to the well-documented resistance of wyosine and derivatives to nuclease P1 digestion used for nucleoside analysis (Noma et al., 2006). This assignment was further confirmed by its UV-Vis absorbance profile showing a $\lambda_{\text{max}}$ at 242 nm and a shouldering peak at 260 nm as has been described elsewhere (Perche-Letuvee et al., 2012). This dinucleoside monophosphate was also corroborated by collision-induced dissociation tandem mass spectrometry (CID-MS/MS) (Figure A.5). No other wyosine derivatives were observed in the cytosol (Figure 2.5D).
Figure 2.5 tRNA nucleoside analyses show the presence of two different wyosine derivatives in *T. brucei* mitochondria. A The left panel shows the HPLC UV (254nm) of the total cytosolic nucleosides. The peak eluting at 42min was further analyzed by UV absorbance showing a spectrum characteristic of hydroxywybutosine/adenosine dinucleoside monophosphate (OHyWpA), the presence of hydroxywybutosine was further confirmed by mass spectrometry m/z measurement (right). All contribute to the identification of the dinucleoside hydroxywybutosine-phosphate-adenosine (OHyWpA). B Mitochondrial hydroxywybutosine. Top-left panel, HPLC measurement at 254nm of peak eluting at 41.9 minutes, absorbance profile (top-middle), and m/z value (top-right) that identify OHyWpA. C Mitochondrial wyosine. Bottom-left panel, HPLC measurement at 254nm of peak eluting at 38.7 min, absorbance profile (bottom-middle), and m/z value (bottom-right) that contributes to the identification of wyosine-phosphate-adenosine (imGpA). D Absence of cytosolic wyosine. HPLC measurement at 254nm shows lack of imGpA in the cytosolic tRNA fraction.
The total mitochondrial tRNA fraction showed a peak with an \( m/z \) value of 854, similar retention time (41.9 min) and comparable UV-Vis absorbance profile as the cytosolic OHyWpA dinucleoside monophosphate (Figure 2.5B). More importantly, an additional peak was detected with an \( m/z \) value of 665 (eluting at 38.6 min retention time), which is consistent with the expected values for a wyosine (imG) adenosine dinucleoside monophosphate (imGpA) (Figure 2.5C). These assignments were confirmed by CID-MS/MS (Appendix Figure A.5) and UV-Vis absorbance showing a \( \lambda_{\text{max}} \) of 237 nm, in line with the published profile for the wyosine dinucleotide (imGpA) (Figure 2.5C, center panel) (Perche-Letuvee et al., 2012). However, isowyosine (imG2) (Figure 1.7E), an isomer of wyosine, has the same \( m/z \) value as wyosine but a different HPLC retention time (de Crécy-Lagard et al., 2010). To distinguish between these two possibilities, we purified total tRNA from \( C.\ utilis \) and used it as a marker for imG. This yielded nucleosides with identical elution profiles and \( m/z \) values to those of the \( T.\ brucei \) mitochondrial fraction (Figure 2.6A and 2.6B). This finding was further confirmed by either analyzing the \( C.\ utilis \) nucleosides by themselves or by mixing equimolar amounts of the \( C.\ utilis \) fraction with the mitochondrial fraction from \( T.\ brucei \) (Figure 2.6C).
Figure 2.6 Positive identification of wyosine, and not isowyosine, in the trypanosome mitochondrion. A imGpA (wyosine-phosphate adenosine) from *Candida utilis* elutes at 38.7 minutes (left panel) and has the absorbance profile seen in the right panel. B Wyosine from *T. brucei* elutes at 38.7 minutes and has a matching absorbance profile to the *C. utilis* wyosine absorbance profile. C Combination of *C. utilis* imGpA fraction and *T. brucei* suspected imGpA fraction co-elute. The combined *C. utilis* and *T. brucei* sample is slightly left-shifted, likely due to the nearly complete lack of other nucleosides that would retard the imGpA dinucleoside as it migrated through the reverse-phased HPLC column.

In our localization analysis in the previous section, we observed that TbTYW2 has dual localization (Figure 2.4). If TbTYW2 was part of the
mitochondrial pathway, a product consistent with its activity (either yW-86 or yW-72) should be found (Figure 2.1A). However, no such product was observed in this analysis and the function that TbTYW2 may be playing in the mitochondrion remains an open question. Regardless, the observation of imG in the mitochondrial tRNA fraction suggests that imG may be the product of the sequential reaction of TbTYW1S and TbTYW3B (Figure 2.1). In T. brucei, all tRNAs are imported into the mitochondrion from the cytosol. The existence of OHyW in the mitochondrion may then be explained by the import of the already modified OHyW-containing cytosolic tRNAs.

2.2.4 Lack of TbTYW1L and TbTYW1S leads to corresponding losses of OHyW and imG

To address the role of the two potential paralogs of TYW1 in T. brucei (one mitochondrial and one cytosolic), we generated individual transgenic RNAi cell lines. A portion of each gene (parts of the coding sequence) were placed in an RNAi plasmid vector under a tetracycline-inducible system, as previously described (Wickstead et al., 2002). Following RNAi induction by tetracycline, we performed growth curves to compare the un-induced and induced cell lines, followed by reverse transcription (RT)-PCR assays using primers specific for each gene (as described in Appendix C.7). By this approach, we were unable to detect products of each target gene following RNAi induction, indicative of a successful down-regulation (Figures 2.7A and 2.7C). In these assays, RNA from wild type cells was used as a positive control, while a mock sample where
reverse transcriptase was left out of the reaction served as a negative control for DNA contamination. To rule out secondary off-target effects between the two genes in question, in each reaction primers specific for the non-targeted TYW1 gene were included. In each case, RNAi induction led to the specific decrease in the levels of one transcript but not the other, while both transcripts were detected in the wild type sample (Figure 2.7).

To assess the effect of RNAi on wyosine and hydroxywybutosine biosynthesis, we then isolated subcellular fractions as before and compared the wyosine-derivative content of mitochondrial and cytosolic tRNA fractions, following RNAi induction (Figure 2.7). We observed a marked reduction in hydroxywybutosine in both fractions when TbTYW1L was down-regulated (Figure 2.7A, right panel and Figure 2.7B, left panel). However, the content of wyosine in the mitochondrial fraction increased in comparison to wild type (Figure 2.7B, right panel). In turn, ablation of TbTYW1S led to a significant reduction in mitochondrial wyosine levels (Figure 2.7D). These data indicate two separate functions for the two TYW1 paralogs: TbTYW1L is important for the presence of hydroxywybutosine in both compartments and TbTYW1S is solely involved in mitochondrial wyosine formation. In turn, the observed reduction in mitochondrial hydroxywybutosine levels in the TbTYW1L RNAi-induced cells, together with the TbTYW1L localization experiments, is in line with its synthesis in the cytosol followed by import of the hydroxywybutosine-containing tRNAs into the organelle.
Figure 2.7 TYW1L is involved in OHyW synthesis and TYW1S is involved in imG synthesis. A A growth curve of RNAi induced and uninduced cells. The inset image shows RT-PCR analysis of TbTYW1L and TbTYW1S in WT and cells where their expression has been down-regulated by RNAi. Both genes are transcribed in wild type cells (WT lanes, showing two bands), where the larger product is specific for TYWS and the smaller product is specific for TbTYW1L. RNAi of TbTYW1L (T1L indicated by arrow) specifically down-regulates the TbTYW1L transcript but does not affect TbTYW1S (T1S indicated by the arrow). The right panel shows that RNAi of TbTYW1L reduces cytosolic hydroxywybutosine/adenosine dinucleoside monophosphate (OHyWpA) to an undetectable level by LC-UV-MS (denoted by dash-line box in the graph). B The same experiment as above was performed with mitochondrial RNA from the TbTYW1L RNAi-induced samples. The left panel shows that RNAi of TbTYW1L also causes the loss of mitochondrial OHyWpA (left panel) but has no effect on mitochondrial imGpA. C Similar experiments as in A, showing growth curves for the RNAi-induced and uninduced TbTYW1S cell line, showing a minor growth phenotype. RNAi of TbTYW1S specifically down-regulates TbTYW1S (T1S) and not TbTYW1L (T1L). WT again refers to RNA from wild type cells used as a positive control. The right panel shows that down-regulation of TbTYW1S has no effect on cytosolic OHyWpA. D Similar experiment with mitochondrial RNA fractions from the TbTYW1S RNAi cells. The left panel shows a 50% reduction in mitochondrial OHyWpA compared to the wild type, a complete disappearance of mitochondrial imGpA. In all panels “Relative abs” refers to the ratio of the observed absorbance of a given peak from the RNAi lines, normalized to the same peak from the wild type, where identical amounts (20 µg) of tRNA were analyzed in all cases. The graphs shown are representative of at least 3 independent experiments. “log cells/ml” refer to cumulative cell counts over the period of the growth curve shown in days. “M” in the inset images refers to a size marker used during electrophoresis.
Figure 2.7

A

Days after induction

log cells/ml

WT
TET-
TET+

M
WT
T1L
TET+
T1S
T1L

B

Time (min)

OHyWpA
imGpA

Relative Abs.

0
1
0.5
0

0.5
1
2
3

41.5
42.0
42.5

38.4
38.9
39.4

C

Days after induction

log cells/ml

WT
TET-
TET+

M
WT
T1S
TET+
T1S
T1L

D

Time (min)

OHyWpA
imGpA

Relative Abs.

0
1
0.5
0

0.5
1
2
3

41.5
42.0
42.5

38.4
38.9
39.4

72
Down-regulation of expression of either gene does not cause a major growth defect (Figure 2.7A and 2.7C). This result is not surprising since modest growth phenotypes are observed with yW mutants in other organisms. However, upon ablation of either TbTYW1S or TbTYW1L, cells showed a growth phenotype in low-glucose media (Figure 2.8). This observation is significant because under low glucose cells are forced to grow primarily by oxidative phosphorylation. Therefore, TbTYW1S and TbTYW1L are important for mitochondrial function. This is supported by an observed increase in mitochondrial membrane potential and a decrease in respiration upon RNAi induction in either cell line. These defects are more pronounced in the TbTYW1S-depleted cells (Figure 2.9). Coincidentally, these phenotypes are analogous to those observed when TbTRM5 is down-regulated (Paris et al., 2013). TbTRM5 is not only important to provide the m^{1}G_{37} precursor for mitochondrial wyosine, but also for additional tRNAs.
Figure 2.8 Effects of A TbTyW1S and B TbTyW1L RNAi on the growth of the procyclic stage of T. brucei in the low glucose media. Growth curves of wild type (WT; triangles), non-induced (TET-; squares) and induced (TET+; circles) knockdown cell line for TbTYW1S and TbTYW1L. The y axis is labeled by a log scale and represents the products of the measured cell densities and total dilutions.
Figure 2.9 Down-regulation of TbTYW1L and TbTYW1S affects mitochondrial function. The top panels in the figure refer to respiration quantification via measurement of oxygen consumption using a Clark electrode. The bottom panels show quantification of membrane potential measured via uptake of the mitochondrial fluorescent dye MitoTracker Red in the various strains as indicated. “wt” refers to wild-type control, “S-” and “L-” refer to uninduced TbTYW1S and TbTYW1L cell lines respectively and “S+” and “L+” refer to cells in which RNAi was induced by tetracycline. A shows values determined when the cells were grown in “Low-glucose” media; these were compared with B the same cell lines grown in normal media “High glucose”. Values shown are the result of at least 3 independent measurements. The measured mean values of red fluorescent intensity are represented as percentages of the WT sample, which was set to 100%. Data were obtained from at least three independent measurements and standard deviations are indicated. A Student’s t-test analysis determined that the results are significant, with P values of less than 0.05 (*).
2.2.5 The evolution of TYW1

Unlike TAW1 of Archaea, most of the known eukaryotic TYW1 proteins possess N-terminal flavodoxin domains. However, as shown here, parasitic trypanosomatids of the group Kinetoplastida encode additional variants that lack this domain and may therefore represent an ancestral Taw1-like gene. Phylogenetic analysis of TAW1/TYW1 proteins, excluding the flavodoxin domain from the alignment supports this hypothesis (Figure 2.9). It suggests that the ancestral Taw1-like gene was duplicated in the eukaryotic lineage prior to the appearance of the last eukaryotic common ancestor (LECA).
Figure 2.10 TbTYW1S is the ancestral enzyme for the first committed step of wybutosine biosynthesis. Maximum likelihood (ML) phylogenetic tree of TAW1/TYW1 protein sequences excluding the flavodoxin domain. Numbers at nodes are ML bootstrap/Bayesian posterior probability values and are shown only for relevant branches representing major taxonomic groups and the relationships among them. The two major evolutionary events in the evolution of the eukaryotic TYW1 are highlighted by arrows.
2.3 Discussion

In many organisms polyuridines in coding sequences may lead to -1 frameshifting, a process by which the translating ribosomes mistakenly "slip" back by one nucleotide and continue protein synthesis in a different reading frame (Brierley and Dos Ramos, 2006; Brierley et al., 1992; Carlson et al., 1999, 2008; Jacks and Varmus, 1985; Urbonavicius et al., 2001; Waas et al., 2007). Although some viruses practice "programmed frameshifting" to their advantage, in most systems it can have undesirable consequences as a source of translational mistakes (Brierley et al., 1992; Jacks and Varmus, 1985; Jacks et al., 1988). The prevailing strategy in Eukarya (and probably Archaea) to deal with potential frameshifting problems in mRNAs rich in phenyalanine codons (i.e. UUU) involves the use of the hypermodified nucleotide wyosine and/or its derivatives. Thus, given the importance of this modification in frameshifting prevention, it is assumed that it was also adopted evolutionarily as an early strategy for translational fidelity.

In eukaryotes, wyosine/wybutosine has only been described in cytosolic tRNAs. Presented here is the first description of an organellar pathway for wyosine biosynthesis. Co-existence of two diverging pathways in \textit{T. brucei} opens a window into the possible evolutionary origin of wyosine/wybutosine biosynthesis. The mitochondrial TbTYW1S resembles the archaeal enzyme, while TbTYW1L is in line with those cytosolic pathways described in other eukaryotes. Molecular phylogeny data supports the idea that mitochondrial
wyosine formation in *T. brucei* is the ancestral pathway dating back to the last common ancestor of Archaea and Eukarya. We propose that the extant cytosolic pathway is the result of a more recent acquisition of the FMN-binding module by TbTYW1L in *T. brucei* and TYW1 in other eukaryotes. By extension, the newly discovered mitochondrial pathway described here, in analogy to Archaea, should require a transacting reductase to provide the essential FMN-domain function.

The phylogenetic data presented here suggests that TYW1 evolution involved a gene duplication event, which enabled one of the TYW1 genes to fuse with the flavodoxin-binding domain. The cis-acting FMN-domain may have provided a catalytic advantage for the cyclization reaction, obviated the need for the second gene, which eventually led to its disappearance. Fusion of the enzymes and the proteins that serve as their electron donors or acceptors (e.g. flavodoxin, cytochrome, etc) is not without precedent, for example, some fatty acid desaturases use soluble cytochrome b5 as electron donor while in others cytochrome b5 is fused to the enzyme (Gostinčar et al., 2010).

These observations lead to several important questions of why trypanosomatids maintain the ancestral pathway and dedicate it to their mitochondria. We suggest that this may be the result of yet another twist of trypanosomatid mitochondrial evolution, going back to the appearance of the robust, highly complex and prevalent U insertion/deletion RNA editing mechanism. Mitochondria-encoded mRNAs in trypanosomatids and related flagellates are transcribed as “scrambled” sequences that undergo, sometimes
extensive, insertions and deletions of uridines (Alfonzo et al., 1997). The obvious end result is the generation of edited translatable messages. A less appreciated facet of editing is that in generating uridines-rich sequences, it may pose the mitochondrion with the danger of potential defects in reading-frame maintenance, as shown for other systems. Hence, a mitochondrial wyosine pathway might have been selected for to avoid translational errors.

Why then the imported wybutosine-containing tRNA$^{\text{Phe}}$ is not sufficient for mitochondrial translation? We posit that in this respect the system is nuanced, in that the very robust mitochondrial tRNA import pathway, in combination with a level of promiscuity by the import machinery, allows for the transport of both fully matured as well as undermodified tRNAs. In a previous report, we presented evidence showing that undermodified tRNAs are imported into the mitochondrion and then “recycled” by mitochondrial modification, making them available for translation (Paris et al., 2013). In the case of tRNA$^{\text{Phe}}$ we suggest that recycling involves wyosine formation, elegantly avoiding potential translational defects but necessitating maintenance of at least a portion of the pathway. Unfortunately, the lack of tractable mitochondrial genetics and an in vitro translation system makes it currently impossible to test these ideas directly. Still, we examined the possible biological significance of hydroxywybutosine/wyosine in mitochondrion by analyzing the effect of ablation of either TbTYW1L or TbTYW1S. Although only minor effects on cell growth and mitochondrial function are observed under standard conditions, when the same cells were grown in low glucose medium,
which demands mitochondrial function, either enzyme becomes important for growth.

More globally in most cases in mutants that do not make yW, m^1G_{37} alone is sufficient to ameliorate frameshifting. This observation is true even for preventing frameshifting at U-rich codons other than phenylalanine. Why then has m^1G not universally replaced yW? The answer may rest in the idea of “frameshifting potential”, whereby, like in many viruses, frameshifting is used in a programmed manner, perhaps to increase coding diversity (Waas et al., 2007). This implies that cells as part of their normal translation programs may exploit frameshifting as a regulatory mechanism and not just avoid it, at all cost, as a source of translational infidelity. In this realm, the levels of either wyosine or derivatives may fluctuate with changes in environmental conditions leading to more or less frequent frameshifting as needed to create protein diversity and alter cell function. This may be especially true in trypanosome mitochondrion where “alternative” editing has been invoked as a possible source of coding diversity (Ochsenreiter et al., 2008), yet the same may be achieved with the newly discovered wyosine pathway presented in this work.

3.1 Introduction

Biologically derived ribonucleic acid (RNA) polymers, such as transfer RNA (tRNA) and ribosomal RNA (rRNA), contain modified nucleotides that are not amenable to sequence determination using the standard RNA sequencing method of RT-PCR followed by Sanger sequencing (Sanger and Coulson, 1975). Instead, these complex molecules, which may be composed with any number of more than 150 naturally occurring modified nucleosides (Cantara et al., 2011; Grosjean, 2009; Jühling et al., 2009; Machnicka et al., 2013), are typically characterized using collision-induced dissociation tandem mass spectrometry (CID MS/MS) (Huber and Oberacher, 2001; Limbach et al., 1995; McCloskey et al., 1999; McLuckey and Habibi-Goudarzi, 1993; McLuckey et al., 1992; Meng and Limbach, 2006) coupled with either matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) or liquid chromatography mass spectrometry (LC-ESI-MS). CID mass
spectrometry of oligoribonucleotides typically yields c-, y-, w-, and a-B-type product ions (Figure 3.1) (McLuckey et al., 1992; Wu and McLuckey, 2004). The data produced by these approaches is inherently difficult to analyze due to the complexity of the resulting spectra and the minimally available software tools to aide in the analysis.

Figure 3.1 Typical RNA oligomer fragmentation products generated by collision-induced dissociation (CID). The most abundant products are typically c- and y-fragment ions, however w- and a-B ions are often commonly observed.

One of the first attempts at computational analysis of mass spectrometric data generated from nucleic acid oligomers focused on the determination of the nucleotide composition of an ion based on the mass of the oligomer and the
masses of the four canonical ribonucleosides: cytidine (I), uridines (U), adenosine (A), and guanosine (G) (Pomerantz et al., 1993). In 2002, Rozenski and McCloskey released the Simple Oligonucleotide Sequencer (SOS), which was a tool capable of assisting in the manual interpretation of oligonucleotide MS/MS data from oligomers up to 20 bases in length (Rozenski and McCloskey, 2002). SOS worked by displaying spectrum peaks that corresponded to a-B– or w–type fragment ions, allowing the user to choose which nucleoside best fit the experimental data. While effective, this program is limited by the manual analysis of data, precluding its use on complex LC/MS/MS data sets, and by the minimal number of modified nucleosides that can be evaluated during data analysis.

Two database search strategies for RNA mass spectrometry data, RRM and Ariadne, have been developed (Matthiesen and Kirpekar, 2009; Nakayama et al., 2009) to approach the analysis of RNA MS and MS/MS data in a similar manner as the polypeptide analysis suite MASCOT and other similar protein-focused software (Kapp et al., 2005). RRM focuses on mass spectral data only and uses the concepts originally described by Pomerantz et al. (1993) to define base compositions that can be searched against genome or RNA sequence databases (Matthiesen and Kirpekar, 2009). While this strategy is effective for sourcing standard RNA, it cannot be used to characterize RNase digestion products containing modified nucleosides. Ariadne takes tandem mass spectrometry data generated from either biologically or in vitro-derived RNA and scores the comparison of the data to an inputted database of theoretical RNase-
digested and CID-fragmented RNA sequences (Nakayama et al., 2009). This software is limited to the databases of modified and unmodified sequences currently known and therefore cannot be used for unknown samples or sequences. Nyakas et al. developed OMA and OPA, which allows analysis of MS and MS/MS data with a customizable database of nucleotides, thus all known RNA modifications can be analyzed (Nyakas et al., 2013). However, this software only compares the predicted fragmentation pattern from an inputted sequence to the MS data, and therefore the sequence to be analyzed must be known in advance.

Automated de novo sequencing of MS/MS data has been attempted with various strategies and degrees of success from samples composed of DNA (Muddiman et al., 1997; Oberacher and Pitterl, 2011; Oberacher et al., 2002, 2004), DNA adducts (Liao et al., 2007, 2009; Sharma et al., 2012), and RNA containing 2’-O-methyl and phosphorothioate linkages (Kretschmer et al., 2010). The global search approach, which involves generating a library of all sequence isomers of a given nucleic acid composition and then scoring these sequence isomers based on the data within the mass spectrum, has returned positive results with oligodeoxyribonucleotides of up to 12 residues (Oberacher et al., 2004). The limiting challenge of such an approach lies in the exponentially increasing number of potential combinations introduced by either long oligomers or large nucleotide pools. A particularly interesting solution to this problem was employed by Oberacher and colleagues, who used a simulated annealing
strategy (Kirkpatrick et al., 1983) for sequence optimization of oligodeoxyribonucleotides as long as 22 residues. Briefly, a random sequence was generated for a nucleotide composition that matched the precursor mass. The positions of two nucleotides are computationally switched, a fitness score that evaluates theoretical fragmentation products compared to the actual spectral data is calculated, and if the new score is higher than the previous score then the new sequence is kept and subjected to additional rounds of variation and selection. In this example, the use of stochastic optimization dramatically reduced the computational load for long oligomers, but the program was only designed to handle the four common DNA bases: C, T, A, and G, and thus cannot be applied to the large number of modified nucleosides found in natural RNAs.

To our knowledge, a global search strategy for de novo sequence analysis has never been attempted with complex, multiple modified base-containing RNA, leaving the field of oligoribonucleotide sequencing via mass spectrometry significantly hindered at analyzing MS/MS datasets generated from RNAs containing multiple modified nucleosides. To address this need, we report on the development of RoboOligo – a user-interactive program equipped with abilities for the analysis of negative ion mode MS/MS data generated by CID. We show that an automated local search paradigm maintains the robustness of the global search paradigm and can efficiently handle modified bases. Additional data analysis flexibility is provided by manual and variable sequence capabilities that allow for user-controlled examination of MS/MS data.
3.2 Results

The main objective in the development of this program was to establish an algorithm that was robust and accurate in the evaluation of CID MS/MS data from moderately sized RNA oligomers containing modified bases (i.e., endonuclease digestion products from tRNA and rRNA). There are three main functions of RoboOligo: 1) Automated de novo sequencing via a local search paradigm with nucleotide pool, RNase digestion context, and composition constraints; 2) Manual sequencing with real-time spectrum labeling and cumulative intensity scoring; 3) A hybrid approach, coined ‘variable sequencing’, which combines the user intuition of manual sequencing with the higher throughput of automated de novo sequencing. To handle modified nucleosides, molecular mass information for 108 uniquely massed nucleotides is available from an internal database within the program. This information can be user edited by appropriate changes to the ‘nucleotides.txt’ file. All mass spectral data analyzed by RoboOligo can be graphically represented and appropriately labeled with sequence assignments.

3.2.1 Automated De Novo Sequencing

The design approach begins with a simple calculation to generate the mass of the oligonucleotide based on the precursor ion mass-to-charge value. This target mass is then used to calculate potential nucleotide compositions. In this process, the program finds all nucleotide combinations that, if combined as a single oligomer, would fall within the total mass range. The total mass range is
defined by the calculated mass of the oligonucleotide and the user-selected total mass tolerance. The unmodified and modified nucleotides to include in the composition analysis are also chosen by the user. The de novo portion of the software then uses a local search approach to create potential sequences from this information.

The principle concept of the local search approach is that sequences are built one nucleotide at a time, and after each nucleotide addition the fitness of the oligomer is evaluated by the abundance of \(-c\) and \(-y\) type product ions (Rozenski and McCloskey, 2002). Lack of evidence for a product ion will cause the incomplete sequence to fail, along with all of the sequences that would have been tested had it not failed. This ability to ignore entire sequence trees that lack data to support their validity provides an efficient way to logically reduce the number of oligonucleotide sequences to be tested.

The algorithm begins by attempting to match expected product ions beginning with the 5'-terminus (Figure 3.2). Once a potential \(-c_1\) ion is identified, the program will begin building a theoretical sequence in the 5’ to 3’ direction. To do so, another nucleotide from the composition is added to the theoretical sequence and the \(-c_2\) ion is calculated and compared to the data. Once a \(-c_2\) ion matches the data, that nucleotide is added and the program continues until all \(-c\) ions are identified for the calculated compositions. To improve sequence reconstruction accuracy, the program then calculates all expected \(-y\) ions for this potential sequence and compares these calculated values against the
experimental data to eliminate incorrect sequences. For a sequence that passes this step, the \(-w\) and \(-(a-B)\) ions are then calculated and all \(-c\), \(-y\), \(-w\) and \(-(a-B)\) product ions are used for theoretical sequence scoring. The entire process is then repeated for additional sequence possibilities.

Each composition is independently passed through the de novo sequencing algorithm, providing the identity and precise count of each nucleotide allowed within a growing oligonucleotide sequence. The benefit of confining sequence attempts through isolated compositions is that all permutations of the compositions lead to total oligonucleotide masses that fit within the target mass range and therefore reduces computational effort by never attempting sequences that would lead to an invalid total mass (Oberacher et al., 2002). The final list of sequences that potentially match the MS/MS data thus have all \(-c\), \(-y\), \(-w\) and \(-(a-B)\) product ions identified including multiply charged ions if the analyzed spectrum has a charge state greater than 1. Currently, the program supports charge states up to -3.
Figure 3.2 The logical design of the *de novo* sequencing algorithm. Potential compositions are produced for a calculated oligonucleotide mass and each composition is used for the *de novo* sequence analysis. The algorithm builds sequences one nucleotide at a time and tests the fit of the data to the expected products. This is done in the 5’ to 3’ direction using –c ions and in the 3’ to 5’ direction using –y ions. Once a potential sequence is found –w and –a-B ions are calculated and the sequences are scored based on the total absolute intensities of the fragment ions for that sequence.
The scoring algorithm used here is similar to that used originally in the SOS program (Rozenski and McCloskey, 2002). Those potentials sequences generated by the algorithm described above are characterized by their calculated sets of –c, –y, –w and –(a-B) product ions. The experimental data from the MS/MS spectrum is evaluated by summing the ion abundances at each of the m/z values for the set of product ions calculated for each potential sequence. These ion abundances are then summed to produce the cumulative score for a sequence.

The assumption is that the cumulative product ion abundance for a correct sequence should be greater than the cumulative product ion abundance for a sequence that does not represent the oligomer being analyzed. Oberacher et al. (2002) implemented a more sophisticated scoring scheme, which calculates the fitness of each sequence by applying penalties for incomplete fragmentation coverage and weighing of ion abundances that reduce the confidence in peaks that deviate more drastically from the predicted m/z value. This fitness equation performed well for their global-search automated de novo sequencing algorithm of DNA, but performed poorly with our modified base-containing RNA MS/MS data (data not shown).

A central difficulty in de novo sequencing of oligonucleotides when considering modified nucleotides lies in the sheer number of potential sequence combinations given the length and unique entities with which these molecules may be composed. For this reason, restrictions on the possible sequence
combinations to be analyzed can be beneficial. Constraints in the analysis can be applied by adjusting for the mass accuracy of the instrument in the MS or the MS/MS data. Mass constraint can have a profound impact on the number of potential compositions and on the match of experimental data to the predicted \( m/z \) as has been reported previously (Meng and Limbach, 2004, 2006). Pomerantz et al. (1993) noted that oligonucleotide base compositions could be constrained by using endonucleases of high selectivity or specificity. For example, ribonuclease T1 (RNase T1) cleaves RNA at all unmodified guanosine residues and at the modified nucleoside 2-methylguanosine. Thus, RNase T1 digests can be restricted to a single unmodified guanosine or 2-methylguanosine residue, which will also be the 3'-nucleotide in the sequence.

RoboOligo incorporates such constraints within the mass tolerance options as well as the RNase digest selection. Endonuclease restrictions for RNase T1 (‘strict’ – digestion at unmodified guanosine only and ‘broad’ – digestion at either unmodified guanosine or methylated guanosine), RNase A (digestion at unmodified pyrimidine), and RNase U2 (digestion at unmodified purine) are supported. For example, if the ‘RNase T1 strict’ selection was used during the composition analysis then any composition not containing only one guanosine would be considered invalid. The user is also able to bypass this constraint for scenarios in which an RNase digestion context is inappropriate; however, the quantity of valid but incorrect sequences could drastically increase, along with an increase in total computation time.
A further constraint is imbedded within the program by limiting the occurrence of particular modified nucleosides. The majority of RNA samples containing modified nucleosides should yield RNase digestion products that are not significantly large (Grosjean, 2009). Under such conditions, it is unlikely that multiple instances of the same modified nucleotide will be present within a single oligomer. This property is exploited such that only C, U, A, G, dihydrouridine (D), and singly-methylated C, U, A, and G are allowed to occur more than once within a composition. However, these default settings can be edited by the user if the sample to be analyzed dictates such a change.

In addition to the constraints mentioned above, a tolerance for the number of missing product ions can also be set within the program. This feature accounts for empirical aspects of typical MS/MS data, wherein occasionally a product ion may be missing in the data, especially for longer oligonucleotides, or the mass range of the experimentally acquired data may not include the product ion that is expected. The user has the ability to define the number of skips that the program can use to compensate for this missing data by allowing the placement of the nucleotide being tested, and then continue to iterate through and test nucleotides at the next position. In a case where all of a developing sequence’s permitted skips are exhausted, the algorithm will change the direction of the local search (from 5’ → 3’ to 3’ → 5’) and will begin to examine the −y ion series for the fitness evaluation. The composition context is conserved after this sequencing direction change, ensuring that all 3’ to 5’ nucleotide additions would eventually meet the
progress made in the initial –c ion series sequencing attempt and sum to a total mass that falls within the target mass range. Skips can also compensate for modified bases that fragment in unusual ways by inferring the position of the base and then building evidence for its placement. This is accomplished by the successful incorporation of subsequent nucleotides or, if the growing oligomer reaches the precursor mass range and is evaluated as a valid complete sequence, via m/z values that correspond to predicted –c, –y, –w and –(a-B) product ions for the nucleotide at that inferred position.

3.2.2 The Automated De Novo Sequencing User Interface

The automated de novo sequencing algorithm requires the following user input (Figure 3.3):

1. Spectra to be analyzed: A single spectrum or multiple spectra may be chosen.

2. Nucleotide pool: The nucleotides to be included in the sequencing attempt. Due to an exponentially increasing computational workload (Figure 3.4), it is advised to limit the nucleotides to be used in the analysis.

3. CID product m/z tolerance: defines the range around a theoretical CID fragmentation product in which an m/z data point would be considered as a match. The quality of the data and accuracy of the mass spectrometer used to obtain the data should be used as a guideline when setting this
tolerance. A smaller value is ideal, as it would reduce the number of false-positive sequences; however, larger m/z tolerances may be required to find data points that deviate substantially from their theoretical values.

4. Precursor m/z tolerance: For a ‘complete’ sequence to be evaluated as valid, its total mass must fall within the range defined by the precursor mass and the precursor tolerance. A lower value will reduce the number of false-positives, but it also risks missing the correct sequence if the tolerance is set too low.

5. Skips: The purpose of this parameter is to provide some leeway for less than ideal data that may be missing important CID product ion m/z values. One skip is set as the default value. This skip would be used at the beginning of any sequencing attempt if the minimum m/z setting on the mass spectrometer is greater than the m/z of the potential –c₁ fragment ion.

6. 5’ and 3’ ends: The phosphate status on the 5’ and 3’ termini (linear phosphate, cyclic phosphate or no phosphate) are user-defined and employed in calculating the total mass of the oligonucleotide.

7. RNase digest context: If the RNA used to generate the data was first digested by an RNase, then it is suggested that this parameter be set to match the digestion condition. This selection not only limits the number of nucleotide compositions that fit a target mass, but also restricts sequences to those that are expected from the RNase used in the digestion.
Figure 3.3 Automated de novo sequencing. **A** MS/MS data can be selected and moved to the ‘To be sequenced’ box to be included in the analysis. **B** MS/MS data with less than a user selected number of $m/z$ reads and/or maximum ion abundance can be removed from the ‘Scan collection’ box. **C** Nucleotides to be included in the automated de novo sequencing attempt. New entries in the ‘Saved pools’ dropdown list can be added by editing the ‘NucleotidePools.txt’ file. **D** The $m/z$ tolerance, target mass tolerance, RNase digestion context, 5’ and 3’ ends, and number of allowed skips are all modifiable by the user.
Figure 3.4 Effects of increasing nucleotide pool size on automated de novo sequencing efficiency of the 6mer ‘U-Um-U-cmmn5s2U-U-Gp’ and the 9mer ‘U-U-t6A-A-U-C-A-U-Gp’. Sequences that are at least minimally supported by the MS² data are considered ‘Valid Sequences’. A The computational time of each analysis. For the 9mer, a pool size 26 nucleotides was the maximum condition before encountering issues with system memory usage. B The number of nucleotide compositions that fit within the target mass range and pass the restriction digest confinement filter. C The number sequences returned that are supported by the given data. The correct sequence of the 9mer was maintained with all nucleotide pools tested while the 6mer sequence was tied with ‘UCCU1mUlp’ after the addition of inosine (I) to the nucleotide pool (step 14).

3.2.3 Accuracy of automated de novo sequencing

Two isolated tRNA isoacceptors, *E. coli* ΔqueC ΔqueF pGAT-queC tRNA\(^{\text{Asp}}\)\(_{\text{GUC}}\) (Phillips et al., 2011) and *E. coli* tRNA\(^{\text{Gln}}\)\(_{\text{UUG}}\) (Rodriguez-Hernandez et al., 2013)
were separately analyzed to assess the accuracy of the automated de novo sequencing algorithm. The nucleotide pools used for each sequencing attempt were determined by independent LC-MS/MS nucleoside analysis and found to be: C, U, A, G, 1sU, 1mA, 1mG, 1mU, D and cmnm\(^5\)s\(^2\)U for tRNA\(^{\text{Gln}}\)\(_{\text{UUG}}\); and C, U, A, G, 1mA, 1mG, 1sU, preQ\(_0\), D, and G+ for Asp-tRNA\(^{\text{Asp}}\)\(_{\text{GUC}}\). The position of the methyl groups cannot be determined with this data, so the general designation for a singly methylated nucleotide is ‘1m[nucleotide symbol]’. For instance, m\(^1\)G and m\(^2\)G are both represented by ‘1mG’. All attempts were allowed one skip, a precursor mass tolerance of 1.0 and a CID product m/z tolerance of 0.3, and the RNase digestion context of ‘RNase T1 strict’ unless otherwise stated.

For tRNA\(^{\text{Asp}}\)\(_{\text{GUC}}\) RNase T1 digest experimental data, previous manual analysis of the LC-MS/MS spectra identified 16 unique oligomer sequences (Phillips et al., 2011). The automated sequencing attempt by RoboOligo identified 14 out of 16 of these oligomers correctly, each with corresponding sequences that achieved the highest cumulative intensity scores (Table 3). The most complex sequence evaluated, based on length (9 nucleotides) and the number of modified nucleotides, was CCU[preQ\(_0\)]UC[m\(^2\)A]CGp. This sequence had a cumulative ion abundance score that was 5% higher than the next highest sequence, which was the isomer CUC[preQ\(_0\)]UC[m\(^2\)A]CGp that differs only by a C and U base flip between positions two and three. This result highlights a general challenge with RNA sequencing and that is that C and U isomers are
particularly troublesome as the mass difference between the two is one $m/z$ if singly charged and 0.5 and 0.33 if doubly or triply charged, respectively. The 3' end of tRNA$^{\text{Asp}}$, CCA-OH, was one of the incorrectly sequenced oligomers and differed from the top scorer, CAC, by 1.5%. The other failed sequencing attempt occurred with m$^7$GUCGp, which is likely due to the unusual fragmentation pattern of m$^7$G (Wong et al., 2013). When m$^7$G is present in an oligonucleotide the weak glycosidic bond fragments producing a large peak that is 165 Da smaller than the precursor and often very few –c and –y product ions are observed. Analysis of tRNA$^{\text{Gln}}_{\text{UUG}}$, conducted in the same manner as tRNA$^{\text{Asp}}_{\text{GUC}}$, correctly identified 13 out of 13 oligomers (Table 3). The sequence with the most modified nucleotides, UU[mU][cmnm$^5$-s²U]Ugp, had a cumulative intensity score 12% higher than that of the second ranked sequence. The longest sequence in this data set was the 9-mer CAUUCCCUGp, which was correctly identified.
Figure 3.5 Analysis of independently verified *E. coli* ΔqueC ΔqueF pGAT-queC tRNA\[^{Asp}\]\[^{GUC}\] and *E. coli* tRNA\[^{Gln}\]\[^{UUG}\] using the automated de novo sequencing function of RoboOligo. Black letters indicate verified sequences that were corroborated as top scorers by RoboOligo automated de novo sequencing. 

A Sequence coverage of *E. coli* tRNA\[^{Asp}\]\[^{GUC}\]. Nucleotide Pool: C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmnm5s2U

B Sequence coverage of *E. coli* tRNA\[^{Gln}\]\[^{UUG}\]. Nucleotide Pool: C, U, A, G, 1mA, 1mG, 1sU, preQ0, D, G+
Table 3 Summary of results generated by the automated sequencing of \textit{E. coli} ΔqueC ΔqueF pGAT-queC tRNA$^{\text{Asp}_{\text{GUC}}}$ and \textit{E. coli} tRNA$^{\text{Gln}_{\text{UUG}}}$. The T1 restriction digest confinement was used to match the sample preparation. Unless otherwise noted, all automated sequencing attempts were allowed one skip, a target mass tolerance of 1 dalton, and a mass-to-charge tolerance of 0.3. ‘Rank’ refers to the relative position of the independently verified sequence in relation to all of the other valid sequences found via automated sequencing. When the correct sequence has the highest cumulative intensity then it is ranked as 1 and the spectrum is considered to have been successfully solved. ‘% Diff.’ is the cumulative intensity difference between the correct oligomer sequence and second highest scoring oligomer if the correct sequence is the highest scoring result. Otherwise, it measures the difference between the highest scoring sequence and the independently verified correct sequence. ‘Comp. Sequences’ is the number of sequences which the m/z data supports, given user-defined sequencing parameters. ‘Compositions’ is the number of nucleotide compositions that fit the target mass range, given the user-defined nucleotide pool. ‘Time’ measures the time in milliseconds that the program takes to return the automated sequencing results.
Table 3

<table>
<thead>
<tr>
<th><em>E. coli</em> ΔqueC ΔqueF pGAT-queC tRNA&lt;sup&gt;45&lt;/sup&gt;</th>
<th>Nucleotide Pool: C, U, A, G, 1mA, 1mG, 1sU, preQ0, D, G+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequence</strong></td>
<td><strong>Precursor Mass</strong></td>
</tr>
<tr>
<td>CGp</td>
<td>667.37</td>
</tr>
<tr>
<td>Āgp</td>
<td>691.27</td>
</tr>
<tr>
<td>s&lt;sup&gt;4&lt;/sup&gt;UAGp&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1013.42</td>
</tr>
<tr>
<td>DCp</td>
<td>975.52</td>
</tr>
<tr>
<td>CAGp</td>
<td>996.39</td>
</tr>
<tr>
<td>CCA&lt;sup&gt;b&lt;/sup&gt;</td>
<td>876.26</td>
</tr>
<tr>
<td>T&lt;sup&gt;Ψ&lt;/sup&gt;Cp&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1293.38</td>
</tr>
<tr>
<td>DDSGp</td>
<td>653.53</td>
</tr>
<tr>
<td>ψCp</td>
<td>1278.47</td>
</tr>
<tr>
<td>m&lt;sup&gt;3&lt;/sup&gt;GUCGp&lt;sup&gt;d&lt;/sup&gt;</td>
<td>???</td>
</tr>
<tr>
<td>UCCGp&lt;sup&gt;e&lt;/sup&gt;</td>
<td>791.83</td>
</tr>
<tr>
<td>UUCCGp</td>
<td>791.83</td>
</tr>
<tr>
<td>UUGAgp</td>
<td>803.91</td>
</tr>
<tr>
<td>AAUACUGp&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1285.97</td>
</tr>
<tr>
<td>CCUp&lt;sub&gt;preQ0&lt;/sub&gt;UCm&lt;sub&gt;2&lt;/sub&gt;ACGp&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1453.5</td>
</tr>
<tr>
<td>CCUG+UCm&lt;sub&gt;2&lt;/sub&gt;ACGp&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1461.9</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th><em>E. coli</em> tRNA&lt;sup&gt;45&lt;/sup&gt;&lt;sub&gt;Lys&lt;/sub&gt;</th>
<th>Nucleotide Pool: C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmnm5s2U</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequence</strong></td>
<td><strong>Precursor Mass</strong></td>
</tr>
<tr>
<td>CGp</td>
<td>667.14</td>
</tr>
<tr>
<td>CGGp&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1026.14</td>
</tr>
<tr>
<td>CCA</td>
<td>876.14</td>
</tr>
<tr>
<td>DAAp</td>
<td>663.73</td>
</tr>
<tr>
<td>T&lt;sup&gt;Ψ&lt;/sup&gt;Cp&lt;sup&gt;c&lt;/sup&gt;</td>
<td>646.64</td>
</tr>
<tr>
<td>UAs&lt;sup&gt;4&lt;/sup&gt;UCGp&lt;sup&gt;a&lt;/sup&gt;</td>
<td>811.68</td>
</tr>
<tr>
<td>CCAAGp</td>
<td>814.89</td>
</tr>
<tr>
<td>CACGp</td>
<td>802.76</td>
</tr>
<tr>
<td>UUmUcmnm&lt;sub&gt;5s2&lt;/sub&gt;UGp</td>
<td>1004.49</td>
</tr>
<tr>
<td>m&lt;sup&gt;3&lt;/sup&gt;AUCCGp&lt;sup&gt;e&lt;/sup&gt;</td>
<td>974.84</td>
</tr>
<tr>
<td>AAUCCAGp&lt;sup&gt;f&lt;/sup&gt;</td>
<td>1132.35</td>
</tr>
<tr>
<td>UACCCCAp</td>
<td>1272.99</td>
</tr>
<tr>
<td>CAUUCCUGp&lt;sup&gt;f&lt;/sup&gt;</td>
<td>942.91</td>
</tr>
</tbody>
</table>
An additional 48 MS/MS spectra from various RNA samples (Appendix Table 6) were analyzed by RoboOligo in the same manner as the two tRNA isoacceptors described above. Regarded as a whole, the RoboOligo automated *de novo* sequencing algorithm correctly identified the sequences of 73 out of 77 (94%) independently verified oligomers. Of the four incorrect sequence identifications, two of these sequences were the second highest scoring choices, one sequence was the fifth highest scoring result, and one sequence could not be determined (Figure 3.6). Two of the longest oligomers tested were the 14-mers ACUCUU[t^6A]AUCUAUGp and ACU[cmmn^{5}s^{2}U]UU[t^6A]AUCAAGp from *L. lactis* total tRNA. The former was correctly identified as the top scorer while the latter scored as the second highest and had a cumulative abundance that differed from the top result by 1%. The automated sequencing trials of the *L. lactis* 14mers were not included in the aggregate results because we did not attempt to analyze the other oligomers within the data set.
Figure 3.6 Sequencing efficiency in relation to oligomer length. **A** Accuracy as a function of RNA oligomer length. **B** Relative score of all tested oligomers. 1\(^{st}\) 95% (73/77) 2\(^{nd}\) 3% (2/77), 5\(^{th}\) 1% (1/77), and uninterpretable 1% (1/77)

3.3 Manual Sequencing

The manual sequencing portion of RoboOligo is the conceptual successor of SOS (Rozenski and McCloskey, 2002) in that it provides a platform for a researcher to build RNA sequences \textit{ab initio} in either the 5’ or 3’ direction using the –c, –y, –w and –a-B product ions series’ alone or in any combination. Upon selection of the spectrum to be analyzed, the program generates a graphical representation of the data and automatically interprets the charge and mass of the precursor to calculate a theoretical oligomer target mass. This number, when coupled with the target mass tolerance, determines the mass range in which the oligomers built by the user will be evaluated as a valid potential sequence.

The \textit{m/z} tolerance is the range used to determine if a spectral data point fits a theoretical CID product ion. The required size of this value depends on the mass
spectrometer technology and the data quality. Testing of this software was performed with data generated by a Thermo LTQ-XL in normal scan mode during LC-MS/MS analysis of tRNAs digested with RNase T1 or RNase A. The minimum MS/MS m/z tolerance tested with this data was 0.3, which was sufficient for finding product ions without significant false positives. Choosing too small of an MS/MS tolerance could miss relevant data points that differ slightly from theoretical fragmentation values. It is recommended that various precursor mass tolerances and MS/MS m/z tolerances be tested using known RNA or oligoribonucleotide standards to identify which combination of values produces the most accurate results while also limiting the number of valid but incorrect sequences.

A useful starting point when analyzing an oligomer is to determine the possible nucleotide compositions by using the ‘Composition Analysis’ form. The composition analysis requires a nucleotide pool, target mass, target mass tolerance, the nature of the 5’ and 3’ ends of the oligomer, and the RNase digestion context if applicable. All of these factors can be edited by the user to fit the profile of the sample and the instrumentation used. The nucleotide pool can contain any number of standard and modified nucleotides, but the interpretation of the results is simplified if the pool is limited to only the expected nucleotides. In a case where the ‘RNase T1strict’ digestion context is selected, only compositions containing one unmodified guanosine will be evaluated as valid and compositions that do not fit the context will be filtered away. The resulting
compositions are transferred back to the manual sequencing form where they can be individually selected, limiting the usable nucleotide list to only those within the composition. Alternatively, if a composition is not chosen, then all of the nucleotides within the program’s database will be accessible for manual sequencing operations.

Manual sequencing enables the user to build sequences one position at a time by clicking on the symbol representation of the nucleotide with the “Nucleotide” list (Figure 3.7). The program will calculate the theoretical CID product ion $m/z$ values and then attempt to find and label the matching $m/z$ values within the mass spectrum data displayed in the graph. Sequences can be built in either the 5’ to 3’ direction (generating $–c$ and $–a$-B fragments) or the 3’ to 5’ direction (generating $–y$ and $–w$ fragments). Sequences may be saved in any state of progress to the “5’ Series” or the “3’ Series” tabs in the workbench (depending on sequencing direction chosen by the user). Built sequences with total masses that fit within the range of the target mass can be saved to the “Complete” tab of the workbench. Once there, all theoretical $–c$, $–a$-B, $–y$, and $–w$ product ions are calculated and labeled in the spectrum. Each workbench list is sorted in order of high to low cumulative intensities, with the idea that sequences with higher summed intensities are more likely to be correct. Manual sequencing in RoboOligo not only allows for quick prototyping of putative sequences, but also serves as a capable tool for the detailed analysis of long oligomers that may prove troublesome for the automated de novo sequencing algorithm.
Figure 3.7 The RoboOligo primary interface - for manual sequencing and automated sequencing results analysis. A. The MSMS data analyzed in this figure was generated from a mutant E. coli tRNA_Gln_UUG and the returned sequences are the product of RoboOligo’s ‘variable sequencing’ function when given the seed sequence of ‘x-x-x-cmnm5s2U-x-x’. The top score, ‘U-1mU-U-cmnm5s2U-U-Gp’ is the correct sequence. B. Nucleotide composition analysis results that fall within the target mass range (2009.988 +/- 1) and that obey the RNase T1 digestion confinement, which limits sequences to one G at the 3’ end. Nucleotides used in the analysis are those found in WT E. coli tRNA_Gln_UUG: C, U, A, G, D, 1mG, 1mU, 1mA, 1sU, and cmnm5s2U. C. The nucleotide pool contains 108 unique masses of normal and modified nucleotides. Clicking on a nucleotide will add it to the ‘Current Sequence’ and attempt to find the corresponding CID ionization products. D. The ‘Current Sequence’ contains the cumulative intensity of all found CID fragments, sequence orientation selection, 5’ and 3’ end selections, and buttons to add sequences to the appropriate workbench. E. Clicking on a nucleotide in the ‘Current Sequence’ will display the m/z data points that fall within the range of that nucleotide’s theoretical CID fragmentation products (-/+ tolerance). F. The three workbenches store sequences that result from automated de novo sequencing, variable sequencing, and manual sequencing; and are sorted from high to low cumulative intensity. The relative intensity (RI) is the ratio of a sequence’s cumulative intensity compared to the sequence with the highest cumulative intensity in the workbench.
3.4 Variable Sequencing

The variable sequencing technique is a hybrid of the manual sequencing approach and the automated *de novo* sequencing algorithm, and has the benefit of automated sequence investigation while providing more direct control of the sequences being tested. This technique introduces the variable nucleotide pools that the user is able to define. In addition to the variable nucleotides, the defined nucleotides, which behave in the same manner as those in the manual sequencing form, can be thought of as constants. Using variable and defined nucleotides, the user can generate a sequence of any combination of the two. This sequencing mode is useful when a portion of the oligonucleotide sequence is known, but may contain modifications that are not known. For instance if the 5’ and 3’ of the oligonucleotide sequence is known, then the sequence CUXXXAG can be entered where CU at the 5’ end and AG at the 3’ end are defined and the XXXX in the middle are unknown nucleotides. The position and identity of the defined nucleotides will be conserved, while each variable nucleotide will iterate through its user-defined nucleotide pool and attempt to find m/z values within the data that match the theoretical CID fragmentation of each sequence generated. The variable sequencing approach is especially useful when parts of the oligonucleotide are known, or at least suspected, and the user wants to quickly test unknown residues. It also works well for generating the first or last few nucleotides of sequences, which can be used as a foundation for the user to manually sequence the rest of the oligonucleotide.
The variable sequencing algorithm, like the automated de novo sequencing algorithm, utilizes the local-search (Rozenski and McCloskey, 2002) paradigm for checking whether the data in the spectrum supports the growing polynucleotide. As such, if the variable nucleotide iterates to a nucleotide whose primary fragment m/z (5’ to 3’: –c ions and 3’ to 5’: –y ions) is not supported then the algorithm will attempt to skip to the next nucleotide. This process involves incorporating the unsupported nucleotide and moving on to the next residue. If the user-allotted skips for a sequence are exhausted then that sequence will fail, along with all of the potential sequences that would have arisen if the data had supported its placement.

As an example, we analyzed MS/MS data with a precursor m/z of 1004.49 and a charge of 2− generated from RNase T1 digested E. coli tRNA\textsuperscript{Gln\textsubscript{UUG}}. Figure 3.8 shows the ‘variable sequencing’ interface. The sequence X-X-X-cnm\textsuperscript{5}s\textsuperscript{2}U-X-X, where each unidentified sequence position (X) could be C,U,A, G, 1mU, 1mA, 1mG, 1sU, or cmnm\textsuperscript{5}s\textsuperscript{2}U (all of the unique masses in the published sequence of E. coli tRNA\textsuperscript{Gln\textsubscript{UUG}}), was attempted with a CID product m/z tolerance of 0.3 and a precursor mass tolerance of 1.0. The results of variable sequencing are returned after summation of m/z ion abundances and are sorted from high to low cumulative ion abundances. The top score corresponded to the known sequence of U-Um-U-cnm\textsuperscript{5}s\textsuperscript{2}U-U-Gp, which is found in the anticodon loop of tRNA\textsuperscript{Gln\textsubscript{UUG}} (Machnicka et al., 2013).
Figure 3.8 Variable sequencing of an MSMS spectrum with a precursor m/z of 1004.49. In this scenario, the fourth nucleotide (cmnm\textsuperscript{5} s\textsuperscript{2} U) is defined, while the three preceding and two following user-defined ‘x’ variable nucleotides will be determined using a local search method that is similar to the automated \textit{de novo} sequencing algorithm. \textbf{A.} The variable nucleotides x, y, and z can be individually assigned different nucleotide pools. \textbf{B.} The ‘seed’ sequence can contain any number and combination of defined and variable nucleotides. \textbf{C.} User-defined CID product m/z tolerance, precursor m/z tolerance, skips, 3’ and 5’ ends.
3.5 Discussion

The analysis of tandem mass spectrometry data of complex RNA oligomers to generate sequence information has relied on the time consuming efforts of experts within the field. RoboOligo was created to serve as a tool for the analysis of such data, where it may serve as a productivity-increasing aide to those with expertise and also as an entry point for researchers with rudimentary knowledge of the data produced by collision induced dissociation tandem mass spectrometry of complex RNA.

We introduce for the first time a method coined as ‘variable sequencing’ that combines the hands-on approach of manual sequencing with the high-throughput sequencing analysis of the automated de novo sequencing algorithm. This technique allows for quick sequence prototyping and specific nucleotide position evaluation given fixed sequence constraint. Additionally, RoboOligo dramatically simplifies and expedites manual sequencing by providing the user with a simple interface that searches for and automatically labels m/z values that correspond with the inputted oligonucleotide data.

The automated de novo sequencing algorithm showed 94% accuracy for the 77 tested RNA oligomers of lengths from two to twelve bases. We show that the local-search paradigm is effective for sequencing RNase digestion products and with the support of 108 standard and modified bases, RoboOligo is a suitable tool for analyzing biologically derived tRNA, rRNA, and other modified base-containing RNA data. Oligomers greater than nine nucleotides will often require
smaller nucleotide pools or higher mass accuracy as the number of valid compositions and sequences increases exponentially (Figure 3.4), potentially creating computational workloads that push the capabilities of the modern desktop computer. Nucleotide pool sizes should be tempered to the oligomer length and reasonable expectations of the modified nucleosides within the sample. Furthermore, although most of the oligomers were corroborated by external evaluation of the data, it should be noted that not all MS/MS scans with similar precursor masses and charge states will produce the same results. Incomplete MS/MS data, improper charge designation, and the presence of different oligomers with similar precursor mass and charge all contribute to explain this discrepancy. Ultimately, the user will have to make the final judgment in sequence determination while RoboOligo can serve as a tool to assist with the process.
Chapter 4: ADAT2 and ADAT3 – Roles Beyond A to I tRNA Editing?

4.1 Introduction

As stated in Chapter 1.2.1, we have shown that adenosine to inosine deamination editing ($A_{34}$ to $I_{34}$) of tRNA$^{Thr}_{AGU}$ in *Trypanosoma brucei* (Figure 4.1) is catalyzed by the heterodimer ADAT2/ADAT3. tRNAs with $I_{34}$ form favorable base pairs with uracil, adenine, and cytosine and results in expanded decoding capability, reducing the number of required tRNA isoacceptors for translation, since a single tRNA can decode more than one codon (Curran, 1995). $A_{34}$ to $I_{34}$ editing occurs in the cytoplasm, but immunofluorescence and Western blot analysis of epitope-tagged ADAT2 and ADAT3 has revealed that the two proteins localize not only to the cytoplasm, but are also found within the nucleus (Gaston, 2009; Gaston et al., 2007; Rubio et al., 2006). The nuclear localization of these proteins is curious and implies some unknown function.
Figure 4.1 The three *Trypanosoma brucei* tRNA<sup>Thr</sup> isoaecceptors undergo deamination RNA editing in the anticodon loop. A tRNA<sup>Thr<sub>CGU</sub></sup> C<sub>32</sub> to U<sub>32</sub> (19 / 36 sequenced clones) B tRNA<sup>Thr<sub>AGU</sub></sup> A<sub>34</sub> to I<sub>34</sub> and C<sub>32</sub> to U<sub>32</sub> (18 / 30 sequenced clones) C tRNA<sup>Thr<sub>UGU</sub></sup>: C<sub>32</sub> to U<sub>32</sub> (2 / 33 sequenced clones)
We predict one of three possible activities for nuclear ADAT2/3. The first possibility is the changing of a gene’s function or its regulation via nucleic acid editing. In fact, transcription via RNA Polymerase II in Trypanosomes is a largely unregulated event, creating large polycistronic mRNA that must be processed in order to be translated (Ginger et al., 2002; Horn, 2001). We wonder if nuclear ADAT2/3 could possess the ability to edit these mRNA in such a way as to regulate their translation or affect the translated product via code change. As discussed in Chapter 1.2, apolipoprotein B editing complex 1 (APOBEC1) in vertebrates deaminates a specific cytidine of the apolipoprotein B mRNA to uridine. This converts a codon for an amino acid into an early stop codon and, when translated, results in a truncated protein (Navaratnam et al., 1993). In Schizosaccharomyces pombe, the ADAT3 homolog TAD3 plays a role in cell cycle progression. A TAD3 mutant that destabilized the TAD2/TAD3 heterodimer and was incapable of adenosine to inosine editing at the wobble base of a subset of tRNAs. Whether the cell cycle arrest was due to a ‘secondary’ role of TAD3 or as a result of decreased translation efficiency of cell cycle genes was not determined (Tsutsumi et al., 2007).

The second possible role of nuclear ADAT2/3 is in antigenic variation via hypermutation of variable surface glycoprotein (VSG) genes (Rubio et al., 2007). VSG is a molecule that coats the outer membrane of trypanosomes and allows them to evade host immune detection (David Barry and McCulloch, 2001). The mechanism of diversification of VSG genes is remarkably similar to the
diversification of antibodies by activation-induced deaminases (AID) in mammals. The vertebrate AID family of proteins have a similar enzymatic activity as the APOBECs but differ dramatically in their biological role. Activation-induced deaminases function in the nucleus where they are central to antigen-driven antibody diversification, either by somatic hypermutation or class-switch recombination (Muramatsu et al., 2000). Somatic hypermutation contributes to antibody diversification by introducing point mutations to the immunoglobulin variable (IgV) segment gene in developing B cells; a required process as the number of potential antigens far exceeds antibody combinations encoded by the mammalian genome. Transcription of the immunoglobulin genes provides the single-stranded DNA substrate for AID that targets deoxycytidine for deamination to form deoxyuridine. A uracil-DNA glycosylase (UNG) then removes the uracil, creating an abasic residue that, once repaired, will create either a transition or transversion of that nucleotide (Maul and Gearhart, 2010).

Additionally, ADAT2/3 has the Zn\(^{2+}\) coordinating motif that is typical of cytidine deaminase: (H/C)…CX\(_2\)C and can perform C to U conversion on single-stranded DNA (Rubio et al., 2007). With this fact and the nuclear localization of ADAT2 and ADAT3 in mind, it was proposed that the nuclear ADAT2 and/or ADAT3 may play a role VSG diversification (Rubio et al., 2007).

Finally, nuclear ADAT2 and ADAT3 may catalyze the C\(_{32}\) to U\(_{32}\) conversion of tRNA\(^{Thr}\). Our group has previously shown that RNAi of ADAT2 showed a reduction in C\(_{32}\) to U\(_{32}\) conversion of tRNA\(^{Thr}\) (Rubio et al., 2006);
however, attempts to recreate this reaction \textit{in vitro} have not been successful. It is possible that nuclear ADAT2 or ADAT3 associate with different components that enable it to deaminate C$_{32}$ or that tRNA$^\text{Thr}$ requires modifications before it can be recognized as substrate for deamination.

Our aim was to investigate these possibilities with the principle in mind that ADAT2 and/or ADAT3 would need to associate with other protein factors in order to facilitate such activities. Our approach involved the affinity purification of epitope-tagged ADAT2 and ADAT3 constructs followed by the analysis of any associating polypeptides. We believe that the results of such an experiment, if they did not support our predictions, could at least provide a significant lead in future hypotheses and experimental direction in elucidating the role of nuclear ADAT2 and ADAT3.

4.2 Results

4.2.1 Affinity purification of epitope-tagged ADAT2 and ADAT3

We began by constructing C-terminally tagged \textit{T. brucei} ADAT2 and ADAT3 with the protein C epitope, which were placed into the tetracycline-inducible plasmid pLew83 - a derivative of pLew82 that confers puromycin resistance. This construct was specifically designed for affinity purification using the commercially available Anti-Protein C Affinity Matrix (Roche Applied Sciences).
Several attempts of overexpressing an epitope-tagged ADAT2 for affinity purification have produced mixed results. Previously, expression of a His-tagged ADAT2 had been successfully probed via α-His antibodies. However, after several generations the same cell line no longer expressed the epitope-tagged ADAT2, despite retaining puromycin resistance. We believe that this could be a result of widespread hypermutation caused by overexpression of ADAT2 and, in turn, the selection of cells that dramatically limit the expression of ADAT2-His. For this reason, we focused on ADAT3, which did not behave in this manner.

A preliminary affinity purification of the C-terminally tagged ADAT3 from *T. brucei* total cell lysate (crude extract) showed the copurification of approximately fifteen different proteins (*Figure 4.2A*). ADAT2, which is known to form a heterodimer with ADAT3 in the cytoplasm, coeluted with ADAT3 in nearly equal molar proportion – a proof of principle demonstration. Antibodies specific for *T. brucei* enolase was used as a negative control to show that there was minimal nonspecific binding after the washing steps (*Figure 4.2B*).
Figure 4.2 Affinity purification of ADAT3 epitope tagged with Protein C. 

A Silver-stain of SDS-PAGE: total cell extract (crude), flow-through after binding to column (FT), column wash, and elution. 

B Western of each fraction. Anti-protein C antibodies detected high concentration of ADAT3-ProC in the elution and an undetectable quantity in the wash. Prevalence of ADAT2 in the elution confirms the suitability of this approach in addressing our aim. Anti-Enolase antibodies were used to test for nonspecific binding.

4.2.2 Mass Spectrometric Analysis of ADAT3-ProC Affinity Purification

Elution

We performed a second affinity purification of ADAT3-ProC from total *T. brucei* cell lysates. In hindsight, it may have been more appropriate to utilize a protein fraction from isolated nuclei, but we reasoned that any nuclear-specific interactions with ADAT3-ProC would be retained in the total cell lysate. The eluant was separated by denaturing SDS-polyacrylamide gel (SDS-PAGE) and stained with Coomassie blue to visualize protein bands. Seven protein-containing
sections of the gel were excised and sent for mass spectrometric analysis at the Mass Spectrometry and Proteomics Facility in the Institute for Materials Research at The Ohio State University Mass Spectrometry Center (Figure 4.3). All protein mass spectrometric results were considered ‘identified’ if they reached a MASCOT (Cottrell and London, 1999) score greater than 75 (p < 0.5 — given this data set and the predicted *T. brucei* proteome database that was used to analyze the results).

![Coomassie stained SDS-PAGE showing the elution from the affinity purification of ADAT3-ProC. Boxed areas indicate slices of the gel that were sent for mass spectrometric analysis. The numbering corresponds to the significant results presented in Tables 4.](image)

**Figure 4.3** Coomassie stained SDS-PAGE showing the elution from the affinity purification of ADAT3-ProC. Boxed areas indicate slices of the gel that were sent for mass spectrometric analysis. The numbering corresponds to the significant results presented in **Tables 4**.
In bands 5 and 6 (Table 4.E and Table 4.F), significant polypeptide matches were found that correspond to the *T. brucei* receptor for activated C kinase 1 (RACK1). The MASCOT scores of 235 and 204, respectively, are moderately low, but the portion of the gel from which they were excised corresponds to the expected molecular weight of RACK1 (34.7 kD). RACK1 homologs have been extensively studied in various organisms and are involved in a wide range of activities. First discovered in rat brains, RACK1 serves as an anchoring protein that bound protein kinase C (PKC) after PKC was activated by its substrate, which led to downstream gene regulatory events (Ron et al., 1994). Since that time, RACK1 has been implicated in interacting with kinases, phosphatases, ion channels, signal proteins, receptors, gene regulatory proteins, and ribosomal subunits (Regmi et al., 2008). In *T. brucei*, down-regulation of RACK1 caused cells to become unable to complete cytokinesis, resulting in procyclic trypanosomes with multiple flagella and nuclei (Rothberg et al., 2006). It has been postulated that this phenotype is due to negatively affected translation elongation – *T. brucei* RACK1 RNAi cells are more susceptible to anisomycin, a translation elongation inhibitor (Regmi et al., 2008). Bolstering this hypothesis, the same study demonstrated that RACK1 associates with polysomes and an affinity purification of a TAP-tagged RACK1 pulled down elongation factor 1A (eEF1A).

Our mass spectrometric data from the ADAT3-ProC affinity purification supports the association between RACK1 and eEF1A, but with low scores;
eEF1A scored significantly in two bands: 3 and 7 with scores of 77 and 309, respectively (Table 4.C and 4.G). However, neither band corresponds to the expected mass of eEF1A (53.6 kD), which makes us question the legitimacy of the results. If ADAT3 associates with RACK1 as our results indicate, then the converse experiment involving the affinity purification of a tagged RACK1 should pull down ADAT3. As shown in Regmi et al., 2008, evidence of ADAT3 was not reported. It could be argued, however, that the three regions of the protein gel that they selected for mass spectrometric analysis did not include the portion to where ADAT3 would migrate. To sow more doubt, conceptualizing the functional relationship between ADAT3 (and potentially ADAT2) and RACK1 is unclear. The appearance of RACK1 could be a result of nonspecific binding and, based on these results, the association of ADAT3-ProC with RACK1 within trypanosomes should be met with skepticism.

An uncharacterized protein with RNA binding motifs was identified, but with a low MASCOT score of 84 (GeneDB: Tb11.01.5570). A BLAST search revealed homologs in kinetoplastids (lowest E-value: $2e^{-80}$ – *Leishmania braziliensis*) and no similar proteins were found in other organisms (highest E-value: $6e^{-7}$ - *Tribolium castaneum*). Considering the potential substrates of ADAT3, its association with an RNA binding motif is reasonable; however, with no other information about the gene product of Tb11.01.5570 it is challenging to speculate any further.
In band 4 (Table 4.D), the *Trypanosoma brucei* serine-arginine protein (TRS1) was identified with a low score of 82. In trypanosomes, all mRNA receive a spliced leader RNA at their 5’ end. The spliced leader provides the 5’ terminal cap and, after polyadenylation at the 3’ end, enables the mRNA to participate in translation (Günzl, 2010). TRS1 interacts with the SL RNA and is required for trans-splicing onto nuclear mRNA. Earlier, we proposed that nuclear ADAT2 and ADAT3 may play a role post-transcriptional gene regulation, so perhaps it could do this by deaminating bases in the SL RNA, making the edited mRNAs unsuitable for translation. At this point, this line of thought is conjecture; more convincing data of TSR1’s association with ADAT3 would have to be provided before continuing past the initial observation.

4.3 Conclusions

After careful analysis of the mass spectrometric data, two potential leads were identified: TSR1 and the protein product of Tb11.01.5570 (GeneDB). Several proteins involved in metabolism were identified: glyceraldehyde 3-phosphate dehydrogenase, 2-oxoglutarate dehydrogenase E2, and 2-oxoglutarate dehydrogenase E1. Binding immunoglobulin protein (BiP), a glucose-regulated protein, heat shock proteins, and cytoskeletal proteins: alpha tubulin, beta tubulin, and l/6 autoantigen (Detmer et al., 1997) were also identified. It is likely that these are simply abundant proteins that bound nonspecifically.
At the time of publication, our laboratory is investigating the relationship between ADAT2/3 and a methyltransferase (temporarily termed m3B) that methylates position 32 of tRNA\textsuperscript{Thr} – the same position that undergoes C\textsubscript{32} to U\textsubscript{32} editing. Initial results show that ADAT2/3 and m3B are together required for the methylation event to take place. This suggests that ADAT2/3 and m3B may form a complex; however, m3B was not identified in the bands represented by this mass spectrometry data. It is possible that the portion of the gel that m3B migrates was not chosen for mass spectrometry analysis, but at 36.5 kDa it would be expected in band 5 (Table 4.E). As stated previously, this experiment may yield more fruitful results if only isolated nuclei were used for the analysis.
**Table 4** MS analysis of ADAT3-ProC affinity purification

<table>
<thead>
<tr>
<th>Score</th>
<th>GeneDB</th>
<th>Product</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>461</td>
<td>11.47.0004</td>
<td>2-oxoglutarate dehydrogenase E1 component, putative</td>
<td>1008</td>
</tr>
<tr>
<td>346</td>
<td>11.01.1740</td>
<td>2-oxoglutarate dehydrogenase E1 component, putative</td>
<td>1005</td>
</tr>
<tr>
<td>342</td>
<td>11.01.7010</td>
<td>hypothetical protein, conserved</td>
<td>1072</td>
</tr>
<tr>
<td>232</td>
<td>11.01.6930</td>
<td>ADAT3</td>
<td>365</td>
</tr>
<tr>
<td>89</td>
<td>927.1.2330</td>
<td>beta tubulin</td>
<td>442</td>
</tr>
</tbody>
</table>

**A Band 1**

<table>
<thead>
<tr>
<th>Score</th>
<th>GeneDB</th>
<th>Product</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1368</td>
<td>11.01.3110</td>
<td>heat shock protein 70</td>
<td>690</td>
</tr>
<tr>
<td>582</td>
<td>11.01.6930</td>
<td>ADAT3</td>
<td>365</td>
</tr>
<tr>
<td>560</td>
<td>927.6.3740</td>
<td>heat shock 70 kDa protein, mitochondrial precursor, putative</td>
<td>657</td>
</tr>
<tr>
<td>535</td>
<td>927.8.4180</td>
<td>ADAT2</td>
<td>225</td>
</tr>
<tr>
<td>411</td>
<td>927.1.2330</td>
<td>beta tubulin</td>
<td>442</td>
</tr>
<tr>
<td>143</td>
<td>927.1.2340</td>
<td>alpha tubulin</td>
<td>451</td>
</tr>
<tr>
<td>79</td>
<td>11.02.5450</td>
<td>glucose-regulated protein 78, putative,luminal binding protein 1 (BiP)</td>
<td>653</td>
</tr>
</tbody>
</table>

**B Band 2**

Continued
### Table 4 continued

<table>
<thead>
<tr>
<th>Score</th>
<th>GeneDB</th>
<th>Product</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>767</td>
<td>927.8.4180</td>
<td>ADAT2</td>
<td>225</td>
</tr>
<tr>
<td>519</td>
<td>11.01.6930</td>
<td>ADAT3</td>
<td>365</td>
</tr>
<tr>
<td>504</td>
<td>927.1.2330</td>
<td>beta tubulin</td>
<td>442</td>
</tr>
<tr>
<td>257</td>
<td>927.1.2340</td>
<td>alpha tubulin</td>
<td>451</td>
</tr>
<tr>
<td>119</td>
<td>927.10.6880</td>
<td>glyceraldehyde 3-phosphate dehydrogenase, cytosolic</td>
<td>331</td>
</tr>
<tr>
<td>113</td>
<td>927.10.6400</td>
<td>chaperonin HSP60, mitochondrial precursor</td>
<td>562</td>
</tr>
<tr>
<td>77</td>
<td>927.10.2090</td>
<td>elongation factor 1-alpha, EF-1-alpha</td>
<td>348</td>
</tr>
</tbody>
</table>

#### C Band 3

<table>
<thead>
<tr>
<th>Score</th>
<th>GeneDB</th>
<th>Product</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>3565</td>
<td>11.01.6930</td>
<td>ADAT3</td>
<td>365</td>
</tr>
<tr>
<td>403</td>
<td>927.1.2340</td>
<td>alpha tubulin</td>
<td>451</td>
</tr>
<tr>
<td>328</td>
<td>927.8.4180</td>
<td>ADAT2</td>
<td>225</td>
</tr>
<tr>
<td>176</td>
<td>927.1.2330</td>
<td>beta tubulin</td>
<td>442</td>
</tr>
<tr>
<td>105</td>
<td>11.01.3550</td>
<td>2-oxoglutarate dehydrogenase E2 component, putative</td>
<td>383</td>
</tr>
<tr>
<td>82</td>
<td>927.8.900</td>
<td>splicing factor TSR1</td>
<td>328</td>
</tr>
</tbody>
</table>

#### D Band 4

Continued
Table 4 continued

<table>
<thead>
<tr>
<th>Score</th>
<th>GeneDB</th>
<th>Product</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1553</td>
<td>11.01.6930</td>
<td>ADAT3</td>
<td>365</td>
</tr>
<tr>
<td>1012</td>
<td>927.1.2330</td>
<td>beta tubulin</td>
<td>442</td>
</tr>
<tr>
<td>794</td>
<td>927.8.4180</td>
<td>ADAT2</td>
<td>225</td>
</tr>
<tr>
<td>520</td>
<td>927.1.2340</td>
<td>alpha tubulin</td>
<td>451</td>
</tr>
<tr>
<td>235</td>
<td>11.01.3170</td>
<td>RACK1</td>
<td>318</td>
</tr>
<tr>
<td>111</td>
<td>927.7.3440</td>
<td>I/6 autoantigen</td>
<td>246</td>
</tr>
<tr>
<td>84</td>
<td>11.01.5570</td>
<td>RNA-binding protein</td>
<td>262</td>
</tr>
</tbody>
</table>

E Band 5

<table>
<thead>
<tr>
<th>Score</th>
<th>GeneDB</th>
<th>Product</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>1825</td>
<td>11.01.6930</td>
<td>ADAT3</td>
<td>365</td>
</tr>
<tr>
<td>940</td>
<td>927.8.4180</td>
<td>ADAT2</td>
<td>225</td>
</tr>
<tr>
<td>710</td>
<td>927.1.2330</td>
<td>beta tubulin</td>
<td>442</td>
</tr>
<tr>
<td>662</td>
<td>927.1.2340</td>
<td>alpha tubulin</td>
<td>451</td>
</tr>
<tr>
<td>204</td>
<td>11.01.3170</td>
<td>RACK1</td>
<td>318</td>
</tr>
<tr>
<td>143</td>
<td>11.02.1085</td>
<td>40S ribosomal protein 54, putative</td>
<td>273</td>
</tr>
</tbody>
</table>

F Band 6

Continued
Table 4 continued

<table>
<thead>
<tr>
<th>Score</th>
<th>GeneDB</th>
<th>Product</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>3558</td>
<td>927.8.4180</td>
<td>ADAT2</td>
<td>225</td>
</tr>
<tr>
<td>1610</td>
<td>11.01.6930</td>
<td>ADAT3</td>
<td>365</td>
</tr>
<tr>
<td>368</td>
<td>927.1.2330</td>
<td>beta tubulin</td>
<td>442</td>
</tr>
<tr>
<td>309</td>
<td>927.10.2100</td>
<td>elongation factor 1-alpha, EF-1-alpha</td>
<td>449</td>
</tr>
<tr>
<td>256</td>
<td>927.1.2340</td>
<td>alpha tubulin</td>
<td>451</td>
</tr>
<tr>
<td>132</td>
<td>09.160.4450</td>
<td>40S ribosomal protein S3, putative</td>
<td>266</td>
</tr>
<tr>
<td>81</td>
<td>927.8.6150</td>
<td>40S ribosomal protein S8, putative</td>
<td>220</td>
</tr>
</tbody>
</table>

**G Band 7**
Chapter 5: Concluding Remarks and Future Directions

5.1 Automated *de novo* sequencing of modified RNA oligomers

Chapter 3 discusses the development of the computer program RoboOligo that was designed to aid in the analysis of tandem mass spectrometry (MS/MS) data of modified nucleotide-containing RNA oligomers. Our automated *de novo* sequencing algorithm, using the local search paradigm that builds sequences one nucleotide at a time based on data within an MS/MS spectrum, correctly identified 94% of oligomers from 2 to 11 nucleotides in length (*Figure 3.6*). This approach, along with RNase restriction digest rules that logically constrain sequence possibilities, limit the computational workload and reduce the number of false positive identifications. The end result is a robust algorithm (over 100 modified nucleotides are supported) that can determine the sequence of a 14mer in 44 seconds (using a pool of 9 nucleotides) (*Appendix Table 6*).

RoboOligo is designed for future compatibility. The database of RNA nucleotides used by RoboOligo can be customized by the user — a useful ability given that new natural RNA modifications are still being discovered and a new field of research studying synthetic nucleotides is currently developing. Parallel
processing is a modern technology that facilitates the use of multiple processing cores for efficient computation. The Microsoft .NET 4.5 programming language features a robust parallel processing architecture that works well with the current norm of two to four cores, but is also able to accommodate additional cores, such as those found in low-end super computers. These design features should allow RoboOligo to be a viable choice for the analysis of modified nucleotide-containing RNA oligomers for years to come.

5.2 A novel eukaryotic pathway for wyosine biosynthesis in \textit{T. brucei} mitochondria

In light of the results presented in Chapter 3, we propose a model for wyosine biosynthesis in \textit{T. brucei} and perhaps in all kinetoplastids. Regardless of the intracellular compartment, the pathway starts with a tRNA^{Phe} substrate containing m\(^1\)G\(_{37}\), which is produced by TbTRM5 (Paris et al., 2013). At this point, this substrate has two possible fates; a portion of it is retained in the cytosol and further modified by TbTYW1L, TbTYW2, TbTYW3A and TbTYW4/5, yielding OHyW\(_{37}\) as the end product. Some of this modified tRNA is then imported into the mitochondrion, explaining the presence of this modification in both compartments. Alternatively, either an m\(^1\)G\(_{37}\)-containing or the unmodified tRNA\(^{Phe}\) are imported into the mitochondrion; the former can directly partake in wyosine formation, the latter can be methylated to m\(^1\)G\(_{37}\) once in the mitochondrion. This last proposal is in agreement with our previous description of
a pathway for $m^1G_{37}$ methylation in the *T. brucei* mitochondrion for tRNAs that, because of transport dynamics, may escape cytosolic methylation. Once in the organelle, the $m^1G_{37}$-containing tRNA$^{\text{Phe}}$ is sequentially modified by TbTYW1S and TYW3B yielding imG$_{37}$ as the final product; mitochondrial biosynthesis of wTosine is thus far a uniquely kinetoplastid reaction (Figure 5.1).

Figure 5.1 Proposed biosynthetic pathways of hydroxywybutosine and wTosine in trypanosomes. Following transcription, tRNA$^{\text{Phe}}$ can enter one of two processing pathways. The cytosolic pathway begins with the methylation of G$_{37}$ by TRM5 to form $m^1G_{37}$ and then becomes further modified by TYW1L, TYW2, TYW3A, and TYW4/5 to form hydroxywybutosine (OHyW$_{37}$). OHyW$_{37}$-containing tRNA$^{\text{Phe}}$ is then either used in cytoplasmic translation or imported into the mitochondrion. The second pathway involves the mitochondrial import of tRNA$^{\text{Phe}}$. Once methylated by TRM5, TYW1S and TYW3B further modify $m^1G_{37}$ containing-tRNA$^{\text{Phe}}$ to produce imG$_{37}$. The imG$_{37}$-containing tRNA$^{\text{Phe}}$ is likely retained in the mitochondrion, as imG was not observed in cytoplasmic tRNA nucleosides. The cytoplasmic localization of TYW2 fits this model and potential roles for mitochondrial TYW2 are discussed in this chapter.
The mitochondrial localization of the C terminally Myc-tagged TYW2 (Figure 2.4) is a curious observation. It is unlikely that it is involved in mitochondrial hydroxywybutosine formation since the last enzyme in the pathway, the fused TYW4/5, is cytosolic. Also, attempts to identify other wyosine derivatives, such as the potential product of TYW1 – TYW3 (yw-72) and the product of TYW1 & TYW2 (yw-86) (Figure 2.1), were not successful. We propose three possibilities that could explain the mitochondrial localization of TYW2-Myc. First, TYW2 may be associated with the outer mitochondrial membrane where it, facing the cytoplasm, may catalyze the \(\alpha\)-amino-\(\alpha\)-carboxypropyl group addition to tRNA\(^{\text{Phe}}\) containing imG\(_{1437}\) or imG\(_{37}\) (the order of these modification steps has not been established). Second, the RNA modification \(3-(3\text{-amino-3-carboxypropyl})-5,6\text{-dihydrouridine (acp}^3\text{D})\) was first observed in \(T. brucei\) tRNA\(^{\text{Lys}}\_\text{UUU}\) where U\(_{47}\) is modified to acp\(^3\text{D}_{47}\) (Krog et al., 2011). The enzyme responsible for this reaction has not been identified, but given the activity of TYW2 in catalyzing the addition \(\alpha\)-amino-\(\alpha\)-carboxypropyl group, we propose that it may be involved in the synthesis of the acp\(^3\text{D}_{47}\). However, the difference in substrate structure, dihydrouridine has a base composed of a single ring whereas the wyosine derivatives are composed of a triple ring, challenge the likelihood of this scenario. Perhaps by associating with another protein the TYW2 could act on a different substrate. Finally, the localization may be an artifact of the overexpression, causing the protein to be
imported into the mitochondrion where it plays no biological purpose. These possibilities require additional investigation.

Upon analysis of the HPLC mitochondrial tRNA nucleoside data, we noticed an unidentified peak that eluted at 34.9 minutes – 3 minutes and 45 seconds before the elution of the dinucleotide wyosine-phosphate-adenosine (imGpA). The spectrum profile has a peak at 237nm (the same as imG) and a shoulder peak at 210nm (similar to cytidine). The mass to charge value for this eluting molecule corresponds to the expected mass of a dinucleoside wyosine-phosphate-cytidine (imGpC). Additionally, the nucleoside analysis of mitochondrial tRNA from the TYW1S RNAi-induced cell line showed the loss of this peak. The two nucleotides adjacent to imG$_{37}$ of mitochondrial tRNA$^{\text{Phe}}$ are both adenosine, A$_{36}$ and A$_{38}$. If this peak at 34.9 minutes is imGpC, it means that wyosine is found on both tRNAPhe as well as another, yet to be identified, tRNA within the trypanosome mitochondrion. The wyosine derivatives are currently known to be exclusive to tRNA$^{\text{Phe}}$. Plans are underway to isolate wyosine-containing tRNAs using a benzyl DEAE-cellulose column, which selectively retards the elution of tRNAs with hydrophobic bases, such as wyosine. tRNA isolated in this manner will be sequenced via tandem mass spectrometry.

We hypothesized that the trypanosome mitochondrion requires tRNA$^{\text{Phe}}$ with a wyosine-related RNA modification at G$_{37}$ to prevent translation frameshifting of uridine-rich mRNAs. Based on this hypothesis, we predict that the efficiency of mitochondrial translation may be reduced in either TYW1L or
TYW1S RNAi-induced cell lines, but attempts at assessing this in vivo has produced mixed results (data not shown). Instead, we show that these RNAi lines have reduced viability in low glucose media (TYW1L & TYW1S – Figure 2.8), decreased cellular respiration (TYW1L & TYW1S – Figure 2.9), and an increase in mitochondrial membrane potential (TYW1S – Figure 2.9). These results can be interpreted as effects of reduced mitochondrial maintenance and support our hypothesis, but fall short of excluding other explanations of these phenomena. The ideal experiment would compare the amount of ribosomal frameshifting between wildtype and TYW1L and TYW1S RNAi-induced cell lines, but the design of such an experiment remains a challenge.

5.3 Concluding Remarks

Beyond the argument for accuracy during protein synthesis, the observations presented here offer an interesting tale about the evolution of translational systems and cells in general. It is clear that modifications such as wyosine/hydroxywybutosine originated in the common ancestor of Archaea and Eukarya to the exclusion of the bacterial lineage. Surprisingly, the bacterial translation system of the trypanosomatid mitochondrion has adopted this purely non-bacterial modification to perhaps ensure reading-frame maintenance in the organelle; highlighting the vagaries that cellular systems endure once endosymbiosis and natural selection ensues. I hope that the findings presented here interest and inform the reader as much as we, my colleagues and I, were interested and informed in the daily pursuit of them.
Appendix A: Supplementary Figures
**Figure A.1** TYW1 alignment.

<table>
<thead>
<tr>
<th>T. brucei (TYW1 N-term)</th>
<th>L. major (TYW1 N-term)</th>
<th>K. t. brucei (TYW1 N-term)</th>
<th>S. cerevisiae (TYW1 N-term)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 MELL SSIPSEG DGSVMT1</td>
<td>1 MSSI IDPGV Y0AASD4</td>
<td>1 . . . . . . . . . . . . . . .</td>
<td>1 . . . . . . . . . . . . . . .</td>
</tr>
<tr>
<td>2 T. brucei (TYW1 N-term)</td>
<td>L. major (TYW1 N-term)</td>
<td>K. t. brucei (TYW1 N-term)</td>
<td>S. cerevisiae (TYW1 N-term)</td>
</tr>
<tr>
<td>1 . . . . . . . . . . . . . .</td>
<td>1 . . . . . . . . . . . . . .</td>
<td>1 . . . . . . . . . . . . . .</td>
<td>1 . . . . . . . . . . . . . .</td>
</tr>
</tbody>
</table>

**Flavodoxin-1 domain**

T. brucei (TYW1 N-term)

<table>
<thead>
<tr>
<th>L. major (TYW1 N-term)</th>
<th>K. t. brucei (TYW1 N-term)</th>
<th>S. cerevisiae (TYW1 N-term)</th>
</tr>
</thead>
<tbody>
<tr>
<td>252 YTTKYSQAPD PVY0 . . . . . . . . . . . . . . .</td>
<td>195 VTHRTAIP E1 . . . . . . . . . . . . . . .</td>
<td>147 SWSVAGKL VSRK RFAQ RPS . . . . . . . . .</td>
</tr>
<tr>
<td>253 EDPVRQ . . . . . . . . . . . . . . . . . . . . .</td>
<td>196 APQFA . . . . . . . . . . . . . . . . . . . . .</td>
<td>148 QAYR . . . . . . . . . . . . . . . . . . . . .</td>
</tr>
</tbody>
</table>

A TYW1 N-terminal alignment. Continued
Figure A.1 continued

Radical SAM

Wysos formation

B TYW1 C-terminal alignment.
Figure A.2 TYW2 alignment.
<table>
<thead>
<tr>
<th>T. brucei (Trypanosoma brucei)</th>
<th>L. major (Leishmania major)</th>
<th>L. major (Trypanosoma brucei)</th>
<th>L. major (Trypanosoma brucei)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFLMDEACAYEDEPSKPIKDIENSTGKQGSSCPHERCVRPKNLAG...</td>
<td>GDDFEKHKMNVSRKSSMCMSKPSGK</td>
<td>MDSAPAPGGAPAASPAPAVDAPP...</td>
<td>LPAKPKPVRKPKPVPRKPRVPRARPV...</td>
</tr>
<tr>
<td>158</td>
<td>92</td>
<td>129</td>
<td>158</td>
</tr>
<tr>
<td>T. brucei (Trypanosoma brucei)</td>
<td>L. major (Leishmania major)</td>
<td>L. major (Trypanosoma brucei)</td>
<td>L. major (Trypanosoma brucei)</td>
</tr>
<tr>
<td>159</td>
<td>92</td>
<td>129</td>
<td>158</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>T. pestis</td>
<td>T. pestis</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>139</td>
<td>32</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>M. ananassi</td>
<td>M. levis</td>
<td>M. levis</td>
<td>M. levis</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>32</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
</tbody>
</table>

**Figure A.3 TYW3 alignment.**
Figure A.4 TYW4 alignment

<table>
<thead>
<tr>
<th>T. brucei (TYW4 Al)</th>
<th>L. major (TYW4 Al)</th>
<th>K. papatasi (TYW4)</th>
<th>S. cerevisiae (TYW4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure A.4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. **Figure A.4 TYW4 alignment**

**Leucine carboxyl methyltransferase (LCM)**

T. brucei (TYW4 Al): 313 GEGSGAARAEGRRKGKFGGRKGEFSGGSLGHEEEERFDHSEAEQGDSVDDPHK | L. major (TYW4 Al): 336 QEGSGAARAEGRRKGKFGGRKGEFSGGSLGHEE4ERFDHSEAEQGDSVDDPHK |
K. papatasi (TYW4): 246 KEGSGAARAEGRRKGKFGGRKGEFSGGSLGHEEEERFDHSEAEQGDSVDDPHK | S. cerevisiae (TYW4): 265 KEGSGAARAEGRRKGKFGGRKGEFSGGSLGHEEEERFDHSEAEQGDSVDDPHK |

**Continued**
Figure A.4 continued

B TYW4/5 C terminal alignment.
Figure A.5 Mass spectral analysis of OHyWpA and imGpA from mitochondrial tRNA. A To confirm the identity of OHyWpA, the base peak from the mass spectra at 41.9 min with an m/z of 854.0 was selected for MS/MS by collision-induced dissociation. The MS/MS produced a predictable fragmentation of this dinucleoside monophosphate including a base loss of OHyW (m/z 462.2), adenosine (m/z 268.08), OHyW nucleoside (m/z 393.2), and OHyWp (m/z 587.3). B The base peak corresponding to imGpA at 38.7 min (m/z 664.9) was also analyzed by MS/MS. Although a more complex fragmentation spectrum was obtained, similar peaks found from the fragmentation of OHyWpA were also observed for imGpA, including the loss of the imG base (m/z 462.1), adenosine (m/z 268.1), and the imG nucleoside (m/z 398.2)
Appendix B: Supplementary Tables
Table 5 Protein accession numbers for **A** *Trypanosoma brucei* and **B** *Saccharomyces cerevisiae* wybutosine synthesizing genes.

### A  
*T. brucei* TYW genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYW1Large</td>
<td>XP_846333</td>
</tr>
<tr>
<td>TYW1Small</td>
<td>XP_803801</td>
</tr>
<tr>
<td>TYW2</td>
<td>XP_822560</td>
</tr>
<tr>
<td>TYW3A</td>
<td>XP_829329</td>
</tr>
<tr>
<td>TYW3B</td>
<td>XP_829699</td>
</tr>
<tr>
<td>TYW4/5</td>
<td>XP_827514</td>
</tr>
</tbody>
</table>

### B  
*S. cerevisiae* TYW genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYW1</td>
<td>CAY86753</td>
</tr>
<tr>
<td>TYW2</td>
<td>CAY81818</td>
</tr>
<tr>
<td>TYW3</td>
<td>CAY79710</td>
</tr>
<tr>
<td>TYW4</td>
<td>CAY86151</td>
</tr>
</tbody>
</table>
### Table 6 Automated de novo sequencing results

<table>
<thead>
<tr>
<th>Oligomer Length: 2</th>
<th>Source</th>
<th>Sequence</th>
<th>Pos. % Diff.</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trm14 tRNA&lt;sup&gt;6&lt;/sup&gt;</td>
<td>CGp</td>
<td>1</td>
<td>16</td>
<td>2C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>2</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UGp</td>
<td>1</td>
<td>NA</td>
<td>1C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>1</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGp</td>
<td>1</td>
<td>94</td>
<td>2C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>2</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>E. coli tRNA&lt;sup&gt;5&lt;/sup&gt;</td>
<td>CGp</td>
<td>1</td>
<td>17</td>
<td>2C, U, A, G, 1mA, 1mG, 1mU, preQ, D, G+</td>
<td>T1 strict</td>
<td>2</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGp</td>
<td>1</td>
<td>NA</td>
<td>1C, U, A, G, 1mA, 1mG, 1mU, preQ, D, G+</td>
<td>T1 strict</td>
<td>1</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGp</td>
<td>1</td>
<td>14</td>
<td>2C, U, A, G, D, 1mG, 1mU, 1mA</td>
<td>T1 strict</td>
<td>2</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>S. subtilis tRNA&lt;sup&gt;6&lt;/sup&gt;</td>
<td>UGp</td>
<td>1</td>
<td>NA</td>
<td>1C, U, A, G, D, 1mC, 1mU, 1mA</td>
<td>T1 strict</td>
<td>1</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGp</td>
<td>1</td>
<td>NA</td>
<td>1C, U, A, G, D, 1mC, 1mU, 1mA</td>
<td>T1 strict</td>
<td>1</td>
<td>34</td>
<td></td>
</tr>
</tbody>
</table>

#### A Oligomer length: 2 nucleotides

<table>
<thead>
<tr>
<th>Oligomer Length: 3</th>
<th>Source</th>
<th>Sequence</th>
<th>Pos. % Diff.</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trm14 tRNA&lt;sup&gt;6&lt;/sup&gt;</td>
<td>CGp</td>
<td>1</td>
<td>38</td>
<td>3C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>2</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UAGp</td>
<td>1</td>
<td>66</td>
<td>2C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>1</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAGp</td>
<td>1</td>
<td>20</td>
<td>3C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>2</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GGCp</td>
<td>1</td>
<td>88</td>
<td>5C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>3</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGAp</td>
<td>1</td>
<td>40</td>
<td>4C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>2</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AG Cp</td>
<td>1</td>
<td>27</td>
<td>4C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>2</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>E. coli tRNA&lt;sup&gt;6&lt;/sup&gt;</td>
<td>CCA</td>
<td>2</td>
<td>3.5</td>
<td>8C, U, A, G, 1mU, 1mA, 1mU, D, cmm6s2U,</td>
<td>None</td>
<td>2</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C(1mG)Gp</td>
<td>1</td>
<td>14</td>
<td>6C, U, A, G, 1mU, 1mA, 1mU, D, cmm6s2U,</td>
<td>T1 strict</td>
<td>4</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1mU)AGp</td>
<td>1</td>
<td>37</td>
<td>4C, U, A, G, 1mA, 1mG, 1mU, preQ, D, G+</td>
<td>T1 strict</td>
<td>2</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>E. coli tRNA&lt;sup&gt;6&lt;/sup&gt;</td>
<td>DCGp</td>
<td>1</td>
<td>40</td>
<td>4C, U, A, G, 1mA, 1mG, 1sU, preQ, D, G+</td>
<td>T1 strict</td>
<td>2</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGA p</td>
<td>1</td>
<td>20</td>
<td>4C, U, A, G, 1mA, 1mG, 1sU, preQ, D, G+</td>
<td>T1 strict</td>
<td>2</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCA</td>
<td>2</td>
<td>15</td>
<td>8C, U, A, G, 1mA, 1mG, 1sU, preQ, D, G+</td>
<td>None</td>
<td>2</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>S. subtilis tRNA&lt;sup&gt;6&lt;/sup&gt;</td>
<td>CGp</td>
<td>1</td>
<td>24</td>
<td>4C, U, A, G, D, 1mG, 1mU, 1mA</td>
<td>T1 strict</td>
<td>2</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UAGp</td>
<td>1</td>
<td>64</td>
<td>2C, U, A, G, D, 1mG, 1mU, 1mA</td>
<td>T1 strict</td>
<td>2</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

#### B Oligomer length: 3 nucleotides

Continued
### Table 6 continued

#### Oligomer Length 4

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos</th>
<th>% Diff</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trm14 tRNA</td>
<td>CUAAGp</td>
<td>1</td>
<td>23</td>
<td>9C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>2</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AUCGp</td>
<td>1</td>
<td>22</td>
<td>9C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>2</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGGCp</td>
<td>1</td>
<td>17</td>
<td>6C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>2</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GGACp</td>
<td>1</td>
<td>27</td>
<td>6C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>2</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGAUUp</td>
<td>1</td>
<td>21</td>
<td>3C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>2</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GGGUp</td>
<td>1</td>
<td>NA</td>
<td>1C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>2</td>
<td>106</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AAAUp</td>
<td>1</td>
<td>NA</td>
<td>1C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>1</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>E. coli tRNA</td>
<td>DAAGp</td>
<td>1</td>
<td>18</td>
<td>4C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmnn6s2U</td>
<td>T1 strict</td>
<td>3</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TUCGp</td>
<td>1</td>
<td>23</td>
<td>14C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmnn6s2U</td>
<td>T1 strict</td>
<td>6</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>E. coli tRNA</td>
<td>gUCGp</td>
<td>1</td>
<td>12</td>
<td>12C, U, A, G, 1mA, 1mG, 1sU, prsQ, D, G+</td>
<td>T1 strict</td>
<td>6</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDAGp</td>
<td>1</td>
<td>33</td>
<td>3C, U, A, G, 1mA, 1mG, 1sU, prsQ, D, G+</td>
<td>T1 strict</td>
<td>1</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UCCGp</td>
<td>1</td>
<td>36</td>
<td>5C, U, A, G, 1mA, 1mG, 1sU, prsQ, D, G+</td>
<td>T1 strict</td>
<td>2</td>
<td>73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CACGp</td>
<td>1</td>
<td>29</td>
<td>9C, U, A, G, D, 1mG, 1mU, t6A</td>
<td>T1 strict</td>
<td>2</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCUGp</td>
<td>1</td>
<td>19</td>
<td>6C, U, A, G, D, 1mG, 1mU, t6A</td>
<td>T1 strict</td>
<td>2</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TCCGp</td>
<td>1</td>
<td>9</td>
<td>9C, U, A, G, D, 1mG, 1mU, t6A</td>
<td>T1 strict</td>
<td>2</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DDAp</td>
<td>1</td>
<td>33</td>
<td>4C, U, A, G, D, 1mG, 1mU, t6A</td>
<td>T1 strict</td>
<td>2</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCGGp</td>
<td>1</td>
<td>10</td>
<td>4C, U, A, G, D, 1mG, 1mU, t6A</td>
<td>T1 strict</td>
<td>2</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>m7GU0Gp</td>
<td>1</td>
<td>NA</td>
<td>C, U, A, G, D, 1mG, 1mU, t6A</td>
<td>T1 strict</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

#### Oligomer Length: 5

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos</th>
<th>% Diff</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trm14 tRNA</td>
<td>GUCCA</td>
<td>1</td>
<td>2</td>
<td>39C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>None</td>
<td>2</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UCUAGp</td>
<td>1</td>
<td>9</td>
<td>18C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>2</td>
<td>116</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AUCCGp</td>
<td>1</td>
<td>14</td>
<td>19C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>2</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GGG[1mG]GUp</td>
<td>1</td>
<td>20</td>
<td>4C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>A</td>
<td>2</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>E. coli tRNA</td>
<td>UAs[1U]UCGp</td>
<td>1</td>
<td>8</td>
<td>68C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmnn6s2U</td>
<td>T1 strict</td>
<td>6</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCAGGp</td>
<td>1</td>
<td>10</td>
<td>14C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmnn6s2U</td>
<td>T1 strict</td>
<td>4</td>
<td>207</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAGCGp</td>
<td>1</td>
<td>17</td>
<td>11C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmnn6s2U</td>
<td>T1 strict</td>
<td>3</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td>E. coli tRNA</td>
<td>UCCCCGp</td>
<td>1</td>
<td>3</td>
<td>30C, U, A, G, 1mA, 1mG, 1sU, prsQ, D, G+</td>
<td>T1 strict</td>
<td>6</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UCCCGp</td>
<td>1</td>
<td>8</td>
<td>14C, U, A, G, 1mA, 1mG, 1sU, prsQ, D, G+</td>
<td>T1 strict</td>
<td>3</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UCCAGp</td>
<td>1</td>
<td>8</td>
<td>25C, U, A, G, 1mA, 1mG, 1sU, prsQ, D, G+</td>
<td>T1 strict</td>
<td>3</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B. subtilis tRNA</td>
<td>1</td>
<td>20</td>
<td>11C, U, A, G, D, 1mG, 1mU, t6A</td>
<td>T1 strict</td>
<td>3</td>
<td>76</td>
<td></td>
</tr>
</tbody>
</table>

#### D Oligomer length: 5 nucleotides

Continued
### Table 6 continued

#### Oligomer Length: 6

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos.</th>
<th>% Diff.</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Context</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli tRNA^Gin</td>
<td>U[mU]U[U]cmmm6s2U</td>
<td>1</td>
<td>15</td>
<td>41 C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmmm6s2U.</td>
<td>T1 strict</td>
<td>24</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T[mA]UACCCGp</td>
<td>1</td>
<td>1</td>
<td>91 C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmmm6s2U.</td>
<td>T1 strict</td>
<td>12</td>
<td>231</td>
<td></td>
</tr>
</tbody>
</table>

**E** Oligomer length: 6 nucleotides

#### Oligomer Length: 7

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos.</th>
<th>% Diff.</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tm14 tRNA^Gin</td>
<td>COUACGp</td>
<td>1</td>
<td>9</td>
<td>59 C, U, A, G, D, 1mC, 1mU, 1mA, 1mG</td>
<td>T1 strict</td>
<td>2</td>
<td>152</td>
<td></td>
</tr>
<tr>
<td>E. coli tRNA^Gin</td>
<td>AAUCCAGp</td>
<td>1</td>
<td>12</td>
<td>12 C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmmm6s2U.</td>
<td>T1 strict</td>
<td>40</td>
<td>588</td>
<td></td>
</tr>
<tr>
<td>L. lactis tRNA</td>
<td>CDAA1mACGp</td>
<td>1</td>
<td>19</td>
<td>263 C, U, A, G, D, m05U, 1sU, 1mG, 1mA, 1mC, 6A, 6A, 1mU, cmmm6s2U, ac1C, k2</td>
<td>T1 strict</td>
<td>77</td>
<td>798</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1mAACCAAAGp</td>
<td>1</td>
<td>11</td>
<td>1037 C, U, A, G, D, m05U, 1sU, 1mG, 1mA, 1mC, 6A, 6A, 1mU, cmmm5s2U, ac1C, k3</td>
<td>T1 strict</td>
<td>110</td>
<td>1407</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Au6AAACCGp</td>
<td>1</td>
<td>7</td>
<td>3789 C, U, A, G, D, m05U, 1sU, 1mG, 1mA, 1mC, 6A, 6A, 1mU, cmmm5s2U, ac1C, kM</td>
<td>T1 strict</td>
<td>246</td>
<td>3952</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mGAAUCUGp</td>
<td>1</td>
<td>9</td>
<td>1070 C, U, A, G, D, m05U, 1sU, 1mG, 1mA, 1mC, 6A, 6A, 1mU, cmmm5s2U, ac4C, k4</td>
<td>T1 strict</td>
<td>109</td>
<td>1070</td>
<td></td>
</tr>
</tbody>
</table>

**F** Oligomer length: 7 nucleotides

#### Oligomer Length: 8

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos.</th>
<th>% Diff.</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli tRNA^Gin</td>
<td>UACCCAGp</td>
<td>1</td>
<td>8</td>
<td>58 C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmmm6s2U.</td>
<td>T1 strict</td>
<td>19</td>
<td>137</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AAUACCUGp</td>
<td>1</td>
<td>1</td>
<td>367 C, U, A, G, 1mA, 1mG, 1sU, proQ, D, G+</td>
<td>T1 strict</td>
<td>15</td>
<td>891</td>
<td></td>
</tr>
<tr>
<td>L. lactis tRNA</td>
<td>CUCUUCGp</td>
<td>1</td>
<td>2</td>
<td>852 C, U, A, G, D, m05U, 1sU, 1mG, 1mA, 1mC, 6A, 6A, 1mU, cmmm6s2U, ac1C, k4</td>
<td>T1 strict</td>
<td>120</td>
<td>1613</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AACCCAGp</td>
<td>1</td>
<td>9</td>
<td>1197 C, U, A, G, D, m05U, 1sU, 1mG, 1mA, 1mC, 6A, 6A, 1mU, cmmm6s2U, ac4C, k4</td>
<td>T1 strict</td>
<td>184</td>
<td>2134</td>
<td></td>
</tr>
</tbody>
</table>

**G** Oligomer length: 8 nucleotides

Continued
### Table 6 continued

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos</th>
<th>% Diff</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em> tRNA^a</td>
<td>CAUCCUCUGp</td>
<td>1</td>
<td>11</td>
<td>423C, U, A, G, 1sU, 1mG, 1mA, 1mU, D, cmm6m6U</td>
<td>T1 strict</td>
<td>12</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em> tRNA^b</td>
<td>CCGU preQ0 UC 1mACGp</td>
<td>1</td>
<td>5</td>
<td>720C, U, A, G, 1mA, 1mG, 1sU, preQ, D, G+</td>
<td>T1 strict</td>
<td>34</td>
<td>491</td>
<td></td>
</tr>
<tr>
<td><em>B. subtilis</em> tRNA^c</td>
<td>UCCACUCAAGp</td>
<td>1</td>
<td>5</td>
<td>253C, U, A, G, D, 1mG, 1mU, 16A</td>
<td>T1 strict</td>
<td>16</td>
<td>178</td>
<td></td>
</tr>
<tr>
<td><em>L. lactis</em> tRNA</td>
<td>UUAUCAAUGp</td>
<td>1</td>
<td>5</td>
<td>2494C, U, A, G, D, mo5U, 1sU, 1mG, 1mA, 1mC, 6A, t6A, 1mU, cmm6m6U</td>
<td>T1 strict</td>
<td>81</td>
<td>7459</td>
<td></td>
</tr>
<tr>
<td><em>D. am cases</em> tRNA</td>
<td>UUAUCAAUGp</td>
<td>1</td>
<td>2</td>
<td>1807C, U, A, G, D, mo5U, 1sU, 1mG, 1mA, 1mC, 6A, t6A, 1mU, cmm6m6U</td>
<td>T1 strict</td>
<td>503</td>
<td>4635</td>
<td></td>
</tr>
<tr>
<td><em>C. sepm</em> tRNA</td>
<td>UUAUCAAUGp</td>
<td>1</td>
<td>6</td>
<td>37618C, U, A, G, D, mo5U, 1sU, 1mG, 1mA, 1mC, 6A, t6A, 1mU, cmm6m6U</td>
<td>T1 strict</td>
<td>500</td>
<td>30345</td>
<td></td>
</tr>
</tbody>
</table>

### H Oligomer length: 9 nucleotides

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos</th>
<th>% Diff</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. lactis</em> tRNA</td>
<td>UCCACUCAAGp</td>
<td>1</td>
<td>5</td>
<td>11056C, U, A, G, 1mA, 1mG, 1mA, 1mG</td>
<td>T1 strict</td>
<td>10</td>
<td>9808</td>
<td></td>
</tr>
</tbody>
</table>

### I Oligomer length: 11 nucleotides

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos</th>
<th>% Diff</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. lactis</em> tRNA</td>
<td>UCCACUCAAGp</td>
<td>5</td>
<td>4</td>
<td>3375C, U, A, G, 1mA, 1mG, 1mA, 1mG</td>
<td>T1 strict</td>
<td>13</td>
<td>2631</td>
<td></td>
</tr>
</tbody>
</table>

### J Oligomer length: 12 nucleotides

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos</th>
<th>% Diff</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. lactis</em> tRNA</td>
<td>AACACAACUGp</td>
<td>1</td>
<td>1</td>
<td>66030C, U, A, G, 1mA, 1mG, 1mA, 1mG</td>
<td>T1 strict</td>
<td>81</td>
<td>99222</td>
<td></td>
</tr>
</tbody>
</table>

### K Oligomer length: 14 nucleotides

<table>
<thead>
<tr>
<th>Source</th>
<th>Sequence</th>
<th>Pos</th>
<th>% Diff</th>
<th>Sequences</th>
<th>Nucleotide Pool</th>
<th>Digest</th>
<th>Compositions</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. lactis</em> tRNA</td>
<td>ACUCUUCU6AARUCUAUGp</td>
<td>1</td>
<td>1</td>
<td>11971C, U, A, G, 1mA, 1mG, 1mA, 1mG4, 16A</td>
<td>T1 strict</td>
<td>133</td>
<td>44221</td>
<td></td>
</tr>
<tr>
<td><em>L. lactis</em> tRNA</td>
<td>ACUCUcmm6m5m6U6AARUCUAAANGp</td>
<td>2</td>
<td>1</td>
<td>235C, U, A, G, 16A, cmm6m6U</td>
<td>T1 strict</td>
<td>5</td>
<td>1015</td>
<td></td>
</tr>
</tbody>
</table>
Appendix C: Materials and Methods
C.1 DNA Miniprep

(Bimboim and Doly, 1979)

Solution 1:
25 mM Tris-HCl (1.25 mL of 1M) and 10 mM EDTA (1 mL of 0.5M). Up to 50 mL ddH₂O

Solution 2:
0.2 M NaOH (1 mL of 10 M) and 1% SDS (2.5 mL of 20% SDS). Up to 50 mL ddH₂O

Solution 3:
14.76g potassium acetate, 5.75 mL glacial acetic acid. Up to 50 mL ddH₂O

1. Harvest 2 mL of mid to late log *E. coli* in 2 mL microcentrifuge tubes
2. Centrifuge for 1 minute at maximum speed and dump supernatant
3. Add 100 µl of ‘Solution 1’ to each tube
4. Resuspend each pellet (raking on microcentrifuge rack 6 – 10 times works well)
5. Add 200 µl of ‘Solution 2’ and invert tubes 3 – 5 times
6. Add 150 µl of ‘Solution 3’ and invert 3 – 5 times
7. Centrifuge the tubes at maximum speed for 10 minutes
8. Collect ~400 µl of the supernatant into a new microcentrifuge tube and add 2x – 2.5x volume ethanol
9. Centrifuge at maximum speed for 30 minutes.
10. Dump the supernatant and resuspend the pellet in 50 µl of ddH₂O.
C.2 Column Purification of tRNA

**Equilibration Buffer – 200 mL (make fresh)**
<table>
<thead>
<tr>
<th>Volume (mL)</th>
<th>Component</th>
<th>Concentration (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>10 mL 1M MOPS pH7</td>
<td>0.05 M</td>
</tr>
<tr>
<td></td>
<td>30 mL isopropanol</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>2 mL Triton</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>158 mL ddH2O</td>
<td></td>
</tr>
</tbody>
</table>

**Wash Buffer – 200 mL per run**
<table>
<thead>
<tr>
<th>Volume (mL)</th>
<th>Component</th>
<th>Concentration (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>10 mL 1M MOPS pH7</td>
<td>0.05 M</td>
</tr>
<tr>
<td></td>
<td>8 mL 5M NaCl</td>
<td>0.2 M</td>
</tr>
<tr>
<td></td>
<td>182 mL ddH2O</td>
<td></td>
</tr>
</tbody>
</table>

**Elution Buffer – 50 mL (make fresh)**
<table>
<thead>
<tr>
<th>Volume (mL)</th>
<th>Component</th>
<th>Concentration (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>2.5 mL 1M MOPS pH7</td>
<td>0.05 M</td>
</tr>
<tr>
<td></td>
<td>7.5 mL 5M NaCl</td>
<td>0.75 M</td>
</tr>
<tr>
<td></td>
<td>7.5 mL EtOH</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>32.5 mL ddH2O</td>
<td></td>
</tr>
</tbody>
</table>

**High Salt Wash**
<table>
<thead>
<tr>
<th>Volume (mL)</th>
<th>Component</th>
<th>Concentration (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>2.5 mL 1M MOPS pH7</td>
<td>0.05 M</td>
</tr>
<tr>
<td></td>
<td>20 mL 5M NaCl</td>
<td>2 M</td>
</tr>
<tr>
<td></td>
<td>7.5 mL EtOH</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>20 mL ddH2O</td>
<td></td>
</tr>
</tbody>
</table>

**Procedure:**

1. Equilibrate column with 50 mL fresh equilibration buffer 2 times

2. Bring nucleic acids (resuspended in water) to 0.1 M MOPS pH7 using 1 M MOPS

3. Spin the resuspended sample again and retrieve the supernatant
   a. This is done to remove any particulates that could clog the column

4. Load the sample into the equilibrated column
   a. Save 20 µl for gel analysis

5. Reload flow-through onto the column two more times
a. Save the flow-through for gel analysis

6. Wash the column with 50 mL wash buffer 4 times

7. Elute with 10 mL elution buffer
   a. This acts as a more stringent wash. It may be wise to hold onto this fraction

8. Elute with 15 mL of elution buffer 2 times in separate tubes
   a. These are tRNA-containing fractions

9. Precipitate with two elutions of 1 equal volume of isopropanol

10. Centrifuge at 13,000 rpm for 30 minutes at 10°C

11. Discard the supernatant and wash the pellet with 70% EtOH and let pellet dry

12. Resuspend in ddH2O or TE

Samples can be analyzed on a 6% acrylamide/urea gel

**Quantify:** Measure absorbance at 260nm in a spectrophotometer.

**Column Storage Procedure**

1. Rinse column with ddH2O

2. Run column volume of high salt wash buffer

3. Run column volume of ddH2O

4. Store in 20% EtOH
C.3 RNA Preparation – acid phenol-guanidinium thiocyanate-chlorophorm

(Chomczynski and Sacchi, 1987)

Solution D
- 4 M guanidinium isothiocyanate
- 25 mM sodium citrate pH7
- 0.5% sarcosyl
- 0.1 M beta-mercaptoethanol

Procedure: (For 10 mL culture)
1. Centrifuge cells and wash with 1xPBS two times
2. Resuspend washed pellet in 250 µl of Solution D
3. Add 25 µl of 2 M sodium acetate (NaOAc) pH4 and vortex
4. Add 250 µl of water saturated phenol and vortex
5. Add 75 µl of chloroform / isoamyl alcohol (49:1)
6. Vortex for 1 minute and place on ice for 10 minutes
7. Centrifuge for 15 minutes
8. Retrieve the supernatant and add an equal volume of isopropanol
   a. Cloudy? A good sign for RNA
   b. Store at -20°C for 15 minutes or indefinitely
9. Centrifuge at maximum speed for 30 minutes
10. Discard supernatant
    a. Optional wash with 70% EtOH
    b. Let partially dry (‘too dry’ is difficult to resuspend)
11. Resuspend in 100 µl TE
12. Extract with 50 µl of Tris-phenol
13. Extract with 50 µl chloroform
14. To supernatant add:
    1/10th volume 3M NaOAc
    2 times volume EtOH
1-3 μl glycogen

15. Set at -20°C for 30 minutes to overnight

16. Centrifuge for 30 minutes at maximum speed

17. Discard supernatant and resuspend pellet in TE or ddH₂O

**Quantify:** Measure absorbance at 260nm in a spectrophotometer.
C.4 Protein-C Purification

(Roche Applied Sciences - Anti-Protein C Affinity Matrix)

**PC-150 buffer**

200 mL: Final Conc.

<table>
<thead>
<tr>
<th>Volume</th>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mL</td>
<td>3M KCl</td>
<td>150 mM</td>
</tr>
<tr>
<td>4 mL</td>
<td>Tris-HCl, pH7.7</td>
<td>20 mM</td>
</tr>
<tr>
<td>600 µl</td>
<td>1M MgCl₂</td>
<td>3 mM</td>
</tr>
<tr>
<td>200 µl</td>
<td>1M CaCl₂</td>
<td>1 mM</td>
</tr>
</tbody>
</table>

**Autoclave, then add:**

<table>
<thead>
<tr>
<th>Volume</th>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 µl</td>
<td>10% Tween 20</td>
<td>0.1%</td>
</tr>
<tr>
<td>100 µl</td>
<td>1M DTT</td>
<td>0.5 mM</td>
</tr>
</tbody>
</table>

**Elution buffer (make ~10 mL)**

Final Conc.

- 5 mM Tris-HCl pH7.7
- 5 mM EDTA
- 10 mM EGTA
- 0.5 mM DTT

**Note:** Buffer must contain at least 1mM Ca\(^{2+}\) for binding to occur. The metal chelators EDTA and EGTA cannot be in the buffer for this reason.

**Procedure:**

1. For 4L of cells place 0.4 mL of Roche Protein C Affinity Matrix into a chromatography column.

2. Equilibrate with 50 mL of PC-150
   a. Do not let the matrix dry

3. Lyse cells by sonication in PC-150 (Previously used 15 mL of PC-150 for 4L of *T. brucei* )

4. Centrifuge at maximum speed for ~20 minutes to precipitate insoluble particulates

5. Cap the bottom of the column and add the lysate

6. Cap the top of the column and apply Parafilm to both ends to prevent leaking
7. Rotate at 4°C for 2 hours

The next steps are all done by setting up the column on a ring stand in the refrigerator (4°C)

8. Let the matrix settle, uncap both ends and collet the flow-through

9. First wash: add 10mL of PC-150, cap, and rotate again for 10 minutes at 4°C

10. Let matrix settle and collect the first wash

11. Wash 2 times with 10 mL PC-150 and collect by gravity flow

12. Add 10mL PC-150, rotate for 10 minutes, and collect flow-through

13. Add 0.6 mL of 5mM elution buffer and rotate at room temperature for 30 minutes

14. Collect flow-through and repeat 4 times with 5 minute incubations

**Storage:**

1. Wash 2 times with 10 mL PC-150

2. Wash with 10 mL of PC-150 with 0.9% sodium azide

3. Cap the column and add 5 mL of PC-150 with 0.9% sodium azide – store at 4°C
C.5 Trypanosome Mitotracker Immunofluorescence Protocol

**Procedure:**
1. Collect 1 mL of cells at 5 – 10 x 10^6 cells/mL in 1.5 mL microcentrifuge tube

2. Add 1 µL of 200 µM Mitotracker and incubate at 27°C for 30 minutes
   a. 20 nM final concentration
   b. shaking not required

3. Spin down cells at 3000rpm for 5 minutes

4. Wash 1 time with 1x PBS

5. Discard supernatant and resuspend cells in 4% paraformaldehyde
   a. 5 mL of 4% paraformaldehyde: 0.2g paraformaldehyde, 1.25 µL of 10 M NaOH, add 1x PBS up to 5 mL. Dissolve in 65°C bath

6. Fix cells on slide by placing a single large drop of the resuspended cells and allowing to dry for ~10 minutes

7. Wash slides with 1x PBS 1 time

8. Permeabilize cells by placing in ice cold methanol for 20 minutes.

9. Wash 2 times in 1x PBS

From this point, all incubations are done in a humidity chamber
- Nontransparent box with damp paper towel works

10. Block for 45 to 60 minutes in 1x PBS-Tween-Milk
    a. PBS-Tween-Milk: 2.5 g dry milk in 50 mL of 1x PBS and 0.05% Tween

11. Dump milk and wash 1 time in 1x PBS

12. Add 250 µL of desired concentration of 1° antibody + PBS-Tween-milk on slide and incubate for several hours to overnight at 4°C

13. Wash 3 times with 1x PBS

14. Incubate slide with 250 µL 2° antibody at desired concentration in 5% PBS-Tween-Milk at 4°C for 1hr (keep dark!)
   a. Cover with Parafilm
15. Wash three times with 1x PBS

Optional: DAPI stain
1. Add 1 µL DAPI to 1 mL 1x PBS
2. Incubate slide with 250 µL of DAPI-PBS for 1 minute
3. Wash 1 time with 1x PBS

16. Let dry (keep dark) and then add ProLong Gold antifade reagent
   a. Gently place coverslip on top and absorb excess at edges of the cover with paper towel

17. Outline the coverslip with nail polish, let dry in dark environment, and visualize
C.6 Reactive Oxygen Species Assay (T. brucei)

The day before:
1. Count cells and dilute to $2 \times 10^6$ in 10 mL media
   - Need cells on the following day at mid-log
2. ROS positive control: a separate 10mL of culture with 10 µL of stock paraquat (1/1000)

Procedure:
1. For each cell line, collect 4 mL of cells in separate 2 mL microcentrifuge tubes
   a. 1 blank (no dye) and 3 experimental replicates
2. Centrifuge at 3000rpm for 5 minutes and discard supernatant
3. Wash pellet by resuspending in 1 mL 1x PBS, spin again, and dump supernatant
4. Resuspend pellet in 1 mL of 1x PBS. Add 1 µL of dihydrorhodamine and invert tubes 3 – 4 times
   a. Do not add dihydrorhodamine to the blank culture!
5. Incubate in a nontransparent box at 27°C for 30 minutes
6. Centrifuge at 3000 rpm for 5 minutes, dump supernatant, and resuspend in 1 mL of 1x PBS
7. Keep dark as much as possible and quickly begin flow cytometry work
   a. Measure via FITC

Note: The procedure (not including the flow cytometry work) takes approximately 1.5 hours with 4 cell lines (4 cell lines = 16 samples for analysis)
C.7 Reverse Transcription – Polymerase Chain Reaction (RT-PCR)

(Invitrogen SuperScript® III First-Strand Synthesis – modified)

RNA for analysis can be isolated using the acid phenol-guanidinium thiocyanate-chlorophorm protocol.

RQ1 DNase treatment:

25 µL reaction
2.5 µL 10x DNase RQ1 buffer
1 µL / 1 µg RNA DNase RQ1
↑ 25 µL ddH2O

1. Incubate at 37°C for 1 hour to overnight

2. Phenol / chloroform extract
   a. Bring RQ1 reaction up to 100 µL with ddH2O then phenol extract with 100 µL tris-phenol followed by 100 µL extraction with chloroform

3. EtOH precipitate (1/10th volume 3M NaOAc, 2 x volume EtOH, 1 µL glycogen

RT reaction

1. Anneal primer to RNA – 2x volume – one half for RT+ and one half for RT - (DNA contamination negative control)

   2x volume
   RNA  1—5 µg (2µg worked for Tyw1L RT)
   Primer 2 µL 2 µM stock (final concentration 2 pM)
   dNTPs 2 µL 10 mM dNTPs
   ddH2O↑ 24 µL

2. Heat for 5 minutes @ 70°C

3. Cool on ice for

4. Make RT- and RT+ master mix

<table>
<thead>
<tr>
<th></th>
<th>RT-</th>
<th></th>
<th>RT+</th>
</tr>
</thead>
<tbody>
<tr>
<td>5x 1st Strand</td>
<td>4 µL</td>
<td>4 µL</td>
<td></td>
</tr>
<tr>
<td>0.1 M DTT</td>
<td>2 µL</td>
<td>2 µL</td>
<td></td>
</tr>
<tr>
<td>ddH2O</td>
<td>2.5 µL</td>
<td>2 µL</td>
<td></td>
</tr>
<tr>
<td>SS III RT enzyme</td>
<td>-</td>
<td>0.5 µL</td>
<td></td>
</tr>
</tbody>
</table>
5. Divide the 24 µL of the annealed reaction from step 3 into two PCR tubes. 12 µL each.

6. Pipet 8 µL of the RT- mix into one tube (Label well – this is the negative control that checks for DNA contamination)

7. Pipet 8 µL of the RT+ mix into the other tube (Label)

8. Incubate both tubes at 50°C for 1 – 3 hours

9. Use 2 µL for the PCR reaction

**PCR (Using the Taq polymerase PCR protocol)**

1. Thermocycler conditions for an expected product size 200 – 600bp:

   **1st cycle**
   - Denaturation at 94°C for 1 minute
   - Annealing at 50°C for 30 seconds (may need to be adjusted)
   - Elongation at 72°C for 30 seconds

   **Cycles 2 – 29**
   - Denaturation at 94°C for 30 minute
   - Annealing at 50°C for 30 seconds (may need to be adjusted)
   - Elongation at 72°C for 30 seconds
C.8 Preparing tRNA for Nucleoside Analysis via High Performance Liquid Chromatography (HPLC)

Starting with 20 µg of tRNA

Digest to nucleosides using nuclease P1

Procedure:

For 20 µg of tRNA:

10 µL 10x P1 buffer
X µL Nuclease P1 (Sigma-Aldrich)
85 µL RNA in ddH2O
100 µL

1. Incubate at 37°C overnight. Incubate for a few hours, if trying to increase the amount of dinucleosides (if looking for wyosine-phosphate-adenosine (imGpA), for example)

Remove phosphate with Calf Intestine Alkaline Phosphotase (CIAP)

Procedure:

Directly to the P1 digest:

100 µL P1 digest
15 µL 10x CIAP buffer
2 µL CIAP enzyme (New England Biolabs)
33 µL ddH2O
150 µL

1. Incubate at 37°C for 2 hours.

2. Increase volume to 200 µL and filter the sample using the smallest 0.2 micron filter and a 1 mL syringe. The sample is now ready for HPLC analysis

Note: If this sample will be used for HPLC-MS/MS then use bacterial alkaline phosphatase (BAP) for better salt compatibility.
C.9 Nucleoside Analysis via High Performance Liquid Chromatography (HPLC)

(McCloskey & Pomerantz, 1990; modified and shared by Dr. Pat Limbach)

All buffers must be filtered and degassed before HPLC use! Except acetonitrile, as it destroys filters, so use HPLC grade acetonitrile.

**Buffer A**

For 500 mL

<table>
<thead>
<tr>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaOAc pH5.3 5mM</td>
</tr>
</tbody>
</table>

**Buffer B**

For 500 mL

<table>
<thead>
<tr>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 mL acetonitrile 40%</td>
</tr>
<tr>
<td>300 mL ddH₂O (filtered)</td>
</tr>
<tr>
<td>500 mL</td>
</tr>
</tbody>
</table>

**HPLC Conditions**

Column: Supelcosil LC-18-S 250 × 2.1 mm column (Sigma-Aldrich)

Flow-rate: 0.3 mL / minute

Temperature: - If able to control: 30°C otherwise, room temperature

**The gradient:**

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>%B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5.8</td>
<td>1</td>
</tr>
<tr>
<td>9.2</td>
<td>2</td>
</tr>
<tr>
<td>10.9</td>
<td>3</td>
</tr>
<tr>
<td>12.7</td>
<td>5</td>
</tr>
<tr>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>38</td>
<td>50</td>
</tr>
<tr>
<td>43.5</td>
<td>75</td>
</tr>
<tr>
<td>45</td>
<td>75</td>
</tr>
<tr>
<td>50</td>
<td>99</td>
</tr>
<tr>
<td>55</td>
<td>99</td>
</tr>
<tr>
<td>60</td>
<td>1</td>
</tr>
</tbody>
</table>
C.10 Silver Stain of Protein Gels

(Swain and Ross, 1995) Courtesy of Dr. Chad Rappleye

Note: Make A, B, and C fresh

**Fixative (ie. Coomassie Destaining Solution)**
40% EtOH
10% Acetic Acid

<table>
<thead>
<tr>
<th>Solution A</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 mg</td>
<td>Na$_2$S$_2$O$_3$·5H$_2$O</td>
</tr>
<tr>
<td>250 mL</td>
<td>ddH$_2$O</td>
</tr>
<tr>
<td></td>
<td>0.02% Na$_2$S$_2$O$_3$·5H$_2$O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solution B</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 mg</td>
<td>AgNO$_3$</td>
</tr>
<tr>
<td>50 mL</td>
<td>ddH$_2$O</td>
</tr>
<tr>
<td></td>
<td>0.1% AgNO$_3$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solution C</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 g</td>
<td>Na$_2$CO$_3$</td>
</tr>
<tr>
<td>50 µL</td>
<td>37% formaldehyde</td>
</tr>
<tr>
<td>2 mL</td>
<td>Solution A</td>
</tr>
<tr>
<td>↑ 100 mL</td>
<td>ddH$_2$O</td>
</tr>
<tr>
<td></td>
<td>6% Na$_2$CO$_3$</td>
</tr>
<tr>
<td></td>
<td>0.02% formaldehyde</td>
</tr>
<tr>
<td></td>
<td>0.0004% Na$_2$S$_2$O$_3$·5H$_2$O</td>
</tr>
</tbody>
</table>

**Procedure:**

Note: Never touch the gel, always perform in very clean containers, and always use milli-Q ddH$_2$O. Any and all contaminants will also be stained.

1. **Fix** proteins in gel by incubating gel in Fixative solution for 30 min – O/N
2. Rinse gel for 5 minutes x 4 in milli-Q H$_2$O
3. **Sensitize** gel by incubating in Solution A for 5 minutes
4. Rinse gel for 5 minutes x 3 in milli-Q ddH$_2$O
5. **Stain** gel in Solution B for 20 – 40 minutes
   - Dump silver stain in silver-waste container
6. Rinse the gels briefly 3 times in milli-Q ddH$_2$O
   - Washing too long will remove the silver stain
7. **Develop** stain by incubating in Solution C until staining develops
   - Watch closely, until desired darkness is achieved, then quench

8. **Quench** by dumping staining solution and add Fixative solution
   - Rinse 2 times in Fixative
C.11 *Trypanosoma brucei* Cyto and Mito Fraction Isolation Protocol

This protocol details the procedure of obtaining cytoplasmic and mitochondrial fractions from 2L of cells at high density.

**Expected tRNA yields**

- Cyto: 750ug/L
- Mito: 10ug/L

### DTE (Dilute Tris + EDTA) (2L)

<table>
<thead>
<tr>
<th>Final Conc</th>
<th>2 mL 1 M Tris pH 8</th>
<th>4 mL 0.5 M EDTA pH 8</th>
<th>1.996 L ddH2O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 mM</td>
<td>1 mM</td>
<td>&gt;autoclave</td>
</tr>
</tbody>
</table>

### SucTM (Sucrose-Tris-MgCl2) (1L)

<table>
<thead>
<tr>
<th>Final Conc</th>
<th>85.6 g Sucrose</th>
<th>20 mL 1 M Tris-HCl pH 8</th>
<th>↑ 1 L ddH2O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25 M</td>
<td>20 mM</td>
<td>&gt;autoclave</td>
</tr>
</tbody>
</table>

### SucTE (Sucrose-Tris-EDTA) (1L)

<table>
<thead>
<tr>
<th>Final Conc</th>
<th>85.6 g Sucrose</th>
<th>20 mL 1 M Tris-HCl pH 8</th>
<th>4 mL 0.5 M EDTA</th>
<th>↑ 1 L ddH2O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.25 M</td>
<td>20 mM</td>
<td>2 mM</td>
<td>&gt;autoclave</td>
</tr>
</tbody>
</table>

### Procedure:

1. Spin @ 6krpm for 10min.

2. Wash 2x in 1x PBS, spinning between each at 6krpm for 10min.

3. Resuspend pellet in 200 mL to 300 mL DTE. Check for proper cell swelling.

4. Pressurized needle: ~100psi, 30g needle

5. Run small portion of the cells through the pressurized needle. Check for complete lysis of whole cells. You should see black dots (mitos) and cell debris/flagella.
• If you don’t, add more DTE, run through the needle, and observe again until you do see complete lysis.

6. Pass everything through the needle. Add 60% sucrose throughout this period.
   • 60% sucrose amount: 1/5\textsuperscript{th} volume of cell lysate

7. Spin the cell lysate at 9krpm for 15min.
   • Use 225 mL centrifuge bottles for better ppt.

Keep the supernatant! (this is the cyto fraction)
Keep the pellet! (this is the mito fraction)

Jump to the ‘Crude Mitochondrial Preparation’ or the ‘Cytosolic Fraction Purification’

Mitochondrial Fraction Preparation
1. Resuspend the mitochondria pellet in 100 mL SucTM

2. DNase treatment:
   • Add 300 µL of 1 M MgCl\textsubscript{2}
   • Add 300 µL of DNase 1
   • Incubate on ice for 1hr then add 300 µL 0.5M EDTA

3. Spin @ 9krpm for 15min in 225 mL centrifuge tube. (you probably DNased’d in this)

4. Dump supernatant and resuspend in SucTE. (~20 – 30 mL)

5. Spin @ 9krpm for 15min.

6. Dump supernatant and resuspend in SucTCa. (5mM CaCl\textsubscript{2})

7. Spin @ 9krpm for 15min.

8. Resuspend in 20 mL SucTCa.


10. Incubate @ 37C for 30min.

11. Spin @ 9krpm for 15min and dump supernatant.
Note: You can stop at this point for a crude mitochondrial preparation and proceed to step 12, otherwise see ‘Mitochondria Gradient Purification Protocol’.

12. Proceed to isolate tRNA by following the acid phenol-guanidinium thiocyanate-chlorophorm RNA preparation protocol

Cytosolic Fraction Preparation

Note: I would always work with half of the cytosolic fraction as it produces a high amount of tRNA and is much easier to work with. Leave the other half in DTE and keep frozen.

1. Add ½ volume of H2O-Phenol, 1/10th volume 2 M NaOAc, and 1/3rd volume 49:1 chloroform/isoamyl alcohol

2. Vortex for 1 minutes, put on ice for 10 minutes

3. Spin for 10 minutes at 12krpm
   - Use the smallest centrifuge bottles that you can for better precipitation

4. Resuspend the pellet in 1 to 2 mL of TE

5. Perform typical Tris-phenol / chloroform / EtoH precipitation
C.12 Mitochondria Gradient Purification Protocol

(Alfonzo Lab protocol)

It is assumed at this point that you have a pellet of DNase and Micrococcal nuclease treated mitochondria (see C.11 Trypanosoma brucei Cyto and Mito Fraction Isolation Protocol)

<table>
<thead>
<tr>
<th>76% RSTE (100 mL)</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.56 g surose</td>
<td></td>
</tr>
<tr>
<td>2 mL 1 M Tris-HCl pH 8</td>
<td>20 mM Tris-HCl 8</td>
</tr>
<tr>
<td>20 mL 0.5M EDTA pH 8</td>
<td>100 mM EDTA 8</td>
</tr>
<tr>
<td>69.4 mL Reno</td>
<td></td>
</tr>
<tr>
<td>↑ 100 mL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20% RSTE (100 mL)</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.56 g surose</td>
<td></td>
</tr>
<tr>
<td>2 mL 1 M Tris-HCl pH 8</td>
<td>20 mM Tris-HCl 8</td>
</tr>
<tr>
<td>20 mL 0.5M EDTA pH 8</td>
<td>100 mM EDTA 8</td>
</tr>
<tr>
<td>26.3 mL Reno</td>
<td></td>
</tr>
<tr>
<td>↑ 100 mL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>35% RSTE (100 mL)</th>
<th>Final Conc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.56 g surose</td>
<td></td>
</tr>
<tr>
<td>2 mL 1 M Tris-HCl pH 8</td>
<td>20 mM Tris-HCl 8</td>
</tr>
<tr>
<td>20 mL 0.5M EDTA pH 8</td>
<td>100 mM EDTA 8</td>
</tr>
<tr>
<td>46.2 mL Reno</td>
<td></td>
</tr>
<tr>
<td>↑ 100 mL</td>
<td></td>
</tr>
</tbody>
</table>

To prepare the gradient:
1. Pipet 17 mL of 35% RSTE into ultracentrifuge tube
2. Freeze @ -20°C
3. Pipet 17 mL of 20% RSTE into ultracentrifuge tube
4. Wrap the top with parafilm and freeze @ -20°C

Procedure:
1. Resuspend the mitochondria pellet in 4 mL 76% RSTE
2. Using a long metal needle and syringe, inject the 4 mL to the bottom of the gradient. Add 3 mL of 76% RSTE to rinse the tube/syringe and inject to the bottom.

3. Use analytical scale to balance the tube for high speed centrifugation.

4. Ultra centrifuge @ 24krpm (double check this) for 1.5 hours

5. Should see to dominant bands. The top is cell junk/vesicles (often contains mitochondrial contamination. The bottom is the mitochondrial fraction.

6. Poke needle at a diagonal on the bottom and collect fractions

7. Add STE – 3x the volume of the fraction

8. Centrifuge for 15min at 9krpm

9. Resuspend in cold water and sonic

10. Spin and collect the supernatant and the pellet
C.13 Mitochondrial Membrane Potential Assay (*T. brucei*)

**The day before:**
- Count cells and dilute to $2 \times 10^6$ in 10 mL media
  - Need cells on the following day at mid-log

**Procedure:**
1. For each cell line, collect 4 mL of cells in separate 2 mL microcentrifuge tubes
   a. 1 blank (no dye) and 3 experimental replicates
2. Add 2 µL of MitoTracker Red and invert tubes 3 – 4 times
3. Incubate in a nontransparent box at 27°C for 30 minutes
4. Centrifuge at 3000 rpm for 5 minutes, dump supernatant, and resuspend in 1 mL of 1x PBS
5. Repeat this washing process 1 more time and resuspend in 1 mL 1x PBS
6. Keep dark as much as possible and quickly begin flow cytometry work
   a. Measure via PE-A

**Note:** The procedure (not including the flow cytometry work) takes approximately 1.5 hours with 4 cell lines (4 cell lines = 16 samples for analysis)
References


trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. Cell 46, 819–826.


tRNA\textsubscript{Lys} anticodon loop: the effect of nucleosides s\textsuperscript{2}U, mcm\textsuperscript{5}U, mcm\textsuperscript{5}s\textsuperscript{2}U, mnm\textsuperscript{5}s\textsuperscript{2}U, t\textsuperscript{6}A, and ms\textsuperscript{2}t\textsuperscript{6}A. Biochemistry 44, 8078–8089.


Helm, M., Giegé, R., and Florentz, C. (1999). A Watson-Crick base-pair-disrupting methyl group (m\textsuperscript{1}A\textsubscript{9}) is sufficient for cloverleaf folding of human mitochondrial tRNA\textsuperscript{Lys}. Biochemistry 38, 13338–13346.


measurements from electrospray ionization mass spectrometry. Analytical Chemistry 69, 1543–1549.


Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T., and Yokoyama, S. (1988). Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification.


Orellana, O., Cooley, L., and Söll, D. (1986). The additional guanylate at the 5'terminus of Escherichia coli tRNA^His is the result of unusual processing by RNase P. Molecular and Cellular Biology 6, 525–529.


Young, A.P., and Bandarian, V. (2011). Pyruvate is the source of the two carbons that are required for formation of the imidazoline ring of 4-demethylwyosine. Biochemistry 50, 10573–10575.