CHAPTER ELEVEN

Pseudo-Seq: Genome-Wide Detection of Pseudouridine Modifications in RNA

Thomas M. Carlile, Maria F. Rojas-Duran, Wendy V. Gilbert

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

1Corresponding author: e-mail address: wgilbert@mit.edu

Contents

1. Introduction 220
   1.1 RNA Modifications 220
   1.2 Detection of Pseudouridine 222
2. Sample Preparation and RNA Isolation 223
   2.1 RNA Isolation from S. cerevisiae 224
   2.2 Poly(A) Selection 225
3. Pseudo-Seq Library Preparation 227
   3.1 RNA Fragmentation 228
   3.2 CMC Modification and Reversal 228
   3.3 3’ End Healing 229
   3.4 RNA Size Selection 230
   3.5 3’ Adapter Ligation 230
   3.6 Reverse Transcription and Size Selection 231
   3.7 Circularization 232
   3.8 PCR Amplification 233
4. Pseudo-Seq Data Analysis 235
   4.1 Library Demultiplexing 235
   4.2 Trimming of 3’ Adapter Sequences 236
   4.3 Mapping Reads 237
   4.4 Computational Identification of \( \Psi \) Residues 238
   4.5 Genetic Assignment of \( \Psi \) s to Pseudouridylation Factors 239
5. Experimental Considerations 239
   5.1 Read Coverage 239
   5.2 Biological Replication 240
   5.3 Computational Parameters 240
6. Solutions, Reagents, and Common Protocols 241
   6.1 Solutions 241
   6.2 Library Oligonucleotide Sequences 242
   6.3 Common Protocols 242
Acknowledgments 243
References 243

Methods in Enzymology, Volume 560
ISSN 0076-6879
http://dx.doi.org/10.1016/bs.mie.2015.03.011
© 2015 Elsevier Inc. All rights reserved.
Abstract

RNA molecules contain a variety of chemically diverse, posttranscriptionally modified bases. The most abundant modified base found in cellular RNAs, pseudouridine (Ψ), has recently been mapped to hundreds of sites in mRNAs, many of which are dynamically regulated. Though the pseudouridine landscape has been determined in only a few cell types and growth conditions, the enzymes responsible for mRNA pseudouridylation are universally conserved, suggesting many novel pseudouridylated sites remain to be discovered. Here, we present Pseudo-seq, a technique that allows the identification of sites of pseudouridylation genome-wide with single-nucleotide resolution. In this chapter, we provide a detailed description of Pseudo-seq. We include protocols for RNA isolation from *Saccharomyces cerevisiae*, Pseudo-seq library preparation, and data analysis, including descriptions of processing and mapping of sequencing reads, computational identification of sites of pseudouridylation, and assignment of sites to specific pseudouridine synthases. The approach presented here is readily adaptable to any cell or tissue type from which high-quality mRNA can be isolated. Identification of novel pseudouridylation sites is an important first step in elucidating the regulation and functions of these modifications.

1. INTRODUCTION

1.1 RNA Modifications

In addition to the four canonical bases, RNA molecules contain a wide array of posttranscriptional modifications. Across species, more than 100 chemically diverse modified bases have been identified within RNAs, primarily in stable noncoding RNAs (ncRNAs, e.g., tRNA and rRNA) (Cantara et al., 2011). Until recently, only three modified bases were known to occur within mRNAs—inosine, \(N^6\)-methyladenosine (m\(^6\)A), and 5-methylcytosine (m\(^5\)C). The most abundant modified base, pseudouridine, is found at multiple positions in both tRNAs and rRNAs (Ge & Yu, 2013), but was not discovered in mRNAs until new genome-wide methods for pseudouridine profiling enabled detection of this modification in low-abundance transcripts (Carlile et al., 2014; Lovejoy, Riordan, & Brown, 2014; Schwartz et al., 2014).

Ψ formation is catalyzed by two classes of enzymes, each with distinct modes of substrate recognition. The first class, Box H/ACA snoRNAs, found in archaea and eukaryotes, uses small noncoding RNAs as guides to target the catalytic protein Cbf5/dyskerin to its substrates by base pairing between the guide RNA and the target RNA (Ganot, Bortolin, & Kiss, 1997). This class primarily targets modifications within rRNA. The second
class of pseudouridine synthase (Pus) proteins modifies tRNAs as well as additional ncRNAs. The Pus proteins, which are conserved in all domains of life, do not require guide RNAs to bind their targets. Instead, Pus proteins directly recognize various structural and/or sequence elements in their RNA substrates (Arluison, Buckle, & Grosjean, 1999; Urban, Behm-Ansmant, Branlant, & Motorin, 2009). Both types of pseudouridine synthases form $\Psi$, which is the C5-glycoside isomer of uridine, by breakage of the N1-glycosyl bond followed by 180° base rotation and formation of a C5-glycosyl bond (Fig. 1A). This modification leaves the Watson–Crick edge of uridine unchanged, but frees the hydrogen at position N1 to act as a hydrogen bond donor, which is implicated in the stabilizing effects of $\Psi$ on RNA structure (Charette & Gray, 2000; Hudson, Bloomingdale, & Znosko, 2013).

Dynamic pseudouridylation may play a regulatory role in response to cellular stressors given that the pseudouridine landscape changes substantially in different growth states (Carlile et al., 2014; Courtes et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014; Wu, Xiao, Yang, & Yu, 2011). Although the functions of endogenous mRNA pseudouridylation events are not yet known, the established effects of $\Psi$ on RNA structure suggest many possibilities for posttranscriptional regulation through the impact of mRNA structure on translation initiation efficiency, ribosome pausing, RNA

![Figure 1 Structures of (A) uridine and pseudouridine, (B) CMC, and (C) $\Psi$-CMC adduct after CMC treatment and reversal.](image-url)
localization, and regulation by RNA interference (Jambhekar & DeRisi, 2007; Kudla, Murray, Tollervey, & Plotkin, 2009; Shah, Ding, Niemczyk, Kudla, & Plotkin, 2013; Somogyi, Jenner, Brierley, & Inglis, 1993; Tan et al., 2012). Furthermore, mRNA pseudouridylation may provide a mechanism for dynamically altering the genetic code. Artificially targeted pseudouridylation of stop codons in vivo leads to noncanonical base pairing of Ψ in the decoding center of the ribosome and efficient nonsense suppression (Karijolich & Yu, 2011). Structural studies of this surprising phenomenon suggested the possibility of more widespread effects of pseudouridine on decoding (Fernández et al., 2013).

Though regulated pseudouridylation of mRNA has the potential to profoundly affect both protein production and protein function, most of the pseudouridylated sites in mRNAs remain to be discovered. The Pseudo-seq method described here will enable investigations of regulated pseudouridylation events in a variety of contexts, including developmental changes and disease states linked to defects in pseudouridine synthase activity (Anderson, Brewer, Singh, & Boothroyd, 2009; Bykhovskaya, Casas, Mengesha, Inbal, & Fischel-Ghodsian, 2004; Heiss et al., 1998; Mei et al., 2011).

1.2 Detection of Pseudouridine

A variety of techniques have been used to identify and map the locations of Ψ residues within RNAs. Pseudouridine was initially found as an unknown “fifth nucleotide” in preparations of yeast RNA using paper chromatography (Cohn, 1960; Davis & Allen, 1957). Sequencing of tRNAs using RNase digestion, column chromatography, and paper electrophoresis and chromatography led to the first identification of specific sites of pseudouridylation (Holley, Everett, Madison, & Zamir, 1965). Similar techniques were subsequently used to map Ψ positions in rRNAs (Choi & Busch, 1978; Gupta & Randerath, 1979; Tanaka, Dyer, & Brownlee, 1980), but these approaches were limited to very abundant RNAs that could be purified to homogeneity.

The ability to map the positions of Ψ residues with single-nucleotide resolution was greatly improved with the method introduced by Bakin and Ofengand (1993). In this assay, isolated RNA is treated with the carbodiimide N-cyclohexyl-N’-(2-morpholinoethyl)carbodiimide metho-p-toluenesulfonate (CMC) in vitro (Fig. 1B) to form covalent adducts at the N1 position of guanidine, N3 position of uridine, and N1 and N3 positions of pseudouridine. Because only the CMC-N3 adduct of pseudouridine is resistant to alkaline cleavage, Ψ residues can be specifically labeled by
incubating CMC-derivatized RNA in sodium carbonate buffer, pH 10.4 (Fig. 1C). The bulky Ψ-CMC adduct acts as a strong stop to reverse transcriptase (RT) one nucleotide 3′ of the Ψ, allowing Ψ locations to be mapped using gene-specific primer extension assays (Bakin & Ofengand, 1993). More recently, techniques have been developed to allow site-specific identification and quantification of Ψ using thin-layer chromatography (Zhao & Yu, 2004), as well as mass spectrometry-based techniques that allow detection of normally mass silent Ψ (Durairaj & Limbach, 2008; Popova & Williamson, 2014). While powerful in some contexts, these assays are low throughput and are best suited to identification of Ψs in highly abundant RNAs.

A growing interest in the biological functions of RNA modifications has created a need for new methods for their transcriptome-wide identification. Recent studies have successfully used next-generation sequencing techniques to map various RNA modifications genome-wide, including m6A (Dominissini et al., 2013; Meyer et al., 2012), m5C (Squires et al., 2012), inosine (Li et al., 2009), and pseudouridine (Carlile et al., 2014; Lovejoy et al., 2014; Schwartz et al., 2014). In this chapter, we describe Pseudo-seq, a technique that allows efficient, single-nucleotide mapping of Ψs genome-wide (Carlile et al., 2014). Pseudo-seq combines high-throughput sequencing with the ability to specifically derivatize Ψ residues with CMC. Deep sequencing of abortive RT products identifies sites where RT stops, and CMC-dependent RT stops represent sites of pseudouridylation. Here, we provide protocols for isolation of total or poly(A)+ RNA, Pseudo-seq library preparation, and sequencing data analysis. While these procedures have been optimized for performing Pseudo-seq on Saccharomyces cerevisiae, they can easily be adapted to other organisms, provided a sufficient quantity of mRNA can be obtained.

### 2. SAMPLE PREPARATION AND RNA ISOLATION

The first step in Pseudo-seq library preparation is the isolation of high-quality total RNA from a cell type of interest. Total RNA is sufficient for the detection of Ψs in highly abundant RNA species, such as tRNAs, rRNAs, or other very highly expressed ncRNAs. However, Ψ detection in less abundant RNAs, including mRNAs, requires their enrichment from the pool of total RNA. The specific protocols used for isolation of RNA will vary widely depending on the biological system being used and can be found elsewhere. Here, we provide a protocol for the isolation of total RNA from...
exponentially growing cultures of *S. cerevisiae* using hot acid phenol (Collart & Oliviero, 2001). However, any method of RNA isolation that yields high-quality total RNA will be compatible with Pseudo-seq library preparation. RNA should be prepared from multiple independent biological replicates, as these replicates are essential for high-confidence pseudouridine identification, especially in transcripts of lower abundance. A discussion of the number of biological replicates required can be found in Section 5.2. Additionally, we provide a cost-effective protocol for the enrichment of polyadenylated RNAs using oligo (dT) cellulose beads (Sambrook & Russell, 2001). Other strategies for mRNA enrichment may be substituted as desired (e.g., rRNA depletion may be used for the purification of mRNA from prokaryotes or from tissue samples with significant RNA degradation).

### 2.1 RNA Isolation from *S. cerevisiae*

Inoculate a 10-ml starter culture in YPAD and grow overnight at 30 °C. The OD₆₀₀ should reach approximately 7.0. Use this starter to inoculate 750 ml of prewarmed YPAD to an OD₆₀₀ of 0.05. Grow this culture at 30 °C in a baffled flask with shaking at 200 rpm to a final OD₆₀₀ of 1.0. Harvest the cells by centrifugation at 16,000×g for 5 min at 4 °C. Resuspend the cell pellet in 25 ml ice-cold water and transfer to a 50-ml conical tube. Pellet cells by centrifugation at 3400×g for 5 min at 4 °C and pour off supernatant. The pellets should be approximately 5 ml and can be either used directly for RNA isolation or snap-frozen in liquid N₂ and stored at −80 °C. The growth conditions described here can be varied to examine pseudouridylation under different conditions, provided a sufficient quantity of cells is harvested.

Isolation of total RNA from yeast cells using hot acid phenol extraction is efficient and yields RNA that is relatively free of DNA contamination. Add 5 ml acid phenol (Sigma P4557, without alkaline buffer) and 5 ml AES Buffer to the cell pellet. Incubate for 30 min in a 65 °C water bath, vortexing for 10 s every minute. Place the tube on ice for 10 min. Add 5 ml chloroform (Sigma C2432) and centrifuge at 3400×g for 5 min at room temperature. Transfer the upper, aqueous phase to a new 15-ml conical tube, add 5 ml of acid phenol:chloroform:isoamyl alcohol (Ambion AM9732; without alkaline buffer), and vortex briefly. Centrifuge and transfer aqueous phase to a new 15-ml conical tube as above. Repeat this extraction step until the interface between the organic and aqueous phases is free of protein (two to three times total). Perform a final extraction by
transferring the aqueous phase to a new 15-ml conical tube, adding 5 ml chloroform, and spinning as above. Transfer the aqueous phase to an Oakridge Tube (Thermo 3114-0030) and add 1/9th volume 3 M NaOAc, pH 5.3, and 1 volume isopropanol. Spin at 14,000 \times g for 30 min at 4 °C. Wash the RNA pellet twice with 10 ml ice-cold 70% ethanol, spinning at 14,000 \times g for 10 min at 4 °C. Air dry the pellet for approximately 3 min. If proceeding directly to RNA fragmentation (Section 3.1), resuspend RNA in 1–2 ml H2O. If performing poly(A) selection (Section 2.2), resuspend in 1–2 ml TES Buffer. A yield of at least 10 μg of RNA per milliliter of cells at an OD600 of 1.0 is expected, and a final RNA concentration greater than 1 μg/μl is desired.

### 2.2 Poly(A) Selection

This protocol uses oligo (dT) cellulose beads (NEB S1408). However, other commercially available beads for poly(A) selection are suitable for use with Pseudo-seq. It should also be possible to use other available methods for mRNA enrichment, such as those based on subtractive hybridization of rRNA. Such techniques should be used if Ψ detection in lower abundance, nonpolyadenylated transcripts is desired. Yeast mRNAs contain relatively short poly(A) tails compared to other eukaryotic organisms (Subtelny, Eichhorn, Chen, Sive, & Bartel, 2014). We have found that poly(A) selection using oligo (dT) cellulose beads coupled with relatively long incubations yields greater recovery of yeast mRNA than rapid, magnetic bead-based capture methods (Dynabeads), which may be used for organisms with longer poly(A) tails.

A sufficient quantity (1.5 ml of 50% slurry per sample) of oligo (dT) cellulose beads for all samples should be washed in batch in 15 ml conical tubes. Pellet the beads at 3000 \times g for approximately 30 s at room temperature and remove the supernatant, taking care to minimize disturbance of the beads. Wash three times in two slurry volumes of water, followed by two washes with two slurry volumes of TES + NaCl, pelleting as above between washes. Prior to pelleting the final wash, distribute equal volumes (~750 μl bead bed volume) to individual 15 ml conical tubes for each sample and pellet as above.

Bring 7.5–10 mg of total RNA up to a volume of 4.5 ml in TES (without NaCl) in a 15-ml conical tube. Denature at 65 °C for 15 min and then place on ice for 2 min. To bind poly(A) RNA to the beads, add 563 μl 5 M NaCl to the denatured RNA and transfer to the oligo (dT) cellulose pellet.
Vortex to mix and incubate at room temperature for 15 min with rotation. To perform a second round of binding, which increases mRNA yield, pellet the beads as above, transfer the supernatant to a 15-ml conical tube, denature at 65 °C for 10 min, and place on ice for 2 min. Then, add the supernatant back to the beads, vortex to mix, and incubate at room temperature for 15 min with rotation. Pellet the beads as above and discard the supernatant. Wash the beads three times with 5 ml TES + NaCl, incubating each wash at room temperature for 2 min with rotation. Wash once with 2 ml ice-cold water, vortex briefly, pellet as above, and discard the supernatant. This wash should be performed quickly to avoid losing RNA bound to the beads. To elute poly(A) RNA, add 2 ml 55 °C water to the beads and incubate at 55 °C for 5 min. Pellet as above and transfer supernatant to a 15-ml conical tube. Repeat elution as before and pool eluates. Wash the beads once in 5 ml water and once in 5 ml TES + NaCl for reuse in the second round of poly(A) selection.

A second round of poly(A) selection can be performed by rebinding the eluates to oligo (dT) beads; this will decrease rRNA contamination of the purified mRNA. Bring the pooled eluates up to 5 ml total volume in TES (without NaCl) by adding 50 μl 1 M Tris, pH 7.6; 10 μl 0.5 M EDTA, pH 8.0; 25 μl 20% SDS; and water to 5 ml total. Denature RNA at 65 °C for 10 min and place on ice for 2 min. To bind poly(A) RNA, add 626 μl 5 M NaCl to the denatured RNA, transfer to the washed beads, and incubate at room temperature for 15 min with rotation. To perform a second round of denaturation, pellet the beads as above, transfer the supernatant to a 15-ml conical tube, denature at 65 °C for 5 min, and place on ice for 2 min. To rebind, add the supernatant back to the beads, vortex to mix, and incubate at room temperature for 15 min with rotation. Perform the washes and elutions as above, except elute twice in 1.8 ml 55 °C water, and pool. To remove residual oligo (dT) cellulose beads, pass the eluates through a 0.45-μm cellulose acetate filter (VWR 28145-481) with a syringe and transfer in 900 μl aliquots to 2 ml microcentrifuge tubes. To each aliquot, add 1/9th volume 3 M NaOAc, pH 5.3; 2 μl GlycoBlue (Invitrogen AM9516); and one volume isopropanol and precipitate at −20 °C for at least 30 min. Spin in a microcentrifuge at max speed at 4 °C for 30 min. Wash the pellet in 750 μl ice-cold 70% ethanol, spin at max speed at 4 °C for 10 min, and air dry for 2 min. Resuspend each pellet in 6 μl of water and pool into a single PCR tube. The expected yield is approximately 2 μg poly(A)+ RNA per sample.
3. PSEUDO-SEQ LIBRARY PREPARATION

Pseudo-seq is derived from the primer extension-based method for ψ detection described earlier (Bakin & Ofengand, 1993), which we have adapted to the Illumina sequencing platform. This is accomplished by incorporating CMC modification into a sequencing library preparation protocol, and deep sequencing of abortive rather than full-length RT products from both CMC-treated (+CMC) and -untreated (−CMC) RNA samples. Subsequent computational analyses (Section 4) are used to identify reproducible CMC-dependent RT stops, which correspond to ψ residues (Fig. 2).

RNA is first randomly fragmented to ensure uniform coverage of the transcriptome. After RNA fragmentation, ψ residues must be specifically derivatized with CMC to allow for their identification. However, since ψ, U, and G residues are CMC reactive, specific reversal of the U-CMC and G-CMC adducts is needed. This is accomplished by reversal of CMC modification under alkaline conditions, which exploits the resistance of the CMC-N3-ψ linkage to alkaline hydrolysis. Following CMC modification, a narrow range of RNA fragment sizes is then selected, which is essential for subsequent separation of truncated cDNAs from full-length cDNAs following the RT step. After size selection, a 3′ adapter is ligated onto the RNA fragments, which provides a binding site for a single RT primer, thereby avoiding the requirement for gene-specific primers used

![Diagram of Pseudo-seq library preparation](image)

Figure 2 A schematic of Pseudo-seq library preparation. CMC-dependent RT stops correspond to sites of pseudouridylation. See Section 3 for details.
in traditional primer extension assays. After cDNA synthesis, truncated RT products are selected by gel purification, enriching for cDNAs whose 3' ends correspond to sites at which RT stops. Stops due to the presence of Ψ-CMC adducts will be present in the +CMC libraries, while natural stops, such as those caused by RNA secondary structure, will be present in both −CMC and +CMC libraries. Intramolecular ligation with a ssDNA ligase circularizes the cDNA, providing a 5' binding site for primers for subsequent PCR amplification, as well as for primers for Illumina sequencing. This chapter assumes that quality control steps to assess library concentration and library size distribution will be carried out by the facility performing the sequencing.

3.1 RNA Fragmentation

Randomly fragmented RNA is needed to ensure even coverage of potential pseudouridine sites. RNA can be fragmented in a relatively sequence-independent fashion using divalent zinc cations. However, the efficiency of RNA fragmentation differs depending upon the source and composition of the RNA sample. Thus, optimization of fragmentation time and temperature may be necessary to obtain a sufficient quantity of fragments in the desired size range. For Pseudo-seq, fragments in the range of 60–150 nt range are acceptable, though a narrow size distribution (e.g., 60–70 nt) should be used for a given experiment. The following fragmentation protocols yield sufficient fragments for both +CMC and −CMC libraries.

Fragment all of the pooled poly(A) RNA from Section 2.2. This should be approximately 2 μg of RNA, but up to 25 μg can be efficiently fragmented in this volume. To the poly(A) RNA (in 24 μl), add water to a final volume of 54 μl and place on ice. Add 6 μl of 100 mM ZnCl₂ (10 mM final) and fragment in a thermocycler for 55 s at 94 °C. Quench the reaction by quickly placing on ice, and adding 60 μl of 40 mM EDTA (20 mM final). Add 1/9th volume 3 M NaOAc and one volume isopropanol, and precipitate at −20 °C for at least 30 min. GlycoBlue does not need to be added to the precipitation since it carries over from the poly(A) selection. Spin and wash as described in Section 2.2. Resuspend fragmented RNA in 30 μl of water. Yeast total RNA can be fragmented in 10 mM ZnCl₂ for 5 min at 94 °C.

3.2 CMC Modification and Reversal

RNA should be modified with freshly made 0.5 M CMC in BEU Buffer (212 mg/ml). CMC-treated and mock-treated (−CMC) libraries should be prepared in parallel. To account for the reduced recovery of
CMC-modified RNA by precipitation, transfer 18 and 12 μl of fragmented RNA to microcentrifuge tubes for the + CMC and −CMC libraries, respectively, and bring the total volume of each to 20 μl with water. Add 2.9 μl of 40 mM EDTA to each sample (5 mM final), denature at 80 °C for 3 min, and place on ice. Add 100 μl of 0.5 M CMC in BEU Buffer (0.4 M CMC final) to the + CMC sample, and add 100 μl of BEU Buffer to the −CMC sample. Incubate at 40 °C for 45 min at 1000 rpm in a Thermomixer (Eppendorf). To precipitate, add 2 μl GlycoBlue; 50 μl 3 M NaOAc, pH 5.3; and 1 ml ethanol and chill at −20 °C for at least 30 min. Spin in a microcentrifuge at max speed at 4 °C for 30 min, and wash twice in 500 μl ice-cold 70% ethanol, spinning at maximum speed at 4 °C for 10 min after each wash. Air dry the RNA pellets for 2 min after washes. These precipitation conditions give a high yield of CMC-modified RNA compared to isopropanol precipitation. The CMC-treated RNA pellets may be more diffuse than the mock-treated pellets.

To reverse the CMC modifications on U and G residues, resuspend the RNA pellet in 30 μl sodium carbonate pH 10.4 Buffer and incubate at 50 °C for 2 h at 1000 rpm in a Thermomixer. Both CMC-modified and mock-treated RNA should be treated in parallel. To precipitate, add 2 μl GlycoBlue, 1/9th volume 3 M NaOAc, pH 5.3, and 2.5 volumes ethanol and chill at −20 °C for at least 30 min. The precipitation should be spun and washed as above. Resuspend the RNA pellets in 8 μl of 10 mM Tris, pH 8.0.

Careful execution of the CMC modification and reversal steps is important to ensure specific derivatization of Ψ residues by CMC. Thus, fresh CMC and properly stored sodium carbonate pH 10.4 Buffer should be used.

Pseudouridine identification by Pseudo-seq is robust to small changes in modification conditions. We have used CMC concentrations of both 0.2 and 0.4 M and have observed similar Ψ signal at both concentrations. However, increasing the extent of CMC modification with higher concentrations may interfere somewhat with the detection of 5′ Ψs in closely spaced groups of modifications due to a shadowing effect of the 3′ Ψs. Thus, for monitoring pseudouridylation in rRNA and tRNA, which contain many closely spaced Ψs, lower concentrations of CMC are recommended.

### 3.3 3′ End Healing

RNA fragmentation with divalent zinc cations leaves a 2′,3′ cyclic phosphate, which must be removed to make the RNA fragments suitable substrates for subsequent 3′ adapter ligation. These cyclic phosphates can be
converted to 3′ phosphates by T4 polynucleotide kinase (PNK), which can then be removed by the action of Calf Intestinal Alkaline Phosphatase (CIP). To 8 μl of RNA, add 0.5 μl of RNasin Plus (Promega N2615), 1.25 μl 10 × PNK Buffer, 1.25 μl T4 PNK (NEB M0201), and 1 μl CIP (NEB M0290) and incubate at 37 °C for 1–2 h. Add 12.5 μl of 2 × RNA Loading Dye to prepare for size selection (Section 3.4).

3.4 RNA Size Selection

After CMC treatment and 3′ end healing, the desired range of RNA fragment sizes is selected. Selection of a narrow range of fragment sizes, spanning 10–20 nt, at this step allows reliable separation of truncated from full-length cDNAs after RT. We have performed Pseudo-seq on a variety of RNA fragment sizes ranging from 60–70 to 120–140 nt and have found that our protocol is robust to changes in RNA fragment size, provided the size range taken is narrow. RNA fragments are size selected by excising and eluting the desired range of RNA fragment sizes from a denaturing PAGE gel.

Prepare 8% TBE/Urea/Polyacrylamide mini-gels (8 × 10 cm), and prerun for 20 min at 200 V. While the gel is prerunning, prepare RNA fragments and 10 bp Ladder (Life Technologies 10821-015) for loading. The 10 bp Ladder should be loaded in both gel lanes flanking RNA fragments to facilitate size selection. For each lane, prepare 20 μl of 10 bp Ladder (0.5 μl 10 bp Ladder, 9.5 μl water, 10 μl 2 × RNA Loading Dye). Denature the RNA fragments and 10 bp Ladder at 95 °C for 2 min, and then place on ice until loading. Load the gel and run for 60 min at 200 V for the fragment sizes indicated below (the bromophenol blue dye front will run off of the gel). Stain the gel for 5 min with SYBR Gold (Invitrogen S-11494) diluted 1:10,000 in 0.5 × TBE, and then visualize by UV transillumination. Excise several ranges of RNA fragment sizes (80–100, 100–120, and 120–140) guided by the 10 bp Ladder (Fig. 3). Proceed with one range of fragment sizes and keep the other size ranges for backup. Backup RNA fragments can be stored as gel slices at −80 °C or can be eluted from the gel slices, and the precipitating eluates can be stored in isopropanol at −20 °C. Elute the RNA fragments from the gel slices as described in Section 6.3.2 and resuspend in 5.5 μl water.

3.5 3′ Adapter Ligation

Ligation of an adenylated DNA adapter to the 3′ ends of RNA fragments provides a uniform primer binding site for cDNA synthesis. Adenylated
adapters are available commercially, but are quite expensive. A protocol for adapter adenylation is provided in Section 6.3.3. To the RNA fragments, add 0.5 μl adenylated adapter (100 μM), 1.2 μl 10 × T4 RNA Ligase Buffer, 1 μl RNasin, 3 μl PEG 8000, and 1 μl T4 RNA Ligase (NEB M0204). The absence of ATP in this reaction ensures that only the adenylated adapters are ligated to the 3' ends of RNA fragments. Incubate at 22 °C for at least 2.5 h. To precipitate, add 30 μl 3 M NaOAc, pH 5.3, 260 μl water, 2 μl GlycoBlue, and 300 μl isopropanol. Precipitate at −20 °C for at least 30 min and spin as described in Section 2.2. Resuspend the pellet in 7 μl water. The efficiency of ligation can be checked by running 0.8 μl of the reaction out on an 8% TBE/Urea/Polyacrylamide gel and should be in the range of 70–90%.

### 3.6 Reverse Transcription and Size Selection

Size selection of truncated cDNAs allows for the sites of RT stops to be determined by sequencing. We have performed Pseudo-seq using both AMV RT (Promega M5108) and ssIII RT (Life Technologies 18080093) enzymes with similar results.
Prepare an annealing mix for each sample by transferring 6.2 μl of each of the ligation reactions to a PCR tube, and adding 1 μl RT Primer, and 0.8 μl 10 × RT Buffer w/o Mg²⁺. In parallel, prepare a no RNA reaction using 6.2 μl of H₂O. To anneal the RT primer to the ligated RNA fragments, incubate in a thermocycler at 65 °C for 4 min, then 55 °C for 2 min, 45 °C for 2 min, and 42 °C for 2 min. Briefly centrifuge the tubes to collect any condensation and then place on ice. Prepare an extension master mix by mixing 0.6 μl 10 × RT Buffer w/o Mg²⁺, 1.4 μl 10 mM dNTPs, 0.7 μl 100 mM MgCl₂, 1.3 μl H₂O, 1.0 μl RNasin Plus, and 1.0 μl AMV RT per sample. Add 6 μl of this master mix to each of the annealing mixes. Incubate in a thermocycler at 42 °C for 1 h. Remove the RNA by adding 1.5 μl 1 N NaOH, and incubating in a thermocycler at 98 °C for 15 min. Neutralize the pH by adding 1.5 μl 1 N HCl. Prepare the samples for size selection by adding 17 μl of 2× RNA Loading Dye.

Prepare 8% TBE/Urea/Polyacrylamide mini-gels, and prerun as described in Section 3.4. Prepare 10 bp ladder as above (Section 3.4). Denature the cDNAs at 95 °C for 2 min, and then place on ice. Load the gel, with each cDNA sample split across two lanes, and run at 200 V for 65 min. The RT primer runs at 85 nt and should be run as close to the bottom of the gel as possible to maximize separation of truncated from full-length cDNAs. For reference, the xylene cyanol dye front runs at approximately 75 nt. Disassemble, stain, and visualize the cDNA gel as described in Section 3.4. Excise gel slices corresponding to truncated cDNAs. These should have at least 25 nt added to the RT primer, since shorter products are nonspecific and should avoid the 105 nt nonspecific product and the full-length cDNA band (Fig. 4). Extract cDNAs from the gel slices as described in Section 6.3.2. Resuspend the cDNA pellets in 15 μl 10 mM Tris, pH 8.0.

### 3.7 Circularization

Intramolecular ligation of cDNAs can be used in lieu of 5′ adapter ligation to provide a 5′ PCR primer binding site. Prepare a circularization master mix. For each sample, mix 2 μl 10 CircLigase Buffer, 1 μl 1 mM ATP, and 1 μl 50 mM MnCl₂. Add 4 μl of this mix to each sample, then add 1 μl 0.5 × CircLigase ssDNA Ligase (Epicentre CL4115K) diluted 1:1 in 1 × CircLigase Buffer, and mix well. Incubate in a thermocycler at 60 °C for at least 1 h, and then heat inactivate at 80 °C for 10 min. Reactions can be used immediately for PCR or can be stored at −20 °C.
Circularized cDNAs are then PCR amplified and gel purified to yield a library suitable for sequencing on the Illumina platform. Several PCR reactions are performed, each using a different number of amplification cycles, to ensure that a reaction with an ideal level of amplification is obtained. For the purposes of this chapter, it is assumed that libraries will be submitted to a facility that will perform quality control and sequencing.

Prepare a PCR Master Mix for each library. Per library add 15 μl HF Buffer, 1.5 μl 10 mM dNTPs, 3.78 μl Forward PCR Primer, 3.78 μl Barcoded Reverse PCR Primer, 52.6 μl H2O, and 0.75 μl Phusion High-Fidelity DNA Polymerase (NEB M0530L); these volumes are sufficient for 4.5 PCR reactions per library. Each library that will be sequenced on the same HiSeq lane should be amplified with a different Barcoded Reverse PCR primer with a unique barcode sequence. To the PCR Master Mix, add 4.5 μl of the circularized cDNA sample, and mix well. Transfer 16.7 μl of the Master Mix to each of four PCR tubes. Perform PCR with the following cycles: (1) 98 °C, 30 s, (2) 98 °C, 10 s, (3) 60 °C, 20 s, (4) 72 °C, 40 s, (5) repeat steps 2–4 17 × (18 cycles total). Remove PCR tubes from the thermocycler and place on ice after the extension phase (step 4) of cycles 12, 14, 16, and 18. Add 3.4 μl of DNA Loading Dye to each PCR reaction. Prepare an aliquot of 10 bp Ladder for each gel by mixing 10 bp Ladder,
15.7 μl H$_2$O, and 3.3 μl 6 × DNA Loading Dye. Load the reactions on an 8% TBE/Polyacrylamide mini-gel and run at 200 V for 40 min. Disassemble, stain, and visualize the PCR gel as described in Section 3.4. Excise PCR products of the appropriate size, based on the range of RNA fragment sizes used. The 105 bp no insert PCR product should be avoided, and bands should not be cut from lanes with saturated reactions, or those with higher molecular weight species (Fig. 5). Extract PCR reactions from the gel slices as described in Section 6.3.2. Resuspend the PCR products in 10 μl 10 mM Tris (pH 8.0). These gel-purified PCR products are suitable for Illumina sequencing and can be submitted to a facility for sequencing. Should insufficient material be recovered, PCR reactions from appropriate cycle numbers can be scaled up to 50 μl.

Avoiding PCR amplification bias during library preparation is important to ensure that composition of library accurately reflects the RNA fragment/cDNA pool, avoiding over or under representation of certain sequences after PCR. This can be done by avoiding saturated PCR products, or those with higher molecular weight bands. The effects of PCR bias

![Figure 5](image.png)

**Figure 5** PCR product gel purification. Excised PCR products are indicated by bounding boxes. The lanes for 16 and 18 cycles are too saturated.
can be assessed and minimized by adding at least eight random nucleotides to the RT primer. These random nucleotides serve as barcodes for RNA fragments. Identical reads with identical random barcodes likely arise from PCR amplification bias and can be collapsed into a single read.

4. PSEUDO-SEQ DATA ANALYSIS

Sites of pseudouridylation can be identified through computational analysis of data obtained through Illumina sequencing of Pseudo-seq libraries. The 5′ ends of Pseudo-seq reads, which correspond to the 3′ ends of cDNAs, represent sites of RT stops (Section 3.6). Those RT stops that are reproducibly enriched in +CMC libraries rather than −CMC libraries represent sites of pseudouridylation. This section describes the computational identification of these CMC-dependent RT stops from Pseudo-seq data. The analyses described are carried out using publically available tools designed for the analysis of next-generation sequencing data, and custom scripts that must be prepared by the user. Preparation of these custom scripts requires familiarity with Python or another similar scripting language.

First, the sequencing reads are processed into a form suitable for mapping to the target genome. If barcoded PCR primers were used to multiplex several libraries, allowing multiple libraries to be sequenced on the same flow cell lane, then the reads must be separated by barcode, or index sequence. Barcode parsing allows individual reads to be assigned to the appropriate library. Some fraction of reads will contain some 3′ adaptor sequence that will interfere with mapping, and therefore must be removed from these demultiplexed reads. Reads processed in this manner can then be mapped to the target genome. These steps can easily be combined into an automated pipeline using bash scripts.

Custom scripts are then employed to identify Ψ sites from the 5′ positions (3′ cDNA ends) of these mapped reads. These scripts calculate a Pseudo-seq peak value for each U in the transcriptome, and those sites whose peak values reproducibly exceed a specified threshold in multiple biological replicates are identified as sites of pseudouridylation.

4.1 Library Demultiplexing

The reads generated from an Illumina sequencing run will be output in a FASTQ file. Within this file, each read is represented by four sequential lines, as illustrated below.
The first line, beginning with @, contains the name of the read, which includes an identifier for the sequencing machine and location of the read on the flow cell. Importantly, it also contains the barcode sequence located between the # and / characters, which is the reverse complement of the barcode sequence in the Barcoded Reverse PCR Primer. The second line contains the read itself, and the third and fourth lines contain an alternate name for the read and quality scores, respectively.

To demultiplex FASTQ files, the index for an individual read should be matched to the reverse complements of the barcodes used in PCR amplification (Section 3.8). This matching can be carried out using a custom script that uses simple string comparisons to assign barcodes, allowing one mismatch. This script should write reads to new FASTQ files specific to each barcode used, including a FASTQ file for unmapped barcodes. Alternately, demultiplexing can be performed using Illumina’s CASAVA software or with other publically available tools. In some cases, the facility used for sequencing may demultiplex sequencing reads.

These FASTQ files can be quite large and may be compressed to save storage space. The downstream analyses described in Sections 4.2 and 4.3 are compatible with sequence data compressed with the gzip utility. In the command line, a FASTQ file can be compressed with the following command in a UNIX shell to yield a compressed file sequence_reads.fastq.gz.

gzip sequence_reads.fastq

4.2 Trimming of 3’ Adapter Sequences

Some significant fraction of reads will be derived from molecules that are shorter than the read length of the Illumina machine and will contain sequences derived from the 3’ adapter. This adapter sequence can interfere with the mapping of reads to the genome and must be removed from reads prior to mapping. The publically available Cutadapt tool can be used to trim adapter sequence from next-generation sequencing data (Martin, 2011). For a given FASTQ file, adapter sequences can be trimmed with the following command in a UNIX shell:

cutadapt -a ADAPTER_SEQ --overlap 3 --minimum-length 18 -o trimmed_reads.fastq.gz input_reads.fastq.gz.

236

Thomas M. Carlile et al.
This command should be run for each individual FASTQ file of reads. The -a ADAPTER_SEQ command specifies the adapter sequence to trim, the --overlap 3 command requires at least three bases of overlap with the adapter for trimming, the --minimum-length 18 command discards reads shorter than 18 nt, the -o trimmed_reads.fastq.gz command specifies the output file name, and input_reads.fastq.gz specifies the input file name.

4.3 Mapping Reads

Reads trimmed of adapter sequences can then be mapped to the target genome using Bowtie2 and to annotated splice junctions using TopHat (Kim et al., 2013; Langmead, Trapnell, Pop, & Salzberg, 2009). Read mapping using these packages requires a bowtie index for the target genome, and a file containing annotated splice junctions in a TopHat readable format. Bowtie indices are available for many commonly used model organisms or can be easily built from a FASTA file of the genomic sequence with the Bowtie2 package, and splice junction annotations for S. cerevisiae can be found at SGD (http://www.yeastgenome.org).

To map reads to the target genome and annotated splice junctions, enter the following command in a UNIX shell. TopHat will use BowTie2 to map reads to the genome.

tophat2 --no-novel-juncs --no-novel-indels --raw-juncs splice_juncs bowtie_index trimmed_reads.fastq.gz

This command should be run for each individual FASTQ file of trimmed reads. The --no-novel-juncs and --no-novel-indels commands specify that TopHat should not map reads to novel splice junctions or indels. The locations of the splice junction file and bowtie index are indicated by splice_juncs and bowtie_index, respectively. The file trimmed_reads.fastq.gz contains the reads trimmed of adapter sequence (Section 4.2).

This command will generate several output files containing mapping information. The relevant file for further analysis is called accepted_hits.bam, which contains the mapped reads and information about their mapped positions. This file must be parsed for subsequent analysis. This can be accomplished using the SAMtools package, and a custom script to parse the SAMtools output. This script should determine the number of reads whose 5’ ends (3’ cDNA ends) map to each position in the genome and store this output in a format suitable for subsequent steps (Section 4.4). Additionally, determine the total number of mapped reads and the total number of
reads mapped to the rRNA for each library. These values will be used to scale +CMC and −CMC library pairs as described in Section 4.4.

4.4 Computational Identification of Ψ Residues

Ψ identification involves the calculation of a Pseudo-seq peak value for each U in the transcriptome and the identification of sites with reproducibly high peak values. This analysis should be performed with user prepared scripts.

First, the +CMC and −CMC library pairs for each sample should be scaled to the same size. Since Pseudo-seq peak values involve the subtraction of −CMC reads at a given position from +CMC reads, this calculation is sensitive to differences in library size. A larger +CMC library will lead to an increase in calculated peak values, while a larger −CMC library will lead to a decrease in peak values. To scale, the −CMC library multiply the reads at each position in this library to the ratio of reads in the +CMC library to the −CMC library. This ratio should be calculated from the total number of mapped reads (Section 4.3) for Ψ identification in mRNAs. Alternately, the number of rRNA mapping reads can be used for Ψ identification in ncRNAs.

Next, transcripts without sufficient read coverage should be removed from further analysis. Only consider transcripts with a read coverage that exceeds a specified threshold of average reads per nucleotide (rc). Empirical determination of a suitable read threshold is described in Section 5.3. With sufficient biological replication, an rc value of 0 may be used, allowing analysis of transcripts with low expression levels.

For each U in the transcriptome, calculate the Pseudo-seq peak value for each +CMC and −CMC library pair. Determine the number of reads whose 5′ ends map one nucleotide 3′ of the U (3′ position), which is the expected position of an RT stop induced by a Ψ-CMC adduct, in both +CMC (r+) and −CMC (r−) libraries. Additionally, determine the number of reads whose 5′ ends map to a window of size ws nucleotides centered at the 3′ position, but exclusive of the reads at that position in both the +CMC (wr+) and −CMC (wr−) libraries. Calculate Pseudo-seq peak values according to the equation:

\[
\text{peak} = ws \times \frac{r^+ - r^-}{wr^+ + wr^-}
\]

Sites of pseudouridylation can be identified in two steps using the calculated peak values. First, for each +CMC and −CMC library pair flag
positions with a peak value exceeding a threshold, \( p \). These flagged positions will contain real sites of pseudouridylation amid a large number of false positives. Second, filter out false positives by determining which positions are reproducibly flagged in multiple library pairs. For \( N \) biological replicate library pairs, the position should be flagged in at least \( n \) replicate pairs. Discussion of biological replication and computational parameters (\( ws, p \), and \( n \)) will be discussed in Sections 5.2 and 5.3.

4.5 Genetic Assignment of \( \Psi \)s to Pseudouridylation Factors

In \( S. \ cervisiae \), or another genetically tractable organism, \( \Psi \)s identified by Pseudo-seq can be assigned to known, nonessential pseudouridylation factors by performing Pseudo-seq on strains lacking the factor of interest. Pseudo-seq libraries should be prepared (Section 3) from two biological replicates of a given deletion strain and analyzed as in Sections 4.1–4.4.

For each \( \Psi \) identified by Pseudo-seq, the median peak height for all \(+CMC\) and \(-CMC\) library pairs should be calculated, as well as the median total reads in the window centered at the site (Section 4.4, \( wr^+ + wr^- \)) for all library pairs. Assign a \( \Psi \) to a given factor if the peak heights for both replicate deletion libraries are less than 25% of the median peak height for that \( \Psi \), and if the total window reads in at least one of two replicates is greater than 25% of the median total window reads for that \( \Psi \).

5. EXPERIMENTAL CONSIDERATIONS

There are several factors that should be considered when planning Pseudo-seq experiments and during data analysis. Prior to carrying out experiments, both the sequencing depth needed and the number of biological replicates required should be determined. These factors will affect the transcripts for which \( \Psi \) identification is reliable, as well as the computational parameters chosen.

5.1 Read Coverage

The read coverage required for successful identification of \( \Psi \) residues will depend upon both the transcriptome size and the abundance of target transcripts. For \( S. \ cervisiae \), we have found that 12 million reads per \(+CMC\) and \(-CMC\) library pair (\(~6\) million per library) mapping to coding sequences is sufficient for identification of \( \Psi \)s in most expressed mRNAs. The number of reads required can be scaled up or down depending on the size of the
transcriptome. The read coverage required is also dependent on the abundance of the desired transcripts. For highly abundant transcripts such as tRNAs and rRNAs, the read number can be scaled down substantially.

5.2 Biological Replication

The use of replicate data for \( \Psi \) identification is important, since biological replicates allow for efficient filtering of false positives (Section 4.4). Biological replicates are especially important for eliminating false positives in transcripts of low abundance. We have found that use of 14 biological replicates allows robust \( \Psi \) identification in mRNAs of a wide range of abundances. While this number of replicates is high, it should be noted that use of strains deleted for known pseudouridylation factors can serve as replicates for control strains. If fewer replicates are used for \( \Psi \) identification, more stringent computational parameters can be used to compensate.

5.3 Computational Parameters

The computational parameters used for calling \( \Psi \)s should be empirically determined from Pseudo-seq data and will depend on both the read coverage attained and the number of biological replicates. These parameters are discussed in Section 4.4 and include the read coverage cutoff for filtering transcripts (\( r_c \)), the window size used for Pseudo-seq peak calculation (\( ws \)), the peak value threshold (\( p \)), and the number of replicate library pairs in which a position must be above \( p \) (\( n \)).

The use of a larger number of biological replicates will allow more permissive cutoffs to be used, as false positives are more reliably filtered out with a higher number of replicates. Conversely, with fewer biological replicates, more stringent cutoffs should be used. For \( \Psi \) calling in an experiment performed in \textit{S. cerevisiae} with 14 biological replicates, the following parameters can be used: \( r_c = 0.0 \), \( ws = 150 \), \( p = 1.0 \), and \( n = 10 \).

The appropriate parameters for \( \Psi \) calling can be determined empirically. \( \Psi \) identification should be performed with several variations of these parameters, and the read distribution surrounding the sites called as pseudouridylation events should be examined. These sites should have a strong enrichment of reads above background one nucleotide 3' of the identified \( \Psi \). The majority of \( \Psi \) sites called with reliable parameters should show this strong, position-specific enrichment of reads. Parameters that yield a large fraction of sites that do not display this enrichment are insufficiently stringent. Known \( \Psi \)s in rRNAs, tRNAs, or snRNAs will provide examples
of reliable Ψ peaks in abundant transcripts (Fig. 6). Random downsampling of the reads in rRNA followed by Ψ calling from these downsampled reads can also be used to assess the computational parameters chosen.

6. SOLUTIONS, REAGENTS, AND COMMON PROTOCOLS

6.1 Solutions

**YPAD:** 1% (w/v) Bacto Yeast Extract, 2% (w/v) Bacto Peptone, 2% (w/v) Glucose, 0.004% (w/v) Adenine Sulfate.

**AES Buffer:** 50 mM NaOAc (pH 5.3), 10 mM EDTA (pH 8.0), 1% (w/v) SDS.

**TES Buffer:** 10 mM Tris (pH 7.6), 1 mM EDTA (pH 8.0), 0.1% (w/v) SDS.

**TES + NaCl Buffer:** 0.5 M NaCl, 10 mM Tris (pH 7.6), 1 mM EDTA (pH 8.0), 0.1% (w/v) SDS. The SDS may come out of solution over time. If this occurs, redissolve by warming buffer.

**BEU Buffer:** 50 mM Bicine (pH 8.5), 4 mM EDTA (pH 8.0), 7 M Urea. The final pH will be approximately 9.0.

**Sodium Carbonate pH 10.4 Buffer:** 50 mM Na₂CO₃ (pH 10.4), 2 mM EDTA (pH 8.0). Prepare from 1 M Na₂CO₃ (pH 10.4) and 0.5 M EDTA (pH 8.0). Adjust the pH to 10.4, filter sterilize, and store aliquots at −20 °C.

**10 × RT Buffer w/o Mg²⁺:** 500 mM Tris (pH 8.6), 600 mM NaCl, and 100 mM DTT. Store at −20 °C.
2 x RNA Loading Dye: 95% formamide, 5 mM EDTA (pH 8.0), 0.025% (w/v) SDS, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF. Store aliquots at –20 °C.

6 x DNA Loading Dye: 30% (v/v) Glycerol, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF.

RNA Elution Buffer: 300 mM NaOAc (pH 5.3), 1 mM EDTA (pH 8.0), 100 U/ml RNasin Plus. The RNase inhibitor should be added immediately prior to use.

DNA Elution Buffer: 300 mM NaCl, 10 mM Tris (pH 8.0).

6.2 Library Oligonucleotide Sequences

3' Adapter:
/5Phos/TGGAATTCTCGGGTGCCAAGG/3ddC/

RT Primer:
/5Phos/GATCGTCGGACTGTAGAACTCTGAACCTGTCGG TGGTCGCGGTATCATT/iSp18/CACTCA/iSp18/GCCTTGG CACCGAGAATTCCA

It is important that this oligo be gel purified to ensure that full-length and truncated cDNAs are reliably separated.

Forward PCR Primer:
AATGATAACGGCGACCACCCA

Barcoded Reverse PCR Primer:
CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCCCTTGGCACCACCGAGAATTCCA

The XXXXXX sequence indicates a unique barcode sequence.

6.3 Common Protocols

6.3.1 Regeneration of Oligo (dT) Cellulose Beads

Oligo (dT) cellulose beads can be reused multiple times. To regenerate, rotate beads at room temperature for 1 h in greater than two bead volumes of 0.1 N NaOH. Wash beads with water until the eluate reaches a neutral pH, resuspend in TES+NaCl, and store at 4 °C.

6.3.2 Gel Extraction of Nucleic Acids

Place the gel slice in a microcentrifuge tube and add 400 μl of RNA Elution Buffer or DNA Elution Buffer and elute overnight 4 °C or room temperature for RNA and DNA fragments, respectively. Remove the supernatant, and filter through a Spin-X cellulose acetate column (Corning 8162) for 1 min at max speed in a microcentrifuge. Add 2 μl GlycoBlue and 1 volume
isopropanol and precipitate at $-20^\circ C$ for at least 30 min. Spin in a microcentrifuge at maximum speed at 4°C for 30 min. Wash the pellet in 750 μl ice-cold 70% ethanol, spin at maximum speed at 4°C for 10 min, and air dry for 2 min. Resuspend the pellet in an appropriate buffer.

### 6.3.3 Adapter Adenylation

Adenylated oligos with blocked 3′ ends are used as adapters because they reduce sequence bias in RNA fragment capture, and the absence of ATP only allows for the ligation of the adapter to the 3′ ends of RNA fragments (England, Gumport, & Uhlenbeck, 1977; Unrua & Bartel, 1998). Adapter adenylation requires the reagent adenosine 5′-phosphorimidazolide (ImpA), which can be obtained by a straightforward chemical synthesis described elsewhere (Pfeffer, Lagos-Quintana, & Tuschl, 2005). To 420 μl of 50 mM ImpA, add MgCl$_2$ to 25 mM, 3′ Adapter to 0.2 mM, and water to 500 μl. Incubate at 50°C for 3 h. Run the adenylation reaction out on a 20% denaturing PAGE gel, and gel purify the upper, adenylated species.

### ACKNOWLEDGMENTS

We thank members of the Gilbert laboratory for helpful discussions regarding this protocol. This work was supported by grants to W.V.G. from National Institutes of Health (GM094303, GM081399), and the American Cancer Society—Robbie Sue Mudd Kidney Cancer Research Scholar Grant (RSG-13-396-01-RMC). T.M.C. was supported by the American Cancer Society New England Division (Ellison Foundation Postdoctoral Fellowship).

### REFERENCES


