View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Bern Open Repository and Information System (BORIS)

cDNA library generation for the analysis of small RNAs by high throughput sequencing

Jennifer Gebetsberger[1,2], Roger Fricker[1,2], Norbert Polacek[1]

[1]Department of Chemistry and Biochemistry, University of Bern, Bern, Switzerland; [2]Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland;

Keywords: cDNA library, reverse transcription, small RNA, non-coding RNA

Running Head: cDNA library of small RNAs

Abstract

The RNome of a cell is highly diverse and consists besides messenger RNAs (mRNAs), transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) also of other small and long transcript entities without apparent coding potential. This class of molecules, commonly referred to as non-protein-coding RNAs (ncRNAs), is involved in regulating numerous biological processes and thought to contribute to cellular complexity. Therefore much effort is put into their identification and further functional characterization. Here we provide a cost effective and reliable method for cDNA library construction of small RNAs in the size range of 20 to 500 residues. The effectiveness of the described method is demonstrated by the analysis of ribosome-associated small RNAs in the

eukaryotic model organism *Trypanosoma brucei*.

1. Introduction

Applications of high throughput sequencing technologies are tremendously raising and allow deep insights into many aspects of cellular biology. Therefore it is not surprising that reverse transcription of RNA into cDNA is one of the most frequently used applications in molecular biology. Although many methods for cDNA library construction are commercially available, they are either rather laborious or expensive. Furthermore, many of these protocols were primarily established for mRNAs only, demanding alternatives for the investigation of other RNA entities, such as primary transcripts and small ncRNAs. There is accumulating evidence that such small ncRNAs derive from diverse transcript classes and possess indispensable functions (reviewed in (1,2)). It appears that we have thus far only scratched the tip of the ncRNA iceberg, emphasizing the need for alternative approaches that allow the detection of yet overlooked ncRNA regulators. Therefore, the method described herein is amenable to construct high quality cDNA libraries of such small RNAs in a straightforward and cost effective manner.

A noteworthy advantage of the herein described method is that any kind of RNA starting material can be used for successful cDNA library generation. Therefore it was not only applied for the investigation of ncRNA transcriptomes (3) and RNomes of specific cellular compartments (4), but also allowed the detection of a so far undescribed population of ribosome-associated ncRNAs (5-7). To this end, specialized cDNA libraries from ribosome-bound small RNAs were generated from organisms spanning all domains of life, including *Staphylococcus aureus*, the archaeon *Haloferax volcanii* (5), the unicellular parasite *Trypanosoma brucei*, and *Saccharomyces*

cerevisiae (7). Besides already known ribosome-associated RNAs, such as mRNAs, transfer RNAs (tRNAs), the single recognition particle RNA, these cDNA libraries revealed hundreds of putative novel ncRNAs that target the ribosome and thereby potentially regulate protein biosynthesis (as an example see **Figure 1** for *T. brucei* deep sequencing results). Although the described method for cDNA library construction is highly reliable and straight-forward, attention has to be taken in the quantitative interpretation of sequencing results, since (i) more structured or post-transcriptionally modified RNAs might be less easily reverse transcribed and therefore underrepresented in the sequencing data, (ii) due to the requirement of PCR amplification **(Figure 2)** smaller RNA species will be more abundant than longer ones, (iii) the anchored primers used for first strand cDNA synthesis will not only anneal to the synthetic 3' C-tail **(Figure 2)**, but also to natural occurring internal C-stretches of RNAs (a problem especially relevant for organisms with a high GC content), and (iv) biases of the used enzymes (e.g. TAP, T4 RNA ligase, reverse transcriptase) cannot be excluded (also discussed in (8)).

2. Materials

2.1 RNA preparation

1. RNA extraction was performed with a Phenol:Chloroform:Isoamyl alcohol (25:24:1) solution (Roti®-Aqua-P/C/I, Roth) (**see Note 1**).

2.TBE buffer (10x): 108 g Tris, 55 g boric acid, and 9.3 g EDTA are dissolved in 1 l water and stored at room temperature (RT) (**see Notes 2 and 3**).

3. Running buffer (1x TBE): The 10x buffer is diluted to 1x and stored at RT.

4. Denaturing 8% (v/v) polyacrylamide solution with 7 M urea: Required amounts of 30%

polyacrylamide-M-Bis 29:1 solution and urea are dissolved in 1x TBE and stored at 4°C. For polymerization 1% (v/v) ammonium persulfate (APS) and 0.1% (v/v) of TEMED (N,N,N',N'-tetramethylethylenediamine) are added.

5. RNA loading dye (2x): 95% Formamide, 0.5 mM EDTA, Bromphenol blue, and Xylene cyanol.

6. Gel staining solution is made by using the running buffer (1x TBE) containing ethidium bromide (f.c. 0.4 µg/ml).

7. Elution buffer (2x): 0.3 M NaOAc (pH 5.4) and 1 mM EDTA in water.

2.2 cDNA library generation

1. Poly(A) Polymerase kit, including 10x reaction buffer: 0.5 M Tris/HCl (pH 8.0), 2.5 M NaCl, 100 mM $MgCl_2$ (Epicentre).

2. Tobacco Acid Pyrophosphatase, including 10x reaction buffer: 500 mM NaOAc (pH 6.0), 10 mM EDTA, 1% β-mercaptoethanol, 0.1% Triton X-100 (Epicentre).

3. T4 RNA Ligase, including 10x reaction buffer: 500 mM Tris/HCl (pH 7.5), 100 mM $MgCl_2$, 100 mM DTT, 10 mM ATP (Thermo Fisher Scientific).

4. Superscript™ II Reverse Transcriptase, including 5x First strand buffer: 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$ (Invitrogen).

5. Taq-DNA polymerase, including 10x PCR buffer: 100 mM Tris/HCl (pH 9.5), 200 mM $(NH_4)_2SO_4$, 15 mM $MgCl_2$, 10 mM DTT, 0.05% (v/v) Nonidet P-40 (9).

6. 5' adapter primer sequence: 5'-GTCAGCAATCCCTAACGAG***TAGT***A-3'

DNA primer harbors three RNA nucleosides at the 3' end (underlined letters). Italic, bold letters are depicting the four letter bar code to discriminate between RNAs originating from different conditions.

7. Achorerd oligo(dG) primer sequences: 5'-AGGACGCATCGTATGTCGGGGGG**X 3'**

X = equal mix of T, C, A

8. PCR primer sequences: forward primer 5'-GTCAGCAATCCCTAACGAG-3'
reverse primer 5'-AGGACGCATCGTATGTCG-3'

2.3 cDNA preparation

1. Native 8% (v/v) polyacrylamide solution: Required amounts of 30% polyacrylamide-M-Bis 29:1 solution is dissolved in 1x TBE and stored at 4°C. For polymerization 1% (v/v) ammonium persulfate (APS) and 0.1% (v/v) of TEMED (N,N,N',N'-tetramethylethylenediamine) are added.

2. DNA loading dye (6x): 60 % Glycerol, 60 mM EDTA, Bromphenol blue, and Xylene cyanol.

2.4 Diagnostic Sanger Sequencing

1. pGEM®-T Easy Vector System, including 2x Rapid ligation buffer: 60 mM Tris/HCl (pH 7.8), 20 mM $MgCl_2$, 20 mM DTT, 2 mM ATP, 10% polyethylene glycol (MW8000) (Promega).

2. LB medium: 1% (w/v) tryptone, 1% (w/v) NaCl, and 0.5% (w/v) yeast extract in water. Autoclave and store at 4°C.

3. LB agar: 1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, and 1.5% (w/v) agar in

water. After autoclaving add required supplements, such as X-gal (f.c. 0.065 mg/ml), ampicillin (f.c. 0.1 mg/ml), and IPTG (f.c. 0.6 μM) if transforming One Shot® TOP10 Electrocomp™ *E. coli* cells (Invitrogen).

4. M13 forward primer: 5'-CCCAGTCACGACGTTGTAAAACG-3'

   M13 reverse primer: 5'-AGCGGATAACAATTTCACACAGG-3'

5. BigDye terminator cycle sequencing reaction kit (PE Applied Biosystems).

3. Methods

The herein described method allows the construction of a cDNA library from all kinds of RNA templates. However, in order to increase the biological relevance of the starting material and to avoid amplification of putative cellular degradation products, isolation of ncRNA/protein complexes via density gradients (10) or affinity purification (7) is highly recommended. Based on our interest in small RNAs, a size-separation step via denaturing polyacrylamide gel electrophoresis (PAGE) is included at the beginning of the protocol **(Figure 2)**. This step allows the exclusion of longer RNAs, such as rRNAs and mRNAs and therefore enriches smaller RNA molecules. Since many small regulatory RNAs are commonly processing products of longer precursors (1,2,5,6,11) we usually attempt to retain the natural ends of RNA molecules for downstream bioinformatic analyses (**see Chapter Zywicki**). Therefore the 3' ends of the RNAs are extended by poly(C)-tailing using the poly(A) polymerase (12). For 5' adaptor ligation, the T4 RNA ligase requires monophosphorylated RNA termini. In order to remove 5' triphosphate groups from primary transcripts (or eukaryal methylguanosine-cap structures), a tobacco acid pyrophosphatase (TAP) treatment is included in the protocol, generating RNAs with 5'-monophosphorylated termini (**Figure 2**). The subsequent 5' adaptor ligation by T4 RNA ligase aims to introduce a specific sequence for final PCR amplification of the cDNA and additionally offers the opportunity of barcode insertion. This enables parallel sequencing of samples originating from different RNA preparations in one deep-sequencing run. T4 RNA ligase uses two RNA strands as substrate. Therefore the last three nucleotides at the 3' end of the 5’ adaptor oligonucleotides should be ribonucleotides to ensure efficient ligation (see Materials). The 3' C-tailed and 5' adaptor-ligated RNA is reverse transcribed into cDNA using oligo(dG) anchored primers. The cDNA is finally amplified via PCR and purified performing native PAGE.

3.1. Size Separation of the RNA

1. For RNA extraction, add 1 x volume of Roti®-Aqua-P/C/I (**Note 1**) vortex for 1 min. Then, centrifuge for 1 min at 16,000 x g, and transfer upper phase into a new precooled 1.5 ml reaction tube. Add 1 x volume of chloroform, vortex 1 min, and centrifuge again for 1 min at 16,000 x g. Transfer upper phase again into a new precooled 1.5 ml reaction tube. Add 2.5 x volume EtOH abs (-20°C) and 300 mM NaOAc (pH 5.4), invert several times, and put the samples for 20 min at -80°C for precipitation. Afterwards, centrifuge 20 min at 16,000 x g at 4°C followed by pellet washing with 70 % EtOH (-20°C). Remove EtOH and let the pellet dry by air for 10 min on ice.

2. The RNA (5 to 200 µg are recommended) is mixed with RNA loading dye and loaded in single slot of a denaturing 8% polyacrylamide gel containing 7 M urea (gel thickness 2 mm). Also load an RNA molecular weight marker, enabling the correct size detection of the RNA of interest. The PAGE is run with 200 Volts for 2.5 hours.

3. The gel was stained with EtBr (f.c. 0.4 µg/ml) in 1 x TBE buffer for 10 min.

4. After gel staining, the RNA in the size-range of interest is excised from the gel, by cutting it out with a sterile scalpel (**see Note 4**).

5. Dice the excised gel piece into 1 mm^3 cubes with a sterile scalpel, enhancing subsequent elution of the RNA from the gel.

6. Carefully transfer the gel pieces into a fresh 1.5 ml reaction tube. Passively elute the RNA into elution buffer containing 0.3 M NaOAc (pH 5.4) and 1 mM EDTA. Therefore add elution buffer until all gel slices are submerged and incubate over night with constant shaking at 4°C (**see Note**

4).

7. Remove the supernatant using a pipette and transfer the solution into a fresh 1.5 ml reaction tube.

8. Centrifuge the sample for 2,300 x g at 4°C for 5 min to pellet potential gel pieces. Transfer the elution buffer, containing the eluted RNA, into a clean reaction tube (**see Note 5** and **Note 6**).

9. Precipitate the eluate with 2.5 x volumes of 99.8% (v/v) ethanol (**see Note 7**) for at least 20 min at -80°C.

10. Ethanol precipitated RNA is centrifuged at 16,000 x g in an Eppendorf centrifuge at 4°C for 30 min, the pellet washed with 70% ethanol and subsequently dissolved in 11 µl water. 1 µl thereof is used for spectrophotometrically determination of the RNA concentration.

3.2 3' C Tailing of the RNA

1. 1 to 5 µg size-separated RNA is pre-incubated at 65°C for 5 min to resolve RNA secondary structures.

2. The tailing reaction containing the pre-incubated RNA is performed in a final volume of 50 µl with 1x reaction buffer, in the presence of 2 mM $MgCl_2$, 2 mM CTP (**see Note 8**), 40 U of ribonuclease inhibitor and 6.25 U of poly(A) polymerase (PAP) at 37°C for 1.5 h. Before addition of PAP, a pre-incubation step at 37°C for 5 min is recommended.

3. The 3' tailed RNA is subsequently Roti®-Aqua-P/C/I extracted, ethanol precipitated (see section 3.1, step 1) and resuspended in 7 µl (if proceeding with section 3.3) or 10 µl (if proceeding with section 3.4) water.

3.3 Decapping of the RNA

1. This reaction is required to remove 5' triphosphate groups from primary transcripts or 5’ methylguanosine-cap structures prior to 5' adaptor ligation. Otherwise continue directly with section 3.4

2. The whole amount of 3' tailed RNA is incubated with 10 U of Tobacco Acid Pyrophosphatase (TAP) (**see Note 9**) at 37°C for 1 h in a total volume of 10 µl containing 1 x reaction buffer and 40 U of ribonuclease inhibitor.

3. TAP- treated C-tailed RNA is Roti®-Aqua-P/C/I extracted, ethanol precipitated (see section 3.1, step 1), and resuspended in 10 µl water (**see Note 10**)

3.4. 5' Adaptor Ligation

1. The aim of the 5' adaptor ligation is the introduction of a specific sequence allowing subsequent PCR amplification. Furthermore barcodes (**see Note 11**) for different conditions can be integrated at this step, enabling sequence assignment after deep sequencing.

2. Since T4 RNA ligase can only join RNA ends, chimeric DNA/RNA oligonucleotides, carrying three RNA nucleosides at their 3' end, are used as 5' adaptors (see Materials).

3. The ligation reaction is performed in a final volume of 20 µl, containing 1x reaction buffer, and 20 µM 5' adaptor. The reaction is pre-incubated at 65°C for 5 min prior addition of 40 U ribonuclease inhibitor and 10 U of T4 RNA ligase. After incubation at 4°C over-night, extra 10 U of T4 RNA ligase are added and incubated for an additional hour on ice.

4. 5' adaptor-ligated RNA is Roti®-Aqua-P/C/I extracted, ethanol precipitated (see section 3.1, step 1), and resuspended in 10 µl water.

3.5. Reverse Transcription

1. The 5' adaptor-ligated and 3' C-tailed RNA is reverse transcribed into cDNA employing anchored oligo(dG) primers.

2. 10 µl of RNA from step 3.4 are pre-incubated in the presence of 5 µM anchored primer and 2 mM dNTPs (f.c.) at 65°C for 5 min and subsequently chilled on ice.

3. After addition of 4 µl 5x first strand buffer, 10 mM DTT and 40 U of ribonuclease inhibitor, the reaction is incubated at 42°C for 2 min.

4. 200 U of reverse transcriptase are added and the reaction is incubated at 42°C for 1 h.

3.6. PCR amplification of the cDNA

1. The cDNA is amplified via PCR.

2. The PCR is performed in a 50 µl reaction volume containing 4 µl of cDNA from step 5 (section 3.5), 1 µM forward primer and 1 µM reverse primer (**Note 12**), 0.1 mM dNTPs and 2 U Taq-DNA polymerase in 1x PCR buffer.

3. The PCR reaction is cycled using the following conditions: 2 min at 94°C; 20 cycles of 1 min 94°C, 1 min 54°C and 1 min 72°C; 5 min at 72°C.

3.7. Purification of PCR Product via PAGE

1. The purification of the amplified cDNA is achieved via PAGE under native conditions.

2. A 50 µl PCR reaction is mixed with 6x DNA loading dye and run on 8% polyacrylamide gel.

3. After gel staining with EtBr (f.c. 0.4 µg/ml) in 1 x TBE buffer for 10 min, the cDNA is excised

(**Note 13**) and eluted in elution buffer over night with constant shaking at 4°C (as performed in section 3.1.3 and 3.1.4).

4. After centrifugation at 2,300 x g at 4°C for 5 min, the eluate is transferred into a fresh reaction tube and precipitated with 2.5 volumes of absolute ethanol for 20 min at -80°C.

5. Ethanol precipitated cDNA is centrifuged at 16,000 x g at 4°C for 30 min, washed with 70% ethanol and dissolved in 20 ul water.

3.8. Preparation for deep-sequencing

1. Diagnostic Sanger sequencing is recommended as cDNA quality control step (**Note 14**).
2. Add adaptors required for deep-sequencing via PCR (**Note 15**). Contact the deep-sequencing platform of your choice for appropriate primer design.
3. PAGE-purify amplified cDNA, carrying the attached sequencing adaptors and precipitate with ethanol (as described in 3.7).
4. Measure cDNA concentration spectrophotometrically and dilute according to the company's requirements.

4. Notes

1. The pH of the P/C/I mixture used for the first RNA extraction should be acidic (optimal 4.5-5) in order to denature DNA that otherwise could possibly be present as a contamination. At pH 7, RNA and DNA will be in the water phase. At pH 5, DNA is partly denatured and will be directed to the organic phase, while RNA remains in the water phase.

2. All solutions should be prepared in deionized water.

3. The buffers are usually stored at -20°C if not otherwise mentioned.

4. Always clean the scalpel with ethanol after cutting out a gel piece with the desired RNA before cutting out the next piece to avoid RNA cross-contaminations. Note that the gel slices should neither be squeezed into the reaction tube, nor should they swim in the elution buffer.

5. Take care that no gel slice is taken along with the elution buffer.

6. If a second round of elution is desired, freeze the gel slices at -20°C, before adding elution buffer. Freezing and thawing will enhance the amount of eluted RNA.

7. Since the amount of precipitated RNA can be little, GlycoBlue (Ambion) can be used as nucleic acid co-precipitant. This increases the pellet visibility.

8. The CTP concentration depends on the amount and length of size-separated RNA. In general 75 pmol CTP per pmol of RNA is recommended.

9. 1 U TAP per pmol RNA is recommended.

10. Spectroscopic determination of RNA concentration at this step is not recommended due to the presence of unincorporated CTP from the C-tailing step (3.2.).

11. Barcodes can be introduced at the 3' end of the 5' adaptor. Due to possible deep sequencing errors, a minimum of four nucleotides is recommended.

12. Alternatively, primers specific for subsequent deep-sequencing can be used. Since they should be suitable for the sequencing platform of choice, contact the sequencing facility for primer design.

13. Consider that the size-range of the cDNA corresponds to the length of size-excluded RNA of

step 3.1 plus the length of the 5' and 3' adaptors.

14. In order to check for the cDNA quality, including e.g. correct adaptor ligation and sufficient insert lengths, diagnostic Sanger-sequencing of ~50 to 100 clones can be performed. Therefore cDNA is ligated into pGEM-T vectors and transformed into electro-competent *E. coli* cells. Here it is advantageous to use cells carrying a lacZ deletion mutant (e.g. One Shot® TOP10 Electrocomp™ *E. coli* are provided by Invitrogen), facilitating rapid and convenient detection of recombinant cells via blue white screening. After colony-PCR and subsequent purification of the resulting PCR products, Sanger sequencing can be performed, e.g. by using the BigyDye terminator cycle sequencing reaction kit (PE Applied Biosystems).

15. To omit the second PCR step, the use of appropriate deep-sequencing adaptors at step 3.6 is recommended.

Acknowledgement

This research was partly supported by the NCCR ‘RNA & Disease’ funded by the Swiss National Science Foundation. Additional grant support derives from the Swiss National Science Foundation (31003A_143388/1) and the Austrian Science Fund FWF (project number: Y315) to N.P.

References

1. Tuck A.C. and Tollervey D. (2011) RNA in pieces. *Trends Gen* **27**, 422-432
2. Gebetsberger, J. and Polacek, N. (2013) Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol*, **10**, 1798-1806.
3. Madej, M.J., Alfonzo, J.D. and Huttenhofer, A. (2007) Small ncRNA transcriptome analysis from kinetoplast mitochondria of Leishmania tarentolae. *Nucleic Acids Res*, **35**, 1544-1554.
4. Lung, B., Zemann, A., Madej, M.J., Schuelke, M., Techritz, S., Ruf, S., Bock, R. and Huttenhofer, A. (2006) Identification of small non-coding RNAs from mitochondria and chloroplasts. *Nucleic Acids Res*, **34**, 3842-3852.
5. Gebetsberger, J., Zywicki, M., Kunzi, A. and Polacek, N. (2012) tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in Haloferax volcanii. *Archaea*, **2012**, 260909.
6. Pircher, A., Bakowska-Zywicka, K., Schneider, L., Zywicki, M. and Polacek, N. (2014) An mRNA-Derived Noncoding RNA Targets and Regulates the Ribosome. *Mol Cell*, **54**, 147-155.
7. Zywicki, M., Bakowska-Zywicka, K. and Polacek, N. (2012) Revealing stable processing products from ribosome-associated small RNAs by deep-sequencing data analysis. *Nucleic Acids Res*, **40**, 4013-4024.
8. Huttenhofer, A. and Vogel, J. (2006) Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res*, **34**, 635-646.
9. Erlacher, M.D., Chirkova, A., Voegele, P. and Polacek, N. (2011) Generation of chemically engineered ribosomes for atomic mutagenesis studies on protein biosynthesis. *Nature Protocols*, **6**, 580-592.

10. Rederstorff, M. and Huttenhofer, A. (2011) cDNA library generation from ribonucleoprotein particles. *Nature Protocols*, **6**, 166-174.

11. Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Human Mol Gen*, **15 Spec No 1**, R17-29.

12. Martin, G. and Keller, W. (1998) Tailing and 3'-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA*, **4**, 226-230.

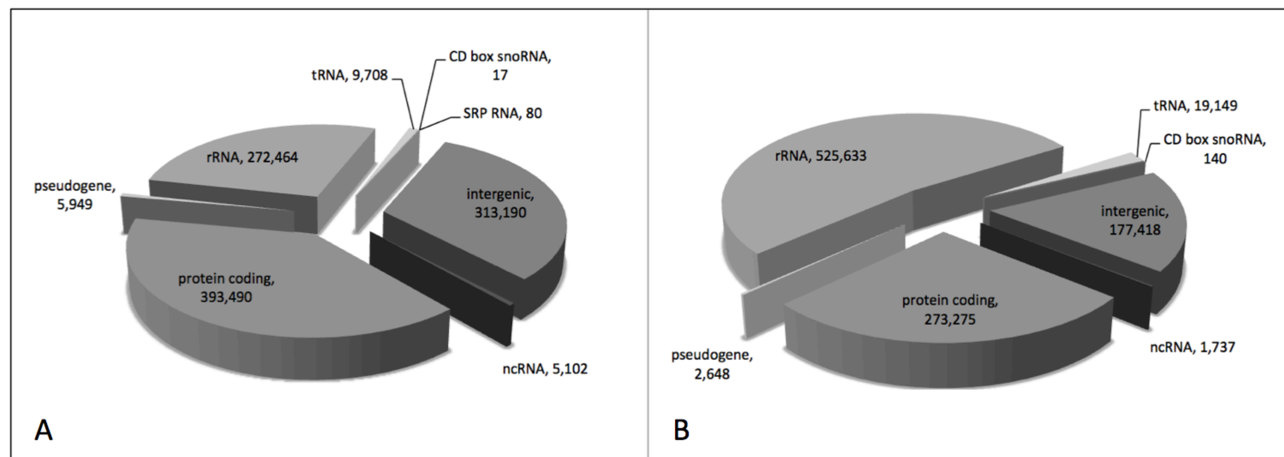

Figure 1: Pie chart of deep sequencing results. A specialized cDNA library was constructed for small ribosome-associated RNAs in *Trypanosoma brucei.* The cDNA library was deep-sequenced using the Illumina platform for single end reads at a maximum read length of 100 base pairs. Sequence reads were grouped according to their genomic origin (snoRNA for small nucleolar RNA, ncRNA for non-coding RNA, rRNA for ribosomal RNA, tRNA for transfer RNA, and SRP RNA for signal recognition particle RNA) and their relative distribution is indicated in reads per million (RPM). Sequencing results of ribosome-bound small RNAs extracted from (A) exponentially growing cultures are shown and compared to (B) RNAs extracted from heat stressed *T. brucei* cells. According to these results, different conditions led to a different ribosome-bound small RNA RNome.

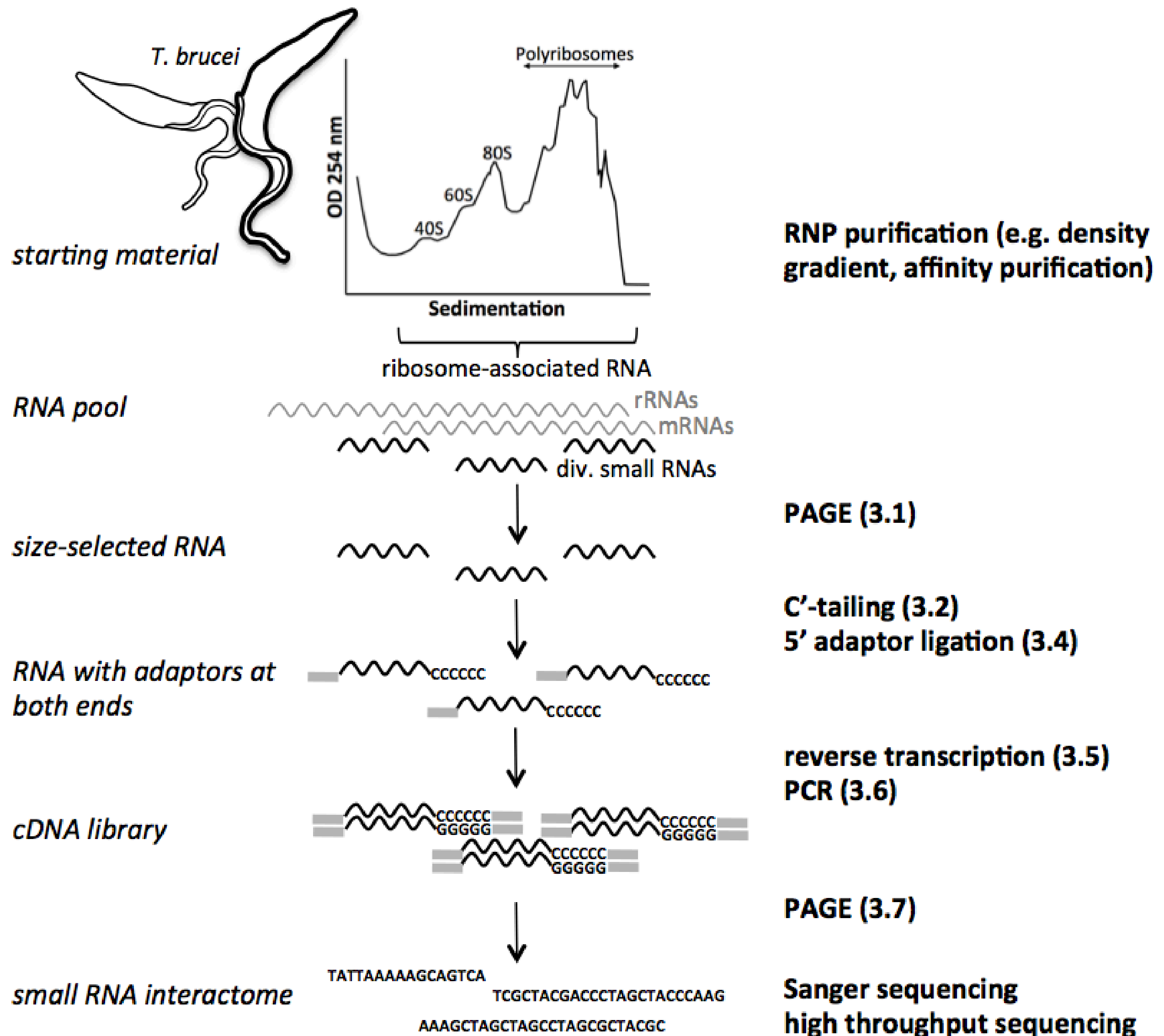

Figure 2: Flowchart for the cDNA library construction method reported. As starting material for cDNA library preparation any RNA entity can be used. In the presented case, ribosomes were isolated from *Trypanosoma brucei* via density gradient centrifugation yielding non-translating subunits and monosomes (40S, 60S, 80S) as well as actively translating polyribosomes. RNA that co-purified with these fractions was used for cDNA library construction, allowing the analysis of the small RNA interactome of *T. brucei* ribosomal particles. In a first step RNA is size-separated by polyacrylamide gel electrophoresis (PAGE) in order to enrich for small RNAs (black). Size-selected RNA is 3' C-tailed and 5' adaptor ligated prior to reverse transcription. The resulting

cDNA is amplified by PCR reaction and purified via native PAGE before sequencing. In order to check the quality of the cDNA library, diagnostic Sanger sequencing can be performed prior to high throughput sequencing. Whereas the type of working material is listed on the left side of the figure, the certain steps of cDNA library construction are given on the right together with the corresponding method section.