Efficient microRNA capture and bar-coding via enzymatic oligonucleotide adenylation

Francois Vigneault, A Michael Sismour & George M Church

Here we report a highly efficient and simplified strategy to preadenylate bar-coded oligonucleotides designed for microRNA (miRNA) capture and multiplex analysis. Using this approach, we enzymatically preadenylated bar-coded oligonucleotides with high efficiency when compared to the chemical method currently used by miRNA investigators. As a case study, we used these oligonucleotides in an ATP-independent ligation to miRNAs, suggesting the utility of our method in end-capture protocols and high-throughput sequencing applications.

miRNAs are endogenous 21–23-nt noncoding RNAs that post-transcriptionally repress gene expression by binding to 3′-untranslated regions of target mRNAs in a sequence-specific fashion to impair mRNA translation and/or stability[1–3]. Although most studies rely on the detection of previously reported miRNAs, the most powerful approach to identify and quantify expression of new miRNAs remains direct cloning and sequencing. To do so, miRNAs need to be extracted from the total RNA sample and then appended with 3′ and 5′ single-strand oligonucleotide adaptors by ligation to allow sequencing[4,5]. As only 678 miRNAs have been reported to be expressed in human cells (mirBase 11.0), we reasoned that in analyses of such a small number of miRNAs multiplexing would considerably minimize the per-sample cost of next-generation DNA sequencing and improve experimental design.

miRNAs are generated by Dicer processing and have 5′-phosphate and 3′-hydroxyl termini. This property, coupled with their short length, poses a substantial challenge during ligation-based capture as circularized miRNAs tend to be the dominant product. Many methods have recently been developed to circumvent this obstacle. One method relies on a dephosphorylation step to prevent self-circularization and/or concatemerization of the miRNAs; this process, however, also converts partly degraded RNA products into substrates for T4 RNA ligase. A favored strategy relies on the ligation of a pre–5′,5′-adenylated adaptor to the 3′ end of the miRNAs using T4 RNA ligase in absence of ATP[4–6]. Because the 5′ phosphate on the miRNA cannot be adenylated in the absence of ATP, miRNA circularization cannot occur, and the dominant reaction product is the desired miRNA–3′ adaptor conjugate. A 5′ adaptor is then ligated to this miRNA–3′ adaptor molecule using T4 RNA ligase in the presence of ATP.

Although this ligation reaction is simple, obtaining the pre-adenylated oligonucleotide needed for the method is not. Until recently, it required chemical synthesis of the adenosine 5′-phosphorimidazolide followed by chemical adenylation of the 5′ phosphate of the oligodeoxynucleotide[4,5]. In addition to not being trivial for most molecular biologists, the published chemical synthesis in this procedure is a slow process and has been reported with only 10–20% yields[4,5]. More recently, preadenylated oligonucleotides have become commercially available, but at such a high cost that only four are available, precluding the use of the technique for high-throughput experiments that necessitate bar-coding, which can require dozens to thousands of bar codes. To solve this problem, we developed a method for the production of preadenylated bar-coded oligonucleotides of any sequence using simple techniques achievable by the average laboratory technician.

T4 DNA ligase proceeds by a reaction mechanism in which 5′-adenylated DNA is an intermediate[7]. Therefore, interrupting the ligase in the middle of the reaction would allow for accumulation of an adenylated intermediate. Enzymatic adenylation of DNA and RNA has been reported in the literature as early as 1974 (refs. 8–10), and more recently for *in vitro* selection of ribozymes as well as 5′ capping and labeling applications[11–13].

Using a similar approach for enzymatic adenylation of DNA[11,13] and RNA[12,14], we developed a simple and highly efficient method for the production of preadenylated oligonucleotides that are now

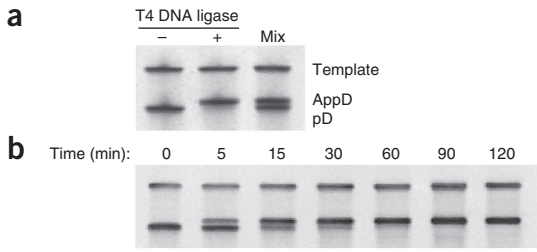

Figure 1 | Adenylation of the 3′ adaptor oligonucleotide by T4 DNA ligase. (**a**) Reaction converting the 5′-phosphorylated donor oligonucleotide (pD) to the adenylated form (AppD) in the presence (+) and absence (–) of T4 DNA ligase. The right lane is a mixture of the two reaction products to show the difference in sizes. The 40-mer oligonucleotide complementary template used in the adenylation reaction is indicated. (**b**) Time-course analysis of the adenylation reaction shown in **a**.

Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. Correspondence should be addressed to F.V. (fvigno@hms.harvard.edu) or G.M.C. (http://arep.med.harvard.edu/gmc/email.html).

RECEIVED 2 JUNE; ACCEPTED 23 JULY; PUBLISHED ONLINE 17 AUGUST 2008; DOI:10.1038/NMETH.1244

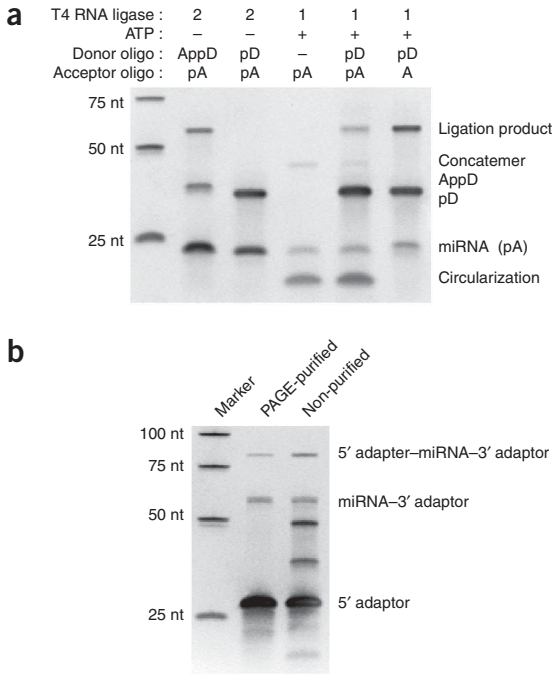

Figure 2 | miRNA capture by ligation. (**a**) Ligation of 3′ preadenylated (AppD) or nonadenylated (pD) adaptor to the synthetic miRNA using T4 RNA ligase 2 in absence of ATP. Control reactions using T4 RNA ligase 1 with ATP demonstrate self-circularization and concatemerization of the synthetic phosphorylated (pA) and dephosphorylated (A) acceptor miRNA. (**b**) Ligation of the 5′ adaptor to the miRNA–3′ adaptor ligation product. The ligation was conducted on a PAGE-purified ligation or directly on the reaction fragment mixture without prior PAGE purification.

used in the cloning and/or sequencing of miRNAs. To do so, we designed a 34-mer oligonucleotide (typically used in next-generation sequencing) phosphorylated at the 5′ end and blocked with a 3′-amino modifier (5′-phosphorylated donor). Blocking the 3′ termini is critical to avoid self-circularization or ligation to the 3′ end in subsequent steps of miRNA capture. We first annealed this 5′-phosphorylated donor oligonucleotide to a 45-nt-long complementary 'template' oligonucleotide to create an 11-nt-long 3′ overhang (see **Supplementary Methods** online for oligonucleotide sequences and experimental procedures). We then added T4 DNA ligase and ATP to the annealed oligonucleotides, incubated the mixture overnight and analyzed the products by denaturing polyacrylamide gel electrophoresis (PAGE). We observed a clear shift in the migration of the oligonucleotide suggesting the successful addition of a 5′,5′-adenyl pyrophosphoryl cap structure by the T4 DNA ligase in the interrupted reaction (**Fig. 1a**). A time course experiment revealed that conversion of the 5′-phosphorylated donor oligonucleotide to the adenylated oligonucleotide (AppD) product was completed by 90 min (**Fig. 1b**). The proportion of nonadenylated 5′-phosphorylated donor oligonucleotide was very low, allowing easy gel purification of the desired product. Alternatively, the oligonucleotide can be purified using various capture methods such as biotin or affinity chromatography (**Supplementary Methods**).

To confirm that the newly formed AppD products were preadenylated at their 5′ termini, we tested their efficiency of ligating to the 3′ end of an miRNA (**Fig. 2a**). To monitor the reaction, we used a synthetic 21-mer acceptor RNA oligonucleotide 5′-phosphorylated with 3′ hydroxyl termini to mimic an actual miRNA. Using T4 RNA ligase 2 (a mutant version of T4 RNA ligase 1 that can only use adenylated oligonucleotides) without ATP, we observed formation of a 55-nt-long ligation product resulting from ligation of synthetic miRNA and AppD. This ligation was specific to the preadenylated donor oligonucleotide as using T4 RNA ligase 2 with the non-adenylated 5′-phosphorylated donor did not result in formation of the ligation product. As expected, using T4 RNA ligase 1 with ATP allowed for ligation of the nonadenylated oligonucleotide, but resulted in a substantial reduction in the ligation owing to self-circularization and concatemerization of the synthetic miRNA. Although use of dephosphorylated miRNA acceptor resulted in good ligation efficiency, it is a poor strategy if one wants to avoid ligating partially degraded RNA fragments as explained above. As the objective is to achieve maximum capture of miRNA, we briefly analyzed the efficiency of 3′ adaptor ligation by testing key variables of this reaction (**Supplementary Fig. 1** online). A 60-min reaction was sufficient to achieve maximum ligation, and 200 units of T4 RNA ligase 2, 12% polyethylene glycol and a 10:1 ratio of 3′ adaptor to miRNA proved to be optimal. Similarly to previous reports[11], we achieved comparable adenylation using oligonucleotides of various sequences and sizes (data not shown).

We followed this first ligation with the ligation of the 5′-end adaptor (a 26-mer DNA-RNA chimera oligonucleotide) and observed that skipping gel purification of the initially ligated product resulted in a higher yield of the final 5′ adaptor–miRNA–3′ adaptor (**Fig. 2b**). We then tested the 5′ adaptor of different composition (DNA, RNA or DNA-RNA chimera), and observed that the RNA and chimera adaptor successfully ligated to the 5′ end of the synthetic miRNA whereas the DNA oligonucleotide did not (**Supplementary Fig. 2** online). Another approach known as 5′-ligation–independent cloning[6], which uses a second preadenylated 5′ adaptor on the reverse-transcribed strand instead of direct 5′ adaptor ligation to the miRNA, will also greatly benefit from simple preadenylation of oligonucleotides demonstrated herein.

To validate our adenylation approach for multiplex analysis, we designed four bar-coded 3′ adaptor oligonucleotides (BC1–BC4) to be used in multiplex sequencing of miRNAs. The barcodes maintain sample identity in a pooled mixture. We adenylated these bar-coded oligonucleotides using a complementary template that accommodates the degenerate nature of the bar codes. We then used the purified bar-coded 3′ preadenylated oligonucleotides to ligate two synthetic miRNAs (an 18-mer and a 21-mer) combined at various concentrations (as indicated in **Table 1**) in four independent reactions, using a different barcode for each reaction,

Table 1 | Distribution of bar-coded miRNAs in a mixed sample

miRNA:	18-mer				21-mer			
Bar code:	BC1	BC2	BC3	BC4	BC1	BC2	BC3	BC4
Expected	25	15	35	45	25	35	15	5
Actual	24	16	33	52	29	31	12	3

We performed ligation-based capture of two miRNAs, combined in different ratios in 4 reaction mixtures, using 3′ preadenylated oligonucleotides containing a different bar code for each reaction (BC1–BC4), then pooled the samples, cloned the ligation products and sequenced inserts from 200 randomly selected colonies. Expected and actual numbers of sequenced clones containing each of the bar codes are shown.

Table 2 | Multiplex analysis of a human brain small RNA sample

		Clones obtained
Distribution of captured small RNAs by type	miRNA	88 (44%)
	rRNA	61 (31%)
	mRNA or contig	32 (16%)
	Small nuclear RNA	19 (9%)
Distribution of small RNAs captured by each bar code	BC1 (ATAT)	46 (23%)
	BC2 (GCGC)	52 (26%)
	BC3 (TAGC)	54 (27%)
	BC4 (CCAA)	48 (24%)

We performed ligation-based capture of small RNAs (15–40 nt, purified from human brain total RNA), in 4 independent reactions, using 3′ preadenylated oligonucleotides containing a different barcode for each reaction (BC1–BC4), then pooled the samples, cloned the ligation products and sequenced inserts from 200 randomly selected colonies. Distribution of sequenced clones is shown.

followed by ligation of the 5′ adaptor. We then combined these four samples and reverse-transcribed the ligation products in a single reaction. After amplification, we cloned the resulting pooled fragments and sequenced DNA from 200 colonies that contained an insert (**Table 1**). Analysis of the resulting sequences revealed that this approach could efficiently be used to achieve multiplex analysis of miRNAs from mixed samples.

To validate the use of these bar-coded adaptors in a biological context, we used human brain total RNA, from which we extracted the small RNA fraction (15–40 nt) and used it as starting material in a similar ligation-based capture of miRNAs (**Supplementary Methods** online). Analysis of the resulting sequences revealed that a large proportion of miRNAs were efficiently captured by this approach, and the relative distribution of the bar-coded adaptors was maintained throughout the samples (**Table 2**). We validated the identity of the sequenced miRNAs using the mirBase 11.0 database (see **Supplementary Table 1** online for miRNA-library sequences).

The yield of adenylation achieved by our approach and the simplicity of this method is a considerable improvement compared to chemical synthesis. miRNA and RNA end-capture experiments will greatly benefit from the speed, convenience and accessibility of the procedure as well as the possibility of using adaptors of any sequence. Aiming for a 100-fold dynamic range from the highest to lowest miRNA expression level, one could easily combine up to 150 different samples, depending on the cyclic array technology used, in a quantitative expression profiling of miRNAs. This would result in a considerable cost reduction associated with the use of next-generation sequencing and will facilitate studies involving multiple conditions and/or time-course experiments. Moreover, the optimized ligation conditions and the use of bar-coded adaptors described here favors the design of more complex experiments and the achievement of higher yield of miRNA capture, which will likely result in the identification of undiscovered miRNAs and a better understanding of their implication in cellular processes.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

This work was supported by the Center for Excellence in Genome Sciences grant from the US National Human Genome Research Institute. We thank A.Z. Fire for helpful discussions and assistance in this project. F.V. is supported by a Canadian Institutes of Health Research Fellowship and A.M.S. is supported by the US National Institutes of Health Ruth L. Kirschstein Fellowship.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Ambros, V. *Nature* **431**, 350–355 (2004).
2. Bartel, D.P. *Cell* **116**, 281–297 (2004).
3. Ambros, V. & Chen, X. *Development* **134**, 1635–1641 (2007).
4. Lau, N.C. *et al. Science* **294**, 858–862 (2001).
5. Pfeffer, S., Lagos-Quintana, M. & Tuschl, T. *Curr. Protoc. Mol. Biol.* 26.4 (2005).
6. Pak, J. & Fire, A. *Science* **315**, 241–244 (2007).
7. Wood, Z.A., Sabatini, R.S. & Hajduk, S.L. *Mol. Cell* **13**, 455–456 (2004).
8. Lehman, I.R. *Science* **186**, 790–797 (1974).
9. Ohtsuka, E. *et al. Nucleic Acids Res.* **3**, 1613–1623 (1976).
10. McLaughlin, L.W., Piel, N. & Graeser, E. *Biochemistry* **24**, 267–273 (1985).
11. Patel, M.P., Baum, D.A. & Silverman, S.K. *Bioorg. Chem.* **36**, 46–56 (2008).
12. Wang, Y. & Silverman, S.K. *RNA* **12**, 1142–1146 (2006).
13. Chiuman, W. & Li, Y. *Bioorg. Chem.* **30**, 332–349 (2002).
14. Silverman, S.K. *RNA* **10**, 731–746 (2004).