A PCR-enhanced Method for Determining the 5' End Sequence of mRNAs

Martin A. Hofmann and David A. Brian

Department of Microbiology, The University of Tennessee, Knoxville, TN 37996-0845

A method was developed to amplify ligated cDNA copies of mRNA 5' ends before cloning and sequencing. The method facilitates the study of the 5' end sequence on many molecules of an individual mRNA species. It can also be used to study the 5' end sequence on other kinds of single-stranded RNA molecules.

Successful cloning of the 5' end of mRNAs by standard procedures (reviewed in ref. 1) requires either (1) dependence on hairpin formation at the 3' end of the first-strand cDNA molecule for successful priming of second-strand DNA synthesis,⁽²⁾ (2) dependence on an RNase H-generated RNA fragment at the 5' end of the mRNA to prime second-strand DNA synthesis,⁽³⁾ or (3) dependence on homopolymeric tailing of the first-strand cDNA for direct ligation into a vector containing a complementary tail.⁽⁴⁾ With each of these methods, it is often difficult to establish the sequence at the very 5' end of an mRNA molecule since, in the first procedure, bases can be removed in the hairpin loop by the single-strand-specific nuclease,⁽¹⁾ and in the second procedure, 5'→3' exonuclease activity in DNA polymerase I can destroy the RNase H-generated fragment used for priming second-strand DNA synthesis.⁽⁵⁾ In the third procedure, it is unknown whether the base nearest the mRNA in a homopolymeric tail is derived from the tailing procedure or from the mRNA template.

To circumvent these problems and to facilitate the study of a large number of individual molecules of a single mRNA species, we have developed a method that employs head-to-tail ligation of single-stranded, first-strand cDNA with RNA ligase, followed by PCR amplification, to yield double-stranded molecules in sufficient abundance for blunt-end cloning into a sequencing vector (Fig. 1).

METHOD

The steps in this method are as follows:

1. 20- to 30-mer oligodeoxynucleotides with a G + C content of ≥50% and complementary to a region begin-

ning 20–200 nucleotides downstream from the 5' end of the mRNA species to be examined (primer 1) are annealed at a concentration of 50 pmoles primer per 25 µg total cytoplasmic RNA.⁽⁶⁾

2. Primer is extended with reverse transcriptase.⁽⁶⁾ Some γ -³²P-labeled primer is used in parallel to generate a size marker for gel purification in the next step.

3. Extended product is separated from RNA and unextended primers by electrophoresis in a 6% polyacrylamide sequencing gel containing 50% urea, eluted by diffusion,⁽⁶⁾ ethanol-precipitated, and dissolved in 20 µl of water.

4. Ligation of ssDNA is done in a 20-µl reaction mix containing 10 µl of dissolved ssDNA, 10 µg of bovine serum albumin per milliliter, 50 mM Tris-HCl (pH 8.0), 10 mM MgCl₂, 10 mM dithiothreitol, 1 mM ATP, and 10 units of T4 RNA ligase (New England Biolabs) for 4 hr at 37°C.

5. Second-strand cDNA synthesis and amplification by PCR are done in a 50-µl reaction mix containing 1 µl of head-to-tail ligation mix, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5 mM MgCl₂, 0.01% (wt/vol) gelatin, 0.1% (vol/vol) Triton X-100, 200 µM of each dNTP, 1.0 µM primer 2, 1.0 µM primer 3, and 1.25 units of *Taq* DNA polymerase (Promega). The PCR is carried out for 30 cycles of 1 min at 94°C, 1 min at 55°C, and 10 sec at 72°C. (Primer 3 rather than primer 1 is used in this amplification step to increase the chances of obtaining 5' end-specific clones; i.e., primer 1 may have caused some random priming in step 1.)

6. PCR-amplified DNA is purified by electrophoresis in a 4% NuSieve agarose gel (FMC Biologicals) and the product is electroeluted,⁽⁶⁾ ethanol-

recipitated, and dissolved in water. Depending on the annealing positions of primers 2 and 3 in step 5, PCR-amplified products of greater than unit length are also obtained that can be likewise purified and studied). Amplified DNA is ligated into *Sma*I-linearized, non-dephosphorylated pGEM3Z (Promega) DNA and used to transform *Escherichia coli* strain JM109 cells following standard procedures.⁽⁶⁾

7. Colonies containing 5' end-specific clones are identified by colony

hybridization to oligodeoxynucleotide 4, which binds to a region between the 5' end of the mRNA and primer 3. Selected clones are sequenced by the dideoxytermination method on asymmetrically amplified DNA prepared directly from bacterial colonies as previously described.⁽⁷⁾

RESULTS AND DISCUSSION

With this method, we have demonstrated heterogeneity in the 5'-terminal five nucleotides of what is an

otherwise common leader sequence on each of the bovine coronavirus mRNA species (ref. 8, and in prep.). We have established that the heterogeneity is probably not a result of the experimental method used, since T7 RNA polymerase-generated transcripts of a single synthetic construct having the 5'-terminal sequence 5'-GATTGT, and producing the capped sequence 5'-^{7m}GpppGATTGT when synthesized in vitro,⁽⁹⁾ yields only a 5' sequence of 5'-GATTGT (six out of six clones)

FIGURE 1 Strategy for cDNA synthesis, PCR amplification, cloning, and sequencing of the 5' end of an mRNA species. (▨) Primer 1; (▩) primer 2; (▧) primer 3; (■) oligonucleotide 4. (⊙) First base on the mRNA (excluding the inverted methylated cap); (+) plus strand or nRNA (and virus) sense; (−) minus or anti-mRNA-sense.

when analyzed by this cloning and sequencing method. Furthermore, heterogeneity did not appear to be caused by a cellular factor since all cDNA clones derived from capped transcripts that had been transfected into uninfected human rectal tumor cells (18 out of 18 clones) likewise had 5'-GATTGT as in their 5'-terminal sequence.

We do not know the lower limit of mRNA concentration required for the method to be successful. We have obtained clones for mRNAs known to be present at the rate of around 100 molecules per cell in persistently infected cells (ref. 8, and in prep.). In our experience, if primer 1 is radiolabeled to a specific activity of 10^6 cpm/pmole, and a radiolabeled extended product is visible after an overnight exposure without a screen in step 3, then ligation and cloning have been invariably successful. In all cases, the mRNA of interest was detectable by Northern analysis in 10 μ g of total cellular RNA when an oligodeoxynucleotide probe of 20–26 nucleotides in length with the G + C content of $\geq 50\%$ and a specific activity of labeling of 10^6 cpm/pmole was used.⁽⁸⁾ All clones of mRNA (337 in all) that have been identified by probe 4 in step 7 have yielded 5'-end-specific sequence.

The method described here has not only proven useful in our laboratory for the analysis of several mRNA species, but also for the determination of 5' end sequences of minus-strand copies of coronavirus mRNAs (unpublished). Therefore, it should prove to be a general method for the analysis of any single-stranded RNA species for which an internal sequence is known.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant number AI14367. M.A.H. was supported by a postdoctoral fellowship from the Swiss National Science Foundation.

REFERENCES

1. Kimmel, A.R. and S.L. Berger. 1987. Preparation of cDNA and the generation of cDNA libraries: Overview. *Methods Enzymol.* **152**: 307–316.
2. Gubler, U. 1987. Second-strand cDNA synthesis: Classical method. *Methods Enzymol.* **152**: 325–329.
3. Gubler, U. 1987. Second-strand cDNA synthesis: mRNA fragments as primers. *Methods Enzymol.* **152**: 330–335.
4. Deininger, P.L. 1987. Full-length cDNA clones: Vector-primed cDNA synthesis. *Methods Enzymol.* **152**: 371–389.
5. D'Alessio, J.M. and G.F. Gerard. 1988. Second-strand cDNA synthesis with *E. coli* DNA polymerase I and RNase H: The fate of information at the mRNA 5' terminus and the effect of *E. coli* DNA ligase. *Nucleic Acids Res.* **16**: 1999–2014.
6. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
7. Hofmann, M.A. and D.A. Brian. 1991. Sequencing PCR DNA amplified directly from a bacterial colony. *BioTechniques* **11**: 16–17.
8. Hofmann, M.A., P.B. Sethna, and D.A. Brian. 1990. Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture. *J. Virol.* **64**: 4108–4114.
9. Theus, S.A. and C.D. Liarakos. 1990. A simple assay for determining the capping efficiencies of RNA polymerases used for *in vitro* transcription. *BioTechniques* **9**: 610–615.

Received April 29, 1991; accepted in revised form May 24, 1991.

A PCR-enhanced method for determining the 5' end sequence of mRNAs.

M A Hofmann and D A Brian

Genome Res. 1991 1: 43-45

Access the most recent version at doi:[10.1101/gr.1.1.43](https://doi.org/10.1101/gr.1.1.43)

References

This article cites 8 articles, 1 of which can be accessed free at:
<http://genome.cshlp.org/content/1/1/43.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Affordable, Accurate
Sequencing.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>