Determination of 5' Ends of Specific mRNAs by DNA Ligase-dependent Amplification

Wolf M. Bertling,¹ Frank Beier,² and Ernst Reichenberger³

¹Paul-Ehrlich-Institute, Unit of Molecular Pathology, 63225 Langen, Germany; ²Max-Planck Clinical Research Units, University of Erlangen, 91054 Erlangen, Germany; ³Harvard Medical School, Department of Anatomical Cell Biology, Boston, Massachusetts 02115

We established a novel way to clone 5' ends of unknown length and sequence of individual cDNAs. T4 DNA ligase is employed to ligate an annealed duplex of complementary primers, one of them with a 4-nucleotide-long randomized overlap, to first-strand cDNA, generating a new 5' end. Subsequent PCR with a downstream primer and a primer with specificity for this new 5' end leads to products that can easily be cloned and sequenced. Considerations for the choice of primers for ligation and amplification are given. We have used this method to determine the 5' sequences of three independent mRNAs: the human collagen type-X gene, the chicken anchorin CII gene, and the human cytidine deaminase gene. We will discuss this method in comparison with other methods published for the amplification of unknown 5' ends of mRNA species.

Standard procedures for cDNA cloning depend on RNase H-generated RNA fragments⁽¹⁾ or the formation of hairpin structures⁽²⁾ at the 3' end of first-strand cDNAs for priming of second-strand synthesis. This may also be done by using random primers similar to the protocol of Feinberg and Vogelstein.⁽³⁾ Alternatively, tailing of the 3' end of the first-strand cDNA with a homopolymer can be used for direct cloning of the resulting cDNA into a suitable vector.⁽⁴⁾

A procedure for rapid amplification of cDNA ends (RACE) or anchored PCR can be performed using oligonucleotides complementary to the homopolymer tail, together with a second primer of given specificity for amplification by PCR.⁽⁵⁾ Recently, two independent publications suggested the use of T4 RNA ligase. T4 RNA ligase is used to tail the 3' end of the first-strand cDNA with a primer. This primer is complementary to an upstream primer to be used in the subsequent PCR assay.⁽⁶⁾ Alternatively, first-strand cDNA molecules are ligated head to tail, resulting in a defined sequence at the 3' end of the cDNA, suitable for priming in a subsequent PCR.⁽⁷⁾

Amplification-driven approaches for cloning of specific sequences are very attractive and have been used successfully. However, RNA ligase has a preference to ligate short fragments, and the efficiency in a head-to-tail reaction is rather low. Oligonucleotide primers are not only ligated to cDNA but also to RNA and short fragments. Homopolymer tailing often causes problems in efficiency and requires large amounts of RNA. Furthermore, unspecific products accumulate as a result of the use of primers comprising homopolymer sequences in the PCR.

These problems should be avoided by the use of DNA ligase. T4 DNA ligase cannot ligate single-stranded substrates. Therefore, we used a set of primers with a degenerated overhang of 4 bases. This overhang would anneal to the 3' end of the cDNA. These primers provide a double-stranded region for the subsequent ligation reaction. The products of this ligation assay can readily be used for PCR amplification and subsequently for cloning and sequencing of specific mRNA species.

MATERIALS AND METHODS

Human chondrocyte RNA for the cloning of collagen type X was prepared as described recently.⁽⁸⁾ Total RNA of human granulocytes and RNA of fetal chicken sternal chondrocytes were provided by H. Burkhardt and U. Dietz, respectively (both at the Max-Planck Society in Erlangen). Oligonucleotides were synthesized on a GeneAssembler Plus (Pharmacia), and synthesis products were purified after detritylation by centrifugation through G50 spin columns.⁽⁹⁾

Reverse transcription, 5'-end phosphorylation of primers, and ligation assays were performed according to modified protocols of the manufacturers. For first-strand cDNA synthesis, specific primers (30 pmoles) complementary to the mRNA were used in each 30 μ l reverse transcription assay containing 3 μ g of total RNA, 30 nmoles of each dNTP, 40 units of RNasin, and 50 units of Moloney murine leukemia virus Moloney murine leukemia virus (M-MuLV) reverse transcriptase (U.S. Biochemical) in the supplied buffer.

First-strand cDNA was treated with

RNAse A and T1, phenolized, purified over a Sephadex G50 column, and ethanol precipitated. The 5' end of the primer 5consdw was phosphorylated with T4 polynucleotide kinase (Boehringer Mannheim). Primer 5consdw and its complementary primer 5consupNNNN (Table 1) were annealed in equimolar concentrations (10 mM) in bidistilled water. Ligation of the primer pair (10 μ M) to the first-strand cDNA was carried out in a volume of 10 μ l at 16°C overnight. The 5' consensus primers (5consdw and 5consupNNNN) comprise the sequence of the universal sequencing primer to avoid synthesizing a third PCR primer for subsequent amplification.

Amplification was performed in two steps. In an initial round of PCR the downstream primer used for reverse transcription was employed, together with the universal primer (NEB) (50 pmoles each), in a 100- μ l PCR assay under conditions published previously.⁽¹⁰⁾ The following temperature profile was used for 30 cycles: 94°C for 45 sec, 42°C for 30 sec, and 72°C for 90 sec.

An aliquot of the PCR product (10–20%) was analyzed by gel electrophoresis, and amplified products were recovered from the gel by adhesion to glassmilk (protocol of the manufacturer, Bio101, CA). Amplification products were subjected to a second round of nested PCR amplification, using the universal primer and a nested gene-specific downstream primer. The nested down-

stream primer comprised an *Eco*RI site at its 5' end (or, in the case of AcIIIn47, it contained a *Nru*I site). The 5' consensus primer pair, which was used in the ligation to the first-strand cDNA, contributed a *Bam*HI site. After digestion with *Eco*RI (or *Nru*I) and *Bam*HI, the resulting fragment was cloned into a suitable vector and several clones were analyzed by sequencing.

RESULTS AND DISCUSSION

The strategy of the DNA ligase-dependent amplification (DLDA) cloning of unknown 5' ends of cDNAs is presented schematically in Figure 1.

After first-strand cDNA synthesis with a specific downstream primer, a complementary primer pair (5consupNNNN and 5consdw) is ligated to the single-stranded cDNA. This simple approach used the specificity of DNA ligase for double-stranded DNA; therefore, participation of the remaining RNA fragments in the ligation reaction is avoided, as well as a concatamerization of the first-strand cDNA fragment. Because only one of the primers (5consdw) is phosphorylated, the number of possible side products in the ligation reaction is further restricted. The subsequent PCR is started with two defined primers, one of them being gene specific. The universal primer is also used in a nested PCR (together with a second gene-specific primer) and marks the end of the reverse transcript of the initial cDNA synthesis.

PCR of human collagen type-X cDNA ligated to this primer pair yielded a 361-bp-long fragment, which could be cut by a restriction enzyme (*Xho*I), and liberated a fragment of the expected size (116 bp). The product of a subsequent nested PCR was cloned into pUC19 and analyzed by sequencing (Fig. 2). An aliquot of the initial PCR was blotted and used for Southern hybridization, which gave a strong signal of the expected size. The sequence revealed 100% homology to the available genomic sequence⁽¹¹⁾ and extended over 339 bp. Similar to other protocols involving reverse transcription, our method is also sensitive to stem-loop structures or long stretches of homopolymers. This might explain why, without any chemical denaturing of the RNA before reverse transcription, our fragment terminated 81 bp short of the transcription initiation site. The transcription initiation site of the COL10A1 gene was determined later by independent means (primer extension and S1 analysis) using primers upstream of the endpoint determined by DNA ligase-dependent amplification (DLDA).

We applied our protocol to the mRNA of fetal chicken sternal RNA to clone the transcription start site of the anchorin CII (AcII) gene. Several resulting clones harbored a sequence that was nonhomologous to the genomic sequence and extended homologously only over 21 bp in a very GC-rich region (Fig. 3). We suggest that this deviation from the genomic sequence might be attributable to cloning artifacts. A mispairing of the second AcII-specific primer, as a result of its high GC content of 67%, might have caused the amplification and cloning of a wrong, although similar, transcript. A Southern hybridization of the products of the first PCR showed a positive signal. Alternatively, a strong secondary structure of the RNA in this GC-rich region might have terminated the reverse transcription of the authentic RNA template, thus favoring the amplification of a different RNA with limited homology. (AcII is just one member of a large family of similar proteins, only some of which are sequenced.) Previous chemical denaturation of the RNA might help to overcome this problem, and a verification by independent methods should be applied.

Recently, the human cytidine deaminase (CDD) cDNA was partially cloned in our laboratory.⁽¹²⁾ We therefore tried to

TABLE 1 Sequence of the Primers Used in Different Applications of DNA Ligase-dependent Amplification of Specific mRNAs

Universal primer: 5'-GTAAACGACGGCCAGT-3'		
5consdw: 5'-GGATCCACTGGCCGTCGTTTAC-3'		
5consupNNNN: 5'-GTAAACGACGGCCAGTGGATCCNNNN-3'		
Collagen type X-specific primers		complementary to nucleotides ^a
ERO47:	5'-CTGGTTTCCCTACAGCT-3'	421–405
ERO50:	5'-GCGAATTCAACCCTGGCTCTCCTTGGAG-3'	384–364
AcII-specific primers		
AcIIIdw:	5'-CGGAATTCAACACTTTCTCATTGGTTCTCTG-3'	424–401
AcIIIn47:	5'-GGAGTCGAATCCGAGTCGCGAC-3'	68–47
CDD-specific primers		
cds1dw:	5'-TAGGGGCGAGTAGGCTGACTGCTTG-3'	95–72
cds2dw:	5'-TCAGAAATTCAGGGTGCAGGCAGGACGCTTC-3'	29–6
confA:	5'-ATGATTTGATGAAGGTCC-3'	
confS:	5'-CAATCTTCTCACTGAGC-3'	

^aNumbering of nucleotides of collagen-specific primers,⁽¹¹⁾ anchorin-specific primers,⁽¹⁴⁾ and CDD-specific primers.⁽¹²⁾

FIGURE 1 Principle of DLDA of specific mRNAs. Total RNA is used for reverse transcription with a sequence-specific downstream primer. A pair of complementary primers is annealed to this single-stranded cDNA. One of these primers (5consupNNNN) is 4 nucleotides longer than the other (5consdw). This overhang allows the ligation to the single-stranded cDNA by T4 DNA ligase. Subsequently, a first round of amplification is performed with universal primer (the sequence of universal primer is comprised in 5consupNNNN) as upstream primer and the primer used for first-strand synthesis as downstream primer. The product of this first round is used for a nested PCR. The product of this second round of PCR is cloned and sequences.

obtain information about the translation start and the 5'-untranslated region

by this new technique. Initial amplification yielded a product of ~380 bp, and

FIGURE 2 Analysis of DLDA products resulting from human collagen type-X mRNA. After reverse transcription with the downstream primer ER047 and subsequent ligation and amplification, we obtained a weak band of ~350 bp (I; lane 1, human chondrocyte RNA), which was strongly positive in a subsequent hybridization analysis (II, lane 1, human chondrocyte RNA). Lane 2 in both panels shows fibroblast RNA (cells that do not cleave collagen type-X mRNA) as negative control. The PCR product could be cleaved with *XhoI* as predicted from the genomic sequence (III). A combination of λ /hindIII and ϕ X/HaeIII was used as marker.

nested PCR resulted in a 258-bp *Bam*HI fragment. This PCR product was sequenced (Fig. 4). In this case, no hybridization probe upstream of the primer cds2dw was available; therefore, we confirmed the proposed sequence by a different method. Human T cell RNA, which also contains small amounts of CDD,⁽¹²⁾ was reverse transcribed with cds1dw. Aliquots of this assay were subjected to PCR amplification (Fig. 4, panel 1). For the nested PCR we tested different combinations of primers (cds2dw and confA, as well as cd2dw and confS), as well as all CDD-specific primers alone. Only the primer combination cd2dw and confA yielded a band of the expected length. To determine whether we were close to the transcription initiation site, conventional S1 or primer extension analysis was performed.

DLDA is a useful alternative to other established methods for the cloning of 5' ends of mRNAs. As with other methods, its efficiency depends on the quality of the RNA, its ability to form secondary structure, and the efficiency of cDNA synthesis. In the following text, we will compare some intrinsic advantages and disadvantages of methods available for cloning 5' ends of RNAs. In comparison to the use of T4 RNA ligase, which ligates

FIGURE 3 Analysis of DLDA testing chicken AcII mRNA. A reverse transcription of chicken RNA was ligated to the primer pair (Table 1), and amplification using AcII_{dw} and universal primer was performed. Although no band was visible (1) after this initial round of amplification, hybridization analysis indicated a product of ~600 bp (2). When using DNA fragments fractionated for this size for nested PCR, a smaller band of ~240 bp appeared (3). We used ϕ X/*Hae*III as marker. Sequence comparison of the product of a DNA ligase-dependent amplification of chicken AcII (top) and the sequence published by Genge et al.⁽¹⁴⁾ for this region (bottom) indicate that only part of this amplification product is homologous to the genomic sequence. Lowercase letters indicate those parts of the sequence that were comprised in the primers used for amplification. G and C residues are bold to indicate the high CG content of the published sequence.

single-stranded RNA and DNA templates, T4 DNA ligase is specific for double-stranded DNA molecules. This should

make the ligation reaction specific for reverse transcripts and avoid the incorporation of partially degraded RNAs. A ma-

jor point to consider is the fact that no circularization of nucleic acid chains can occur using T4 ligase, whereas circularization is observed quite frequently using T4 RNA ligase.⁽¹³⁾ We therefore expect that the amount of RNA can be reduced with this method, although a careful purification of cDNAs before ligation is still advisable. Although different from T4 RNA ligase, T4 DNA ligase shows no preference to ligate short nucleotide chains. We expect this method to reduce the chance of getting (mixed) concatamers and their recombination products interfering with this subsequent PCR.

Priming conditions cannot be set very stringently for the cloning method that uses terminal deoxytransferase (RACE), because priming conditions have to allow for the annealing of the PCR primer to a short homonucleotide tail. Compared with homopolymer tailing, the method described here promises higher specificity. This not only reduces the background but also minimizes the amount of RNA needed for determination of specific 5' ends.

ACKNOWLEDGMENTS

We thank Drs. Burkhardt and Dietz for providing us with human granulocyte RNA and chicken sternal chondrocyte RNA. This work was supported by the Bundesministerium Forschung und Technologie (BMFT) (Grant O1VM8819/9) to W.M.B.

REFERENCES

- Gubler, U. 1987. Second-strand cDNA synthesis: mRNA fragments as primers. *Methods Enzymol.* **152**: 330–335.
- Gubler, U. 1987. Second-strand cDNA synthesis: Classical method. *Methods Enzymol.* **152**: 325–329.
- Feinberg, A.P. and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specificity activity. *Anal. Biochem.* **132**: 6–13.
- Deininger, P.L. 1987. Full length cDNA clones: Vector primed cDNA synthesis. *Methods Enzymol.* **152**: 371–389.
- Frohman, M.A., M.K. Dush, and G.R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts. Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci.* **85**: 8998–9002.
- Troutt, A.B., M.G. McHeyzer-Williams, B.

FIGURE 4 Analysis of DLDA results of human CDD mRNA. Although the initial PCR yielded only weak bands above 200 bp (1), we used this size fraction for nested PCR. After the cloning and sequencing of the nested PCR product, we verified the identity by using cds2dw and a primer of the newly defined region (confA and confS) in a reverse transcriptase-PCR assay. confS is complementary to a part of the human CDD 5' untranslated region (UTR). RNA was reverse transcribed with cd1dw, and the resulting cDNA was used for PCR with different combinations of primers: cds2dw alone (lane 1), confA alone (lane 2), cds2dw together with confA (lane 3), cds2dw alone (lane 4), confS alone (lane 5), and cds2dw together with confS (lane 6). The sequence of the 5' UTR and the start of the coding sequence of human CDD-specific mRNA is shown below. The sequence that is complementary to the CDD-specific primer used in the nested PCR reaction (cds2dw) is underlined. The previously known sequence from the downstream primer and the *Bam*HI site is in lowercase letters. In panel 1 we used ϕ X/*Hae*III as marker; in panel 2, a 100-bp ladder.

- Pulendran, and G.J. Nossal. 1992. Ligation-anchored PCR: A simple amplification technique with single-sided specificity. *Proc. Natl. Acad. Sci.* **89**: 9823–9825.
7. Hofmann, M.A. and D.A. Brian. 1991. A PCR-enhanced method for determining the 5' end sequence of mRNAs. *PCR Methods Applic.* **1**: 43–45.
 8. Reichenberger, E., T. Aigner, K. von der Mark, H. Stöss, and W. Bertling. 1991. In situ hybridization studies on the expression of type X collagen in fetal human cartilage. *Dev. Biol.* **148**: 562–572.
 9. Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. 1987. *Current protocols in molecular biology*. Green Publishing/Wiley Interscience, New York.
 10. Krawetz, S.A., R.T. Pon, and G.H. Dixon. 1989. Increased efficiency of the Taq polymerase catalyzed polymerase chain reaction. *Nucleic Acids Res.* **17**: 819.
 11. Reichenberger, E., F. Beier, P. LuValle, B.R. Olsen, K. von der Mark, and W.M. Bertling. 1992. Genomic organization and full-length cDNA sequence of human collagen X. *FEBS Lett.* **311**: 305–310.
 12. Kühn, K., W.M. Bertling, and F. Emmrich. 1993. Cloning of a functional cDNA for human cytidine deaminase (cdd) and its use as a marker of monocyte/macrophage differentiation. *Biochem. Biophys. Res. Commun.* **190**: 1–7.
 13. Brennan, C.A., A.E. Manthey, and R.I. Gumpert. 1983. Using T4 RNA ligase with DNA substrates. *Methods Enzymol.* **100**: 38–52.
 14. Genge, B.R., X. Cao, L.N. Wu, W.R. Buzzi, R.W. Showman, A.L. Arsenault, Y. Ishikawa, and R.E. Wuthier. 1992. Establishment of the primary structure of the major lipid-dependent Ca^{2+} binding proteins of chicken growth plate cartilage matrix vesicles: Identity with anchorin CII (annexin V) and annexin II. *J. Bone Miner. Res.* **7**: 807–819.

Received May 24, 1993; accepted in revised form September 8, 1993.

Determination of 5' ends of specific mRNAs by DNA ligase-dependent amplification.

W M Bertling, F Beier and E Reichenberger

Genome Res. 1993 3: 95-99

Access the most recent version at doi:[10.1101/gr.3.2.95](https://doi.org/10.1101/gr.3.2.95)

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>