SHORT TECHNICAL REPORTS

Manuscripts published in the Short Technical Reports section are shorter and less comprehensive in scope than full Research Reports.

Optimized Rapid Amplification of cDNA Ends (RACE) for Mapping Bacterial mRNA Transcripts

BioTechniques 28:448-456 (March 2000)

ABSTRACT

A simple, efficient and sensitive RACE-based procedure was developed for the determination of unknown 5′ regions from bacterial cDNA. A number of critical modifications were made to the standard RACE method, including the optimization of the RNA extraction, reverse transcription and PCR conditions. This procedure was used to accurately determine the site of transcript initiation and structure of the promoter region of the Helicobacter pylori *aspartate carbamoyltransferase gene* (pyrB*). The technique avoids many of the difficulties associated with established bacterial transcript mapping protocols and can be performed in two days starting with less than 1 mg of total RNA. The modifications described here have significant potential for the identification of transcript start sites of bacterial genes and non-polyadenylated eukaryotic RNA.*

INTRODUCTION

Rapid amplification of cDNA ends (RACE) is a powerful PCR-based technique for the rapid characterization of the 5′ end of mRNAs and the start of transcription (7). RACE avoids many of the difficulties associated with other transcript mapping procedures such as primer extension and the RNase protection assay (6,7). These non-RACE techniques require large amounts of bacterial RNA (>100 μg), use potentially hazardous radioactive probes and require gel electrophoresis of the resulting products, which often results in low specificity and high background. The basic RACE protocol involves the tailing of the 3′ end of the first-strand cDNA with a single nucleotide using terminal transferase (3). The specific transcript is then PCR amplified by using a tail-specific primer and one or more target-specific primers.

Although numerous RACE modifications and improvements have been developed (7), none has been reported for the mapping of bacterial mRNA transcripts. Bacteria provide a number of unique challenges to the RACE procedure. First, bacterial mRNA transcripts do not contain polyadenosine 3′ tails and so require the use of total RNA. Second, bacterial mRNA transcripts are highly unstable with half-lives often under 2 min (6). Third, since bacterial genes do not contain introns, contaminating genomic DNA can result in the generation of false RACE products that are indistinguishable from the correct product. Fourth, bacterial mRNA transcripts do not contain the 5′-attached 7-meGppp cap found on eukaryotic transcripts, making it difficult to select for only full-length mRNA transcripts (4). Finally, many bacteria are slow growing and contain high levels of RNases, making the isolation of poorly expressed transcripts extremely difficult.

We describe the development of a RACE-based procedure optimized for the mapping of bacterial mRNA transcripts. Total RNA is extracted by a simple hot phenol-based method with residual genomic DNA removed by DNase digestion. RNA secondary structure-induced transcript truncations are minimized by the performance of a thermal cycled first-strand cDNA synthesis. The T4 RNA ligase is used to introduce a 3′-modified anchor oligonucleotide to the cDNA transcript, a step that enables PCR amplification of the desired transcript. The resulting PCR product is directly sequenced and allows the precise transcript initiation site to be mapped. Furthermore, all procedural steps were optimized to ensure simplicity and minimize the use of expensive enzymes and chemicals. This procedure was used to precisely map the transcription start site of the *H. pylori pyrB* gene encoding aspartate carbamoyltransferase.

MATERIALS AND METHODS

H. pylori strain RU1 was obtained from the University of New South Wales (UNSW) Culture Collection, and grown and maintained as previously described (1). RNA was isolated from late-exponential phase bacterial cultures. Cells were pelleted and resuspended in 50 μL of 150 mM NaCl. The concentrated cells were added to 1400 μL of preheated (70°C) SSE lysis buffer (700 μL of 2% sodium SDS, 300 mM sodium acetate, 10 mM EDTA and 700 μL of phenol) in a 1.5 mL microcentrifuge tube. Total RNA was extracted twice with 1 volume of hot (55°C) phenol/chloroform/isoamyl alcohol (50:49:1) as previously described (9). The RNA was precipitated by the addition of 1 volume of isopropanol and incubated on ice for 30 min. The RNA was pelleted by centrifugation at 12 000× *g* for 10 min before being resuspended in 100 μL of R-buffer (1× SuperScript™ II buffer, Life Technologies, Gaithersburg, MD, USA), and 10 mM dithiothreitol. Chromosomal DNA was removed by addition of 10 U of RNase-free DNase (Sigma, St. Louis, MO, USA) in 100 μL of R-buffer and incubation at 37°C for 15 min. The DNA-free RNA was phenol/chloroform extracted, precipitated by the addition of 1 vol isopropanol and resuspended in 50 μL of R-buffer.

Reverse transcription (RT) reactions were performed using primer BB4 (5′-ATATCCCTATGCACCGGG-3′) and 200 U SuperScript II reverse transcriptase (Life Technologies). The RT reactions were performed in a 50 μL volume containing 5 μL of 2 mM dNTP, 10 pmol of the BB4 primer and between 0.5 and 1 μg of total RNA. The primer was annealed stepwise

(70°C for 2 min, 65°C for 1 min, 60°C for 1 min, 55°C for 1 min, 50°C for 1 min and 45°C for 1 min) before the SUPERSCRIPT II was added. The RT reaction was performed at 42°C for 30 min, followed by 5 cycles at 50°C for 1 min, 53°C for 1 min and 56°C for 1 min. Two parallel reactions were performed with the reverse transcriptase omitted from one reaction as a negative control.

The RNA was removed by alkaline hydrolysis cleavage with the addition of 1 µL of 0.5 M EDTA followed by 12.5 µL of 0.2 M NaOH before incubation at 68°C for 5 min. The reactions were neutralized by adding 12.5 µL of 1 M Tris-HCl (pH 7.4) and precipitated by the addition of 20 µg of glycogen (Roche Molecular Biochemicals, Mannheim, Germany), 5 µL of 3 M sodium acetate and 60 µL of isopropanol. Both samples were resuspended in 20 µL of 10:1 TE (10 mM Tris-HCl, pH 7.4; 1 mM EDTA, pH 8.0).

The T4 RNA ligase was used to anchor a 5′-phosphorylated, 3′-end cordecypin-blocked (3′ phosphate) anchor oligonucleotide DT88 (5′-GAAGAGAAGGTGGAAATGGCGTTTTGG-3′) to the single-stranded cDNA (11). Two parallel ligations were performed overnight at room temperature with each containing: 1.3 µL of 10× RNA ligase buffer (Roche Molecular Biochemicals), 0.4 µM DT88, 10 µL of either the cDNA or RT control sample and 3 U of T4 RNA ligase. As negative controls, a second set of ligation reactions was performed, omitting the T4 RNA ligase. The resulting ligation-reaction mixtures were used directly without purification in the subsequent PCRs.

The anchor-ligated cDNA was amplified using a combination of hemi-nested, touchdown and hot-started PCR approaches to increase the sensitivity and product yield (2). Four separate first-round PCRs were performed on the RT/ligase reaction and controls, each containing: 3 µL of 10× PCR buffer (Biotechology International, Perth, Australia); 3 µL of 25 mM $MgCl_2$; 3 µL of 2 mM of each deoxynucleotide triphosphate; 1 µL of corresponding RNA ligation reaction, 10 pmol each of primers; and water to 25 µL. The first ligation-anchored PCR (LA-PCR) was performed using 10 pmol of both the anchor-specific primer DT89 (5′-CCAAAACGCCATTTCCACCTTCTCTTC-3′) and the transcript-specific primer BB5-1 (5′-CCTGCGTTAATCAAAGGGC-3′), which is internal to the 3′ end RT primer BB4. The touchdown PCR was performed using an initial 90°C for a 2 min denaturation step, and by 10 cycles at 95°C for 10 s and 70°C for 1 min, reducing one degree per cycle; this step was followed by 15 cycles at 95°C for 10 s and 60°C for 1 min. The PCRs were hot-started by adding 5 µL of water containing 1 U of *Taq* DNA polymerase (Biotechnology International) at the initial 90°C step. The second hemi-nested PCR was performed under the same conditions using primer DT89 and the BB5-1 internal primer, together with 10 pmol of BB6-2 (5′-GTTCGCATAAACGCTCGC-3′) and 1 µL of the first-round PCR. Amplification products were ethanol precipitated with the addition of 3 µL of 3 M sodium acetate and 33 µL of 85% EtOH before electrophoresis on a 3% agarose gel.

The BigDye™ Terminator kit (PE Biosystems, Foster City, CA, USA) was used to directly sequence the RACE PCR products using 5 pmol of primer BB6-2. Sequencing reactions were performed according to the provided instructions and were purified by n-butanol precipitation (8). Sequencing products were separated on a model 377 sequencer (PE Biosystems, Foster City, CA, USA) at the UNSW Automated Sequencing Facility and analyzed using the INHERIT™ package (PE Biosystems).

RESULTS AND DISCUSSION

The start of transcription for *H. pylori pyrB*-mRNA was mapped 25 bp upstream of the ATG start codon using

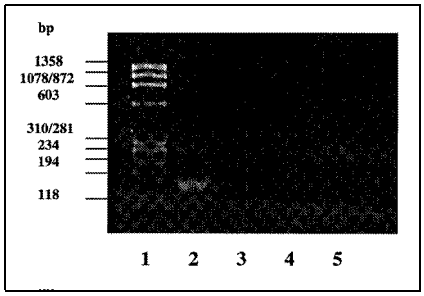

Figure 1. Isolation of 5′ end of *pyrB* transcript from *H. pylori* using RACE. Lane 1: ϕX174 RF DNA *Hae*III cut markers (100 ng). Lane 2: Amplification of anchored first-strand cDNA using anchor-specific primer DT89 and BB6-2; Lane 3: PCR with DT89/BB6-2 primers with reverse transcriptase omitted from cDNA synthesis step; Lane 4: PCR with DT89/BB6-2 primers after T4 RNA ligase is omitted from ligation step.

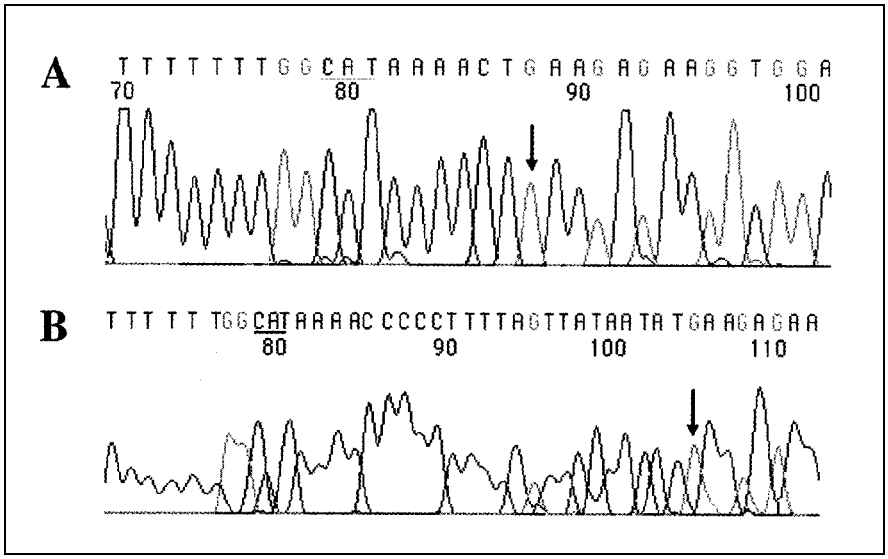

Figure 2. Transcript mapping of the *H. pylori pyrB* gene using direct automated sequencing of the LA-PCR product using primer BB6-2. (A) The false transcription start of *H. pylori pyrB* identified using the original non-cycled RT-PCR conditions. (B) The correct mRNA start site identified after using thermal cycled RT-PCR conditions. Sequences are shown in reverse orientation with the translation initiation codon underlined. Arrow indicates the initial 5′ nucleotide of the cordecypin-blocked anchor oligonucleotide DT88.

the described optimized bacterial RACE procedure. A putative bacterial transcriptional promoter was identified immediately upstream of the transcriptional start site mapped with this technique. This tentative σ^{70}-like promoter matched five out of six bases with the *Escherichia coli* promoter consensus sequences of the proposed -10 region, and four out of six bases for *E. coli* σ^{70}-like promoter -35 consensus sequences (6). In addition, only one PCR product was obtained (Figure 1), which strongly suggests this gene is transcribed from a single initiation site, at least under the culture conditions studied.

The development of the procedure involved the optimization and simplification of many standard RACE procedures. First, a simple and rapid hot phenol-based method was used to extract total RNA. This method provides consistently high yields of RNase-free RNA from a wide range of bacteria, including species that contain high levels of RNases. Also, the resulting RNA proved free of all RNases and so eliminated the need for expensive enzymatic RNase inhibitors. Second, the remaining contaminating genomic DNA was removed by a quick DNase digestion step. Together with the inclusion of an RT-negative control, this step ensured the resulting RACE PCR products were derived from mRNA transcripts alone.

Third, optimization of the RT conditions proved particularly critical. Any initial attempts to map the *H. pylori pyrB* transcript initiation site using standard RT conditions at 42°C resulted in an apparent false termination, with the transcript start site mapped within the putative ribosome binding site (Figure 2). This region contains a 9 bp purine sequence that is likely to form a stable secondary structure, which results in the premature termination of the first-strand cDNA synthesis (5). To minimize problems with false RT termination, we used a thermal cycled RT procedure where the reaction was incubated at 42°C for 30 min before being subjected to 5 cycles of 50°C for 1 min, 53°C for 1 min and 56°C for 1 min. This combination of conditions enables the use of high temperatures to destabilize RNA secondary structure and, at the same time, minimizes the thermal inactivation of the SUPERSCRIPT II RT enzyme. Using these optimized conditions, the *H. pylori pyrB* transcription start was repeatedly mapped at a position 25 bp upstream of the ATG translation-start (Figure 2), at a position more favorable for transcript initiation. In addition, the use of alkaline hydrolysis to remove RNA avoided the use of expensive RNase cocktails (6).

Fourth, the use of the ligation-anchored approach (10) provided a significant benefit because transcript initiation could be determined with less than

1 μg of total RNA, as compared to the significantly larger quantities (>100 μg) required when using other bacterial transcript mapping procedures (6). The substitution of random hexamers for the transcript-specific primers at the RT stage allowed all mRNA transcripts to be oligonucleotide anchored, which in turn allowed multiple transcripts to be studied from one RNA sample. This is decidedly advantageous for the study of mRNA transcription in environmental or slow-growing bacterial species where it is often difficult to obtain large quantities of RNA.

Fifth, optimization of the PCR conditions proved critical. The use of touchdown PCR conditions allowed the successful use of PCR primers with large differences in optimal annealing conditions. Additionally, the use of a hot-start process, in which the *Taq* DNA polymerase is added at the initial denaturation temperature, minimized false product generation. Finally, the combination of both approaches, together with a second hemi-nested PCR, enabled the direct sequencing of the resulting PCR product (Figure 2). This sequencing allowed the identification of the specific site of transcript initiation of *H. pylori pyrB* with no need for PCR product cloning or radioactive reagents.

In summary, our study demonstrates a novel application of the RACE procedure for the precise mapping of transcription initiation in bacteria. The study identified the transcript initiation site of the *H. pylori pyrB* gene and allowed for the putative identification of upstream promoter sequences in this bacterium. The procedure has also been used to successfully map the mRNA transcriptional initiation of several genes from the microcystin synthetase gene cluster identified in *Microcystis aeruginosa* PCC7806 (unpublished data). Finally, the technique may prove useful in the mapping of non-polyadenylated eukaryotic transcripts.

REFERENCES

1. **Burns, B.P., G.L. Mendz and S.L. Hazell.** 1997. *In situ* properties of *Helicobacter pylori* aspartate carbamoyltransferase. Arch. Biochem. Biophys. *347*:119-125.
2. **D'Aquila, R.T., L.J. Bechtel, J.A. Videler, J.J. Eron, P. Gorczyca and J.C. Kaplan.** 1991. Maximizing sensitivity and specificity of PCR by pre-amplification heating. Nucleic Acids Res. *19*:3749.
3. **Frohman, M.A., M.K. Dush and G. Martin.** 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA *85*:8998-9002.
4. **Grovsky, M.A. and X. Liu.** 1993. Mapping the 5′ and 3′ ends of *Tetrahymena thermophila* mRNAs using RNA ligase mediated amplification of cDNA ends (RML-RACE). Nucleic Acids Res. *21*:4954-4960.
5. **Kotewicz, M.L., C.M. Sampson, J.M. D'alessio and G.F. Gerard.** 1988. Isolation of cloned Moloney leukemia virus reverse transcriptase lacking ribonuclease H activity. Nu-

cleic Acids Res. *16*:265-277.
6. **Sambrook, J., E.F. Fritsch and T. Maniatis.** 1989. Molecular Cloning: A Laboratory Manual, 2nd ed. CSH Laboratory Press, Cold Spring Harbor, NY.
7. **Schaefer, B.C.** 1995. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal. Biochem. *227*:255-273.
8. **Tillett, D. and B.A. Neilan.** 1999. Simple n-butanol purification of dye terminator sequencing reactions. BioTechniques *26*:606-610.
9. **Tillett, D. and B.A. Neilan.** 1999. Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. J. Phycol. (In press.)
10. **Troutt, A.B., M.G. McHeyzer-Williams, B. Pulendran and G.J.V. Nossal.** 1992. Ligation-anchored PCR: A simple amplification technique with single-sided specificity. Proc. Natl. Acad. Sci. USA *89*:9823-9825.

This work was supported by grants from the National Health and Medical Research Council and the Australian Research Council. We are grateful to M. Cairns for the kind gift of the cordecypin modified oligonucleotide. *Address correspondence to Dr. Brett Neilan, School of Microbiology and Immunology, University of New South Wales, Sydney 2052, Australia. Internet: b.neilan@unsw.edu.au*

Received 3 May 1999; accepted 25 October 1999.

Daniel Tillett, Brendan P. Burns and Brett A. Neilan
University of New South Wales Sydney, Australia

Monitoring and Purification of Proteins Using Bovine Papillomavirus E2 Epitope Tags

BioTechniques 28:456-462 (March 2000)

ABSTRACT

We describe here the use of two newly mapped bovine papillomavirus type 1 (BPV-1) E2 protein epitopes as tags. We constructed several vector plasmids for overexpression as well as for moderate expression of single- or double-tagged proteins in either Escherichia coli *or eukaryotic cells. The new tags were fused to several proteins, and the activity of the tagged proteins was tested in different assays. The tags were shown not to interfere with the function of these proteins in vivo and in vitro. Interaction of the monoclonal antibodies 3F12 and 1E2 with their respective epitopes was specific and had high affinity in a variety of conditions. We have demonstrated that the 3F12 antibody-epitope interaction tolerates high salt concentrations up to 2 M. This permits immunoprecipitation and immunopurification of the tagged proteins in high-salt buffers and reduction of the nonspecific binding of the contaminating proteins. We also provide a protocol for DNA binding and DNase I footprinting assays using the tagged, resin-bound DNA-binding proteins. The BPV-1 E2-derived tags can be recommended as useful tools for detection and purification of proteins.*

INTRODUCTION

Epitope tagging is a recombinant DNA technique by which a protein is made immunoreactive to a preexisting antibody. This technique simplifies detection, characterization, purification and in vivo localization of proteins and has become a standard method of molecular genetics (3). However, some tags are not useful in certain applications due to high background binding. In some cases, affinity purification and immunoprecipitation of a tagged protein is problematic due to co-precipitation of contaminating proteins.

Here, we describe the use of two recently mapped bovine papillomavirus type 1 (BPV-1) E2 protein epitopes as tags. We constructed several vector plasmids for over-expression as well as for moderate-level expression of either single- or double-tagged proteins in *Escherichia coli* and eukaryotic cells. The new tags were fused to functionally different proteins: a bacterial transcriptional activator, XylS, that agregates and becomes nonfunctional at high levels of expression, several mutants of the tumor suppressor protein p53 for overexpression in *E. coli* and rat glyceraldehyde-3-phosphate dehydrogenase (GAPDH) for expression in eukaryotic cells. Hereby, we report the use of BPV-1 E2-derived tags and respective monoclonal antibodies for the identification of proteins on Western blots, in the immunofluorescence staining of cells and DNA band-shift assay. In addition, we describe an application of the resin-bound tagged protein in DNA binding and DNase I footprinting assays.

MATERIALS AND METHODS

Plasmid Constructions

For the construction of pBR-3F12 and pBR-1E2, we inserted the coding sequence of XylS with the N-terminally fused influenza virus hemagglutinin (HA) epitope (2) from the pET11c-based parent plasmid pETSN117 (4) between the *Nhe*I and *Bam*HI sites of pBR322. The subcloned fragment contained in addition to a ribosome binding site, an extra *Nde*I site and a start codon preceding the tag sequence. The resultant plasmid pBRSN 117 contains *Xba*I and *Bam*HI sites for cloning of a recombinant coding sequence (4). Then, the coding sequence of the HA epitope between *Nde*I and *Xba*I sites was replaced with double-stranded synthetic oligonucleotides encoding peptides GVSSTSSDFRDR and TTGHYSVRD, recognized by anti-BPV-1 E2 monoclonal antibodies 3F12 and 1E2, respectively (7).

For the construction of pBR-NC, we amplified *xylS* sequence by PCR using the 3′-end primer containing a coding sequence for the peptide TSSDFRDR. The peptide recognized by 3F12 Mab was fused in frame to the C-terminus of XylS and flanked by *Kpn*I and *Bam*HI sites. The resultant PCR fragment was cloned into pBR-1E2 generating the expression plasmid with cloning sites *Xba*I and *Kpn*I.

pET-3F12 was generated by cloning the *Bam*HI/*Nde*I fragment from pBR-3F12-*xylS* into the corresponding sites in pET-11c. Then, the *XylS* gene was removed by cleavage with *Xba*I and *Bam*HI and replaced with coding sequences for different mutant p53 proteins bearing *Xba*I and *Bgl*II sites at the ends.

For the construction of pCG-3F12, double-stranded synthetic oligonu-