Mapping of Transcription Start Sites by Direct Sequencing of SMART\textsuperscript{TM} RACE Products

BioTechniques 34:482-486 (March 2003)

The rapid accumulation of transcription data from gene array-based experiments opens the opportunity for wide-scale analysis of promoter sequences. For such studies, precise mapping of the transcription start sites of a large number of genes is a prerequisite. Conventional methods for the determination of transcription start site are quite arduous, which makes them inappropriate for high-throughput application. Primer extension and RNase protection methods (8) are time and labor consuming, requiring significant amounts of mRNA for analysis, and extensive optimization is often necessary for success. In addition, exact information on the exon/intron structure of the 5'-region of the gene is necessary to interpret the results. An alternative approach for mapping transcripts, based on cloning and sequencing of the products of 5'-rapid amplification of cDNA ends (RACE) (3,8), requires minimal amounts of mRNA and does not demand prior knowledge of gene structure. However, approximate localization of the RNA start site calls for sequencing of multiple 5'-RACE clones, thus increasing the cost of procedure.

We developed a fast and reliable method for mapping the transcription start sites, based on direct sequencing of the SMART\textsuperscript{TM} (switching mechanism at 5' end of RNA template) (10) RACE products. As traditional 5'-RACE analysis, our method provides the sequence of the 5' end of the transcript and therefore does not require prior knowledge of exon/intron structure; rather it provides this knowledge by itself. However, our approach is free of the problem of inadequate statistical representation that usually plagues the conventional analysis of several individual cloned RACE products. Direct sequencing of the PCR products allows the analysis of the whole population of molecules produced by 5'-RACE at once; hence, this method greatly improves the reliability of transcription start site mapping as compared to the traditional approach, in addition to saving time and effort usually allocated for cloning.

The SMART technology is based on the terminal transferase activity of reverse transcriptase, which adds several deoxyctydine nucleotides once the enzyme reaches the 5' end of the mRNA template. The oligo(dC) stretch is then annealed with the 3' end of the adaptor, and the sequence complementary to the adaptor is added at the end of cDNA after the d(C) stretch. Thereby, PCR amplification of the cDNA using adaptor-homologous primers in combination with gene-specific primers will yield the complete 5'-end sequence.

Total RNA was prepared from Drosophila melanogaster testes with the TRIZOL\textsuperscript{R} reagent (Invitrogen, Carlsbad, CA, USA), and poly(A)\textsuperscript{+} RNA was purified using the Poly(A)Tract\textsuperscript{R} system (Promega, Madison, WI, USA). SMART cDNA was prepared using the SMART III adaptor (BD Biosciences Clontech, Palo Alto, CA, USA; available at http://www.clontech.com/techinfo/manuals/PDF/PT3529-1.pdf). Two primers homologous to the SMART III adaptor sequence were synthesized, SmartIII-1 (5'-AAGCAGTGGTATC-AACGCAGAGTG-3') and SmartIII-2 (5'-GTATCAAACGCAGATGGCC-ATT-3'). For each gene, two genespecific primers were synthesized, gene-specific primer (GSP) 1 compatible with the SmartIII-1 and GSP2 compatible with the SmartIII-2. Both GSP1 and GSP2 were chosen in noncoding strand of cDNA, GSP1 being located downstream of GSP2. A general outline of the RACE procedure is available at http://www.clontech.com/techinfo/manuals/PDF/PT3269-1.pdf. Primer design was performed using the Oligo 4.0 software (National Biosciences, Plymouth, MN, USA).

Testis cDNA was diluted in water 1:50, and 2 \muL of the diluted sample served as a template for 15 \muL reaction in the presence of GSP1 and SmartIII-1 oligonucleotides (0.4 \muM each). Advantage\textsuperscript{TM} 2 enzyme mixture and reagents (BD Biosciences Clontech) were used. Amplification was performed in a model 9600 thermal cycler (Applied Biosystems, Foster City, CA, USA) as follows: 30 s at 95\degree C and then 33 cycles of 20 s at 95\degree C, 90 s at 58\degree C, and 2 min at 68\degree C, followed by 5 min at 68\degree C. Products of this primary amplification reaction were diluted 1:200 in water, and 2 \muL diluted samples were used as a template for 50 \muL reaction in the presence of GSP2 and SmartIII-2 oligonucleotides. Reaction conditions were as described above, with the exception of the number of cycles (25 instead of 33).

Seven-microliter aliquots of the second reaction were analyzed by electrophoresis in 1% agarose/TBE gel stained with ethidium bromide. PCR products were purified from the rest of reaction mixture on the QIAquick\textsuperscript{R} spin columns (Qiagen, Valencia, CA, USA) and subjected to sequencing in the ABI PRISM\textsuperscript{R} 3100 automated sequencer (Applied Biosystems) using the dye terminator chemistry and GSP2 as a sequencing primer. Sequences and chromatograms were analyzed with the Sequencher 3.0 software (GeneCodes, Ann Arbor, MI, USA).

It appeared that the quantity and purity of SMART RACE products are often sufficient for direct sequencing after purification on QIAquick spin columns (Figure 1). In all cases when the quality of sequence was acceptable, the position of the transcription start site was readily detected as a point of abrupt loss of sequence identity between the RACE product and the 5'-genomic sequence. In many instances, the gene-specific sequence was immediately appended by an oligo(C) stretch followed by the SMART adaptor sequence (Figure 2a). Sometimes, the quality of sequence sharply

Figure 1. SMART RACE products obtained by two consecutive amplifications of cDNA for Drosophila genes des (1), CG7848 (2), boule (3), aly (4), CG15710 (5), twin (6), CG3492 (7), can (8), and cyclinB (9), M, 1-kb ladder (Invitrogen). PCR products from lanes 1–5 and 8 were successfully sequenced.
deteriorated near the end of RACE product. We suggest that this phenomenon results from the presence of more than one strong transcription start sites. In this situation, a significant number of the RACE molecules has the junction with SMART adaptor at or near the downstream transcription start site, while other RACE products are extended further upstream. As a result, the sequence at the end of total RACE product is a superposition of the SMART adaptor sequence and genomic sequence (Figure 2b). In such cases, only the position of the downstream transcription start site could be unequivocally determined.

To evaluate the reliability of the procedure, we first analyzed four genes with known transcription start sites, including adh (2), RpS17 (5), bam (6), and TopoII (9). Our procedure identified start sites at positions +2 (adh), +3 (RpS17), +7 (bam), and +11 (TopoII). Then, we successfully amplified and sequenced RACE products for 14 genes with previously uncharacterized promoters. Conserved promoter elements were found near the mapped transcription start sites. For eight genes, these included a sequence similar to the Drosophila initiator (Inr) consensus T-C-\(\text{A}^{+1}\)-G/T-T/C (1,7). In Drosophila, about two-thirds of the promoters possess discernible Inr motifs (4); this is similar to the proportion observed in our experiment. With the assumption that the A found in Inr in the eight Inr-containing promoters is a first nucleotide in mRNA, the transcription start sites were mapped at positions +2 in two cases, +3 in three cases, and once at each of positions +4, +5, and +6.

Ideally, start sites would map at the position +1. Instead, we observed that mapping at positions +2 and +3 was the most frequent. Perhaps the cap structure interferes with proper recognition of the very 5'-terminal mRNA nucleotide(s) by reverse transcriptase. Therefore, analysis of SMART RACE products does not, as a rule, reveal an exact position of transcription start site. However, the method allows for the localization of transcription start within several nucleotides. This level of accuracy is usually sufficient for further promoter analysis.

Our procedure combines all the advantages offered by the 5'-RACE method with the precision of mapping
similar to classical approaches such as primer extension. The minimal requirement for mRNA template permits mapping of multiple promoters even when the mRNA source is limited. The mapping procedure includes sequencing of the 5′ end of the transcript, thus providing the information on the exon/intron structure of the gene of interest. This feature is very useful for analysis of sequenced eukaryotic genomes because computerized gene prediction often yields erroneous gene structure, especially in the 5′-regions. PCR amplification, purification of RACE products, and sequencing are easily adapted to a standard microplate format for high-throughput application. Thus, we report here an approach for large-scale transcript mapping that, in combination with ample gene expression data available from the microarray experiments, may lead to rapid identification of regulatory sequences in eukaryotic promoters.

REFERENCES


Address correspondence to Dr. Dmitry I. Nurminsky, Tufts University School of Medicine, Department of Anatomy & Cell Biology, 136 Harrison Avenue, Boston, MA 02111, USA. e-mail: dmitry.nurminsky@tufts.edu

Received 26 July 2002; accepted 21 November 2002.

Inna Tabansky and Dmitry I. Nurminsky
Tufts University School of Medicine
Boston, MA, USA

Use of the Formazan Dye Zincon for Staining Proteins in Polyacrylamide Gels

BioTechniques 34:486-488 (March 2003)

The formazan dye 2-carboxy-2′-hydroxy-5′-sulfonamidobenzene (zincon) has been used previously as a colorimetric indicator of calcium in serum (1) and, histologically, for the differentiation of zinc and copper in brain tissue (2). Zincon has been shown to inhibit insulin binding to its receptor (3) and to attenuate the activity of ATPase (4) and other enzymes (5), presumably because of the chelating properties of the dye. Zincon forms blue complexes with zinc at alkaline pH and with copper at neutral or acidic pH; it forms other colored complexes characteristic of the specific transition metal ion with which it coordinates (6,7).

The finding that zincon nonspecifically binds proteins at acidic pH prompted further investigation into the use of the chelating agent as a polyacrylamide gel stain. Preliminary studies showed that in this capacity, zincon staining of SDS protein derivatives was (i) more rapid than Coomassie Brilliant Blue® (CBB) staining, (ii) reversible, and (iii) only slightly less sensitive than the CBB dyes. Moreover, (iv) the sensitivity of zincon staining could be increased by coordination of the dye with metals such as copper, cobalt, and nickel.

A saturated zincon solution was prepared by dissolving 200 mg monosodium salt (Sigma, St. Louis, MO, USA) in 100 mL absolute ethanol. This solution was filtered through Whatman® chromatography paper and was stable at room temperature for several months. SDS-PAGE was performed as described by Laemmli (8) and later modified by Shainoff et al. (9). Representative protein samples included human plasma, purified human ceruloplasmin, and protein standards from New England Biolabs (Beverly, MA, USA) and Promega (Madison, WI, USA). Concentrations of the protein standards were provided by the manufacturer. The concentration of purified ceruloplasmin was determined spectrophotometrically as described by