

CapSelect: A highly sensitive method for 5′ CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs

Wolfgang M. Schmidt[2] and Manfred W. Mueller[1,2,*]

[1]VBC-GENOMICS Research GmbH, PO Box 207, A-1091 Vienna, Austria and [2]Vienna Biocenter, Institute of Microbiology and Genetics, Dr Bohr-Gasse 9, A-1030 Vienna, Austria

Received July 19, 1999; Revised and Accepted September 17, 1999

ABSTRACT

Here we present CapSelect as a novel experimental approach for the selective enrichment of full-length cDNAs in PCR-mediated analysis of mRNA sequences. The method combines the 5′-CAP-dependent addition of specifically three to four non-templated dCMP residues to the 3′-end of full-length cDNAs by reverse transcriptases in the presence of manganese and the controlled ribonucleotide tailing of cDNA ends by terminal deoxynucleotidyl transferase using rATP. By virtue of the generated terminal sequence motif (5′-$dC_{3-4}rA_{3-4}$), full-length cDNAs are selectively anchored to a double-stranded DNA adapter (with a dT_{3-4} dG_3 3′-overhang) by *T4* DNA ligase. The technique described is highly efficient, discriminates premature termination products and enriches full-length cDNAs.

Generation of full-length cDNAs from an mRNA template is a major challenge in biotechnology research. It constitutes the ultimate prerequisite in the construction of cDNA libraries, expression profiling and as a new area in DNA chip technology. Utilising current techniques, the 5′-ends of genes tend to be under-represented in cDNA populations, especially if a poly(dT) primer is used during first-strand synthesis and if the starting material is limited. So far, several approaches have been developed to overcome this problem and focus on employing the cap structure of the mRNAs as a tag for an intact and complete RNA template (1–6). For example, specific selection of capped eukaryotic mRNA can be accomplished by affinity retention using a cap-binding protein coupled to a solid support (1) or through the chemical introduction of a biotin group into the diol residue of the cap structure (2,3). Alternatively, an oligo-capping technique is used, whereby a synthetic oligonucleotide replaces the cap structure and is ligated to the mRNA (5,6).

The major drawback of these techniques is, however, that they are costly and time consuming by requiring multiple enzymatic steps and/or are not pronounced sensitive, since they are adjusted to a rather large amount of starting poly(A)$^+$ mRNA. Furthermore, these methods focus primarily on a selective enrichment of full-length mRNA and do not eliminate premature termination products generated in cDNA synthesis. The CapFinder™ method (7–9) offers a solution to this central problem and also allows the analysis of limited starting material. The technique relies on the terminal transferase and the template switching activity of the reverse transcriptases. The addition of cytosine residues to the 3′-end of full-length cDNAs allows the reverse transcriptase to generate a specific anchor sequence complementary to a template switching oligonucleotide. However, it should be stressed that the nucleotide addition activity of reverse transcriptases pausing at the 7-methylguanosine cap structure of the mRNA template has not been studied in detail. Moreover, structural requirements of the mRNA template and/or the cDNA terminus for template switching activity can lead to inefficient amplification and therefore to a biased representation within the amplified cDNA pool. Another problem associated with this technique, background amplification produced by a single primer, was recently shown to be overcome by step-out PCR utilising the PCR-suppression effect (10).

Here, we present CapSelect (Fig. 1) as a novel technique for the selective enrichment of full-length cDNAs. The method was successfully applied for the specific enrichment of complete mRNA 5′-ends in PCR-mediated sequence analysis, cDNA library construction and direct transcription start site mapping.

In a first step we investigated in more detail the specificity of the cap-dependent terminal transferase activity intrinsic to the M-MLV reverse transcriptase (RNase H deficient, SuperScriptII™ and Expand-RT™) and tested both enzymes with an *in vitro* transcribed RNAs with capped (7-methylG-5′-ppp-5′-G) termini. Under the buffer conditions recommended by the manufacturers, the enzymes preferentially added one additional nucleotide residue to the 3′ terminus of the cDNA. For instance, using SuperScriptII™ (Gibco BRL Life Technologies), only a minor fraction of the reaction products contained more than one additional cytosine (0, 5%; +1, 86%; +2, 6%; +3, 3%) (Fig. 2, lane 1). Increasing the $MgCl_2$ concentration in the reaction buffer to 6 mM as recommended by the CapFinder™ protocol (Clontech Laboratories) (7), the transferase activity was more pronounced (+1, 62%; +2, 10%; +3, 17%; +4, 11%), but still insufficient for generating a 5′-cap-specific sequence tag of three to five cytosine residues. On the contrary, when using $MnCl_2$ as an additional divalent cation (3 mM $MgCl_2$ and

*To whom correspondence should be addressed at: VBC-GENOMICS Research GmbH, PO Box 207, A-1091 Vienna, Austria. Tel: +43 1 31336 1478; Fax: +43 1 31336 1463; Email: mueller@vbc-genomics.com

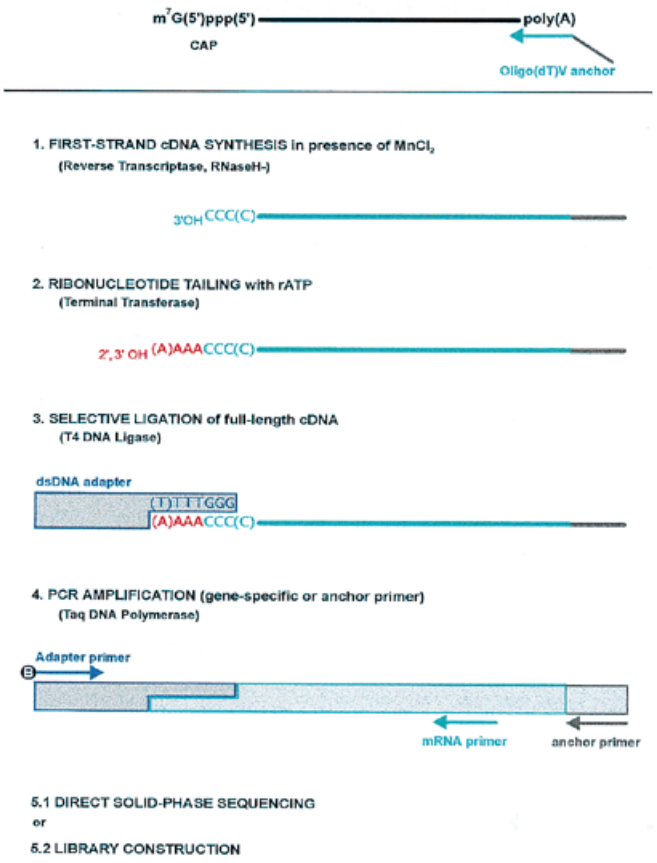

Figure 1. Principle of the CapSelect technique. (1) First-strand cDNA synthesis of poly(A)$^+$ mRNA is performed with an oligo(dT) anchor primer (or with an mRNA-specific downstream primer within the known region of the gene). In the presence of $MnCl_2$ the reverse transcriptase adds three to four dCMP residues to the full-length cDNA when reaching the CAP structure on the mRNA template. (2) The purified cDNA fraction is modified by the addition of three to four rAMP residues to the 3′-end (CRTC), allowing (3) the selective ligation of (full-length) cDNAs terminating with a 5′-CCCAAA(A) motif to an adapter with a complementary (T)TTTGGG 3′ overhang. (4) PCR amplification using an adapter-specific primer with an mRNA-specific (nested) primer yields a unique PCR product suitable for direct solid-phase sequencing (5.1). The amplification using the adapter-specific primer together with the anchor primer generates the double-stranded cDNA pool for the construction of a full-length cDNA enriched library (5.2).

2 mM $MnCl_2$) and bovine serum albumin (BSA) as stabilising agent (0.1 µg/µl) the reverse transcriptase revealed an extraordinarily high cap-dependent transferase activity: typically, the enzyme added preferentially three to four cytosine residues in the presence of 5′ capped mRNA templates (+1, 8%; +2, 2%; +3, 36%; +4, 47%; +5, 7%) (Fig. 2, lane 3). Using mRNA templates with 5′-OH termini (+1, 81%; +2, 6%; +3, 8%; +4, 4%; +5, 1%) (Fig. 2, lane 6) or 5′-phosphate termini (+1, 80%; +2, 13%; +3, 6%; +4, 1%) (not shown), the terminal transferase activity was less pronounced and not cytosine specific. Thus, dCMP tailing by the reverse transcriptase is 10-fold more efficient with capped RNA templates compared to 5′-OH-containing RNA.

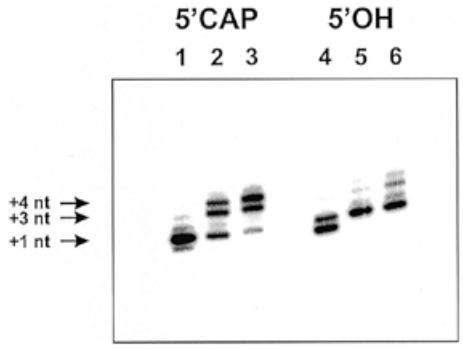

Figure 2. CAP-specific tailing activity of reverse transcriptase in the presence of manganese. SuperScript reverse transcriptase was tested in a primer extension reaction containing an *in vitro* transcribed RNA template with either 5′-capped (lanes 1–3) or -hydroxyl termini (lanes 4–6) using a radioactively labelled oligonucleotide primer. The reactions were performed as described in the text either in a standard reaction buffer (lanes 1 and 4) or supplemented with 1 mM (lanes 2 and 5) or 2 mM $MgCl_2$ (lanes 3 and 6). Reaction products were analysed on a denaturing 10% polyacrylamide/8 M urea gel and quantified using a PhosphorImager.

The CAP-dependent addition of three to five non-templated cytosine residues can be performed also after completion of cDNA synthesis using the standard reaction buffer in a post-incubation in the presence of manganese (+1, 10%; +2, 5%; +3, 57%; +4, 28%).

The following technical strategy (Fig. 1) was designed to exploit this 5′-CAP-dependent transferase activity for the selective enrichment of full-length cDNA molecules. We applied controlled ribonucleotide tailing of cDNA ends (CRTC) (11) to tag the crude cDNA pool with a short ribonucleotide tail in order to facilitate the selective ligation to an adapter DNA. Ribo-tailing is highly efficient and self-limited (11) to the addition of three or four AMP residues (rate of the reaction, 96%; +2, 3%, +3, 64%, +4, 25%, +5, 4%). The double-stranded DNA adapter used in the subsequent ligation was constructed with either TTTGGG or TTTTGGG 3′ overhangs (as a 5:2 molar mixture). Under the reaction conditions described, i.e. the addition of DMSO to a standard DNA ligation buffer, 93% of the ribo-tailed full-length cDNA were selectively anchored to the adapter whereas only 10% of non-compatible molecules were ligated. Thus, the combination of the CAP-dependent addition of three to four cytosine residues in the presence of manganese and CRTC favours the enrichment of complete cDNA copies by elimination of premature termination products, mediated by sequence-specific ligation.

The following outlined protocol was successfully applied to detect the complete 5′-region of the human GAPDH mRNA (12) starting from poly(A)$^+$ RNA and to directly map the transcription start site (Fig. 3). To provide evidence for the reliability of the method the protocol was also applied to other human genes including middle- to low-abundance mRNAs (Fig. 4).

First-strand cDNA synthesis and purification

In a first step, poly(A)$^+$ mRNA (100 ng) prepared from the cell line MOLT-3 (ATCC CRL-1552) using the Oligotex™

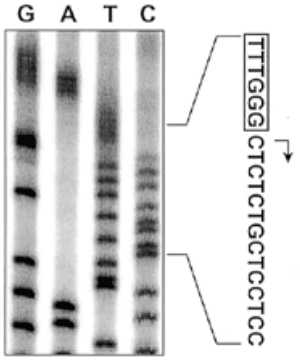

Figure 3. Mapping of the transcription start of the human GAPDH gene using CapSelect. Poly(A)$^+$ mRNA was reverse transcribed with a poly(dT) primer in the presence of manganese, ribo-tailed with rAMP, anchored to a specific adapter (adapter sequence is boxed) and amplified as described. A direct sequencing reaction using a nested primer allowed the exact determination of the GAPDH transcription start site (marked with an arrow) at the nucleotide level avoiding an additional cloning step.

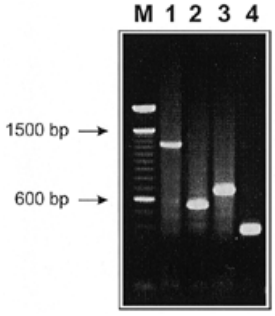

Figure 4. Detection of human mRNAs using CapSelect. The protocol described in the text was applied to detect the complete 5′ region of different human mRNAs: the highly abundant mRNAs of cytoplasmatic β-actin (lane 1) and GAPDH (lane 2) and the low-abundance mRNAs of the eIF-4E gene (lane 3) and of the gene for the RNA polymerase II 13 kDa subunit (lane 4). PCR products were separated on a 1.5% agarose gel. M, molecular weight marker (100 bp ladder).

poly(A)$^+$ mRNA isolation system (Qiagen) was subjected to reverse transcription either with an oligo(dT) anchor primer [SPS-T18V, 5′ GCTAATACGTACGATCGGTCGACAAGT-TTTTTTTTTTTTTTTTV (0.75 µM)] or an oligonucleotide specific for the GAPDH gene [GAPDH-3/M, 5′ GTTGTCAT-ACTTCTCATGGTTCAC (0.25 µM); position 3798–3821 according to Ercolani *et al.* (12)]. The reaction mixture (20 µl) containing 50 mM Tris–HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 5 mM DTT, dNTPs (Pharmacia) at 1 mM each, 0.1 mg/ml BSA, 20 U RNase inhibitor (Roche Diagnostics) and 200 U SuperScriptII™ M-MLV RT RNaseH (Gibco BRL Life Technologies) was incubated for 1 h at 42°C. After the addition of 0.4 µl fresh $MnCl_2$ solution (100 mM) the reaction was further incubated for 15 min at 42°C. Purification of the cDNA was performed using the High Pure (Roche Diagnostics) protocol except that the cDNA fraction was recovered with 40 µl of 10 mM Tris–HCl pH 8.3.

Ribonucleotide tailing of cDNAs

Tailing of the purified cDNA was performed in a 30 µl reaction containing 200 mM potassium cacodylate, 25 mM Tris–HCl pH 6.6, 0.25 mg/ml BSA (TdT-buffer, Roche Diagnostics), 150 µM ATP, 2.5 mM $CoCl_2$, 10 U terminal transferase (Roche Diagnostics) and 20 µl of the purified cDNA sample. The reaction mixture was incubated for 30 min at 37°C and stopped by precipitation with ethanol/sodium acetate (0.3 M).

Anchoring to the adapter

The ribo-tailed cDNA was co-incubated with 3 pmol of the adapter (100 ng) in the presence of 1 U *T4* DNA ligase (Roche Diagnostics) for 5 min at 37°C, 3 min at 24°C and 16 h at 16°C in 66 mM Tris–HCl pH 7.5, 5 mM $MgCl_2$, 1 mM DTT, 1 mM ATP, 7.5% (v/v) DMSO in a 10 µl reaction. The adapters were prepared from hybridisation of a 58mer (5′-GGTCCGAT-GAAGAACGCAGCGCTCGAGTCGACACATGCGCGGC-CGCAGATCTTTTGGG) to a 50mer (and 49mer, mixed at a molar ratio of 5:2 to generate TTTGGG and TTTTGGG 3′ overhangs) with 5′ phosphate and 3′ primary amino groups (5′-p-(A)GATCTC-CCGGGGCATGTGTCGACTCGAGCGCTGCGTTCTTCA-TCGGACC-NH_2) and purified from a non-denaturing 15% polyacrylamide gel. In addition to the (T)TTTGGG 3′ overhangs the DNA adapter contains several restriction sites (*Bgl*II, *Sal*I, *Not*I, *Xho*I, *Xma*I) for convenient cloning of PCR products and library construction.

PCR amplification

An aliquot of the ligation mixture (1 µl) was directly subjected to PCR using an adapter-specific biotinylated primer (STI-2, 5′-biotin-GGTCCGATGAAGAACGCAGC) and a GAPDH-specific minus-strand primer [GAPDH-2/M, 5′-CCTGGAAG-ATGGTGATGGGATTTC; position 3388–3411 according to Ercolani *et al.* (12)]. Other primers used: cytoplasmatic β-actin primer, 5′-AGAAGCATTTGCGGTGGACG (EMBL accession no. M10277); eIF-4E gene primer, 5′-TGTCTGCGT-GGGACTGATAACC (EMBL accession no. M15353); RNA polymerase II subunit gene primer, 5′-CTGATGAGGTCGGT-GATGGC (EMBL accession no. X82385). Reactions (20 s at 94°C, 20 s at 64°C, 40 s at 72°C, for 35 cycles; pre-incubations at 72°C for 30 s and 3 min at 94°C) were performed using the Expand™ High Fidelity PCR system (Roche Diagnostics) under standard conditions recommended by the manufacturer.

Direct dideoxy-sequencing

Purification of PCR products and dideoxy-sequencing using magnetic beads as solid support were performed as described (13). Sequencing primer used: GAPDH 1/M, 5′-GGCAACAA-TATCCACTTTACCAGAG; position 3158–3182 according to Ercolani *et al.* (12).

In this study, we established an alternative strategy in amplification-driven approaches for mRNA sequence analysis and construction of full-length enriched cDNA libraries. The recommended procedure offers several improvements when compared to other protocols developed for the purpose of selectively cloning full-length cDNA. First, in contrast to the oligo-capping and biotinylated cap trapper techniques our protocol does not involve enzymatic or chemical modification of the RNA prior to reverse transcription and therefore minimises the potential risk of RNA degradation and, furthermore,

requires only moderate amounts of starting material. This might be especially useful if the starting material is limited (e.g. as from biopsies) or of poor integrity. Second, when compared to the CapFinder™ protocol (Clontech Laboratories) our novel approach offers higher sensitivity due to the extraordinarily efficient combination of CAP-dependent tailing, CRTC (11) and sequence-specific ligation to a DNA adapter.

ACKNOWLEDGEMENTS

We thank Dr R. Mader and his laboratory for supplying us with RNAs used in this study. This work was supported in part by Roche Diagnostics Boehringer Mannheim GmbH (Penzberg, Germany).

REFERENCES

1. Edery,I., Chu,L.L., Sonenberg,N. and Pelletier,J. (1995) *Mol. Cell. Biol.*, **15**, 3363–3371.
2. Carninci,P., Kvam,C., Kitamura,A., Ohsumi,T., Okazaki,Y., Itoh,M., Kamiya,M., Shibata,K., Sasaki,N., Izawa,M., Muramatsu,M., Hayashizaki,Y. and Schneider,C. (1996) *Genomics*, **37**, 327–336.
3. Carninci,P., Westover,A., Nishiyama,Y., Ohsumi,T., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M., Schneider,C. and Hayashizaki,Y. (1997) *DNA Res.*, **4**, 61–66.
4. Fromont-Racine,M., Bertrand,E., Pictet,R. and Grange,T. (1993) *Nucleic Acids Res.*, **21**, 1683–1684.
5. Maruyama,K. and Sugano,S. (1994) *Gene*, **138**, 171–174.
6. Suzuki,Y., Yoshimoto-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S. (1997) *Gene*, **200**, 149–156.
7. Clontech Laboratories (1996) *Clontechniques*, **11**, 1.
8. Maleszka,R. and Stange,G. (1997) *Gene*, **202**, 39–43.
9. Chenchik,A., Zhu,Y.Y., Diatchenko,L., Li,R., Hill,J. and Siebert,P.D. (1998) In Siebert,P. and Larrick,J. (eds), *Gene Cloning and Analysis by RT–PCR*. Biotechniques Books, Natick, MA, pp. 305–319.
10. Matz,M., Shagin,D., Bogdanova,E., Britanova,O., Lukyanov,S., Diatchenko,L. and Chenchik,A. (1999) *Nucleic Acids Res.*, **27**, 1558–1560.
11. Schmidt,W.M. and Mueller,M.W. (1996) *Nucleic Acids Res.*, **24**, 1789–1791.
12. Ercolani,L., Florence,B., Denaro,M. and Alexander,M. (1988) *J. Biol. Chem.*, **263**, 15335–15341.
13. Hultmann,T., Stahl,S., Hornes,E. and Uhlen,M. (1989) *Nucleic Acids Res.*, **17**, 4937–4939.