Nucleic Acids Research, 1993, Vol. 21, No. 15 3597–3598

Determination of messenger RNA 5′-ends by reverse transcription of the cap structure

J.Hirzmann, D.Luo, J.Hahnen and G.Hobom*
Institut für Mikrobiologie und Molekularbiologie, der Justus-Liebig-Universität Giessen, Frankfurter Straße 107, D-6300 Giessen, Germany

Received April 13, 1993; Revised and Accepted June 17, 1993

A complete description of a eukaryotic transcriptional unit requires the determination of its messenger RNA 5′-end(s) via cDNA sequencing. This also serves as a first step in the analysis of promoter function and may be required to differentiate between two or more overlapping potential promoter regions, e.g. in polyoma viral DNA. As a standard method used for this purpose primer extension along isolated mRNA (1) may not always be possible, since it will require a rather large amount of RNA in high quality. Also, two or more adjacent start sites used alternatively within one promoter region could not reliably be identified as true mRNA 5′-ends in this way. The problem of RNA quantity is overcome by an alternative method, the 5′-RACE technique (2, 3), in its standard version. However, PCR amplification following dA-tailing requires rather low annealing temperatures, which may favour unspecific binding of the oligo-dT primer to internal A-rich sequences, which in general are more common in untranslated regions of mRNAs than G-rich segments. Therefore, we have changed the procedure to dG homopolymeric tailing (4), and have further modified this technique in such a way that the cap-G nucleotide will become reverse transcribed and PCR amplified to show up as a regular G residue in the final sequence ladders obtained from individual RACE clones (see Figure 1). Sequence ladders with an uncoded 5′-terminal cap-G residue give direct proof of their origination from complete messenger RNA molecules, and may also be observed in different locations as resulting from alternative start sites, against a genomic DNA reference. The method described will also tolerate slight degrees of 5′-exonuclease degradation, since any clones resulting from shorter RNA molecules (if not excluded in an insert sizing step) will not carry an extra G residue at their 5′-ends and be discarded, together with full-length clones without a cap-G.

Total RNAs were isolated from the parasitic nematode *Litomosoides carinii* (4), or from chicken cells infected by budgerigar fledgling disease virus (BFDV (5)) according to the method of Chirgwin *et al.* (6), and poly(A)$^+$−RNA was obtained using Hybond-mAP paper (Amersham). 1 μg of poly(A)$^+$−RNA in 9.65 μl of water was heated to 60°C for 3 min, cooled on ice, added to 4 μl of 5×RT buffer (1×RT buffer is 40 mM Tris−HCl, pH 8.3, 5 mM $MgCl_2$, 40 mM KCl, 2 mM DTE), 3 μl dNTPs (10 mM each), 0.25 μl (10 units) of RNasin (Promega), 2.5 μl of *Xho*I-$(dT)_{17}$ oligonucleotide primer (300 pmol, 0.5 μg/μl), and 0.6 μl (16 units) of avian myeloblastosis virus (AMV) reverse transcriptase (Boehringer). The reaction was continued at 42°C for 2 hours, and excess *Xho*I-$(dT)_{17}$ primer and substrates were removed by glass powder purification followed by ethanol precipitation. For dG tailing, an aliquot of the reaction products (equivalent to 50 ng mRNA) in 38 μl water was added to 10 μl of 5×tailing buffer (BRL), 1 μl of 10 mM dGTP, and 1 μl (15 units) of terminal desoxynucleotide transferase (BRL), the mixture was incubated for 20 min at 37°C, and inactivated at 75°C for 10 min. 10 μl of the dG-tailed reverse transcription product was used for PCR amplification. A first round of PCR amplification was performed using an *Eco*RI−*Bam*HI-$(dC)_{12}$- oligonucleotide primer (5′TT-CTAGAATTCGGATC$_{12}$) and a distal specific oligonucleotide (complementary to a sequence in the C-terminal region of the

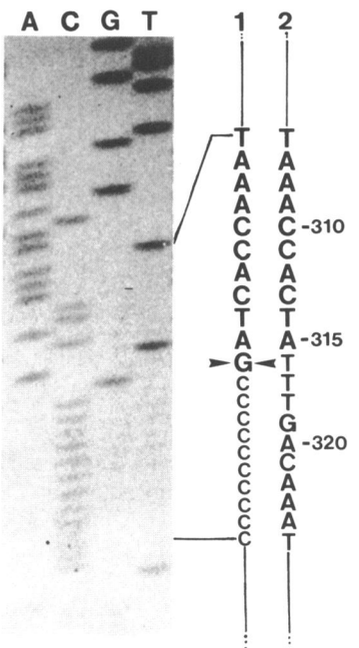

Figure 1. Sequencing reaction of a G-tailed 5′-end cDNA clone from the *L.carinii* gp22 gene showing an extra 5′-G indicated by an arrow adjacent to the primer complementary C_{12} sequence (lane 1). In lane 2 the corresponding genomic sequence (see ref. 4) is shown for comparison.

* To whom correspondence should be addressed

gene). A second round of (nested) PCR reaction was performed following a 1:10 dilution step, and a proximal specific oligonucleotide was used this time (complementary to a more proximal region of the gene), again together with the 5′-terminal dC_{12}-primer. A single major DNA band resulted after PCR amplification and was cloned into standard vectors using unique restriction sites attached to the flanking primer sequences.

In the *L.carinii* gp22 gene analysis 24 out of 50 roughly full-size clones had an identical 5′-end at A^{-315} relative to the ATG start codon, a single one extended further to A^{-320} while other molecules scattered over various distal positions between −309 and −160 in their insert DNA sequences. Eleven of the 24 full-length clones and also one with its 5′-end at position −309 all carried an extra G in their 5′-end sequences which is not present in the sequence of the genomic clone. From these results we conclude that pre-mRNA transcription initiates predominantly at position −315 which is in excellent agreement with a TATAA box sequence at position −343/−339 (4). Similarly, in an analysis of the BFDV late mRNA transcription initiated at promoter p_{L1}-8 cap-G clones have been observed out of 10 cDNA clones starting at position 264, and 9 cap-G out of 12 clones starting at 266, with another 8 sequence ladders beginning in different distal positions, but none of them carrying an extra 5′-residue. From these results we conclude that BFDV P_{L1} transcription starts in approximately 1:1 ratio either at A^{264} or at A^{266} with two alternative, true mRNA 5′-ends.

For confirmation of these *in vivo* results RNA transcripts have been synthesized *in vitro* with or without initiation by cap dinucleotide 7me-GpppA (Boehringer) (7) using as template DNA a T7 promoter sequence fused to the *L.carinii* gp22 cDNA sequence at the A^{-315} start position. Following the procedure as described above none of the full-length cDNA clones obtained from non-capped RNA, but 10 out of 20 full length cDNA clones from the capped *in vitro* RNA transcript contained an extra 5′-G residue in their sequence patterns.

From these results we conclude that the 5′-cap-G can to some extent be reverse transcribed into a 3′ terminal C residue, which will give rise to an un-coded G residue in the complementary strand sequence patterns. Model building using the Quanta molecular modelling package and CHARMM energy calculation programs (8) for energy minimization appeared to allow the formation of hydrogen bonds between a template cap-G and a terminal dC residue in the product DNA strand, and make this reaction plausible.

ACKNOWLEDGEMENTS

We would like to thank N.Pohl for expert technical assistance and wish to recognize helpful suggestions by Dr M.Strauss and S.Bähring, MDC, Berlin. This work has been supported by the DFG via SFB 272.

REFERENCES

1. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Frohmann,M.A., Dash,M.K. and Martin,G.R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998−9002.
3. Loh,E.Y., Elliot,J.F., Cwirla,S., Lanier,L.L. and Davis,P.M. (1989) *Science* **243**, 217−220.
4. Christ,H., Hirzmann,J., Conraths,F., Zahner,H., Stirm,S. and Hobom,G. (1992) *Gene* **121**, 219−226.
5. Stoll,R., Luo,D., Kouwenhoven,B., Hobom,G. and Müller,H. (1993) *J. Virol.* **74**, 229−237.
6. Chirgwin,J.K., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1987) *Biochemistry* **18**, 5294−5299.
7. Nielsen,D.A. and Shapiro,D.J. (1986) *Nucleic Acids Res.* **14**, 5936.
8. Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) *J. Comp. Chem.* **4**, 187−217.