Transcription of a cloned rainbow trout protamine gene is accurately initiated following transfection into HeLa cells but the majority of the transcripts fail to polyadenylate at the correct site.

Niall O. Dillon, Vernon M. Spencer and Peter H. W. Butterworth

Department of Biochemistry, University College London, Gower Street, London WC1E 6BT, UK

Received 29 October 1985; Accepted 22 November 1985

ABSTRACT

The expression of a cloned trout protamine gene transfected into mammalian cells in culture has been studied. This small intronless gene has a consensus TATA-box, a classical AATAAA sequence and the cap and polyadenylation sites are separated by only 228 base pairs (Gregory et al., ref 10). When 1kb of cloned trout genomic DNA containing this sequence was introduced into HeLa cells, S1-mapping showed that transcripts of the protamine gene were accurately initiated at the in vivo cap site but were not polyadenylated at the authentic 3'-site. Replacement of the 3'-end of the protamine transcription unit with a fragment of SV40 containing the small-t intron and early polyadenylation site resulted in only a modest increase in transcript levels over the wild-type gene in HeLa cells. However, transcripts of a fusion gene in which the 5'-end of the protamine gene was replaced by the SV40 early promoter were present at extremely low levels in transfected COS cells. The data are discussed in the context of the involvement of RNA processing events in the stabilisation of eukaryotic gene transcripts.

INTRODUCTION

The expression of eukaryotic protein coding genes is known to involve a number of sequential steps whose ultimate aim is the production of a mature cytoplasmic mRNA. Initiation of transcription by RNA polymerase II occurs under the control of a region whose principal elements are located in the first 120 bp 5' to the cap site although more distant elements both upstream and downstream from the cap site can also play a role [1-4]. The large primary transcript is then subject to a series of processing events which lead to the formation of the mature (usually polyadenylated) 3'-end and the precise excision of intervening sequences. Finally, the mature message is transported to the cytoplasm for translation.

Available evidence suggests that regulation of the expression of protein coding genes can occur at each of these steps. The initial decision to express a particular gene appears to be largely at the transcriptional level but the rate of transcript turnover in the nucleus is now known to be...
important in the regulation of many genes \([5,6]\). The use of different splice and poly-A addition sites can also result in a single transcription unit giving rise to several different mRNAs thus increasing genetic diversity \([7,8]\).

The use of recombinant DNA techniques to isolate and characterise specific eukaryotic genes has allowed considerable progress towards an understanding of these events. In vitro mutagenesis of cloned genes and the introduction of these mutated genes into living cells makes it possible to define some of the sequences involved in the transcription and maturation of eukaryotic mRNA. We are interested in the rainbow trout protamine genes as a model regulated gene expression system. The members of this gene family code for a series of small highly basic proteins whose function is to bind to the DNA during sperm maturation, displacing the histones and condensing the DNA for packaging in the small sperm head [reviewed by Dixon in ref. 9]. We have previously characterised a cloned trout protamine gene TPG3 [Gregory et al., ref. 10]. The gene is extremely short (228bp from cap to poly-A site) and lacks intervening sequences. We describe here the introduction of TPG3 into HeLa cells using an SV40 early region transient expression vector. Transcription was found to initiate accurately; however, the majority of the transcripts were not polyadenylated at the natural site but instead continued downstream into adjacent plasmid sequences.

**MATERIALS AND METHODS.**

**Materials**

Restriction enzymes and T4 DNA ligase were obtained from New England Biolabs. Bacterial alkaline phosphatase and T4 DNA polymerase were from Bethesda Research Laboratories. SI nuclease was from Sigma. Radiochemicals were obtained from New England Nuclear Limited.

**Fig. 1.** Strategy for the construction of pETP in which the protamine gene TPG-3 is incorporated into an SV40 early region expression vector and (inset) a diagram of the fusion gene pETP-R in which the 3'-end of the protamine transcription unit has been replaced with an SV40 fragment containing the small-t splice and early polyadenylation site. Details of the protocols used are provided in Methods. Together with the identity of each construct, the size in base pairs is given. \(\ldots\) bacterial plasmid sequences; \(\ldots\) SV40 sequences; \(\ldots\) TPG-3 5'- and 3'-flanking sequences; \(\ldots\) cap to polyadenylation site of TPG-3.

Abbreviations used: o, origin of replication; pA, polyadenylation site; Apr and Tet, ampicillin and tetracycline resistance genes derived from the plasmid vector (pX3, see ref. 11); E, SV40 early region; L, SV40 late region; A, AvalI; B, BamHI; C, ClaI; H2, HpaII; H3, HindIII; K, KpnI; R, EcoRI; S, SalI; Taq, TaqI; Th, TthIII.
Plasmid Constructions.

Unless stated otherwise, non-compatible overhanging ends were made blunt by end-repair with the Klenow fragment of E.coli DNA polymerase (5'-overhangs) or by removal of 3'-overhangs with T4 DNA polymerase. These were then subjected to blunt-end ligation or, alternatively, cloning was carried out by the attachment of synthetic linkers.

pXSVE2: this SV40 early region expression vector was constructed by attaching an EcoR1 linker at the unique BamH1 site of SV40 (position 2533) and cloning the fragment extending from this point to the HpaI site (position 346) into the EcoR1 and CiaI sites of plasmid pXF3 [11] as illustrated in Fig. 1.

pSTPG6: is a subclone of the trout protamine gene TPG3 of Gregory et al. [10]. It consists of a 950 bp fragment of cloned trout genomic DNA extending from the PstI site at position -450 to the BamH1 site at position +501. A HindIII-linker was inserted at the upstream PstI-site and the protamine gene-containing fragment was cloned into the HindIII and BamH1 sites of pAT153.

pETP: incorporates the trout protamine gene cloned into the expression vector and was constructed by cloning the 950 bp HindIII/BamH1 fragment from pSTPG6 into the KpnI and BamH1 sites of pXSVE2 by blunt-cohesive cloning (see Fig. 1).

pETP-R: was constructed by blunt-cohesive cloning of the 2.3 kb TaqI/BamH1 fragment of SV40 into the TthIII and BamH1 sites of pETPT as illustrated in the inset in Fig. 1; the SV40 fragment contains the small-t intron and sequences extending downstream to the early polyadenylation site.

pSO59-TP: this construct was generated by blunt-ending the AavI site at position -25 of TPG3 (i.e. immediately downstream from the TATA-box) with S1-nuclease and attaching a HindIII linker at the site; the fragment extending from this point to the BamH1 site at +501 was cloned by cohesive-end ligation into the HindIII and BamH1 sites of pSO59 of Chia et al. [12] and is shown in Fig. 2. Sequencing of the junction revealed that treatment of the AavI site with S1-nuclease, as well as removing the 3 base 5'-overhang, had also 'chewed back' a further 4 base pairs into the double-stranded DNA. The 5'-boundary of the TPG3 fragment in pSO59-TP therefore lies at position -17.

pSV2neo-SVTP: was constructed by cohesive-end ligating the large EcoR1-BamH1 fragment of pSV2neo containing the SV40/neomycin-resistance fusion gene [Southern & Berg, ref. 13] to the 982 bp EcoR1-BamH1 fragment from pSO59-TP which contains the SV40/TPG3 fusion gene (see Fig. 2).

Cell culture and transfection procedures.

Cells were cultured in DMEM + 10% calf serum. Transfection of cells was carried out using the dextran procedure [14,15]. Subconfluent cells in 75 cm² bottles were incubated for 30 minutes in serum-free DMEM containing 500μg/ml dextran and 3μg/ml plasmid DNA. This was followed by incubation in DMEM + serum containing 100μM chloroquine [16] for 3 hours (COS cells) or 5 hours (HeLa cells). Cells were harvested after 48 hours and lysed with 5M guanidinium hydrochloride, 0.1M β-mercaptoethanol. The RNA was purified by centrifugation through a 5M caesium chloride pad.

S1-mapping

S1-mapping, using probes labelled with ³²P at the 5' or 3' ends was carried out as previously described [10].
Fig. 2. Strategy for the construction of pSV2neo-SVTP: replacement of the TPG-3 promoter with the SV40 early promoter and the incorporation of this fusion gene into pSV2neo which contains a control gene (neo') fused to the same SV40 early promoter. Details of the sub-cloning protocols are described in Methods; abbreviations are given in the legend to Fig. 1.
Fig. 3. Mapping the 5'- and 3'-ends of TPG-3 transcripts following transient expression in HeLa cells. Each panel shows the autoradiogram from an S1-mapping experiment, under which is a diagram of the probe and the expected and observed fragments of the probe protected by RNA from transfected cells.

A: mapping of 5'-ends of TPG-3 transcripts usigg 20µg aliquots of RNA and hybridising with 181000 cpm/assay of probe at 52°C for 14 hours, followed by digestion with 100 units per ml S1-nuclease at 37°C for 30 minutes; protected fragments were resolved on 8% polyacrylamide/urea gels and the autoradiogram was exposed for 15 hours. Lanes 1 and 2 contain 1/2 and 1/7 loading of 200ng of trout testis poly-A+ RNA: the RNA analysed in the remaining lanes is as shown.

B: mapping of 3'-ends of TPG-3 transcripts using 20µg aliquots RNA and
hybridising with 43000 cpm/assay of probe at 52°C for 16 hours, followed by
digestion with 100 units per ml S1-nuclease at 30°C for 1 hour; protected
fragments were resolved on 6% polyacrylamide/urea gels and the autoradiogram
was exposed for 7 days. Lane 3 shows the analysis of 100ng trout testis poly-A+ RNA.

C: mapping the 5'-termini of SV40 early transcripts as a control for
transfection efficiency. 20μg aliquots RNA was hybridised with 76000
cpm/assay of probe for 13 hours at 48°C, followed by digestion with 100 units
per ml S1-nuclease at 14°C for 3 hours; protected fragments were resolved on 6%
polyacrylamide/urea gels and autoradiogram was exposed for 13 days.

Size markers were derived from pAT153 digested with HpaII.

RESULTS

Mapping the 5'-ends of TPG3 transcripts produced during transient
expression

The construction of the SV40 early-region vector pXSVE2 is described in
the Methods section and is shown in Fig. 1. The vector consists of a 3.1 kb
fragment of SV40, cloned into the plasmid pXF3 [11]: the SV40 fragment
contains the complete SV40 early region and the origin of replication and is
therefore capable of replication in semi-permissive cells such as HeLa cells;
the plasmid pXF3 lacks the poison sequences which are inhibitory to SV40
replication [17].

A 1kb fragment containing the complete TPG3 transcription unit together
with 5'- and 3'-flanking sequences, was cloned into pXSVE2 in the opposite
orientation to the SV40 early region segment to give the plasmid pETP (see
Methods and Fig. 1). The construct was transfected into HeLa cells, total RNA
was isolated 48 hours after transfection and TPG3 transcripts were identified
by S1-mapping using the 5'-end labelled probe illustrated in the diagram under
Fig. 3A. Authentic TPG3 transcripts should protect an 83 bp fragment of the
probe extending from the TthIII site at +83 to the transcription start site
at +1. RNA from cells transfected with pETP (Fig. 3A, lane 5) did indeed give
rise to a band migrating at 83 bp. RNA from mock transfected cells (lane 3) or
cells transfected with the vector pXSVE2 (lane 4) failed to give rise to any
protected fragments.

Mapping of the 3'-ends of TPG3 transcripts produced during
transient expression.

The 3'-ends of transcripts obtained following transfection of pETP into
HeLa cells were mapped using a probe which was 3'-end labelled at the TthIII
site at position +83 and extended to the BamHI site at position +501 (see
diagram under Fig. 3B). Transcripts which are polyadenylated at the correct
site would be expected to protect a 146 bp fragment of the probe. Lane 3
PROBE FOR SV-TP TRANSCRIPTS

PROBE FOR SV-neo<sup>R</sup> TRANSCRIPTS
corresponds to 100ng of trout testis poly-A+c RNA mapped with this probe and contains a strong band sizing at 146 bp. RNA from cells transfected with pETP (lane 2) did not give rise to a co-migrating band. A faint band is visible in lane 2 sizing at 263 bp which is not present in lane 1 (pXSVE2) or lane 3 (testis poly-A+c RNA). This band corresponds to a 3'-end at position +346 which is 118bp downstream from the natural TPG3 polyadenylation site. RNA from cells transfected with pETP (lane 2) also gave rise to a strong band corresponding to protection of the full length probe which is not visible in the tracks corresponding to pXSVE2 (lane 1) or testis poly-A+c RNA (lane 3) suggesting high levels of readthrough into adjoining plasmid sequences.

Transient expression of chimaeric protamine SV40 fusion genes.

The effect of replacing the 3'-portion of TPG3 with a fragment from the SV40 early region containing the small-t intron and early region poly-A site was tested using the construct pETP-R (see inset in Fig. 1 and Methods section). pETP-R was transfected into HeLa cells and total RNA was again isolated after 48 hr. The level of specific transcripts from the chimaeric gene was assayed by S1-mapping using the same 5'-end labelled probe which had been used to map the transcripts from the authentic TPG3 gene in Fig. 3A, lane 5: the result is shown in Fig. 3A, lane 6. Once again, the expected 83 bp-protected fragment was obtained. Although an increase in the levels of transcript was observed relative to that obtained for the native TPG3 gene, the effect does not appear to be very great. The presence of the SV40 early region in the transient expression vector allowed us to assess possible variation in transfection efficiencies by quantitating the transcription products from this region. Although transcripts from the SV40 early promoter are not very abundant in this system due to auto-repression by large T-antigen, we have found it possible to quantitate them by S1-mapping using the probe shown in the diagram under Fig. 3C. The results of this control

Fig. 4. Mapping the 5'-ends of RNA from COS cells transfected by pSV2neo and pSV2neo-SVTP. 20ug aliquots of RNA were mapped using probes (illustrated in the diagrams under the autoradiograms) for the SV-TP (lanes 1-4) or SV-neo (lanes 5-7) fusion genes or a mixture of both probes (lanes 8 and 9). RNA was hybridised with 57000 cpm of either probe or, for the mixed assays, with 57000 of each probe. Specific radioactivities (cpm per picomole of ends) were 1.1 x 106 for the SV-TP probe and 5.43 x 106 for the SV-neo probe. Hybridisation was at 54°C for 14 hours and samples were digested with 100 units per ml of S1-nuclease at 37°C for 30 minutes. Protected fragments were resolved on 5% polyacrylamide/urea gels. Lanes 1 and 2 contain 1/2 and 1/7 loadings of 200ng trout poly-A+c RNA and size markers were derived from a pAT153/HpaII digest.

Panel A shows a 24 hour exposure of the autoradiogram.

Panel B shows an 8-day exposure of the same autoradiogram.
assay are shown in Fig. 3C and indicate that there is little variation in the transfection efficiencies for the different constructs.

In a second fusion gene construct, the TPG3 promoter was replaced by the SV40 early promoter giving rise to the plasmid pS059-TP (see Methods and Fig. 2). Transcripts from this fusion gene (referred to below as SV-TP) would be expected to initiate at the SV40 early start and to possess a 5′-untranslated region of around 100 bp deriving from both the SV40 early and TPG3 5′-untranslated regions. A fragment containing this fusion gene was cloned into the plasmid pSV2neo of Southern and Berg [12] to give pSV2neo-SVTP (see Fig. 2). pSV2neo contains a bacterial neomycin-resistance gene which has also been placed under the control of the SV40 early promoter. This fusion gene will be referred to as SV-neo\(^r\) and is an ideal control gene for comparison with the SV40-protamine fusion gene, SV-TP. The plasmid pSV2neo-SVTP contains two origins of replication but lacks the early region coding for large T antigen which is required for replication. So transient expression from this construct was assayed following transfection of COS cells which express large T antigen constitutively from a stably integrated copy of SV40 [see Gluzman, ref. 18]. RNA was isolated 48 hours after transfection and analysed in a mixed \(S_1\)-mapping assay using probes specific for transcripts from the SV-neo\(^r\) and SV-TP fusion genes (see line drawings under Fig. 4). The results are shown in Fig. 4A and a much longer exposure of the same autoradiogram is shown in Fig. 4B. Since the probe for SV-TP extends past the normal TPG3 cap site, authentic TPG3 transcripts would be expected to protect an 83 bp fragment when mapped with this probe. Lanes 1 and 2 correspond to different loadings of testis poly-A\(^+\) RNA mapped with this probe and contain the expected band sizing at 83 bp. Mapping of RNA from cells transfected with pSV2neo-SVTP using either the SV-TP probe (lane 4) or a mixture of probes for the SV-TP and SV-neo\(^r\) genes (lane 9) gives rise to a faint band sizing at 199 bp (mapping to SV40 position 19) which is visible only after a long exposure (Fig 3B). Bands are also visible after the long exposure which map inside the late transcription unit (positions 187 and 277) as well as some protection of the full length probe. In contrast to the weak signal obtained with the SV-TP probe, \(S_1\)-mapping with the SV-neo\(^r\) probe either in a single probe assay (lane 7) or a mixed assay (lane 9) gives rise to a strong doublet of bands sizing at 388 and 416 bp (mapping to SV40 positions 5230 and 15) which is readily visible after the short exposure (Fig. 4A). Since experimental variation is excluded in the mixed assay, this result indicates that the levels of transcripts from the SV-neo\(^r\) fusion gene are at least 100-fold greater than those from the SV-TP.
fusion gene, although transcription of both genes is driven by the same SV40 early promoter.

DISCUSSION

The data presented in this paper indicate that transcription of a cloned rainbow trout protamine gene TPG3, is accurately initiated following transfection of the gene into HeLa cells in an SV40 early region vector. A region of 5'-flanking sequence extending 450 bp upstream from the TPG3 cap site was sufficient (in the presence of an SV40 enhancer) to give rise to levels of authentically initiated transcripts which were readily detected by S1-mapping.

In contrast, the bulk of the transcripts fail to form 3'-ends at the normal poly-A site which lies 16 bp downstream from the characteristic AAUAAA sequence. This sequence is known to be an essential part of the recognition site for 3'-end processing of primary transcripts [19]. Instead transcripts read through the 3'-flanking region and into adjoining plasmid sequences. A small proportion of the transcripts possess 3'-ends which map to a site 116 bp downstream from the normal poly-A site at position +346. The presence of the sequence AAUAAU 22 bp upstream from this 3'-end suggests that it may arise from a cryptic polyadenylation site.

The failure of the TPG3 transcripts to polyadenylate correctly is unusual. In general it has been found that polyadenylation sites are recognised across wide species gaps; for example the poly-A site of the Drosophila HSP-70 heat shock gene is correctly recognised and utilised in COS cells [20]. There have been reports of inefficient polyadenylation of transcripts from heterologous genes following injection into Xenopus Laevis oocytes [21,22]. It is interesting the latter observations were made when genes normally transcribed in somatic cells were introduced into a germ line cell (the oocyte). In the present work, we report a similar phenomenon when a gene normally transcribed in a germ line cell (the primary spermatocyte) is introduced into somatic cells.

Several of the sea urchin histone genes have been found to give rise to transcripts which are incorrectly processed at the 3'-ends following introduction into HeLa cells [23] and Xenopus oocytes [24]. The failure of correct 3'-end formation for sea urchin histone H3 gene transcripts in Xenopus oocytes has been shown to result from a requirement for a specific sea urchin snRNA (U7) with regions of sequence which are complementary to sequences close to the site of 3'-end formation in the H3 gene. On the basis of sequence
complementarity studies, it has been suggested that U4 snRNA may be involved in the formation of polyadenylated 3'-ends in eukaryotic mRNAs [25]. Evidence for this hypothesis has come from the finding of Moore & Sharp [26] that polyadenylation of adenovirus transcripts in a HeLa whole-cell extract is blocked by antibodies to snRNPs. It is possible that the failure of the TPG3 gene to form correctly polyadenylated transcripts in HeLa cells may be due to species specific differences in snRNAs. The existence of a lesion of this type could be of value in characterising the factors involved in correct poly-A site recognition using complementation assays.

Despite the lack of polyadenylation at the natural site, a significant proportion of TPG3 transcripts are polyadenylated, probably as a result of read-through to the SV40 late poly-A site in the expression vector sequences (data not shown). This might explain in part why replacement of the downstream sequences of TPG3 with an SV40 derived fragment containing a functional poly-A site (in pETP-R) has little effect on transcript levels.

The second fusion gene, SV-TP, in which the TPG3 promoter was replaced with the SV40 early promoter gave rise to only very low levels of transcripts in COS cells compared with the levels obtained for the SV-neo\textsuperscript{r} fusion gene which was also transcribed from the same SV40 early promoter. It is possible that transcription from the early promoter may be inhibited by the downstream TPG3 sequences; however, this is unlikely because this promoter has been used successfully to drive transcription of several different fusion genes [12,27]. Instability of the fusion transcripts in COS cells seems a more likely explanation. Given the failure of the TPG3 poly-A site to be correctly utilised in HeLa cells, it is unlikely that correct polyadenylation occurs efficiently in primate cells. However, the location of the SV-TP gene relative to the SV-neo\textsuperscript{r} gene in the plasmid construct pSV2neo-SVTP means that there is an SV40 late polyadenylation site 140 bp downstream from the BamH\textsuperscript{I} site at position +501 (i.e. 413 bases downstream from the natural TPG3 poly-A site) at which polyadenylation of readthrough transcripts would be expected to occur.

An alternative explanation for the instability of the fusion transcripts is the absence of introns. The presence of introns has been found to be required for the production of stable transcripts from a number of genes transcribing from the SV40 early and late promoters [28-30]. We are currently testing this possibility by introducing a functional intron into the SV-TP fusion gene transcription unit and monitoring the effect on transcript stability.
ACKNOWLEDGEMENTS

The authors are grateful for the help and advice of Drs Steven Gregory and Chris Kontalis. Plasmid pS059 was kindly provided by Peter Rigby. This work was supported by grants from the SERC (GR/C/4417.4) and the Wellcome Trust and the work was completed during the tenure of an MRC Research Fellowship for Academic Staff (G8308275CB) by P.H.W.B.

REFERENCES
