The EMBO Journal vol.10 no.13 pp.4251-4257, 1991

# Isolation and expression of cDNA clones encoding mammalian poly(A) polymerase

**Elmar Wahle, Georges Martin, Emile Schiltz[1] and Walter Keller[2]**

Biozentrum der Universität Basel, Abteilung Zellbiologie, Klingelbergstrasse 70, CH-4056 Basel, Switzerland and [1]Institut für Organische Chemie und Biochemie der Universität Freiburg, Albertstrasse 21, D-7800 Freiburg, FRG

[2]To whom correspondence should be addressed

Communicated by W.Keller

**cDNA clones encoding mammalian poly(A) polymerase were isolated with probes generated by the polymerase chain reaction based on amino acid sequences derived from the purified enzyme. A bovine cDNA clone was obtained encoding a protein of 82 kDa. Expression in *Escherichia coli* resulted in the appearance of a poly(A) polymerase activity that was dependent on the addition of the purified specificity factor CPF and the presence of the polyadenylation signal AAUAAA in the RNA substrate. The activity copurified with a polypeptide of the expected size. A second class of cDNAs encoded a polypeptide of 43 kDa which was closely related to the N-terminal half of the 82 kDa protein. Northern blots showed two mRNAs of 4.2 and 2.4 kb that probably correspond to the two classes of cDNAs, as well as a third band of 1.3 kb. The sequence of the N-terminal half of bovine poly(A) polymerase is 47% identical with the amino acid sequence of the corresponding part of yeast poly(A) polymerase. Homologies to other proteins are of uncertain significance.**

*Key words:* cDNA cloning/polyadenylation/poly(A) polymerase/mRNA processing

## Introduction

The polyadenylate tails present on almost all eukaryotic messenger RNAs are added post-transcriptionally by a multicomponent machinery (reviewed by Manley, 1988; Wickens, 1990). Synthesis of the poly(A) tail is carried out by poly(A) polymerase with the help of the cleavage and polyadenylation factor, CPF (Christofori and Keller, 1988, 1989; Takagaki *et al.*, 1988, 1989; Gilmartin and Nevins, 1989; Wahle, 1991a; Bienroth *et al.*, 1991). CPF [also called SF (Takagaki *et al.*, 1989) or PF2 (Gilmartin and Nevins, 1989)] binds to the essential polyadenylation signal AAUAAA which is located in the RNA 10-30 nucleotides upstream of the polyadenylation site (Gilmartin and Nevins, 1989; Bardwell *et al.*, 1991; Keller *et al.*, 1991). Elongation of the growing tail is assisted by a poly(A) binding protein (Wahle, 1991b; Gershon *et al.*, 1991). The polyadenylation reaction is preceded by endonucleolytic cleavage of the primary transcript at the polyadenylation site. This cleavage reaction requires at least two factors in addition to poly(A) polymerase and CPF (Christofori and Keller, 1988; Takagaki *et al.*, 1989; Gilmartin and Nevins, 1989) and depends on a second sequence element downstream of the cleavage site (reviewed by Proudfoot, 1991).

Poly(A) polymerase has been purified to homogeneity from calf thymus as a protein of 57-60 kDa (Tsiapalis *et al.*, 1975; Wahle, 1991a). The enzyme by itself does not recognize the polyadenylation signal AAUAAA. In fact, it has a poor affinity for any RNA substrate and only a slight preference for poly(A) (Wahle, 1991a). The addition of $Mn^{2+}$ to the reaction nonspecifically increases the affinity of the enzyme for RNA primers and permits efficient polyadenylation independent of AAUAAA and CPF (Tsiapalis *et al.*, 1975; Christofori and Keller, 1988; Takagaki *et al.*, 1988; Wahle, 1991a).

We have used amino acid sequences derived from the purified enzyme to isolate cDNA clones of poly(A) polymerase. Expression of a clone in *Escherichia coli* confirmed that it encoded poly(A) polymerase functional in the AAUAAA-dependent and CPF-mediated polyadenylation of RNA.

## Results

### *Isolation of a cDNA clone encoding poly(A) polymerase*

Amino acid sequences were obtained from the N-terminus as well as from three tryptic peptides of purified poly(A) polymerase (Figure 1). DNA oligonucleotides to be used as primers for the polymerase chain reaction (PCR) were synthesized based on some of these amino acid sequences (Figure 1 and Materials and methods). The templates used for PCR were obtained by reverse transcription of calf thymus RNA, primed either by oligo(dT) or by the same DNA oligonucleotide that was used as a PCR primer (see Materials and methods). A DNA fragment of 700 nucleotides was obtained, among others, in a PCR reaction using primers 1 and 4. After gel purification, this fragment could be reamplified with primers 2 and 4. Direct sequencing of the

| Peptide: | Sequence: |
|---|---|
| | 1 → |
| | 2 → |
| N-terminus | PFPVTTQGSQQTQPXQKXYG |
| | 4 ← |
| 101 | TDEILHLVPNIDNFRLTLRAIKLXAK |
| 96 | THNIYSNILGFLGGVSXAMLVAR |
| 66 | QRLEWVGLVESK |

**Fig. 1.** Peptide sequences of purified poly(A) polymerase. Sequences of the N-terminus and three internal peptides (arbitrary numbering) are presented in the single letter code in their order of appearance in the predicted protein sequence (see Figure 2). X, unidentified amino acid. The primers used for PCR amplification of cDNA fragments are indicated as arrows pointing in the 5′ to 3′ direction (for details, see Materials and methods).

PCR product with primer 2 revealed a nucleotide sequence downstream of the primer that encoded six amino acids known from the sequencing of the purified protein (data not shown). Correct products, identified by similar procedures, were also obtained with other primer combinations (data not shown).

The 700 bp fragment obtained with primers 2 and 4 was used as a probe to screen two cDNA libraries, derived from HeLa cells and calf thymus respectively. Clones were only obtained from the HeLa library. Although these clones encoded the correct amino acid sequences, they had an open reading frame for a protein of only 43 kDa and lacked peptide 66 (see below). Therefore, a fragment from the coding region of these clones was used to screen an additional cDNA library from calf muzzle epithelium. The largest cDNA clone obtained in this screen was 2.5 kb in length (Figure 2). It contained an open reading frame coding for a protein of 739 amino acids with a molecular weight of 82.4 kDa. The N-terminus of the predicted amino acid sequence was identical with the N-terminal sequence obtained from purified poly(A) polymerase except for the absence of the initiating methionine in the protein. The predicted sequence also contained all internal peptide sequences derived from the purified protein.

### *Expression of poly(A) polymerase in E.coli*

A fragment containing the open reading frame of the bovine clone was inserted into a T7 expression vector such that the initiating ATG codon was that encoding the first amino acid of the authentic protein (see Materials and methods). Depending on the induction conditions, the synthesis of a polypeptide of the expected size could be barely or not at all detected by SDS−PAGE of total cell lysates. After induction overnight at 18°C, nonspecific poly(A) polymerase activity, measured by the incorporation of radiolabeled ATP into acid-precipitable material in the presence of a poly(A)

```
GCGGTTGCGGGGGGGAAGTGACTGGGCGGTGCGGCGCAGGAGACGATGCCGTTTCCAGTTACAACACAGGGATCACAGCAAACACAGCCG    90
                                                M  P  F  P  V  T  T  Q  G  S  Q  Q  T  Q  P   15
CCACAGAAGCACTATGGCATTACTTCTCCCATCAGCTTAGCAGCCCCCAAGGAGACTGACTGCCTGCTCACACAGAAGCTGGTGGAGACT   180
P  Q  K  H  Y  G  I  T  S  P  I  S  L  A  A  P  K  E  T  D  C  L  L  T  Q  K  L  V  E  T    45
CTGAAGCCCTTCGGGGTTTTTGAAGAGGAAGAGGAACTGCAGCGCAGGATTTTTATTTTGGGAAAATTAAATAACCTGGTAAAAGAGTGG   270
 L  K  P  F  G  V  F  E  E  E  E  E  L  Q  R  R  I  F  I  L  G  K  L  N  N  L  V  K  E  W   75
ATACGAGAAATCAGTGAAAGCAAGAATCTTCCACAATCTGTAATTGAAAATGTTGGTGGGAAAATTTTTACATTTGGATCTTATAGATTA   360
 I  R  E  I  S  E  S  K  N  L  P  Q  S  V  I  E  N  V  G  G  K  I  F  T  F  G  S  Y  R  L  105
GGAGTACATACAAAAGGTGCTGATATTGATGCATTGTGTGTTGCACCAAGACATGTTGATCGAAGTGATTTTTTCACCTCATTCTATGAT   450
 G  V  H  T  K  G  A  D  I  D  A  L  C  V  A  P  R  H  V  D  R  S  D  F  F  T  S  F  Y  D  135
AAGTTGAAATTACAGGAAGAAGTAAAAGATTTAAGAGCTGTTGAAGAGGCATTTGTACCAGTTATCAAACTGTGTTTTGATGGGATAGAG   540
 K  L  K  L  Q  E  E  V  K  D  L  R  A  V  E  E  A  F  V  P  V  I  K  L  C  F  D  G  I  E  165
ATTGATATTTTGTTTGCAAGATTAGCACTGCAGACTATTCCAGAAGACTTGGACTTAAGAGATGACAGTCTGCTTAAAAATTTAGATATA   630
 I  D  I  L  F  A  R  L  A  L  Q  T  I  P  E  D  L  D  L  R  D  D  S  L  L  K  N  L  D  I  195
AGATGTATAAGAAGTCTTAACGGTTGCAGGGTAACCGATGAAATTTTACATCTAGTACCAAACATTGACAACTTCAGGTTAACCCTGAGA   720
 R  C  I  R  S  L  N  G  C  R  V  T  D  E  I  L  H  L  V  P  N  I  D  N  F  R  L  T  L  R  225
GCTATCAAACTGTGGGCCAAACGCCACAACATCTATTCCAATATATTAGGTTTCCTCGGTGGTGGTTTCCTGGGCTATGCTAGTAGCAAGA   810
 A  I  K  L  W  A  K  R  H  N  I  Y  S  N  I  L  G  F  L  G  G  V  S  W  A  M  L  V  A  R  255
ACTTGCCAGCTTTATCCAAATGCAATAGCATCAACTCTTGTACATAAATTTTTCTTGGTATTTTCTAAATGGGAATGGCCAAATCCAGTC   900
 T  C  Q  L  Y  P  N  A  I  A  S  T  L  V  H  K  F  F  L  V  F  S  K  W  E  W  P  N  P  V  285
CTATTGAAACAGCCTGAAGAATGCAATCTTAATTTGCCTGTATGGGACCCAAGGGTAAACCCCAGTGATAGGTACCATCTTATGCCTATA   990
 L  L  K  Q  P  E  E  C  N  L  N  L  P  V  W  D  P  R  V  N  P  S  D  R  Y  H  L  M  P  I  315
ATTACACCAGCATACCCACAACAGAACTCCACGTACAATGTGTCCGTTTCAACACGGATGGTCATGGTTGAGGAGTTTAAACAAGGTCTT  1080
 I  T  P  A  Y  P  Q  Q  N  S  T  Y  N  V  S  V  S  T  R  M  V  M  V  E  E  F  K  Q  G  L  345
GCTATCACAGATGAAATTTTGCTGAGTAAGGCAGAGTGGTCCAAACTTTTTGAAGCTCCAAACTTCTTTCAAAAGTACCAGCATGATATT  1170
 A  I  T  D  E  I  L  L  S  K  A  E  W  S  K  L  F  E  A  P  N  F  F  Q  K  Y  Q  H  D  I  375
GTACTTCTAGCAAGTGCACCAACTGAAAAACAACGCCTAGAATGGGTGGGCTTGGTGGAATCAAAAATCCGAATCCTGGTTGGAAGTTTG  1260
 V  L  L  A  S  A  P  T  E  K  Q  R  L  E  W  V  G  L  V  E  S  K  I  R  I  L  V  G  S  L  405
GAGAAGAATGAGTTTATGACACTGGCTCATGTGAATCCCCAGTCATTTCCAGCACCCAAAGAAAATCCCGACAAGGAAGAATTTCGCACT  1350
 E  K  N  E  F  M  T  L  A  H  V  N  P  Q  S  F  P  A  P  K  E  N  P  D  K  E  E  F  R  T  435
ATGTGGGTGATTGGGTTAGTGTTTAACAAAACAGAAAACTCTGAAAATCTCAGTGTTGATCTCACCTATGATATTCAGTCTTTCACAGAT  1440
 M  W  V  I  G  L  V  F  N  K  T  E  N  S  E  N  L  S  V  D  L  T  Y  D  I  Q  S  F  T  D  465
ACAGTTTATAGGCAAGCAATAAACAGCAAGATGTTTGAGGTGGACATGAAAATTGCTGCGATGCATGTAAAAAGAAAGCAACTCCATCAA  1530
 T  V  Y  R  Q  A  I  N  S  K  M  F  E  V  D  M  K  I  A  A  M  H  V  K  R  K  Q  L  H  Q  495
CTACTGCCTAGTCATGTGCTTCAGAAAAAGAAAAAGCATTCAACAGAAGGCGTCAAGTTGACACCTCTGAATGATAGCAGCCTCGACTTG  1620
 L  L  P  S  H  V  L  Q  K  K  K  K  H  S  T  E  G  V  K  L  T  P  L  N  D  S  S  L  D  L  525
TCTATGGACAGTGACAACAGCATGTCTGTGCCTTCACCTACTAGTGCTATGAAGACCAGTCCGTTGAACAGTTCTGGCAGCTCTCAGGGC  1710
 S  M  D  S  D  N  S  M  S  V  P  S  P  T  S  A  M  K  T  S  P  L  N  S  S  G  S  S  Q  G  555
AGAAACAGTCCTGCTCCAGCTGTAACAGCAGCATCTGTGACCAACATACAGGCTACTGAAGTTTCTCTGCCACAAATAAATTCCAGTGAA  1800
 R  N  S  P  A  P  A  V  T  A  A  S  V  T  N  I  Q  A  T  E  V  S  L  P  Q  I  N  S  S  E  585
AGCTCAGGGGGTACATCGAGTGAAAGCATTCCTCAAACTGCCACACAACCAGCCATTTCATCACCGCCAAAGCCTACGGTCTCCAGAGTT  1890
 S  S  G  G  T  S  S  E  S  I  P  Q  T  A  T  Q  P  A  I  S  S  P  P  K  P  T  V  S  R  V  615
GTTTCCTCAACACGTTTGGTAAACCCACCACCAAGACCTTCAGGAAATGCAGCAGCAAAGATACCTAATCCTATAGTAGGAGTCAAGAGG  1980
 V  S  S  T  R  L  V  N  P  P  P  R  P  S  G  N  A  A  A  K  I  P  N  P  I  V  G  V  K  R  645
ACATCCTCACCTCATAAAGAAGAGAGCCCCAAGAAAACCAAAACAGAAGAGGATGAAACAAGTGAAGATGCTAACTGTCTTGCTTTGAGT  2070
 T  S  S  P  H  K  E  E  S  P  K  K  T  K  T  E  E  D  E  T  S  E  D  A  N  C  L  A  L  S  675
GGACATGATAAAACAGAAACAAAGGAACAACTTGATACAGAGACAAGTACAACTCAATCAGAAACCATTCAGACAGCGACTTCTCTGTTG  2160
 G  H  D  K  T  E  T  K  E  Q  L  D  T  E  T  S  T  T  Q  S  E  T  I  Q  T  A  T  S  L  L  705
GCCTCTCAGAAAACATCCAGTACAGACCTTTCTGATATCCCTGCTCTCCCTGCAAACCCTATTCCTGTTATCAAGAATTCAATAAAACTG  2250
 A  S  Q  K  T  S  S  T  D  L  S  D  I  P  A  L  P  A  N  P  I  P  V  I  K  N  S  I  K  L  735
AGATTGAATCGGTAAAAACAACCTCAGGGTCCAGAAACAGTGTCTGCCAACTCAACCTGTTGTCTTCAAATGCTAAAAAAGGAGAACGGA  2340
 R  L  N  R                                                                                  739
GGGTGCAAGACTAGACGTGACTGAAAGTGGATTGAGGGTTTTTTTGTGACCTCCCTTACTGGGCTAATCAGCACTTGATCGGAAGTCCAG  2430
GTTAGTATGTGAAGCCAGGAGTACTATTATTATTGTGTTAGCAACAGTTGCA                                        2482
```

**Fig. 2.** Nucleotide sequence and predicted amino acid sequence of a cDNA clone for bovine poly(A) polymerase. The complete nucleotide sequence of the cDNA clone and the translation of its largest open reading frame are shown. Underlined amino acid sequences are those corresponding to the sequenced peptides (Figure 1). Peptides 101 and 96 were contiguous. Numbers of the rightmost nucleotide or amino acid, respectively, in each lane are given on the right. This sequence has been deposited in the EMBL data bank under the accession number X61585.

primer and $Mn^{2+}$, was only 2-fold higher than in a control lysate, 12 800 versus 6700 U/mg. [In the expression of cloned yeast poly(A) polymerase, the background activity present in control lysates of *E.coli* was negligible (J.Lingner, J.Kellermann and W.Keller, submitted for publication). This difference was due to the different reaction conditions used to assay the poly(A) polymerases from yeast and mammalian cells.] However, DEAE chromatography of the lysate divided this activity into two equal portions, one in the flowthrough and one in the eluate. The activity in the control lysate, in contrast, bound almost completely to the DEAE column under the same conditions. The presence of mammalian poly(A) polymerase in the *E.coli* lysate was clearly detected by the specific polyadenylation assay: upon complementation with purified specificity factor CPF, the lysate from the strain harboring the cloned gene, as well as the DEAE flowthrough fraction derived from this lysate, polyadenylated a radiolabeled precursor RNA containing the AAUAAA sequence (Figure 3, lanes 4 and 5). In contrast,

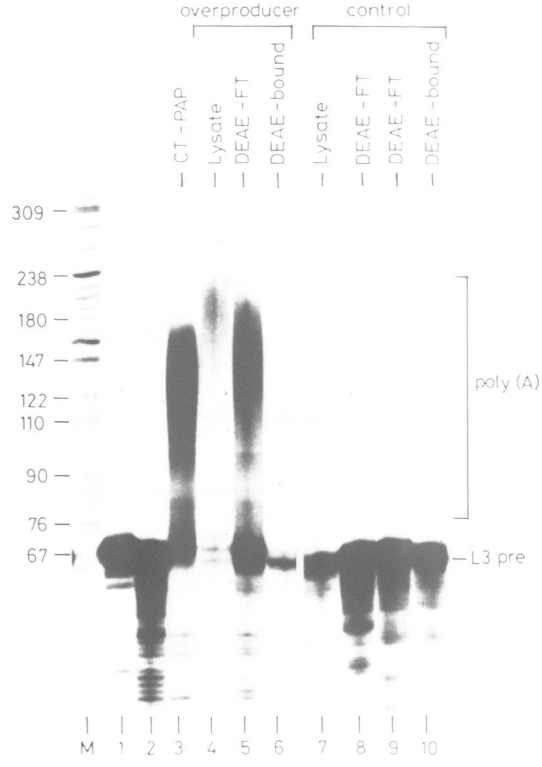


**Fig. 3.** Specific polyadenylation by *E.coli*-produced poly(A) polymerase. Specific polyadenylation reactions were carried out (see Materials and methods) with the L3pre substrate RNA. Lane 1, incubation with purified calf thymus poly(A) polymerase (5 U) in the absence of CPF; lane 2, incubation with CPF alone; lane 3, incubation with purified CPF plus purified poly(A) polymerase. All reactions displayed in lanes 4–10 contained purified CPF plus various *E.coli* fractions. Lanes 4–6, fractions derived from overproducing strain (pT7-PAP82): lane 4, 300 ng of crude lysate (3.5 U poly(A) polymerase); lane 5, 70 ng of DEAE flowthrough peak fraction (1.5 U); lane 6, 150 ng of DEAE-bound peak fraction (2 U). Lanes 7–10, fractions derived from control strain (pJC10): lane 7, 275 ng of crude lysate (1.5 U); lane 8, 35 ng of DEAE flowthrough peak fraction (0.25 U); lane 9, 670 ng of the same DEAE flowthrough peak fraction (5 U); lane 10, 150 ng of DEAE bound peak fraction (1.3 U). For details of the preparation, see Materials and methods. The RNA was run on a denaturing 12% polyacrylamide gel. Positions of substrate and polyadenylated products are indicated on the right. Poor recovery of RNA in lane 4 was very probably due to the presence of ribonucleases in the crude extract. M, DNA size markers.

the control lysate was inactive in this assay, as were both DEAE fractions derived from it as well as the DEAE-bound fraction derived from the overproducing strain (Figure 3, lanes 6–10). The polyadenylation activity present in the DEAE flowthrough of the overproducing strain was partially purified (see Materials and methods). Activities measured by the specific and nonspecific polyadenylation assays copurified and in the final MonoQ column a protein of the anticipated molecular weight was correlated with both activities (Figure 4). At all stages tested, polyadenylation activity in the specific assay was dependent on the presence of both CPF and an intact AAUAAA sequence in the RNA (Figure 4C and data not shown). These results thus confirm that the cDNA described above encodes functional poly(A) polymerase.

### *Multiple messenger RNAs related to poly(A) polymerase*

Northern blot analysis of HeLa and calf thymus poly(A)$^+$ RNA under stringent conditions with a probe derived from the N-terminal half of the poly(A) polymerase coding region revealed two strong bands of ~4.5 kb and 1.3 kb and a weak band of 2.4 kb (Figure 5, lanes 1). The largest RNA appeared as a double band. A probe derived from the C-terminal half of the same clone reacted only with this large RNA (Figure 5, lanes 2). This suggests that the cDNA clone was derived from the 4.5 kb mRNA and thus lacks a substantial amount of untranslated sequence, and that the two other mRNAs are related to the 5′-half of the 4.5 kb species. As mentioned above, we also isolated several independent human cDNA clones encoding a protein of 43 kDa. The first 371 of its 375 amino acids differed from the corresponding part of the long bovine clone by a single amino acid replacement. The sequence deviated from that of the bovine clone in the last four amino acids and the long 3′-untranslated region. The 5′-untranslated regions were also dissimilar with the exception of ~30 nucleotides immediately preceding the coding regions (data not shown). Two arguments suggest that the major differences between the clones were not related to their difference in origin, bovine versus human. First, sequences very similar to those unique for the short human clone could be amplified from bovine cDNA by PCR using appropriate primers (G.Martin and W.Keller, unpublished data). Second, the length of the clones encoding the 43 kDa protein [2.1 kb, including 49 nucleotides of poly(A)] is quite similar to the length of the intermediate mRNA visible in the Northern blot of Figure 5. The probe that detected this band consisted of sequence common to both classes of cDNA clones and the band was present in both human and bovine mRNA. The same band of mRNA was not detected by the probe unique for the long cDNA clone (see above). These data thus suggest that the short cDNA clone represents the 2.4 kb mRNA encoding a protein of 43 kDa that is identical or nearly identical to the first half of poly(A) polymerase. So far, no cDNA clones have been obtained that might represent the smallest class of mRNA visible in the Northern blot.

## Discussion

We have isolated a cDNA clone for bovine poly(A) polymerase. The clone was identified based on the fact that it encoded all amino acid sequences obtained by sequencing

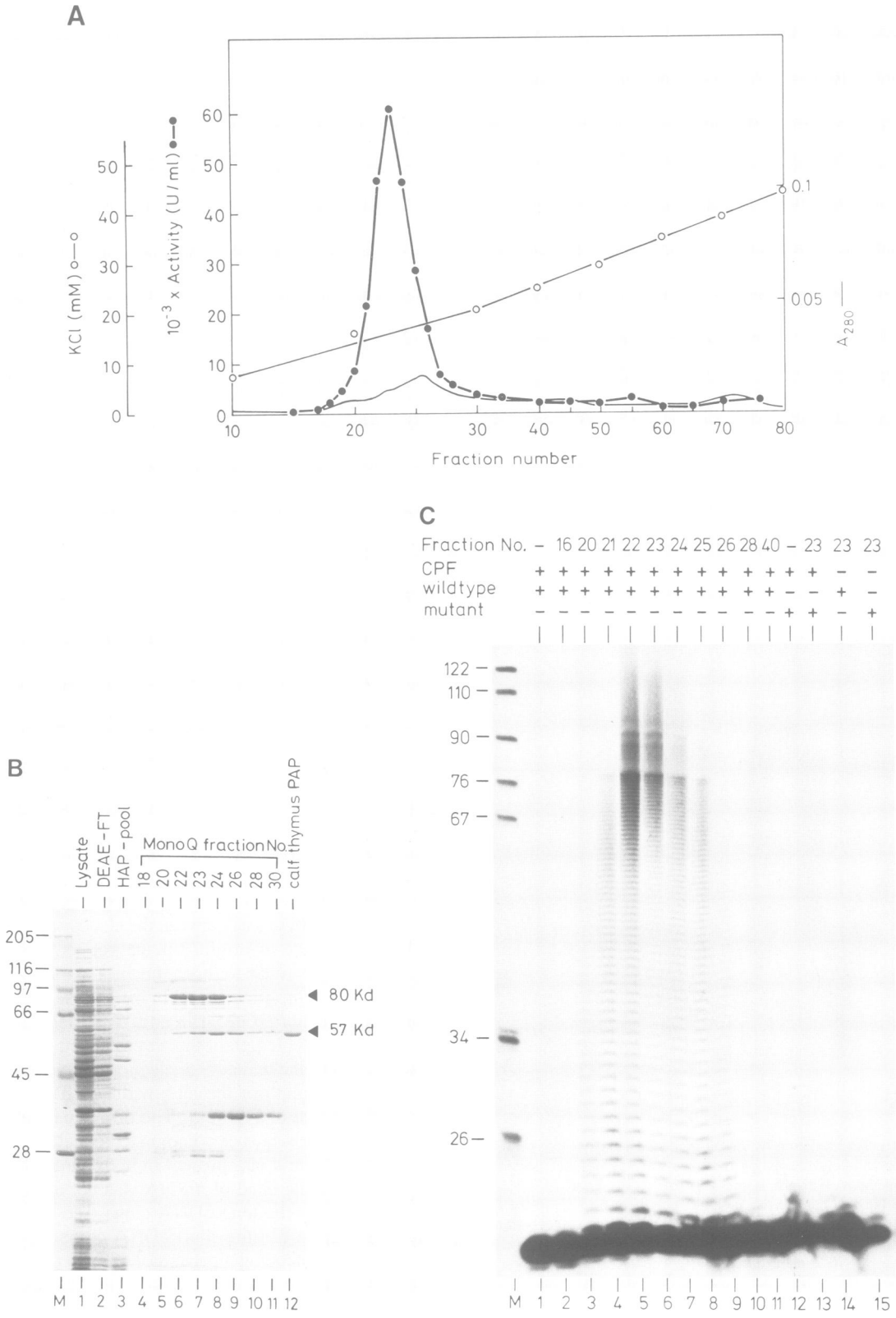


**Fig. 4.** Partial purification of *E.coli*-produced poly(A) polymerase. For details of the purification from *E.coli* carrying plasmid pT7-PAP82 see Materials and methods. **A**. Profile of the MonoQ column showing $A_{280}$, salt gradient and nonspecific poly(A) polymerase activity. Note that the scale in which the peak fractions are presented differs from the scale in the rest of the column profile. **B**. SDS–PAGE of aliquots throughout the purification. Lane M, molecular weight markers; lane 1, 2.5 μl of lysate (375 U); lane 2, 2.5 μl of DEAE flowthrough (175 U); lane 3, 150 μl of the hydroxyapatite pool (750 U); lanes 4–11, 200 μl each of MonoQ fractions 18–30; lane 12, 1.2 μg of purified calf thymus poly(A) polymerase. Electrophoresis was carried on a 10% polyacrylamide gel. Proteins were detected by staining with Coomassie Brilliant Blue. Numbers on the left indicate the molecular weights of marker proteins in kDa. Arrowheads point to the 80 kDa and 57 kDa polypeptides which copurified with poly(A) polymerase activity. **C**. Specific polyadenylation activity in the MonoQ column. Aliquots of column fractions (1 μl of 1:15 dilutions) were assayed with synthetic RNA 18mers, wild type or mutant as indicated. Assays were supplemented with purified CPF as indicated in the top panel. M, DNA size markers. Samples were separated on a denaturing 12% acrylamide gel.

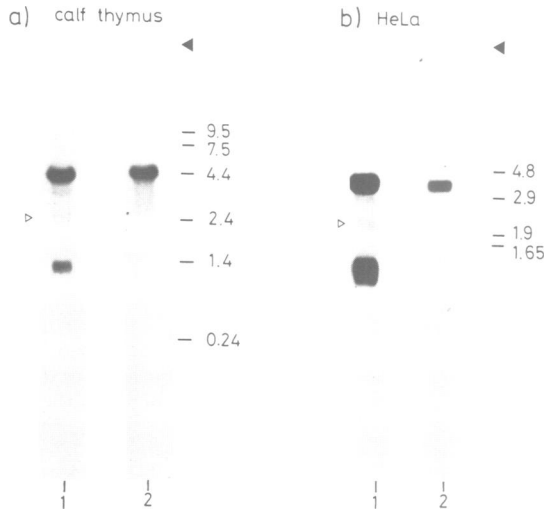


**Fig. 5.** Northern blot analysis of mRNA encoding poly(A) polymerase. Northern blotting was carried out with poly(A)$^+$ RNA from calf thymus or HeLa cells as indicated (see Materials and methods). The amounts of RNA loaded per lane were 4.5 $\mu$g for HeLa RNA and 25 $\mu$g for calf thymus RNA. Exposure times were 2 weeks for HeLa RNA and 2 days for calf thymus RNA. Probes were a fragment extending from a *Hin*dIII site in the polylinker of the vector to the *Kpn*I site at position 975 (lanes 1) and a *Pvu*II–*Eco*RI fragment (positions 1729–2235; lanes 2). Size markers were either those obtained from Gibco/BRL (**a**) or ribosomal RNAs from *E.coli* and calf thymus (**b**). The empty arrowheads point to the weak 2.4 kb bands; full arrowheads indicate the loading wells.

of the purified enzyme. Its identity was confirmed by its expression in *E.coli* and detection of the expected enzymatic activity. Polyadenylation of RNA substrates dependent on the addition of the specificity factor CPF and the presence of the hexanucleotide sequence AAUAAA in the RNA distinguished the activity derived from the cDNA from a high background of endogenous activity. Whether the background is due to *E.coli*'s own poly(A) polymerase (Sippel, 1973) or some other enzyme that incorporates ATP in the crude extract is not known. The background activity could be readily removed by DEAE chromatography.

The open reading frame in the cDNA clone appears to be complete: it starts with an ATG codon, followed by amino acid sequence present at the N-terminus of the purified enzyme, and is preceded and followed by sequences containing multiple stop codons in all three frames. The predicted protein has a size of 82 kDa and a polypeptide of this size copurified with poly(A) polymerase activity after expression in *E.coli*. Although expression and purification of the protein clearly have to be improved, a rough estimate based on the MonoQ column profile suggests that the *E.coli*-made enzyme has a specific activity very similar to that of the enzyme purified from calf thymus. The *E.coli*-made protein was substantially larger than the one purified from calf thymus, suggesting that the latter had suffered proteolysis. Proteolysis is likely to be at least a partial explanation for the heterogeneity of poly(A) polymerase during purification that has been observed by numerous investigators (Ryner *et al.*, 1989; Wahle, 1991a; earlier work reviewed by Edmonds and Winters, 1976). These data also suggest that a C-terminal domain of 20 kDa is dispensable for specific and nonspecific polyadenylation. The activity of the 57 kDa poly(A) polymerase core in the endonucleolytic cleavage preceding polyadenylation is unclear at present (K.Beyer, E.W.Wahle and W.Keller, unpublished data). Interestingly, MonoQ chromatography of poly(A) polymerase expressed in *E.coli* revealed an additional protein that exactly comigrated with the poly(A) polymerase purified from calf thymus and was only slightly displaced from the 82 kDa polypeptide in the column (Figure 4B). This may suggest that *E.coli* proteases clip the protein at the same site that is vulnerable in calf thymus extract. An estimate based on the molecular weight of the major poly(A) polymerase polypeptide purified from calf thymus places the hypothetic protease site near amino acid 520. This coincides very well with the beginning of the serine/threonine-rich end of the protein. Its serine/threonine content is 29.7% compared with 17.1% for the complete protein and 10.9% for the 43 kDa protein encoded by the human clones.

The two probes used in Northern blot analysis detected a mRNA of ~4.5 kb in HeLa and bovine mRNA. This suggests that long untranslated sequences are missing from the clone. The N-terminal probe also detected two additional mRNAs of 2.4 and 1.3 kb. The 2.4 kb message is probably represented by several human cDNA clones encoding a protein of 43 kDa that is almost identical to the N-terminal half of the protein predicted from the long bovine clone. The minute difference between the two proteins may be due to their different origins and the two mRNAs might be the product of alternative splicing. Further work is required to substantiate this and to characterize the 1.3 kb mRNA for which we have not yet found a corresponding cDNA clone. Whatever the origin of the mRNA encoding the 43 kDa protein, a second (and possibly a third) protein related to poly(A) polymerase is likely to exist and may also contribute to the heterogeneity of poly(A) polymerase discussed above. It should be interesting to find out whether these proteins have distinct roles and/or localizations in the cell. Attempts to find an enzymatic activity after expression of the 43 kDa protein in *E.coli* have failed so far (E.Wahle, G.Martin and W.Keller, unpublished data).

We have recently reported the cloning of the *Saccharomyces cerevisiae* poly(A) polymerase gene (J.Lingner, J.Kellermann and W.Keller, submitted for publication). The first 395 amino acids of the yeast gene and amino acids 14–407 of the bovine gene are 47% identical with a single amino acid missing in the bovine protein. The remaining parts of the two sequences are not obviously related. This may suggest that the N-terminus of the protein is responsible for conserved functions, i.e. substrate binding and catalytic activity, while the rest of the protein interacts with other factors, e.g. CPF, that are not conserved between yeast and mammals. A comparison of the bovine gene sequence with sequences in the EMBL data bank revealed no strong similarities to other proteins. Surprisingly, the protein has also no significant sequence similarity to the catalytic subunit of the vaccinia virus poly(A) polymerase that has recently been cloned (Gershon *et al.*, 1991). Potentially interesting are three independent weak similarities to the $\beta$ subunit of chloroplast RNA polymerases (J.Lingner, personal communication). Four lysine residues at positions 503–506 may play a role in nuclear transport, especially since they are followed by a potential casein kinase II phosphorylation site (a serine at position 520, followed by acidic residues three, seven and nine positions downstream; Rihs *et al.*, 1991). Weak similarities to the RNP 1 and RNP 2 consensus

sequences of RNA binding proteins (Kenan *et al.*, 1991) are found in amino acids 96–103 and 62–67, respectively.

The significance of these and other sequences may now be explored by mutagenesis. Moreover, improved overproduction and purification will be useful in studying the biochemical properties of poly(A) polymerase and the enzyme's role in polyadenylation.

## Materials and methods

### Materials

Poly(A) polymerase and CPF were purified from calf thymus as described (Wahle, 1991a; Bienroth *et al.*, 1991). Sequencing grade trypsin was from Boehringer Mannheim, T7 DNA polymerase (Sequenase 2.0) from USB, *Taq* polymerase from Cetus and M-MuLV reverse transcriptase from Pharmacia. *E.coli* ribosomal RNA was from Boehringer Mannheim and RNA size markers were from GIBCO/BRL. Chemically synthesized RNA oligonucleotides used as polyadenylation substrates were those described by Bardwell *et al.* (1991). The enzymatically synthesized substrate for polyadenylation assays, L3pre, contained the L3 polyadenylation site of adenovirus 2 and ended one nucleotide upstream of the natural polyadenylation site (Christofori and Keller, 1988). L3preΔ had a U to C mutation in the AAUAAA sequence. These substrates and poly(A) were obtained as described (Wahle, 1991a). Oligo(dT)–cellulose (type 7) was from Pharmacia, nitrocellulose (BA 85) from Schleicher & Schuell and isopropylthiogalactoside (IPTG) from Bachem. Hydroxyapatite (Ultrogel HA) was purchased from IBF, Villeneuve-la-Garenne, France, and all other resins, including the FPLC equipment, from Pharmacia.

### Manipulation of nucleic acids

Total RNA was prepared from HeLa cells or calf thymus by the guanidinium isothiocyanate procedure (MacDonald *et al.*, 1987). Poly(A)$^+$ RNA was isolated from total RNA by a single run on oligo(dT)–cellulose (Jacobsen, 1987; Sambrook *et al.*, 1989). RNA was separated in 1% agarose–formaldehyde gels and transferred to nitrocellulose by capillary blotting according to Sambrook *et al.* (1989), except that the filters were not washed before baking. Markers were run in neighbouring lanes and visualized by ethidium bromide staining. Blots were prehybridized for 2–3 h at 68°C in 6 × SSC, 5 × Denhardt's solution, 0.5% SDS and 0.1 mg/ml herring sperm DNA. Hybridization was carried out overnight under the same conditions with double-stranded DNA probes, labeled by random priming (Sambrook *et al.*, 1989) and denatured by boiling, at ~5 × $10^6$ c.p.m./ml (Cerenkov counting). Blots were washed twice in 2 × SSC, 0.5% SDS at room temperature (20–40 min with agitation), once in 0.1 × SSC, 0.5% SDS at room temperature, then under the same conditions at 68°C and again at room temperature.

Reverse transcription of RNA was done with 75 U of reverse transcriptase, 2 μg of calf thymus RNA and 1 μg of oligo(dT) or 0.2 μg of specific oligonucleotide primer in 50 μl of 50 mM Tris–HCl, pH 8.3, 100 mM KCl, 3 mM $MgCl_2$, 10 mM DTT and 0.2 mM of each dNTP. The mixture was heated for 5 min at 80°C before the addition of DTT and enzyme and subsequently incubated for 30 min each at 37°C and 43°C. The enzyme was inactivated by heating for 10 min to 90°C.

Manipulations of DNA were carried out according to standard protocols (Sambrook *et al.*, 1989). The host strain for plasmids was DH5 (Hanahan, 1983).

### Amino acid sequencing and PCR

For N-terminal sequencing of purified poly(A) polymerase, ~30 μg of protein were precipitated with 20% trichloroacetic acid, washed with ether, dissolved in 2% SDS and sequenced on an Applied Biosystems model 477A/120A sequencer according to the procedures of the manufacturer. For preparation of tryptic fragments, ~60 μg of purified protein were dialyzed against 10 mM Tris–HCl, pH 8.5, and concentrated to 50 μl in a SpeedVac. The protein was digested with trypsin and peptides were isolated on a PEP-S column (Pharmacia) as described (Schiltz *et al.*, 1991). Four peptides were sequenced as described above.

All PCR primers derived from amino acid sequences were fully degenerate. For some primers this required synthesis in two batches. If necessary, the last nucleotide of the 3′ terminal codon was omitted so that ambiguities in the 3′-terminal nucleotide were avoided. All PCR primers carried restriction sites at their 5′-termini for ease of subcloning.

For PCR, 5 μl of a reverse transcriptase reaction were used as template. The reactions were carried out in 30 μl of 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 2.5 mM $MgCl_2$, 0.01% Triton X-100, 0.2 mM of each dNTP with 200–300 ng of each primer and 2.5 U of *Taq* polymerase. Typically, 30 cycles were carried out consisting of 0.5 min at 94°C, 1 min at 55–59°C and 2 min at 72°C. Products were analyzed on agarose gels and the desired products were purified by a GeneClean kit (obtained from Bio 101).

### Libraries, isolation of clones and sequencing

A HeLa cDNA library in λgt11 was obtained from Peter Nielsen, Max Planck-Institut für Immunbiologie, Freiburg, FRG. An oligo(dT)-primed cDNA library from calf muzzle epithelium in λZAPII (Stratagene) was obtained from Peter Koch and Werner W.Franke, Deutsches Krebsforschungszentrum, Heidelberg, FRG (Koch *et al.*, 1990). Libraries were screened according to standard procedures (Sambrook *et al.*, 1989). Initial screening of the HeLa library was done with a 700 bp DNA fragment, obtained in PCR reactions with primers 2 and 4 (see Results). The DNA fragment was gel-purified, labeled by random priming and used directly for screening. The calf muzzle library was screened with a *PstI–KpnI* fragment derived from the short HeLa cDNA clone (see text), labeled by random priming. This fragment corresponded to nucleotides 572–975 of the clone represented in Figure 2. Subcloning from λZAP was done *in vivo* according to the procedure suggested by Stratagene. The nucleotide sequence of the insert was determined by dideoxy sequencing of double-stranded DNA with Sequenase 2.0 according to Winship (1989). Primers were synthesized to match internal sequences of the insert or standard primers were used in combination with deletions generated by exonuclease III according to a protocol supplied by Stratagene.

DNA sequences were analyzed with GCG software (Devereux *et al.*, 1984). The EMBL data bank was searched with the program TFASTA (Pearson and Lipman, 1988).

### Expression and partial purification of poly(A) polymerase

The coding sequence of the bovine poly(A) polymerase clone was inserted into two related expression vectors, pJC10 (Clos *et al.*, 1990) and pT7-7 (constructed by Stan Tabor, Harvard University, Cambridge MA). Both vectors contain a promoter recognized by the bacteriophage T7 RNA polymerase. The host strain, BL21, carried the phage DE3 with the T7 RNA polymerase gene under the control of the *lacZ* promoter as well as the pLysS plasmid (Studier, 1991). The poly(A) polymerase coding sequence was PCR-amplified. The upstream primer covered the ATG initiation codon, converting it into an *NdeI* site, the other was the Bluescript SK sequencing primer. The amplification was carried out with 1 μg of plasmid DNA and eight PCR cycles and the amplified fragment was cloned into the *NdeI* and *BamHI* sites of the two vectors. For each of the two vectors, two identical constructs were made with inserts from two independent PCR reactions. One representative of each of the four constructions and a control with the empty pJC10 vector were induced with IPTG and lysates were tested for nonspecific poly(A) polymerase activity. All four constructs had 1.5- to 2-fold elevated levels of poly(A) polymerase activity.

The most promising strain, pT7-PAP82, was induced at a larger scale together with the empty pJC10 vector as a control. Single colonies of each strain were inoculated into LB medium with 100 μg/ml ampicillin and grown overnight at 37°C. These cultures were diluted 1:80 into 2 l of the same medium and grown at 37°C to an $A_{600}$ of 0.15. They were further incubated at 18°C and, at an $A_{600}$ of 0.5–0.6, IPTG was added to a final concentration of 1 mM to induce the synthesis of T7 RNA polymerase. 18 h after induction the cells were harvested by centrifugation and each strain was resuspended in 50 ml of 100 mM KCl, 50 mM Tris, pH 7.9, 5 mM EDTA. The suspensions were frozen in liquid nitrogen, thawed, and the following additions were made: 10% glycerol, 0.5 mM DTT, 0.5 mM phenylmethylsulfonylfluoride, 0.4 μg/ml leupeptin, 0.7 μg/ml pepstatin. The cells were sonicated until the viscosity was reduced to that of buffer and the lysate was centrifuged for 30 min at 48 000 *g* and 0°C. The supernatant of pT7-PAP82 contained 754 mg of protein and 9.7 × $10^6$ U of poly(A) polymerase, whereas the lysate of pJC10 contained 708 mg of protein and 4.7 × $10^6$ U of poly(A) polymerase. Both lysates were applied to DEAE–Sepharose FF columns (150 ml) equilibrated in the same buffer as above, including all additions, except that EDTA was reduced to 1 mM. The columns were washed with 100 ml of equilibration buffer and eluted with the same buffer containing 0.6 M KCl. Fractions of 30 ml were collected. The combined flowthrough fractions of pT7-PAP82 contained 329 mg of protein and 5.2 × $10^6$ U of poly(A) polymerase, the eluate contained 506 mg and 5.2 × $10^6$ U. The flowthrough of pJC10 contained 252 mg and 0.6 × $10^6$ U, the eluate contained 569 mg and 3.7 × $10^6$ U. The DEAE flowthrough derived from pT7-PAP82 was applied to a 150 ml hydroxyapatite column in the same buffer as above except that EDTA was omitted. The column was washed with one volume of equilibration buffer and eluted with a 10 volume gradient from equilibration buffer to 0.3 M potassium phosphate, pH 7.2, 10% glycerol, 0.5 mM

DTT, 0.4 μg/ml leupeptin, 0.7 μg/ml pepstatin. The combined peak fractions, eluted at 100 mM phosphate, contained 30 mg of protein and $2.2 \times 10^6$ U of poly(A) polymerase. Two attempts to adsorb this material to a MonoS column under the conditions described previously (Wahle, 1991) failed. A portion of the preparation ($0.8 \times 10^6$ U) was therefore dialyzed overnight against 25 mM Tris−HCl, pH 7.9, 10% glycerol, 1 mM EDTA, 0.5 mM DTT, 0.05% Nonidet P40 and applied to a 1 ml MonoQ FPLC column. After washing with loading buffer, the column was eluted with a 40 ml gradient from 0 to 50 mM KCl in loading buffer. Fractions of 0.5 ml were collected. Binding of poly(A) polymerase was very weak again: 70% of the poly(A) polymerase activity applied was recovered in the flowthrough and 15% was eluted as a single peak at 15 mM KCl.

Protein concentrations were determined according to Bradford (1976) and SDS−polyacrylamide gels were run according to Laemmli (1970). If necessary, proteins were precipitated with 20% trichloroacetic acid.

### *Enzyme assays*

Nonspecific poly(A) polymerase activity was measured in the presence of $Mn^{2+}$ as the primer-dependent incorporation of radiolabeled ATP into acid-precipitable material under the conditions described (Wahle, 1991a). The primer used was poly(A). Poly(A) polymerase units were also as defined by Wahle (1991a). Specific polyadenylation assays were carried out by complementation of poly(A) polymerase fractions with purified CPF as described (Wahle, 1991b; Bienroth *et al.*, 1991; Keller *et al.*, 1991). AAUAAA-containing RNA substrates and their mutant derivatives were either chemically synthesized 18mers or transcripts of SP6 RNA polymerase (see above).

## Acknowledgements

We are grateful to Peter Nielsen, Peter Koch and Werner Franke for gifts of cDNA libraries, to Vivian Bardwell and Angus Lamond for synthetic RNA oligonucleotides, to Joachim Lingner for computer analyses, to Silke Bienroth for supplying purified CPF and for help with some experiments, and to Tobias Raabe for discussing the amino acid sequence of poly(A) polymerase. This work was supported by grants from the Kantons of Basel and from the Schweizerischer Nationalfonds.

## References

Bardwell,V.J., Wickens,M., Bienroth,S., Keller,W., Sproat,B.S. and Lamond,A.I. (1991) *Cell*, **65**, 125−133.
Bienroth,S., Wahle,E., Suter-Crazzolara,C. and Keller,W. (1991) *J. Biol. Chem.*, **266**, 19768−19776.
Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248−254.
Christofori,G. and Keller,W. (1988) *Cell*, **54**, 875−889.
Christofori,G. and Keller,W. (1989) *Mol. Cell. Biol.*, **9** 193−203.
Clos,J., Westwood,J.T., Becker,P.B., Wilson,S., Lambert,K. and Wu,C. (1990) *Cell*, **63**, 1085−1097.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387−395.
Edmonds,M. and Winters,M.A. (1976) *Progr. Nucleic Acid Res. Mol. Biol.*, **17**, 149−179.
Gershon,P., Ahn,B.-Y., Garfield,M. and Moss,B. (1991) *Cell*, **66**, 1269−1278.
Gilmartin,G.M. and Nevins,J.R. (1989) *Genes Dev.*, **3**, 2180−2189.
Hanahan,D. (1983) *J. Mol. Biol.*, **166**, 557−580.
Jacobson,A. (1987) *Methods Enzymol.*, **152**, 254−261.
Keller,W., Bienroth,S., Lang,K.M. and Christofori,G. (1991) *EMBO J.*, in press.
Kenan,D.J., Query,C.C. and Keene,J.D. (1991) *Trends Biochem. Sci.*, **16**, 214−220.
Koch,P.J., Walsh,M.J., Schmelz,M., Goldschmidt,M.D., Zimbelmann,R. and Franke,W.W. (1990) *Eur. J. Cell Biol.*, **53**, 1−12.
Laemmli,U.K. (1970) *Nature*, **227**, 680−685.
MacDonald,R.J., Swift,G.H., Przybyla,A.E. and Chirgwin,J.M. (1987) *Methods Enzymol.*, **152**, 219−227.
Manley,J.L. (1988) *Biochim. Biophys. Acta*, **950**, 1−12.
Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444−2448.
Proudfoot,N. (1991) *Cell*, **64**, 671−674.
Rihs,H.-P., Jans,D.A., Fan,H. and Peters,R. (1991) *EMBO J.*, **10**, 633−639.
Ryner,L.C., Takagaki,Y. and Manley,J.L. (1989) *Mol. Cell. Biol.*, **9**, 4229−4238.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schiltz,E., Kreusch,A., Nestel,U. and Schulz,G.E. (1991) *Eur. J. Biochem.*, **199**, 587−594.
Sippel,A.E. (1973) *Eur. J. Biochem.*, **37**, 31−40.
Studier,F.W. (1991) *J. Mol. Biol.*, **219**, 37−44.
Takagaki,Y., Ryner,L. and Manley,J.L. (1988) *Cell*, **52**, 731−742.
Takagaki,Y., Ryner,L. and Manley,J.L. (1989) *Genes Dev.*, **3**, 1711−1724.
Tsiapalis,C.M., Dorson,J.W. and Bollum,F.J. (1975) *J. Biol. Chem.*, **250**, 4486−4496.
Wahle,E. (1991a) *J. Biol. Chem.*, **266**, 3131−3139.
Wahle,E. (1991b) *Cell*, **66**, 759−768.
Wickens,M. (1999) *Trends Biochem. Sci.*, **15**, 277−281.
Winship,P.R. (1989) *Nucleic Acids Res.*, **17**, 1266.

*Received on August 19, 1991; revised on September 20, 1991*