4034–4041 Nucleic Acids Research, 1995, Vol. 23, No. 20

Isolation of genomic and cDNA clones encoding bovine poly(A) binding protein II

Annette Nemeth[1,+], Sabine Krause[1], Diana Blank[1], Andreas Jenny[1], Paul Jenö[2], Ariel Lustig[3] and Elmar Wahle[1,*,+]

Departments of [1]Cell Biology, [2]Biochemistry and [3]Biophysics, Biozentrum, Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Received July 28, 1995; Revised and Accepted September 14, 1995

EMBL accession no. X89969

ABSTRACT

cDNA clones for bovine poly(A) binding protein II (PAB II) were isolated. Their sequence predicts a protein of 32.8 kDa, revising earlier estimates of molecular mass. The protein contains one putative RNA-binding domain of the RNP type, an acidic N-terminal and a basic C-terminal domain. Analyses of authentic PAB II were in good agreement with all predictions from the cDNA sequence except that a number of arginine residues appeared to be post-translationally modified. Poly(A) binding protein II expressed in *Escherichia coli* was active in poly(A) binding and reconstitution of processive polyadenylation, including poly(A) tail length control. The cDNA clones showed a number of potential PAB II binding sites in the 3′ untranslated sequence. Bovine poly(A)+ RNA contained two mRNAs hybridizing to a PAB II-specific probe. Analysis of a genomic clone revealed six introns in the coding sequence. The revised molecular mass led to a demonstration of PAB II oligomer formation and a reinterpretation of earlier data concerning the protein's binding to poly(A).

INTRODUCTION

The poly(A) tails found at the 3′-ends of nearly all eukaryotic mRNAs are generated by post-transcriptional processing in two universally conserved steps. The RNA is first cleaved endonucleolytically, then the upstream cleavage product receives a poly(A) tail, whereas the downstream fragment is degraded (1–3). The newly synthesized poly(A) tail has a defined length, near 250 nucleotides in mammalian cells (4). Shortening of the poly(A) tail after transport to the cytoplasm gives rise to the heterogeneous steady-state population (4,5).

Whereas the factors responsible for the initial cleavage reaction have only been partially characterized (1–3), poly(A) addition to a 'pre-cleaved' RNA can be reconstituted from three purified proteins. Poly(A) polymerase (6–8), using ATP as a precursor, elongates the 3′-end, normally generated by endonucleolytic cleavage of the pre-mRNA. For its activity and primer-specificity, the polymerase relies on two additional factors. One of these, the cleavage and polyadenylation specificity factor (CPSF) (9,10), binds the primary polyadenylation signal AAUAAA present in nearly all pre-mRNAs just upstream of the cleavage and polyadenylation site (11). The second auxiliary factor, poly(A)-binding protein II (PAB II) binds the growing poly(A) tail once the latter has reached a length of 10–11 nucleotides (12,13). Both factors stimulate poly(A) polymerase by holding it on the primer RNA. A complex formed from the primer and all three proteins synthesizes a full-length tail in one processive event (14). After 250 adenylate residues have been polymerized, processive elongation terminates; further extension is distributive and slow (14,15). In other words, the three proteins show the same length control of poly(A) tail synthesis that is observed *in vivo*. As PAB II appears to bind stoichiometrically to the growing tail, this protein is probably responsible for length control (15).

Here, we report the isolation of bovine cDNA and genomic clones encoding PAB II. A revised molecular weight determination leads to a reinterpretation of earlier data concerning the native structure of PAB II and its binding to poly(A).

MATERIALS AND METHODS

Libraries and reagents

An oligo(dT)-primed cDNA library from calf muzzle epithelium in λ ZAPII was provided by Dr Peter Koch and Dr Werner W. Franke, Deutsches Krebsforschungszentrum, Heidelberg, Germany (16). Oligo(dT)- and random-primed libraries from calf liver and bovine aorta endothelial cells in λ ZAPII were purchased from Stratagene. An oligo(dT)-primed calf thymus plasmid cDNA library (17) was also used. A calf thymus genomic library of a partial Sau3A digest in λ EMBL3 was provided by Dr Sigfried Ruppert, Deutsches Krebsforschungszentrum, Heidelberg, Germany.

Escherichia coli indoleglycerolphosphate synthetase-phosphoribosylanthranilate isomerase was a gift from Halina Szadkowski and Dr Kaspar Kirschner, Biozentrum Basel. Poly(A) polymerase, CPSF and calf thymus PAB II were the preparations described previously (15). Other proteins were obtained commercially. 3,3′-Dithio-bis(propionic acid *N*-hyxdroxysuccinimide ester) [Dithiobis(succinimidylpropionate), DTSP] and suberic

* To whom correspondence should be addressed at +present address: Universität Giessen, Institut für Biochemie, Heinrich-Buff-Ring 58, D-35392 Giessen, Germany

acid bis(*N*-hydroxysuccinimide ester) (disuccinimidyl suberate; DSS), both from Sigma, were used as freshly made 20 mM stock solutions in dimethylsulfoxide. 30 mg of dimethylsuberimidate (DMS) (Sigma) was dissolved immediately before use in 1 ml of 50 mM Hepes-KOH, pH 8.2 and the pH was adjusted by the addition of 15 µl 10 N NaOH. Preparation of 5′-end labeled poly(A) and oligo(A) for binding assays and of L3 pre- and oligo-adenylated L3pre for polyadenylation assays has been described (13,15).

Nucleic acid methods

Small-scale-preparation of plasmid DNA was performed according to the CTAB method (18). For large-scale preparations, the Qiagen Plasmid Maxi Kit was used. Double-stranded DNA was sequenced with the Sequenase 2.0 sequencing kit (USB). GC-rich regions were sequenced with 7-deaza-dGTP (Boehringer Mannheim) replacing dGTP. For particularly difficult regions, *E.coli* single-strand binding protein was used according to the recommendations of USB, except that the proteinase K digestion was extended to several hours. Sequence analysis was done with the DNAStar programs (DNAStar Inc., Madison, WI) and with software from the Genetics Computer Group (Madison, Wisconsin; GCG Package, Version 7.2, April 1991) (19). The EMBL data bank was searched with the GCG programs FASTA and TFASTA (20).

Subcloning and transformation, labelling of DNA by random priming and Southern blotting were all done according to ref. 21. Extraction of total RNA from tissue was performed as described (22). Poly(A)$^+$-RNA was isolated on oligo(dT)-cellulose (21,23). For Northern blots, the RNA was separated in 1.2% agarose-formaldehyde gels and transferred to Hybond N$^+$ membranes (Amersham) by capillary blotting. RNA was cross-linked to the membrane by UV-irradiation. RNA size markers (GIBCO) were detected by methylene blue staining (21). Hybridization and washing were done under standard high stringency conditions (21).

PCR was carried out in 50 µl of buffer (Perkin-Elmer Cetus) containing 5 U Taq DNA polymerase (Perkin-Elmer Cetus), 50 pmol of each primer, 100 ng template, 0.2 mM of each dNTP. 20–30 cycles were performed with times and temperatures adjusted to individual primer-template combinations. Reaction products were separated on agarose gels and recovered by the Qiaex kit (Qiagen).

Protein analysis

Protein concentrations were determined by Bradford assays (24). SDS-polyacrylamide gels were run according to ref. 25. Western blots were prepared as described (26). The ECL system (Amersham) was used for detection.

Trypsin digestion, HPLC purification of peptides and sequencing have been described (12,13). Poly(A) binding protein II was desalted on a C_4 reverse-phase column. CNBr cleavage was carried out with desalted protein in 50 µl of 70% formic acid, 50 mg/ml CNBr for 24 h in the dark. The digest was diluted with 150 µl of water and taken to dryness. Fragments were separated on a Vydac 218TP51 column. Quantitative amino acid analysis was done with the DABS-Cl method (27). For matrix-assisted laser desorption-ionisation time-of-flight mass spectrometry, 1 µl of the desalted PAB II preparation was mixed with 1 µl of 1% sinnapinic acid in 10% ethanol and spotted onto the sample plate of a Vestec Benchtop II mass spectrometer. Spectra were acquired at a laser wavelength of 337 nm and an acceleration voltage of 25 kV.

Analytical ultracentrifugation was carried out in a Beckman XLA centrifuge equipped with absorption optics. Poly(A) binding protein II was dialyzed against 50 mM Tris-HCl, pH 8.0, 50 mM KCl, 1 mM EDTA, 0.5 mM dithiothreitol. The PAB II concentration after dialysis was 8.6 µM. A sedimentation velocity run was carried out at 56 000 r.p.m., 20°C, in a 12 mm double-sector cell. Buffer viscosity and density were corrected to those of water. Sedimentation equilibrium runs were performed at 20 000 and 28 000 r.p.m. in the same cell type filled to a height of 3 mm with sample in addition to FC-43 bottom fluid. Molecular weights were determined by a linear regression computer program that adjusts the base line to obtain the best linear fit of lnA (A = absorption) versus r^2 (r = distance from rotor center). A partial specific volume of 0.73 ml/g was used.

Isolation and analysis of PAB II clones

PAB II tryptic peptide 19 had the sequence IYVGNVDYGATAEE. The following oligonucleotides were derived from it: 5′-CCGAATTCATTAGTNGGNAA-3′, representing the N-terminus, preceded by an *Eco*RI site (underlined) and two irrelevant nucleotides and 5′-CCGGATCCTCTCNGCNGTNGC-3′, representing the C-terminus, preceded by a *Bam*HI site (underlined) and two irrelevant nucleotides. Oligo(dT)-primed cDNA synthesis was carried out (21) with calf thymus total RNA as a template. An aliquot of the cDNA was used as a template for PCR with the two PAB II-specific primers. After 30 cycles of amplification (15 s at 90°C, 20 s at 48 or 50°C and 50 s at 72°C), products were analyzed on a 6% polyacrylamide gel, a band of the expected size was subcloned and the anticipated nucleotide sequence confirmed.

Screening of the calf muzzle epithelium library with this PCR product led to the isolation of clone ZAB6. A *Xho*I–*Sma*I fragment (nucleotides 2466–2960) from ZAB6 was used to screen the muzzle epithelium, liver and thymus cDNA libraries. 30 clones were analyzed by restriction mapping and several were sequenced. The aorta endothelium library was screened with a 5′-end probe generated by PCR from the subcloned genomic 2.1 kb *Eco*RI–*Xho*I fragment (see Fig. 1) with a plasmid primer located upstream of the *Eco*RI site and an internal PAB II-specific primer located at nucleotides 2256–2274. Insert sizes and terminal sequences of six clones were determined and aligned with the sequences of the genomic *Eco*RI–*Xho*I fragment and the previously isolated cDNA clones. Three clones were sequenced completely. The aorta endothelium library was also screened with a synthetic oligonucleotide from the 3′-end of the available cDNA clones (nucleotides 3023–3057). The clones extending the longest distance downstream of the known sequence were identified and one of them, CA11, was sequenced completely.

All cDNA sequences were determined on both strands from exonuclease III deletions (28) and some additional specific primers. Numbering of nucleotides refers to the sequence as deposited in the data base.

The genomic library was screened with the *Xho*I–*Sma*I cDNA fragment from ZAB6 (see above). Seven different clones were analyzed in detail by restriction mapping and Southern hybridizations with probes derived from the cDNA sequence. The

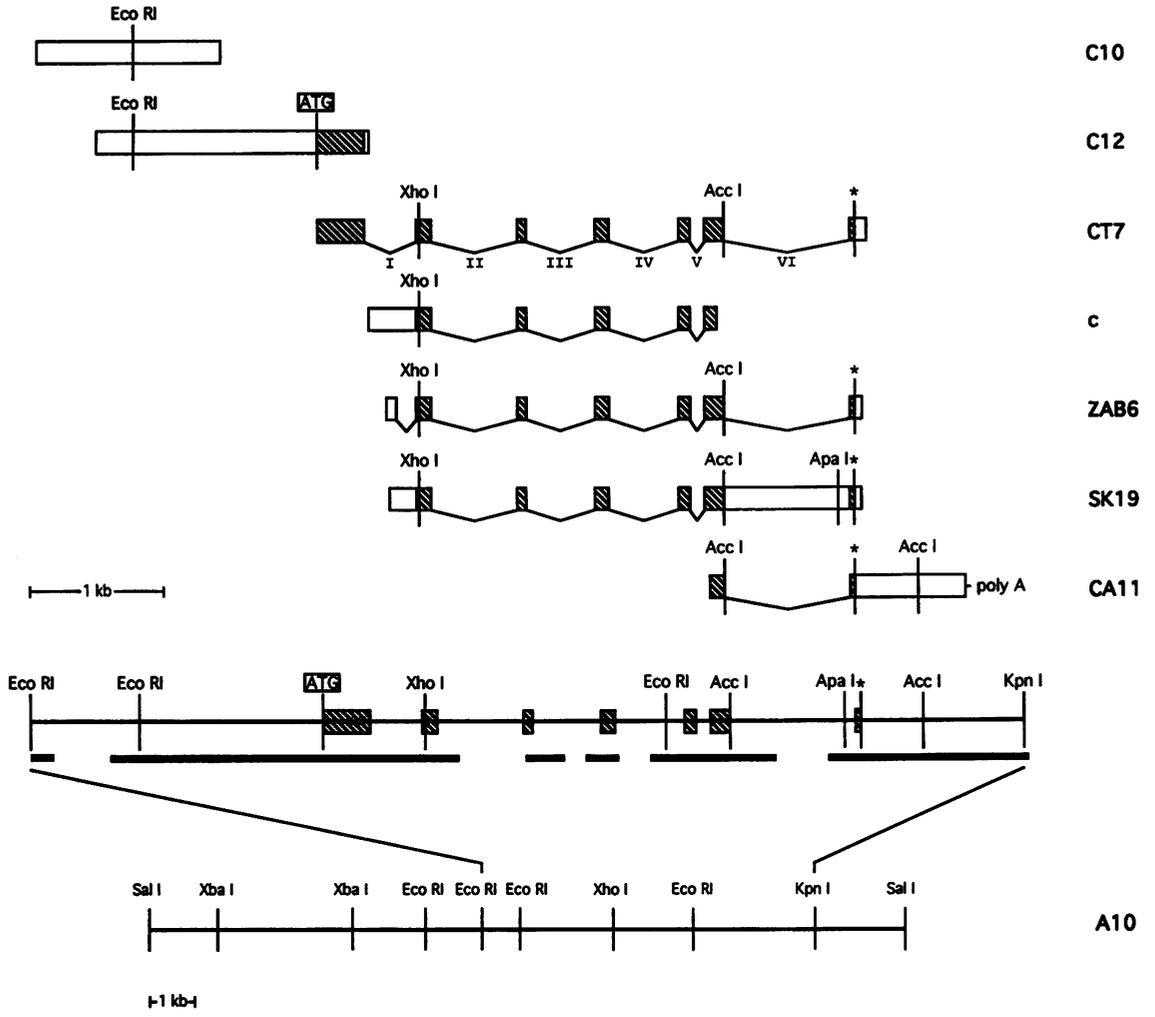

Figure 1. Schematic comparison of different PAB II cDNA clones and the genomic clone A10. Seven different cDNAs (C10–CA11) are shown in the upper part of the figure. Hatched boxes represent the open reading frame, white boxes represent unspliced introns, 5′ UTR and 3′ UTR. Spliced introns are indicated by simple lines. The introns are numbered in Roman numerals below clone CT7. The position of the initiation codon is represented by 'ATG', the stop codon by an asterisk. Clones C10, C12 and CA11 were from the aorta endothelium library, clone CT7 from the calf thymus library, clone c from the liver library and clones ZAB6 and SK19 from the calf muzzle epithelium library. The restriction map of the genomic clone A10 (17 kb) is shown at the bottom. The region covered by the cDNA sequences is enlarged and some of the corresponding restriction sites, initiation and stop codons as well as the open reading frame are marked. The *Kpn*I site at the right boundary of the enlarged section may not be the only one in clone A10: only a subcloned fragment was digested by *Kpn*I. The black bars below the enlarged section represent the sequenced regions of A10. All cDNA sequences were determined on both strands, whereas most genomic sequences were read on one strand only.

intron-exon structure of clone A10 was analyzed by subcloning and partial sequencing (usually only one strand).

Expression and purification of PAB II

The coding sequence of PAB II was inserted into the *Nde*I and *Bam*HI sites of pT7-7 (29) as follows: The DNA was PCR-amplified from cDNA clone CT7. The upstream primer covered the 5′-end of the gene and introduced the ATG initiation codon missing in CT7 as part of an *Nde*I site. The downstream primer introduced a *Bam*HI site 60 nucleotides after the stop codon. Several amplifications under different conditions and ligation to the vector yielded no complete clones. However, one clone contained the coding sequence from nucleotide 2347 to the C-terminus, preceded by 63 bp of irrelevant sequence. The missing N-terminus was generated by PCR with the upstream primer as above and an internal primer (nucleotides 2461–2477). Fusion to the C-terminal fragment in pT7-7 made use of the *Nde*I site at the initiation codon and the internal *Xho*I site at nucleotide 2466. The correct sequence of the complete open reading frame in the expression construct was confirmed.

The plasmid was transformed into BL21(pLysS) (30). Cells were grown at 37°C in Super Broth supplemented with 200 μg/ml ampicillin and 30 μg/ml chloramphenicol to OD_{600} = 3, 0.5 mM isopropyl-β-D-thiogalactoside and additional ampicillin (125 μg/ml) were added and the culture was grown overnight at room temperature. Cells were harvested, stored at –70°C and resuspended in 50 mM Tris-HCl, pH 8.0, 50 mM KCl, 10% (v/v) glycerol, 1 mM phenylmethylsulfonyl fluoride, 2 μg/ml leupeptin, 10 μg/ml DNaseI, 0.1 mM $MgCl_2$, 0.5 mM dithiothreitol. The lysate was sonicated and, after centrifugation, 1 mM EDTA was added. A control lysate from cells containing the empty pT7-7

vector was prepared in the same manner. Western blotting with an antiserum against PAB II (31) showed a reactive band of the anticipated size only in the lysate from the overexpressing strain.

Purification of the overexpressed PAB II was monitored by Western blotting and, at later stages, also by poly(A) binding assays. The lysate (3000 mg protein) was applied to a 500 ml Macroprep Q-column equilibrated in buffer 1 (13) containing 50 mM KCl and eluted with a 5 l gradient from 50–500 mM KCl in buffer 1. Poly(A) binding protein II-containing fractions were pooled according to a Western blot, and batch-adsorbed to 70 ml packed volume of Blue Sepharose. Upon elution (13), ~25% of PAB II were found in the 4 M NaCl step. The major portion, eluted with 2 M guanidinium hydrochloride, was used for further purification as described (13), except that a Mono Q-column replaced Macroprep Q. Approximately 0.08 mg of pure PAB II with a specific poly(A) binding activity of 360 000 U/mg were obtained.

Nitrocellulose filter-binding and polyadenylation assays were carried out as described (13,15).

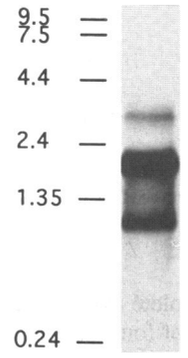

Figure 2. Northern blot analysis of mRNA encoding PAB II. Northern blotting was carried out with 15 µg calf thymus poly(A)$^+$ RNA as described in Materials and Methods. The 495 bp *XhoI–SmaI* fragment of ZAB6 (nucleotides 2466–2960) was used as a probe. The sizes of the RNA markers (in kb) are indicated on the left.

RESULTS

cDNA cloning of PAB II

Purified PAB II was digested with trypsin and four peptides purified by reversed-phase HPLC were sequenced. Fully degenerate DNA oligonucleotides were synthesized based on the N- and C-termini of one peptide. PCR with these oligonucleotides serving as primers and oligo(dT)-primed calf thymus cDNA as template (see Materials and Methods) yielded a product encoding the entire peptide sequence. This PCR product was used to screen a bovine cDNA library. The clone isolated (ZAB6; Fig. 1) encoded two of the peptides sequenced, but was incomplete. A fragment of this clone was used to isolate additional clones from the same library and two other cDNA libraries. Representative clones (CT7, c, SK19) are shown schematically in Figure 1. Although all four peptides were represented in these clones, they were again all incomplete. Clones containing the 5′-end of the open reading frame and the 5′-UTR (e.g. C12 and C10; Fig. 1) were obtained from an additional library in a screen using a genomic 5′-probe (see below and Materials and Methods). In a final screen with a 3′-probe (see Materials and Methods), the same library yielded, among other clones, CA11 (Fig. 1). CA11 had an AAUAAA sequence 15 nucleotides upstream of a poly(A) tail and a second AAUAAA sequence overlapping the beginning of the poly(A) tail.

From the clones described above, a complete open reading frame was reconstructed that contained all four peptides mentioned above as well as additional peptides sequenced later (see below). The sequence predicted a protein of 306 amino acids with a molecular weight of 32.8 kDa. Of 10 cDNA clones analyzed in detail and extending into the coding sequence, six contained either one or both of two introns (introns I and VI, see below). Evidence for the intron character of these sequences is their absence from other cDNA clones, the presence of one or multiple stop codons and of splice site consensus sequences at the junctions to the coding sequence (data not shown). Four additional introns, identified by a comparison to a genomic clone (see below) were not represented in the cDNA clones. The mRNA from which clone ZAB6 was derived had made use of an alternative 5′-splice site in intron I (Fig. 1). ZAB6 ends shortly upstream of this alternative splice site so that the structure of the corresponding mRNA can only be guessed. Assuming that the upstream portion of intron I was retained in the message, the coding sequence was interrupted by an in-frame stop codon.

The 5′-noncoding sequence represented in cDNA-clones C10 and C12 is unusually long, 2095 nt. The 3′-noncoding sequence is ~870 nt, the exact number depending on which A residue is considered to be the beginning of the poly(A) tail (see below). Within the first 140 nucleotides beyond the stop codon, there are six stretches of oligo(A) with lengths of 8, 7, 11, 14, 9 and 10 nucleotides. Internal oligo(dT) priming at these oligo(A) sequences accounted for the poor representation of the genuine 3′-end in the cDNA libraries.

PAB II-related mRNAs

Poly(A)$^+$-RNA purified from calf thymus was separated on a denaturing gel, blotted and hybridized with a probe from the PAB II coding sequence (Fig. 2). The most abundant message had a size of 2.1 kb, corresponding to the sum of open reading frame and 3′-UTR. A minor band near 1.1 kb varied in abundance between different RNA preparations and may thus be a degradation product. An additional weak band of 3.3 kb might be expected to correspond to PAB II mRNAs containing the long 5′-UTR found in cDNA clones C10 and C12. However, this RNA was shorter than expected from the cDNAs and did not hybridize with probes derived from clones C10 and C12. Clones C10 and C12 were from an aorta endothelium library, whereas the mRNA used for the Northern blot came from thymus. Thus, it is possible that the long 5′-UTR is not expressed in thymus. The 3.3 kb mRNA might be the transcript of a related gene (see Discussion).

Genomic clones

Restriction mapping of seven clones, isolated from a bovine genomic library with a PAB II probe, divided these into three classes. Clones from two classes did not hybridize with all PAB II-derived probes, and very limited sequencing of one clone showed deviations from the expected structure (data not shown). The third class, represented by clone A10, hybridized with all probes derived from the PAB II cDNAs. Partial sequencing of

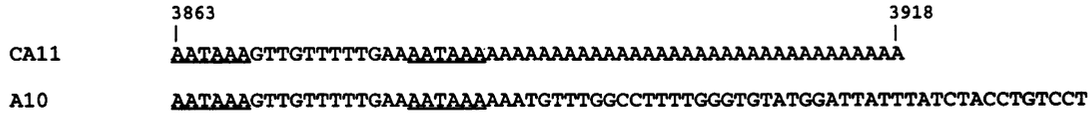

Figure 3. Sequence comparison of cDNA clone CA11 and genomic clone A10 at the two possible polyadenylation signals (underlined). The genomic sequence shown here was determined on both strands.

clone A10 showed complete agreement with the cDNA clones, except for the presence of four additional introns (Fig. 1).

Comparison of the genomic sequence to that of cDNA clone CA11 confirmed the two AAUAAA motifs, but also showed three additional A residues directly following the second motif (Fig. 3). Thus, one cannot exclude the possibility that clone CA11 was generated by internal oligo(dT) priming at this site. However, sequencing (one strand only) further downstream did not reveal additional AAUAAA motifs within the next 350 base pairs. Also, the sequence immediately following the two AAUAAA motifs has similarity to sequences selected as binding sites for CstF, the factor recognizing the downstream element of polyadenylation signals (Katrin Beyer, pers. comm.). Therefore, clone CA11 probably represents the true 3′-end of the PAB II mRNA.

Molecular mass of PAB II

The molecular mass of PAB II has previously been estimated as 50 kDa based on SDS gel electrophoresis and analytical ultracentrifugation (13). Since the cDNA sequence predicted only 32 766 Da, additional analyses of PAB II were performed. The N-terminus of the protein appeared to be blocked. Mass spectrometry of HPLC-purified PAB II gave a molecular mass of 32.8 kDa, in agreement with the prediction. Peptides generated by CNBr cleavage at methionine residues were analyzed by sequencing and mass spectrometry. One peptide corresponded to the N-terminal fragment judged from its mass of 12 549 Da (predicted mass: 12 541 Da) and its blocked N-terminus. The mass of a second peak obtained from the same HPLC column fraction was 14 748 Da, close to that expected for an N-terminal peptide generated by cleavage at the second methionine (14 745 Da). Of the four small internal fragments predicted by the cDNA, two were identified by sequencing. The N-terminal sequence of the predicted C-terminal CNBr peptide was also found. Its mass (16 313 Da) was larger than calculated from the sequence (15 825 Da).

Sequences of tryptic peptides derived from the C-terminal CNBr fragment also agreed with the cDNA. One sequence, XGXAXAT, matched a peptide generated from the extreme C-terminus presumably by chymotryptic cleavage after tyr^{293}. At all three positions marked X, identical sequence signals were obtained that did not match any of the standard PTH-amino acids. Since the cDNA predicts arginines at all three positions, the unidentified amino acids were probably modified arginine residues. The C-terminus of PAB II is very arginine-rich and modifications of these amino acids may account for the discrepancy between predicted and measured molecular mass. The difference between the number of arginine residues predicted by the cDNA sequence (28 residues) and the number found by quantitative amino acid analysis (14 residues) suggests that as many as one-half of all arginines may be modified.

Expression of PAB II

The PAB II cDNA was inserted into an expression vector such that the protein, using the authentic translation initiation codon, was expressed under the control of a bacteriophage T7 promoter. Upon induction, analysis of bacterial lysates revealed a polypeptide that reacted with a PAB II-specific antiserum and comigrated with purified PAB II in SDS gels. This protein was not detectable in control lysates from bacteria carrying the empty vector (data not shown).

Poly(A) binding protein II was purified from the lysate by a series of four chromatographic steps (see Materials and Methods). The purification was monitored by Western blotting with PAB II-antiserum and by poly(A) binding assays. The profile of the final column (Fig. 4A) shows the peak of poly(A) binding activity. In the SDS gel, a protein comigrating with authentic PAB II can be seen in the same fractions (Fig. 4B). This protein reacted with PAB II-antiserum (data not shown). Its specific activity in poly(A) binding was similar to that of authentic protein. The same protein also stimulated the extension of an oligoadenylated substrate RNA with the same efficiency as authentic protein and showed proper poly(A) tail length control (Fig. 5).

Native structure of PAB II

With the monomeric molecular mass of PAB II now established as 32.8 kDa, the previous measurement of 50 kDa in analytical ultracentrifugation (13) is inconsistent with both a monomeric and a dimeric structure. Thus, the measurement was repeated. The sedimentation coefficient was 2.4 S, close to the earlier measurement of 2.2 S (13). In equilibrium runs, measurements at low protein concentration ($A_{273} = 0.05$, ~2.5 μM protein) close to the meniscus of the cell gave a minimum molecular mass of 34 kDa. The molecular mass increased with increasing protein concentration to a value of 70 kDa at $A_{273} = 0.25$ (~12.5 μM) near the cell bottom. This value is close to that expected for a dimer, but the molecular mass appeared to increase further with protein concentration, indicating a further association as noted before (13). A control run in 6 M guanidinium hydrochloride gave a molecular mass of 30.4 kDa. It thus appears that, under native conditions, there was an equilibrium between monomers, dimers and higher oligomers. The agreement between the previously obtained number of 50.3 kDa and the apparent molecular mass in SDS-polyacrylamide gel electrophoresis was fortuitous. Although some sample heterogeneity was noticed previously in sedimentation velocity runs (13), the full extent of heterogeneity was not obvious, suggesting that shape and mass changes compensated for each other.

As an independent assay of the native structure of PAB II, chemical cross-linking experiments were carried out. The highest yield of cross-linked PAB II was obtained with DSS and DTSP

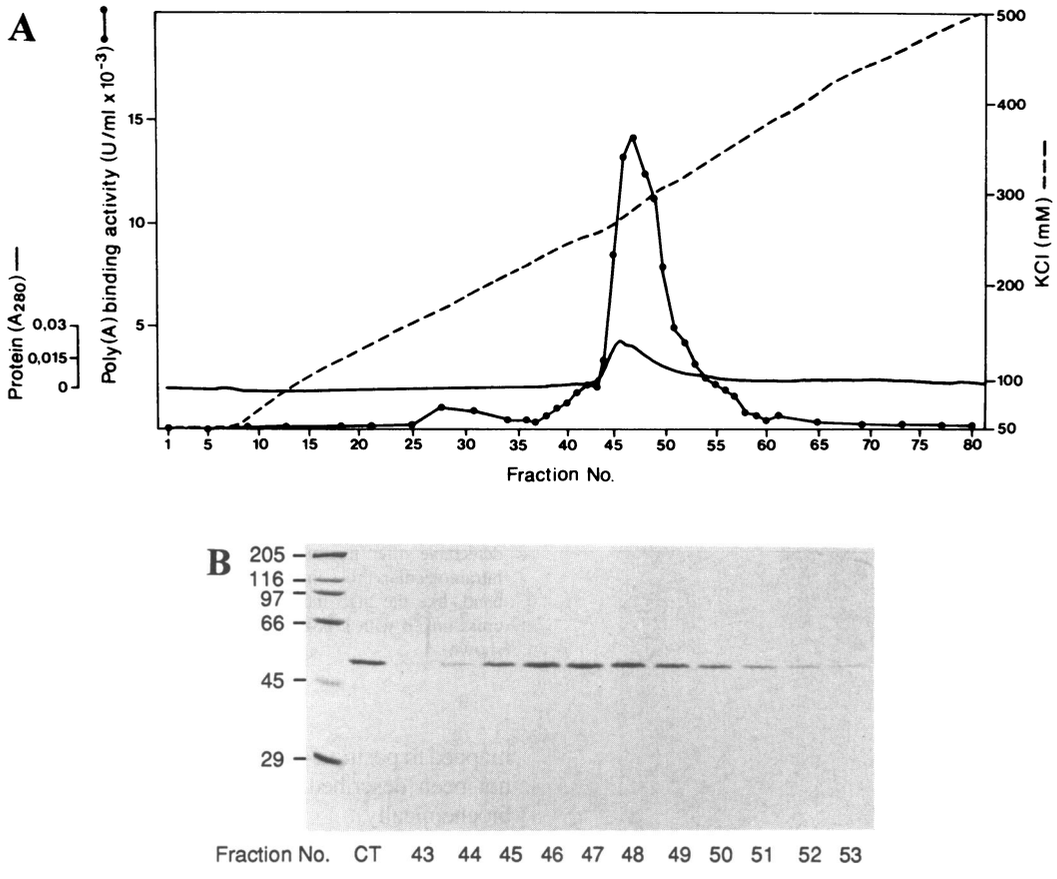

Figure 4. Purification of *E.coli*-expressed PAB II. (**A**) Profile of the final Mono S column. Poly(A) binding activity was measured by nitrocellulose filter binding assays. (**B**) 10% SDS-polyacrylamide gel of the Mono S peak fractions. In the first lane next to the marker (CT), 0.5 µg of PAB II purified from calf thymus was loaded. The following lanes each contain 10 µl of the fractions indicated at the bottom. Proteins were detected by Coomassie Brilliant Blue staining. The molecular masses of marker proteins (in kDa) are indicated on the left.

(Fig. 6), but the same species were also seen with DMS (data not shown). Up to six different cross-linked species were detectable, but this was apparently limited by the resolution of the gel. Crosslinks were observed at 3.4 and 0.34 µM protein. With a monomeric control protein, *E.coli* indoleglycerolphosphate synthetase-phosphoribosylanthranilate isomerase (trpC gene product), no crosslinks were observed under the same conditions, except a minute amount of aggregated material remaining at the top of the gel. The control protein was also not cross-linked to PAB II when the two were mixed. Cross-links generated with the S-S bridge containing, cleavable DSP were fully reversible by boiling in the presence of β-mercaptoethanol (data not shown).

DISCUSSION

We have cloned a cDNA for bovine PAB II. The identity of the encoded protein was confirmed by expression of an active protein in *E.coli*. The amino acid sequence of PAB II contains a clear match to the RNP-type RNA-binding domain (32), extending approximately from glu^{164} to thr^{249}. The highly conserved RNP I and RNP II motifs are found at amino acids 213–220 and 174–179, respectively. They were present in two of the sequenced peptides, as noted earlier (12). The N-terminal part of the protein (amino acids 1–163) is extremely acidic, containing 36 acidic amino acid residues, mostly glutamate, compared to 51 for the total protein. The 12 basic amino acids in the acidic N-terminal domain occur in three patches. The C-terminus (amino acids 250–306), in contrast, is very basic, containing asp^{258} as the single acidic residue, but 14 basic residues, all of them arginines. Many glutamate and arginine residues occur in stretches where one of them alternates with another amino acid, often proline, glycine or alanine. In its overall organization, with an acidic N-terminal domain containing basic patches, followed by an RNA-binding domain and an arginine-rich C-terminal domain, PAB II is reminiscent of the nucleolar RNA-binding protein nucleolin (33,34).

A substantial portion of arginine residues is probably modified. The most common modified form of arginine is dimethylarginine, which has been found in a number of RNA-binding proteins, including the C-terminal RGG-domain of nucleolin (35–37) and hnRNP proteins (38). Attempts to identify dimethylarginine in total hydrolysates of PAB II have been negative so far, and it should be noted that the arginine-rich domain of PAB II does not resemble a standard RGG-domain (32). The preliminary analysis of *E.coli*-expressed PAB II suggests that the arginine modification is not essential for the protein's function.

Six stretches of oligo(A) in the 3′-UTR of the PAB II message are potential binding sites for PAB II and suggest the possibility of autoregulation of PAB II expression.

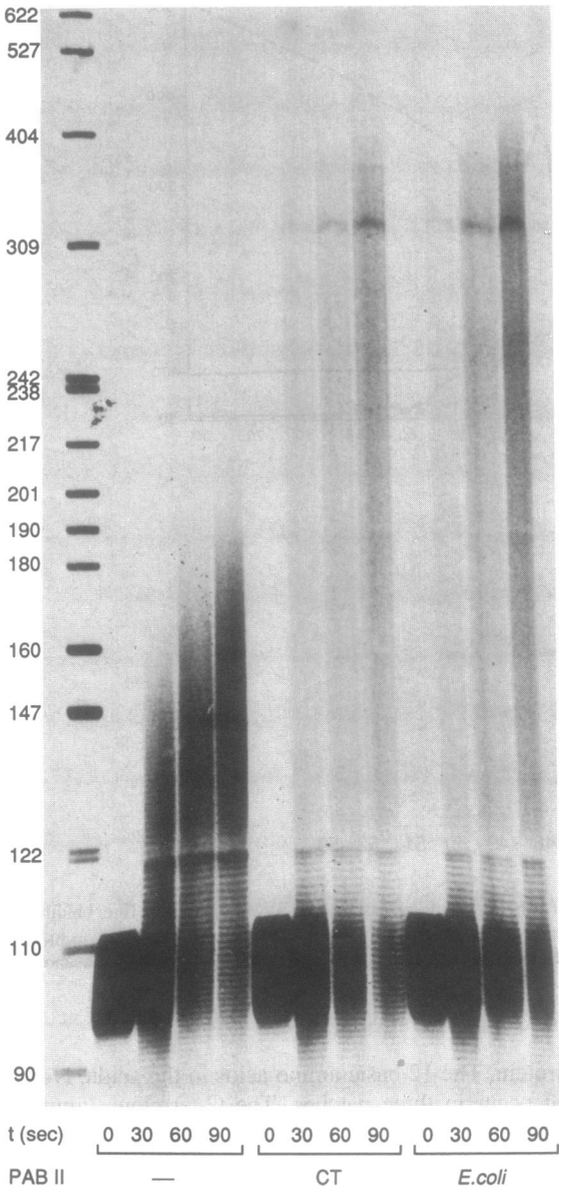

Figure 5. Reconstitution of processive polyadenylation of an oligoadenylated RNA with *E.coli*-expressed PAB II. The assay was carried out with 9 fmol poly(A) polymerase, 100 fmol CPSF, 80 fmol oligoadenylated L3pre-A_{40} substrate and with or without 200 fmol PAB II purified from calf thymus (CT) or *E.coli* (fr. 47 of the column shown in Fig. 4). Reaction mixtures were assembled on ice in the absence of ATP and prewarmed to 37°C before polyadenylation was initiated by the addition of ATP. At the times indicated, samples were withdrawn into SDS-containing buffer. After purification, the reaction products were separated on a denaturing 6% polyacrylamide gel.

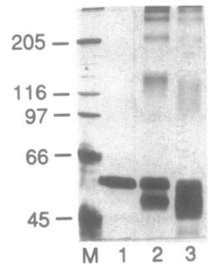

Figure 6. Chemical cross-linking of PAB II. 4 µl of a stock solution (17 µM) of PAB II purified from calf thymus was mixed with 15 µl buffer (50 mM Hepes-KOH, pH 8.2, 50 mM KCl, 10% glycerol) and 1 µl dimethylsulfoxide (negative control; lane 1), with 15 µl buffer and 1 µl DSS (lane 2) or with 14 µl buffer and 2 µl DTSP (lane 3). After incubation at room temperature for 10 min, 2 µl 2 M ethanolamine was added to each sample. One-half of each sample was mixed with SDS gel loading buffer lacking β-mercaptoethanol and analyzed on a 6% SDS-polyacrylamide gel. Markers were run in lane M. Their molecular masses (in kDa) are indicated on the left. The faster migrating species detectable after treatment with DSS or DTSP presumably resulted from intramolecular cross-linking. After treatment with β-mercaptoethanol, this band, like the high molecular weight products, disappeared in the sample cross-linked with DTSP but not in the one cross-linked with DSS (data not shown).

Data base searches revealed several 'expressed sequence tags' encoding the human homologue of PAB II. One of these clones has been mapped to Xq12–q13 and 14q11 (39). The presence of PAB II sequences on two human chromosomes may be related to the fact that three different classes of genomic clones were isolated from a bovine library. One or two additional gene(s) (or pseudogenes) might also explain the 3.3 kb band in the Northern blot (Fig. 2). The data base also contains a *Drosophila* cDNA closely related to bovine PAB II (40). The gene, rox2, was mapped to position 44B on chromosome 2 (41), but mutants have not been described and the protein has not been investigated biochemically.

The PAB II cDNA sequence led to a revision of the protein's molecular mass and native structure. The previous estimate of 50 kDa and a monomeric structure was based on the fortuitous agreement between the protein's anomalous migration in SDS gels and ultracentrifugation experiments under conditions of an association equilibrium. A calculation from sedimentation coefficient and Stoke's radius determined by gel filtration gave the correct molecular mass of 33 kDa, but was considered unreliable since the protein appeared to interact with gel filtration resins (13). Ultracentrifugation now suggests that a large fraction of PAB II may associate to dimers and larger structures. Cross-linking studies confirmed oligomerization of the protein. Electron microscopy of PAB II-poly(A) complexes suggests that oligomerization of the protein is functionally significant (R. Keller, M. Aragon, E. Wahle and D. Bear, unpublished data). The native structure of PAB II and the concentration dependence of dimerization and oligomerization have to be studied further.

A number of calculations (13), based on the wrong molecular mass estimate for PAB II, can now be corrected. The high frictional coefficient of 1.9–2.0 is no longer valid. The dissociation constants for the PAB II-oligo(A) interaction are probably valid since the ultracentrifugation data suggest that PAB II is monomeric at the concentrations used for filter-binding assays. Stoichiometric binding experiments appeared to show the binding of two molecules of oligo(A) per protein monomer (13). With the revised estimate of the molar concentration, this number is reduced to 1.3 mol oligo(A) per mol PAB II, consistent with a 1:1 ratio. A packing density of 23 nucleotides/PAB II monomer, based on fluorescence quenching data, can now be recalculated as 15 nucleotides/monomer. This last number presents a problem, since the number of retarded complexes formed in gel retardation experiments with poly(A) of defined size and limiting amounts of PAB II was not readily compatible with the independent binding

of monomers each covering 15 nucleotides. The data might be reconciled with binding occurring in the form of PAB II dimers. From the dependence of the binding affinity on the length of oligo(A), the binding site size was previously estimated as 12 nucleotides (13). A more rigorous quantitative evaluation (42) of the same set of data leads to an estimate of 9–10 nucleotides. This is similar to site sizes found for other RNP proteins (43).

Knowledge of the primary structure of PAB II and the availability of recombinant protein now offer the possibility of investigating the PAB II–poly(A) interaction and the protein's role in poly(A) synthesis in detail.

ACKNOWLEDGEMENTS

We thank Drs Peter Koch, Werner W. Franke and Sigfried Ruppert for libraries, Halina Szadkowski and Dr Kaspar Kirschner for trpC protein, Thomas Enzler for screening cDNA libraries, Georges Martin for the isolation of ZAB6 and for advice, Dr Walter Keller for support and Dr Jürgen Engel and Ursula Rüegsegger for criticism of the manuscript. This work was supported by the Kantons of Basel and the Schweizerischer Nationalfonds. E. W. was the recipient of a Heisenberg fellowship from the Deutsche Forschungsgemeinschaft, Germany.

REFERENCES

1 Keller, W. (1995) *Cell*, **81**, 829–832.
2 Manley, J. L. (1995) *Curr. Opin. Gen. Dev.*, **5**, 222–228.
3 Wahle, E. (1995) *Biochim. Biophys. Acta*, **1261**, 183–194.
4 Brawerman, G. (1981) *CRC Crit. Rev. Biochem.*, **10**, 1–38.
5 Sachs, A. B. (1993) *Cell*, **74**, 413–421.
6 Wahle, E. (1991) *J. Biol. Chem.*, **226**, 3131–3139.
7 Raabe, T., Bollum, F. J. and Manley, J. L. (1991) *Nature*, **353**, 229–234.
8 Wahle, E., Martin, G., Schiltz, E. and Keller, W. (1991) *EMBO J.*, **10**, 4251–4257.
9 Bienroth, S., Wahle, E., Suter-Crazzolara, C. and Keller, W. (1991) *J. Biol. Chem.*, **266**, 19 768–19 776.
10 Murthy, K. G. K. and Manley, J. L. (1992) *J. Biol. Chem.*, **267**, 14 804–14 811.
11 Keller, W., Bienroth, S., Lang, K. M. and Christofori, G. (1991) *EMBO J.*, **10**, 4241–4249.
12 Wahle, E. (1991) *Cell*, **66**, 759–768.
13 Wahle, E., Lustig, A., Jenö, P. and Maurer, P. (1993) *J. Biol. Chem.*, **268**, 2937–2945.
14 Bienroth, S., Keller, W. and Wahle, E. (1993) *EMBO J.*, **12**, 585–594.
15 Wahle, E. (1995) *J. Biol. Chem.*, **270**, 2800–2808.
16 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.
17 Jenny, A., Hauri, H.-P. and Keller, W. (1994) *Mol. Cell. Biol.*, **14**, 8183–8190.
18 Del Sal, G., Manfioletti, G. and Schneider, C. (1988) *Nucleic Acids Res.*, **16**, 9878.
19 Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
20 Pearson, W. R. and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
21 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22 Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156–159.
23 Jacobson, A. (1987) In Berger, S. L. and Kimmel, A. R. (eds), *Methods in Enzymology. Guide to Molecular Cloning Techniques*. Academic Press, Inc., San Diego, Vol. 152, pp. 254–261.
24 Bradford, M. M. (1976) *Anal. Biochem.*, **72**, 248–254.
25 Laemmli, U. K. (1970) *Nature*, **227**, 680–685.
26 Kyse-Andersen, J. (1984) *J. Biochem. Biophys. Methods*, **87**, 203–209.
27 Jenö, P., Mini, T., Moes, S., Hintermann, E. and Horst, M. (1995) *Anal. Biochem.*, **224**, 75–82.
28 Henikoff, S. (1984) *Gene*, **28**, 351–359.
29 Tabor, S. (1990) In Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (eds), *Current Protocols in Molecular Biology*. Current Protocols, Brooklyn, New York, Vol. 2, pp. 16.2.1–16.2.11.
30 Studier, F. W. (1991) *J. Mol. Biol.*, **219**, 37–44.
31 Krause, S., Fakan, S., Weis, K. and Wahle, E. (1994) *Exp. Cell Res.*, **214**, 75–82.
32 Burd, C. G. and Dreyfuss, G. (1994) *Science*, **265**, 615–621.
33 Olson, M. O. J., Kirstein, M. N. and Wallace, M. O. (1990) *Biochemistry*, **29**, 5682–5686.
34 Lapeyre, B., Bourbon, H. and Amalric, F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1472–1476.
35 Lapeyre, B., Amalric, F., Ghaffari, S. H., Venkatarama Rao, S. V., Dumbar, T. S. and Olson, M. O. J. (1986) *J. Biol. Chem.*, **261**, 9167–9173.
36 Najbauer, J., Johnson, B. A., Young, A. L. and Aswad, D. W. (1993) *J. Biol. Chem.*, **268**, 10 501–10 519.
37 Rajpurohit, R., Ok Lee, S., Ok Park, J., Ki Paik, W. and Kim, S. (1994) *J. Biol. Chem.*, **269**, 1075–1082.
38 Liu, Q. and Dreyfuss, G. (1995) *Mol. Cell. Biol.*, **15**, 2800–2808.
39 Lamartine, J., Wang, Q., Kaneko, K., Tsuji, S., Mattei, M., Lenoir, G. and Sylla, B. S., unpublished (EMBL accession no. U12206).
40 Brand, S., Pichoff, S. and Bourbon, H. M. G., unpublished (EMBL accession no. L34934).
41 FlyBase (1994) *Nucleic Acids Res.*, **22**, 3456–3458.
42 de Haseth, P. L. and Uhlenbeck, O. C. (1980) *Biochemistry*, **19**, 6138–6146.
43 Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. and Nagai, K. (1994) *Nature*, **372**, 432–438.