Nucleic Acids Research, 1996, Vol. 24, No. 14 **2799–2806**

Characterization of the single-strand-specific BPV-1 origin binding protein, SPSF I, as the HeLa Purα factor

Marion Jurk, Florian Weissinger[+], Friedrich Lottspeich[§], Ursula Schwarz and Ernst-Ludwig Winnacker*

Institut für Biochemie der Ludwig-Maximilians-Universität München, Genzentrum, Würmtalstraße 221, D-81375 München, Germany

Received February 12, 1996; Revised and Accepted May 24, 1996

ABSTRACT

SPSF I and II are two cellular proteins which bind specifically to single-stranded DNA. SPSF I and II binding sites are found in the minimal origin of replication of BPV-1 DNA and near the P2 promoter of the cellular *c-myc* gene. DNA-binding properties of the two proteins to single-stranded oligonucleotides of different lengths and sequences were quantified by determination of DNA-binding constants. The binding constant of SPSF proteins to the lower strand of the BPV-1 origin was determined to be 1.5×10^{-10} M^{-1}. Peptide sequences derived from purified SPSF I and II revealed the identity of at least one of the SPSF proteins with the so-called HeLa Purα factor. The HeLa Purα factor was identified previously by virtue of its capacity to bind to purine-rich strands of the PUR element found in initiation zones of DNA replication [Bergemann,A.D., Ma,Z.-W. and Johnson,E.M. (1992) *Mol. Cell. Biol.* 12, 5673–5682]. Expression of the Pur cDNA confirmed the identity of the Purα protein with the 42 kDa SPSF I protein. Analysis of several Purα cDNA clones revealed the existence of an extended 3′-untranslated region in all Pur mRNAs.

INTRODUCTION

The Bovine Papillomavirus (BPV) type 1 genome persists extrachromosomally in transformed rodent cells with a copy number of ~200 per cell. Its replication requires a minimal origin of replication located between nucleotides 7911 and 27 of the viral genome as well as the presence of the two viral proteins defined by reading frames E1 and E2 (1,2,3,4). In addition, replication of BPV-1 DNA also depends on the cellular replication machinery. Therefore, this system is particularly well-suited to obtain additional insights into the mechanisms of eukaryotic DNA replication. Currently the processes controlling the initiation of DNA replication are poorly understood. Factors which are involved in initation of DNA replication are only known in *Escherichia coli, Saccharomyces cerevisiae* and several viral systems (for reviews see 5,6). Binding of these factors to origins of DNA replication seems to be an obligatory step for targeting cellular replication proteins to the initiation zone.

We have recently isolated two closely related cellular proteins which bind specifically to single-stranded DNA oligonucleotides derived from one strand of the minimal origin of BPV-1 DNA replication (7). These two proteins were designated SPSF I and II for specific single-stranded DNA-binding factors. The two proteins also interact specifically with sequences from the so-called Plasmid Maintenance Sequence (PMS) domain 1 of the BPV genome (8). The function of this PMS domain is still uncertain and requires further characterization (9,10). A cellular binding site for SPSF I and II was found in the P2 promoter of the human *c-myc* gene (7).

The presence of protein binding sites in the vicinity of viral and cellular origins of DNA replication is not uncommon (6). Binding sites for single-stranded DNA-binding proteins include the T-rich stretches of the *S.cerevisiae* ARS sequences, which bind the protein ACBP-60 (11,12) and the purine-rich strand of the inverted repeat found in the SV40 origin core sequence, which preferentially binds a cellular factor, IRF-B (13).

Initial experimental evidence suggesting a function of specific single-stranded DNA-binding proteins in DNA replication was obtained through the characterization of the MSSP proteins, which recognize a transcriptional enhancer and DNA replication origin on the human c-*myc* gene. Overexpression of these proteins stimulates the replication of a plasmid containing a mutated SV40 origin in which the A/T rich region was replaced by a MSSP binding site (14,15).

The HeLa cell derived Purα factor binds single-stranded DNA around initiation zones of eukaryotic DNA replication (16,17). It was isolated through its ability to bind to a region of stably-bent DNA located ~1.6 kb upstream of the P1 promoter of the human *c-myc* gene. The Purα protein binds preferentially to the purine-rich single strand of the so called PUR-element. This PUR-element is found in several mammalian zones of initiation of DNA replication, e.g. in the *dhfr*-gene, the *c-myc*-gene and the

* To whom correspondence should be addressed

Present addresses: [+]Medizinische Poliklinik der Justus-Maximilians-Universität Würzburg, Klinikstraße 8, D-97070 Würzburg, Germany and [§]Max-Planck-Institut für Biochemie, Am Klopferspitz 18a, D-82152 München, Germany

β-*globin* gene locus (18,19). Other Pur binding sites are located in the lytic control element (LCE) of the JC polyomavirus as well as in the myelin basic promoter region (20,21). Bergemann *et al.* (17) identified another cDNA fragment coding for a related protein, Purβ, with significant homology to Purα. The exact role of the Pur proteins in the regulation of initiation of DNA replication and/or transcriptional control is still unknown.

Here we report the purification and characterization of the sequence-specific single-stranded DNA-binding proteins SPSF I and II from calf thymus. Sequence determination of SPSF I and II peptides revealed the identitiy of SPSF I with the Purα protein. Expression of the Purα cDNA in heterologous systems and characterization of the DNA-binding properties of the recombinant Purα protein confirmed this identity. We isolated Pur mRNAs containing extended 3′-untranslated regions. The biological significance of these sequences was determined by Northern blot analysis.

MATERIALS AND METHODS

Oligonucleotides used

The oligonucleotides used for DNA-binding and complementation assays are listed in Table 1, except for PGK1 (5′-GATCTAACTTGCATAAATTGGTCAATGCAAGAAAGATAG-3′), U5 (5′-TCGAGTGCATAAATTGTTCGAGTTTTCGCGCTTAATTTGAGAAAGGGCGCGAAACTCGTGTCAATCG-3′) and U6 (5′-TCGAGCATTGACACGAGTTTCGCGCCCTTTCTCAAATTTAAGCGCGAAAACTCGAACAATTTATGCAC-3′). The oligonucleotides PUR6 (nucleotides +444 to +426, 5′-CAGGAACTCGCTTTTGAG-3′) and PUR8 (nucleotides +769 to 751, 5′-TGGAGTTGCGATAGGTGG-3′) used for screening of the cDNA library are derived from the Purα cDNA (GenBank accession no. M96684). The oligonucleotides PUR9 (5′-GATCCTACGCAGCATCATGGCG-3′) and PUR10 (3′-GATGCGTCGTAGTACCG CCTG-5′) were used for cloning as described below. All oligonucleotides used for complementation analyses, gel retardation assays and cloning were purified on denaturing urea polyacrylamide gels.

Gel retardation assay and determination of binding constants

Single-stranded oligonucleotides were 5′-end-labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase. In a standard DNA-binding assay, 5–25 fmol 5′-end-labeled oligonucleotides were incubated with 5–10 fmol of SPSF I and II or ≤2 µg of crude or enriched protein extract. The reaction buffer contained 1 µg poly[d(A–T)], 1 µg BSA (bovine serum albumin), 50 mM HEPES (*N*-2-hydroxymethylpiperazine-*N*′-2-ethanesulfonic acid, pH 7.9), 1 mM DTT (dithiothreitol), 1 mM EDTA and 10% glycerol. The reaction mix was adjusted to 150 mM NaCl and 10 mM $MgCl_2$ and incubated for 5 min on ice. For competition analyses, the competitor DNAs (usually in a 40–100-fold molar excess) and the 5′-end-labeled oligonucleotide were added prior to the addition of the protein extracts. DNA–protein complexes were separated on native 8% polyacrylamide gels, containing 1× TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA).

DNA-binding constants were determined as described (22,23). Briefly, a constant amount of affinity-purified SPSF I and II (10–30 fmol) was incubated with an increasing amount of 5′-end-labeled oligonucleotide. The DNA–protein complexes were separated from uncomplexed DNA on native polyacrylamide gels. The amounts of free and bound DNA were quantified on dried gels by using a PhosphorImager (Fuji BAS 1000 Bio-imaging analyzer).

Purification of SPSF I and II from calf thymus

Purification was carried out as described by Habiger *et al.* (7) with the following modifications. The DEAE fractions containing SPSF I and II were diluted to 250 mM NaCl and applied to a MonoQ column (16/10, Pharmacia, FPLC) which was equilibrated in buffer II (25 mM HEPES–NaOH, pH 7.9, 5 mM DTT, 1 mM EDTA, 10% glycerol) containing 250 mM NaCl (buffer II-250). For optimal results the maximal load on this column was 300 mg protein. The column was washed with 10 vol buffer II-250. A linear gradient (100 ml) from 250 to 1000 mM NaCl was appplied and the fractions collected. The SPSF proteins eluted at 500–600 mM NaCl. Active fractions were pooled and diluted with buffer II (without salt) to a final concentration of 250 mM NaCl. $MgCl_2$ was added to a final concentration of 10 mM. These fractions were applied to a 5 ml HiTrap™-heparin–Sepharose column (Pharmacia) equilibrated with buffer II-250 containing 10 mM $MgCl_2$. After washing the column with 10 vol buffer II-250 (+10 mM $MgCl_2$), the proteins were eluted by applying a linear gradient (75 ml) from 250 to 2000 mM NaCl. SPSF I and II eluted at 900 mM NaCl.

DNA affinity columns were prepared by coupling the single-stranded 5′-biotinylated oligonucleotide P-l to streptavidin–agarose (Sigma). The coupling buffer was 0.2 M ammonium acetate and the coupling reaction was carried out overnight at 4°C. The column was washed extensively with buffer II-250 (+10 mM $MgCl_2$). The pooled heparin–Sepharose fractions were diluted and applied to the DNA affinity column. The column was washed with 10 vol buffer II-250 (+10 mM $MgCl_2$), 3 vol buffer II-600 (+10 mM $MgCl_2$). SPSF proteins were eluted from the column in a single step using buffer II containing 2 M NaCl (+10 mM $MgCl_2$). In order to obtain homogenous preparations of SPSF I and II the DNA affinity step was repeated. Purified SPSF proteins could be stored for several weeks at 4°C without losing their binding activity. For long time storage the proteins were kept at –80°C.

To test for possible posttranslational modifications, ~50 ng of the affinity-purified proteins were precipitated with acetone–methanol (1:1) at –80°C overnight. The pellet was then resuspended in 10 µl water. The proteins were denatured and treated with 0.4 U *N*-glycosidase F (Boehringer-Mannheim) as described previously (24). The *N*-glycosylated protein alkaline phosphatase treated in the same way was used as positive control.

To test for phosphorylation of the SPSF proteins, 20 ng of the purified proteins were incubated with various amounts (1–100 U) alkaline phosphatase (Boehringer-Mannheim) in buffer containing 50 mM Tris–HCl, pH 8.5 and 1 mM EDTA. After incubation at 37°C the DNA-binding properties of the SPSF proteins were tested in a gel retardation assay.

Renaturation of the proteins

Affinity-purified proteins were loaded on a 12% SDS gel (25). An aliquot of the affinity-purified proteins was loaded separately in a marker lane. Following electrophoresis, the marker lane was cut out of the gel and stained with Coomassie Blue. The lanes containing SPSF I and II were then cut into fractions containing (i) material from above the 42 kDa band, (ii) the 42 kDa band, (iii)

the 39 kDa band, and (iv) material from below the 39 kDa band. The gel slices were suspended in elution buffer, denatured with urea, and renatured as described previously (7). 1/25 of the renatured protein fractions were tested in a standard DNA-binding assay.

Cloning of the Purα cDNA

A HeLa cDNA library in pλDR2 (26) was screened with oligonucleotides PUR6 and PUR8 as described previously (27,28). Positive phages were plaque-purified and the vector pDR2 containing the cDNA insert was isolated as described by Murphy *et al.* (26). The 1.1 kb cDNA insert was isolated and partially digested with *Ava*II. This cDNA fragment (containing the nucleotides +6 to +1085 of the Purα cDNA, GenBank accession no. M96684) was ligated to the double-stranded oligonucleotide pair Pur9/10 which contains the 5′-untranslated region from nucleotide –9 to +5 of the Purα cDNA as well as a 5′-overlapping, *Bam*HI compatible end and a 3′-*Ava*II site. The ligation product (designated as Pur 9/10) was cloned in vectors pGEM-4Z (Promega), pVL1393 (Pharmingen) and pGEX-3X (Pharmacia). Correct orientations or open reading frames respectively, were confirmed by DNA sequencing (29).

Expression of the Purα cDNA in the Sf9/baculovirus system

Spodoptera frugiperda cells (Sf9, ATCC CRL 1711) were grown at 27°C as monolayers in TC100 medium (Gibco-BRL) supplemented with 10% fetal calf serum. Recombinant baculoviruses were obtained by cotransfection of the recombinant vector pVL1393 containing the Pur9/10 cDNA and a specifically modified wild-type AcNPV DNA (30) using the BaculoGold™ transfection kit (Pharmingen). Purification of the recombinant viruses as well as infection of Sf9 cells was performed as previously described (31,32). Whole cell extracts of Sf9 cells (48 h post-infection) were prepared by resuspending the cell pellet in five pellet volumes of ice-cold buffer (25 mM HEPES–NaOH, pH 7.9, 5 mM DTT, 1 mM EDTA, 400 mM NaCl, 10% glycerol, 0.1% Triton X-100 and 0.5% Nonidet P-40). Following a 20-min incubation period on ice, the insoluble components were separated by centrifugation and the cell extracts diluted and purified as described above.

Northern blot analysis

A commercially available human multiple tissue Northern blot (Clontech) was hybridized with radioactive probes using the ExpressHyb Hybridization system (Clontech). Hybridization temperature for the 560 bp *Ava*II–*Ear*I fragment was 60°C, for the 777 bp *Pst*I fragment 68°C. The blot was washed as recommended by the supplier and exposed for 1–2 days at –80°C using two intensifying screens.

RESULTS

SPSF I and II are two proteins of 42 and 39 kDa molecular mass which copurify but bind DNA independently of each other

SPSF I and II are ubiquitous proteins found in mammalian cell lines. Crude extracts of calf thymus were shown to be an excellent source of the SPSF I and II proteins. Purification of the two proteins required four different chromatographic steps, including a heparin–Sepharose column purification step. The final purification step was a DNA affinity column carrying the single-stranded oligonucleotide P-l (Fig. 1A). Using this purification scheme we obtained 150 µg SPSF I and II from 500 g of calf thymus, resulting in a 1400-fold enrichment of the two proteins.

Affinity-purified fractions of SPSF I and II always displayed two bands at 39 and 42 kDa upon SDS gel electrophoresis (Fig. 1A, lane 8).

The molar ratio of the two proteins remained constant during the purification process and during several independent purification experiments. Treatment of the two proteins with *N*-glycosidase F or alkaline phosphatase had no influence on DNA-binding

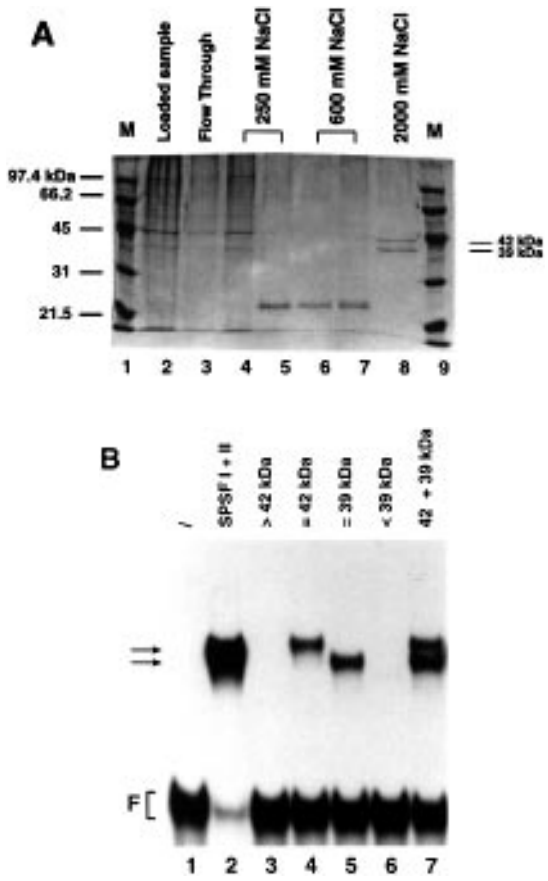

Figure 1. DNA affinity purification and DNA-binding characteristics of SPSF I and II. (**A**) SDS–PAGE of fractions obtained by DNA affinity chromatography in the presence of single-stranded oligonucleotide P-l. Twenty microlitres of each fraction (1/250 of fraction volumes) were separated on a 12% SDS gel and stained with Coomassie Blue (lane 2, loaded heparin–Sepharose fractions; lane 3, flow through; lanes 4–8, fractions eluted with buffer containing sodium chloride, as indicated). The affinity-purified proteins display two bands at 42 and 39 kDa respectively (lane 8). M; protein size marker in kDa (lanes 1 and 9). (**B**) DNA-binding activity of renatured SPSF proteins. Approximately 80 ng of the affinity-purified proteins were separated on an SDS gel, eluted and renatured as described in Materials and Methods. Material from above the 42 kDa band (lane 3) and below the 39 kDa band (lane 6) was eluted and treated analogously as negative controls. The renatured 42 kDa protein (SPSF I, lane 4) and 39 kDa protein (SPSF II, lane 5) were incubated in a gel retardation assay with 25 fmol 5′-end-labeled oligonucleotide ORI-l. In lane 2 the native affinity-purified proteins were incubated with the labeled oligonucleotide. F; free oligonucleotide, lane 1: 5′-end-labeled oligonucleotide in the absence of protein.

or migration behaviour upon SDS gel electrophoresis. Post-translational modifications such as *N*-glycosylation or phosphorylation are therefore unlikely.

In order to prove that the two bands visible in the SDS gel possessed the DNA-binding properties characteristic for SPSF I and II, the two bands were eluted from the SDS gel, renatured and tested for DNA-binding. As shown in Figure 1B, the two renatured proteins bind to 5′-end-labeled single-stranded oligonucleotides in the same manner as the native affinity-purified proteins SPSF I and II (Fig. 1B, lane 2). The two proteins thus interact independently with DNA. The 42 kDa protein SPSF I forms the slower migrating DNA–protein complex (Fig. 1B, lane 4) whereas the faster migrating DNA–protein complex contains the smaller 39 kDa protein SPSF II (Fig. 1B, lane 5). Co-incubation of the two renatured proteins with labeled oligonucleotides does not influence the DNA-binding characteristics (Fig. 1B, lane 7).

DNA-binding properties and DNA-binding constants of proteins SPSF I and II

SPSF I and II bind specifically to single-stranded oligonucleotides. The affinity-purified proteins from calf thymus were analyzed in detail with respect to their binding behaviours towards DNA molecules of different length and composition. In previous work we had shown that SPSF I and II bound to single-stranded oligonucleotides containing the lower strand of the origin of DNA replication of BPV-1 (ORI-l) or to the so-called PMS region [plasmid maintenance sequence, (8)]. In a first set of experiments we studied the specifity of DNA-binding through competition experiments. As shown in Figure 2A, binding of SPSF I and II to labeled oligonucleotide ORI-l could be competed most efficiently with the homologous oligonucleotide ORI-l (Fig. 2A, lane 5). DNA-binding could also be competed with the two complementary oligonucleotides containing the lower and upper strands of the PMS region (Fig. 2A, lanes 3 and 4).

By using heterologous oligonucleotides with different length we noted that longer, unrelated oligonucleotides can compete with the specific binding to ORI-l more efficiently than shorter unrelated oligonucleotides. A comparison of the competition data of oligonucleotides PGK1 and PGK1+1, which differ only by the fact that PGK1+1 is a head-to-tail dimer of PGK1 shows that the longer oligonucleotide appears to compete slightly more efficiently. The same effect can be seen with the complementary oligonucleotides PGK2 and PGK2+2. This preference indeed translates into a difference in DNA-binding constants of one order of magnitude (see below and Table 1).

In competition experiments we tested the ability of the two complementary oligonucleotides, ORI-u and ORI-l, representing the upper and lower strand respectively of the BPV-1 origin of replication to compete for the binding of SPSF I and II to labeled oligonucleotide P-l. As shown in Figure 2B, only ORI-l competed the binding to P-l even with a minimal excess whereas a 100-fold excess of ORI-u competed only with minimal effect. These experiments show that SPSF I and II bind specifically and with high affinity to the lower strand of the origin of BPV.

In order to quantify the DNA-binding properties of SPSF I and II we determined the DNA-binding constants of the two affinity-purified proteins to several oligonucleotides (Table 1).

Due to the fact that the two proteins always copurify, it was not possible to determine DNA-binding constants for the separate proteins. All measured binding constants therefore refer to both proteins. The data can be summarized as follows: (i) binding to the oligonucleotides ORI-l and P-l occurs with binding constants on the order of 10^{10} M^{-1}. Binding to the oligonucleotide derived from the corresponding complementary strand was ~5–10-fold less efficient; (ii) binding constants increase with the length of an oligonucleotide; (iii) purine-rich sequences are preferred strongly, as shown with the PUR1 oligonucleotide carrying the consensus sequence of the Purα protein binding site (16, see below). We conclude that the SPSF I and II proteins bind single-stranded DNA with high affinity in a sequence-specific manner with some preference towards longer oligonucleotides.

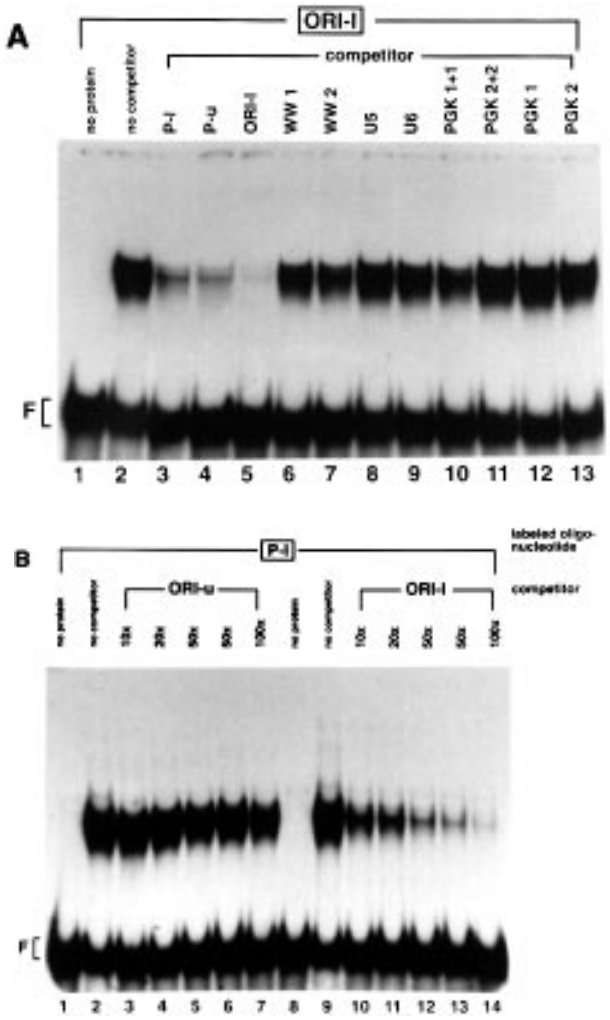

Figure 2. Specific binding of SPSF I and II to the single-stranded oligonucleotides ORI-l, P-l and P-u. (**A**) SPSF I and II enriched heparin–Sepharose fractions (~10 fmol of SPSF I and II) were incubated in a standard binding assay with 25 fmol of 5′-end-labeled oligonucleotide ORI-l in the absence (lane 2) or presence (lanes 3–13) of 1 pmol of non-radioactive oligonucleotides. Sequences of the oligonucleotides are described in Materials and Methods and in Table 1. Lane 1 represents the free unbound oligonucleotide (F). (**B**) Affinity-purified SPSF I and II were incubated with 10 fmol 5′-end-labeled oligonucleotide P-l (lanes 2 and 9). Unlabeled oligonucleotides ORI-u (lanes 3–7) and ORI-l (lanes 10–14) were added in molar excesses as indicated. Lanes 1 and 8 represent the free oligonucleotide P-l.

Table 1. DNA-binding constants of SPSF I and II to various oligonucleotides

Oligonucleotide	Length	Binding constant (M^{-1})
ORI-l		
TTTTTTCCCGCTTGAAAAAACGGTGATGGTGTGATTATTGTTAACAACAATTATTCACTGGGAAAAAATACATA-GTCTTTACTTACCGGTTTCGGTGAGCT	100	1.5×10^{10}
(BPV-1 origin, lower strand, nt 45–7891)		
P-l	96	8.5×10^{9}
AATTCTGTCCTCTAATATCGATGAGGTAGGAGGCTGCACACCTATTTCCCAATTTTCAAGCACAGAGGGCATAA-GTCCTTGCAGATGTGACACAGG		
(BPV-PMS domain 1, lower strand, nt 6849–6760)		
D1	88	4.5×10^{9}
GCGAGGACTTTGTGCTGGCCATCACCGGCAAGAAGGCGCCAGGCTGCGTGCTCTCCAACCCCGACCAGAAG-GGCAAGATGCGCCGCAT		
(pNFI, Dimerization domain, nt 266–353)		
ORI-u	100	3.2×10^{9}
AGCTCACCGAAACCGGTAAGTAAAGACTATGTATTTTTTCCCAGTGAATAATTGTTGTTAACAATAATCACACC-ATCACCGTTTTTTCAAGCGGGAAAAAA		
(BPV-1 origin, upper strand, nt 7891–45)		
P-u	96	1.6×10^{9}
AATTCCTGTGTCACATCTGCAAGGACTTATGCCCTCTGTGCTTGAAAATTGGGAAATAGGTGTGCAGCCTCCTA-CCTCATCGATATTAGAGGACAG		
(BPV-PMS domain 1, upper strand, nt 6760–6849)		
WW1	78	4.0×10^{8}
AATTACATATGGCAGCTAAGCTTGAAGACCGGCCGGCGCGCCATGGTCTCGCGAATTCCGGACGCGTAGCTGC-TCTAG		
(Polylinker-oligonucleotide with various restriction sites)		
PGK1+1	78	3.7×10^{8}
GATCTAACTTGCATAAATTGGTCAATGCAAGAAAGATAGGATCTAACTTGCATAAATTGGTCAATGCAAGAAA-GATAG		
(Dimer of promoter of the phosphoglycerate-kinase gene in S.c., upper strand, nt –121 to –145)		
PUR1	24	3.0×10^{8}
GGAGGTGGTGGAGGGAGAGAAAAG		
[Purine rich motif found by Bergemann and Johnson (1)]		
PGK2+2	78	1.1×10^{8}
GATCCTATCTTTCTTGCATTGACCAATTTAAGCAAGTTAGATCCTATCTTTCTTGCATTGACCAATTTAAGCAAG-TTA		
(Dimer of PGK2)		
BPV2	43	1.0×10^{8}
GATCTTGATGGTGTGATTATTGTTAACAACAATTATTCACTGG		
(BPV-1 origin, lower strand, nt 22–7931)		
PGK2	39	1.8×10^{7}
GATCCTATCTTTCTTGCATTGACCAATTTAAGCAAGTTA		
(promoter of the phosphoglycerate-kinase gene in S.c., upper strand, nt –121 to –145)		
BPV1	43	1.4×10^{7}
GATCCCAGTGAATAATTGTTGTTAACAATAATCACACCATCAA		
(BPV-1 origin, upper strand, nt 7931–22)		

A constant amount of affinity-purified SPSF I and II was incubated with increasing amounts of 5′-end-labeled oligonucleotide. The ratio of bound and unbound DNA was quantified using a PhosphorImager and the binding constants were determined as described (22,23)

SPSF I is identical to the HeLa Purα factor

Approximately 30 µg (each) of the affinity-purified proteins were separated on a preparative SDS gel, eluted and digested with endopeptidase Lys-C. Three peptides derived from the 42 kDa (SPSF I) and two peptides from the 39 kDa protein (SPSF II) were sequenced. Both peptides derived from the 39 kDa protein were identical to those from the 42 kDa protein. All three peptide sequences were present within the amino acid sequence of the HeLa Purα factor described by Bergemann *et al.* (17). The Purα cDNA was originally obtained by screening an expression library from human fetal liver for proteins binding to a purine-rich single-stranded oligonucleotide. This PUR element was found upstream of the *c-myc* gene and in the vicinity of initiation zones of DNA replication. SPSF I and II bind to the oligonucleotide (PUR1) used by Bergemann *et al.* (16,17). The binding is specific

and cannot be competed with heterologous oligonucleotides (data not shown). The measured binding constant was 3×10^8 M^{-1} and is significantly higher than binding constants observed with unrelated longer oligonucleotides (e.g. BPV1, PGK2) indicating the preferred binding of the SPSF or Pur proteins respectively to purine-rich sequences.

In order to establish the identity of the SPSF proteins with the Purα factor, we screened 600 000 clones from a HeLa cDNA library with two oligonucleotides derived from the Purα cDNA sequence as described in Materials and Methods. We isolated and sequenced two identical cDNA clones with a 1.1 kb insert. The isolated cDNAs were identical with the cDNA described previously (nucleotides –49 to +1085) by Bergemann *et al.* (17) except for one nucleotide exchange in the 3′-untranslated region (position +976, T instead of A). To prove the exact relationship between the SPSF and Pur proteins the Purα cDNA was expressed in various heterologous expression systems and the DNA-binding properties of the synthesized proteins were determined. To simplify cloning and expression of the cDNA the 5′-untranslated region of the cDNA was replaced by an oligonucleotide linker as described in Materials and Methods.

In vitro translation of an *in vitro*-transcribed Purα mRNA and subsequent analysis of the ^{35}S-labeled reaction products on SDS–PAGE resulted in a single band at 42 kDa indicating that the Purα factor is identical with the 42 kDa SPSF I protein. DNA-binding studies led to a similar conclusion. Upon incubation of the *in vitro* ^{35}S-labeled Purα protein with unlabeled, single-stranded oligonucleotide ORI-l, a DNA–protein complex was observed in a gel retardation assay comigrating with the SPSF I–DNA complex (data not shown).

The Purα cDNA encodes a protein of 322 amino acids with a theoretical molecular mass of 35 kDa. The fact that the *in vitro*-translated cDNA gave rise to a protein migrating like a protein with a molecular mass of 42 kDa must be due to some peculiar migratory properties of the Purα protein in SDS gels since there is no evidence of posttranslational modifications.

We also expressed the Purα cDNA as a fusion protein with glutathione-*S*-transferase (GST) using the GST-Vector pGEX-3X. The purified fusion protein displayed one main band upon SDS–PAGE at 70 kDa. The fusion protein also bound to ^{32}P-labeled oligonucleotide ORI-l whereas purified GST displayed no DNA-binding activity (data not shown). These data also support the notion that SPSF I and the Purα protein are identical. Furthermore, the expression of a biologically active GST-Pur protein in *E.coli* suggests that protein modifications specific for eukaryotic cells are not required for DNA-binding of these proteins which is in agreement with our experimental findings (see above).

Final confirmation of the identity of SPSF I with Purα was obtained from the expression of the Purα cDNA in the eukaryotic Sf9/baculovirus system. As shown in Figure 3A, only cell extracts from insect Sf9 cells infected with a recombinant virus (designated as Pur42) encoding the Purα cDNA displayed a DNA-binding activity characteristic for the SPSF proteins. Extracts from uninfected cells and from cells infected with wild-type baculovirus showed no specific single-strand DNA-binding activity.

Purification of the recombinant Purα protein was performed according to the purification protocol for the calf thymus proteins SPSF I and II. The recombinant Purα protein eluted at identical

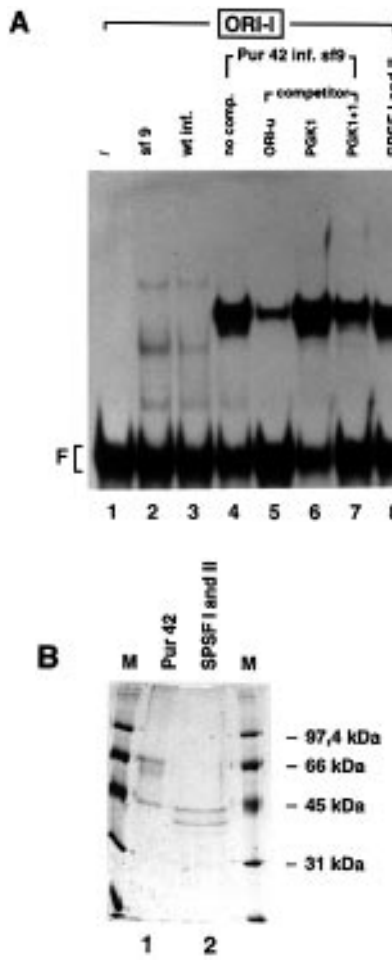

Figure 3. Expression of the Purα cDNA in the Sf9/baculovirus system. (**A**) Gel retardation assay with crude cell extracts. Twenty fmol of 5′-end-labeled oligonucleotide ORI-l was incubated with 0.5 µg of cell extracts from uninfected Sf 9 cells (lane 2), from cells infected with wild-type virus (lane 3) and from cells infected with recombinant Pur42-baculovirus (lanes 4–8). Cell extracts containing the recombinant Purα protein were incubated in the absence (lane 4) and presence of a 100-fold molar excess of unlabeled competitor DNA (lanes 5–8). (**B**) DNA affinity purification of recombinant Purα protein. DNA affinity-purified fraction of recombinant Purα protein (lane 1) and DNA affinity-purified SPSF I and II proteins from calf thymus (lane 2) were separated on a 12% SDS gel and stained with Coomassie Blue. M, protein size standard.

conditions from each column. Affinity-purified fractions displayed a single band upon SDS–PAGE at 42 kDa (with a contaminating band at 67 kDa, probably due to remaining BSA from the fetal calf serum present in the culture medium, Fig. 3B). The second band at 39 kDa always found in preparations of the SPSF proteins from calf thymus was missing. This is in agreement with the gel retardation experiments indicating the presence of a single DNA–protein complex (Fig. 3A, e.g. lane 5). We also determined the DNA-binding constant of the recombinant Purα protein (Pur42) to ORI-l (5×10^9 M^{-1}). This binding constant for Pur42 is in agreement with the theoretical binding constant for one of the two SPSF proteins ($2–3\times10^9$ M^{-1}) calculated under the assumption that both proteins, SPSF I and II, exist in equimolar concentrations and bind DNA equally well.

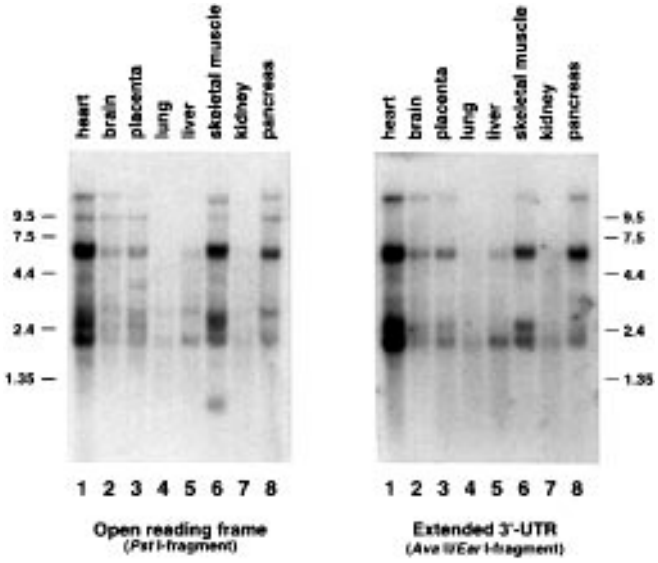

Figure 4. Northern blot analysis using different Pur DNA fragments. The same human multiple tissue northern blot was hybridized either with a 560 bp *Ava*II–*Ear*I fragment specific for the 3′-untranslated region of the Purα cDNA (right) or with a 777 bp *Pst*I DNA fragment derived from the coding region (left). The positions of RNA size marker bands are indicated at the margin of the blot. Lanes 1–8: heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas. Each lane contains 2 µg poly(A)$^+$ RNA.

Purα/SPSFI mRNAs contain 3′-UTRs of different lengths

Analysis of various Purα cDNA clones obtained in screens of different independent HeLa-cDNA libraries [λpDR2, (26) and pJG4–5 (33, 34)] revealed the existence, in several cDNAs, of an extended 3′-untranslated region . cDNA clones possess 3′-UTRs with lengths of either 228 or 282 bp. The longest 3′-UTR found within a Pur cDNA fragment was published under the accession no. X91648 (dated 19 September 1995). This sequence contains 1546 nucleotides in addition to the 115 nucleotides following the stop codon of previously published Purα cDNA (17).

There are five possible polyadenylation signals (AAUAAA or AUUAAA respectively) within this extended 3′-UTR of the Purα cDNA. The finding of multiple poly(A) signals is consistent with the existence of 3′-UTRs of different length.

In order to establish the physiological relevance of this sequence we performed Northern blot analysis using DNA probes derived from different parts of the Purα cDNA. As shown in Figure 4 (left blot) hybridization of a human multiple tissue Northern blot with a 560 bp *Ava*II–*Ear*I fragment derived from the newly found 3′-UTR identified several mRNAs with sizes of 10.6, 5.8, 2.6 and 2.0 kb respectively. We therefore conclude that this extended 3′-UTR is transcribed in all tissues examined. However, the distribution of individual mRNAs seems to differ slightly among the tissues. The ratio of the 2.0 and the 2.6 kb mRNA species is ~1:1 in heart, placenta, skeletal muscle, kidney and pancreas (Fig. 4, lanes 1, 3, 6, 7 and 8) whereas in lung only the 2.0 kb mRNA is present.

To prove that the observed mRNAs are coding for Purα proteins we hybridized the same Northern blot with a DNA probe specific for the coding region of Purα (777 bp *Pst*I fragment). As shown in Figure 4 (right blot) all of the mRNA species hybridizing with the 3′-UTR probe also gave signals with the *Pst*I probe, confirming that these mRNAs code for Pur proteins. Additional bands are detected exclusively at positions 9.5 and 3.0 kb in all the tissues and at 0.8 kb only in skeletal muscle with the labeled *Pst*I fragment.

DISCUSSION

A protein, SPSF I, identified as a factor interacting specifically with a single DNA strand derived from the BPV-1 origin of DNA replication has been shown to be identical with Purα, a protein which binds to the purine-rich strand of the PUR element repeatedly observed to be present in initiation zones of DNA replication. By expressing the Purα cDNA in the Sf9/baculovirus system we confirmed the identitiy of the calf thymus protein SPSF I with the Purα protein.

Characterization of the DNA-binding properties of the SPSF I and II proteins revealed a high affinity of these proteins to purine-rich single strands. The sequences of the oligonucleotides ORI-l and P-l, which are bound with very high affinity and specifity, do not match perfectly with the consensus sequence of the PUR-element. This indicates that the SPSF/Pur proteins also bind to purine-rich DNA with slightly different consensus sequences. The preferred binding of the two SPSF proteins to longer oligonucleotides may indicate a possible affinity of the two proteins to oligonucleotides with particular conformations or secondary structures which can be formed more easily with longer oligonucleotides. The observation that the SPSF proteins bind with lower affinity to short double-stranded oligonucleotides is also in agreement with this hypothesis (7,20).

The relationship between the two copurifying SPSF proteins remains to be determined. They must be closely related since they share identical endopeptidase Lys-C fragments. Although we cannot rule out the possibility that the 39 kDa protein is simply a degradation product of the 42 kDa protein experimental findings argue against this hypothesis. For once, the molar ratio of the two calf thymus proteins remains unaltered during several purification procedures. The two proteins are always obtained in equimolar amounts. Secondly, proteins derived from the Purα/SPSF I cDNA always displayed the 42 kDa band but never the 39 kDa band whether expressed as a recombinant protein from insect cells or whether analyzed as a product from *in vitro* translation sudies.

The existence of extensive posttranslational modifications was suggested by Johnson *et al.* (35). The 39 kDa protein might therefore represent a precursor protein without or with less posttranslational modifications. The nature of these modifications remains to be determined. However, our *in vitro* analysis of the purified calf thymus proteins suggests that posttranslational modification is not important for DNA-binding.

In this study we present additional 3′-UTR sequences of the Purα mRNA. Through Northern blot analysis we confirm that these sequences are actually transcribed in the cell and that they are part of the Purα mRNA. Since there are only very few mRNA species hybridizing exclusively with the Purα *Pst*I fragment but not with the *Ava*II–*Ear*I fragment, it is likely that all of the Purα mRNAs in the cell contain at least parts of the 3′-UTR sequences presented here. Our own analysis is consistent with the studies of Bergemann *et al.* (17), who identified Purα mRNAs of the same size by Northern blot analysis and RACE–PCR.

The extended 3′-UTR sequences explain, in part, the difference in lengths between the 1144 bp long Purα cDNA clones and the

Northern blot studies, in which only larger (>2.0 kb) transcripts were detected. There is a poly(A) signal at position 1062 of the Purα cDNA which could be used for polyadenylation and therefore creates the abundant 1.1 kb Purα cDNA species. Another explanation for the difference could be innappropriate priming during the reverse transcriptase reaction using oligo d(T) primer since there are extensive poly(A) stretches in the vicinity of the first polyadenylation signal as well as in the whole 3′-UTR.

Until now, the complete genomic structure of the Purα gene was not available. Analysis of genomic clones has so far not revealed introns in this region (data not shown). Therefore it seems more likely that the existence of various mRNAs derive from an alternative polyadenylation mechanism. The presence of different cDNA species with variable lengths of the 3′-untranslated regions is not uncommon (36,37).

The 3′-untranslated region of the Purα cDNA has a high percentage of A and T residues (68%). A/U-rich regions present in this 3′-untranslated region are known to influence mRNA regulation. 3′-UTRs may be involved in mRNA stability, translational regulation, subcellular transcript localization and regulation of cell differentiation (for reviews see 38,39,40).

The Purα protein interacts specifically with the hypophosphorylated form of the Rb protein, an important cell cycle regulator protein (35). Different forms of Pur proteins which might be regulated, at least in part, through their 3′-UTRs could interact differently with cell-cycle regulators such as Rb and therefore influence certain cell cycle processes.

The Pur protein family is interesting for several reasons: (i) they interact with DNA at origin regions or initiation zones, as shown by the studies of Bergemann *et al.* (16,17) and independently by our own binding studies with the BPV origin; (ii) they may also be involved in transcriptional regulation, as shown through CAT assays by Haas *et al.* (21); (iii) they interact specifically with the hypophosphorylated form of Rb, an important regulatory protein; and (iv) they must contain a novel DNA-binding motif for the sequence-specific interaction of proteins with single-stranded DNA. So far, functional studies of almost all known specific single-stranded proteins are restricted to their DNA-binding properties. With the existence of a BPV-1 DNA *in vitro* replication system (4) the role of the Pur family proteins can now be addressed in more detail.

ACKNOWLEDGEMENTS

We thank Drs Heidi Feldmann and Akis Zorbas for discussions and critically reading the manuscript. Oligonucleotides were obtained from Dr G. Arnold. The HeLa cDNA library in pλDR2 was kindly provided by Dr M. Müller, the HeLa cDNA library in pJG4-5 by Dr R. Brent. This work was supported by the Deutsche Forschungsgemeinschaft grant Fa 138/6-1.

REFERENCES

1 Ustav, E., Ustav, M., Szymanski, P. and Stenlund, A. (1993) *Proc. Natl Acad. Sci. USA* **90**, 898–902.
2 Ustav, M. and Stenlund, A. (1991) *EMBO J.* **10**, 449–457.
3 Ustav, M., Ustav, E., Szymanski, P. and Stenlund, A. (1991) *EMBO J.* **10**, 4321–4329.
4 Yang, L., Li, R., Mohr, I., Clark, R. and Botchan, M. R. (1991) *Nature* **353**, 628–632.
5 Coverley, D. and Laskey, R. A. (1994) *Annu. Rev. Biochem.* **63**, 745–776.
6 DePamphilis, M. L. (1993) *Annu. Rev. Biochem.* **62**, 29–63.
7 Habiger, C., Stelzer, G., Schwarz, U. and Winnacker, E.-L. (1992) *J. Virol.* **66**, 5988–5998.
8 Lusky, M. and Botchan, M.R. (1984) *Cell* **36**, 391–401.
9 Nallaseth, F. S. and DePamphilis, M. L. (1994) *J. Virol.* **68**, 3051–3064.
10 Yang, L., Mohr, I., Li, R., Nottoli, T., Sun, S. and Botchan, M. R. (1991) *Cold Spring Harbor Symp. Quant. Biol.* **56**, 335–346.
11 Cockell, M., Trutiger, S., Hughes, G. J. and Gasser, S. M. (1994) *Nucleic Acids Res.* **22**, 32–40.
12 Hofmann, J. F.-X. and Gasser, S. M. (1991) *Cell* **64**, 951–960.
13 Carmichael, E. P., Roome, J. M. and Wahl, A. F. (1993) *Mol. Cell. Biol.* **13**, 408–420.
14 Negishi, Y., Nishita, Y., Saëgusa, Y., Kakizaki, I., Galli, I., Kihara, F., Tamai, K., Miyajima, N., Iguchi-Agira, S. M. M. and Agira, H. (1994) *Oncogene* **9**, 1133–1143.
15 Takai, T., Nishita, Y., Iguchi-Ariga, S. M. M. and Ariga, H. (1994) *Nucleic Acids Res.* **22**, 5576–5581.
16 Bergemann, A. D. and Johnson, E. M.(1992) *Mol. Cell. Biol.* **12**, 1257–1265.
17 Bergemann, A. D., Ma, Z.-W. and Johnson, E. M. (1992) *Mol. Cell. Biol.* **12**, 5673–5682.
18 Dobbs, D. L., Shaiu, W.-L. and Benbow, R. M. (1994) *Nucleic Acids Res.* **22**, 2479–2489.
19 Kitsberg, D., Selig, S., Keshet, I. and Cedar, H. (1993) *Nature* **366**, 588–590.
20 Chen, N.N., Chang, C.-F., Gallia, G. L., Kerr, D. A., Johnson, E. M., Krachmarov, C. P., Barr, S. M., Frisque, R. J., Bollag, B. and Khalili, K. (1995) *Proc. Natl Acad. Sci. USA* **92**, 1987–1091.
21 Haas, S., Thatikunta, P., Steplewski, A., Johnson, E.M., Khalili, K. and Amini, S. (1995) *J. Cell Biol.* **5**, 1171–1179.
22 Fried, M. G. (1989) *Electrophoresis* **10**, 366–376.
23 Meisterernst, M., Gander, I., Rogge, L. and Winnacker, E.-L. (1988) *Nucleic Acids Res.* **16**, 4419–4435.
24 Haselbeck, A. and Hösel, W. (1988) *Topics in Biochemistry* 8, Boehringer Mannheim.
25 Laemmli, U. K. (1970) *Nature* **227**, 680–685.
26 Murphy, A. J. M., Kung, A. L., Swirski, R. A. and Schimke, R. T. (1992) *METHODS: A Companion to Methods in Enzymology* **4**, 111–131.
27 Quertemous, T. (1987) 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, vol. 2.* Greene Publishing Associates and Wiley Interscience, New York, pp. 6.1.1.–6.1.4.
28 Strauss, W. M. (1987) 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, vol. 2.* Greene Publishing Associates and Wiley Interscience, New York, pp. 6.3.1–6.3.6.
29 Sanger, F., Micklen, S. and Coulson, A. R. (1977) *Proc. Natl Acad. Sci. USA* **74**, 5463–5467.
30 Kitts, P. A., Ayres, M. D. and Possee, R. D. (1990) *Nucleic Acids Res.* **18**, 5667–5672.
31 O'Reilly, D. R., Miller L. K. and Luckow, V. A. (1992) *Baculovirus Expression Vectors, A Laboratory Manual.* WH Freeman and Company, New York.
32 Summers, M. D. and Smith, G. E. (1987) A manual of methods for baculovirus vectors and insect cell culture procedures. *Texas Agricultural Experiment Station Bulletin* No 1555.
33 Guyris, J., Golemis, E., Chertkov, H. and Brent, R. (1993) *Cell* **75**, 791–803.
34 Zervos, A. S., Gyueris, J. and Brent, R. (1993) *Cell* **72**, 223–232.
35 Johnson, E. M., Chen, P.-L., Krachmarov, C. P., Batt, S. M., Kanovsky, M., Ma, Z.W. and Lee, W.-H. (1995) *J. Biol. Chem.* **270**, 24352–24360.
36 Mallo, M., Steingrimmson, E., Copeland, N. G., Jenkins, N.A. and Gridley, T. (1994) *Genomics* **21**, 194–201.
37 Brocard, M.-P., Rousseau, D., Lawrence, J.-J. and Khochbin, S. (1994) *Eur. J. Biochem.* **221**, 421–425.
38 Wahle, E. and Keller, W. (1992) *Annu. Rev. Biochem.* **61**, 419–440.
39 Chen, C.-Y. and Shyu, A.B. (1995) *Trends Biochem. Sci.* **20**, 465–470.
40 Curtis, D., Lehmann, R. and Zamore, P.D. (1995) *Cell* **81**, 171–176.