

mRNA Polyadenylate-Binding Protein: Gene Isolation and Sequencing and Identification of a Ribonucleoprotein Consensus Sequence

STEPHEN A. ADAM, TERRY NAKAGAWA, MAURICE S. SWANSON, TERESA K. WOODRUFF,
AND GIDEON DREYFUSS*

Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, Illinois 60201

Received 31 December 1985/Accepted 6 May 1986

We identified and produced antibodies to the major proteins that interact with poly(A)⁺ RNAs in the yeast *Saccharomyces cerevisiae*. The major proteins which were cross-linked by UV light to poly(A)⁺ RNA in intact yeast cells had apparent molecular weights of 72,000, 60,000, and 50,000. The poly(A) segment of the RNA was selectively cross-linked to the 72,000-molecular-weight protein (72K protein). Mice immunized with purified UV-cross-linked RNA-protein (RNP) complexes produced antibodies to the three major RNP proteins. A yeast genomic DNA library constructed in the λ gt11 expression vector was screened with the anti-RNP serum, and recombinant bacteriophage clones were isolated. One recombinant phage, λ YPA72.1, bearing a 2.5-kilobase insert, produced a large β -galactosidase-RNP fusion protein. Affinity-selected antibodies from the anti-RNP serum on this fusion protein recognized a single 72K protein which was cross-linked to the poly(A) segment of RNA in the intact cell. Furthermore, the fusion protein of λ YPA72.1 had specific poly(A)-binding activity. Therefore, λ YPA72.1 encodes the 72K poly(A)-binding protein. Immunofluorescence microscopy showed that this protein was localized in the cytoplasm. Hybrid-selected mRNA translated in vitro produced the 72K poly(A)-binding protein, and mRNA blot analysis detected a single 2.1-kilobase mRNA. DNA blot analysis suggested a single gene for the poly(A)-binding protein. DNA sequence analysis of genomic clones spanning the entire gene revealed a long open reading frame encoding a 64,272-molecular-weight protein with several distinct domains and repeating structural elements. A sequence of 11 to 13 amino acids is repeated three times in this protein. Strikingly, this repeated sequence (RNP consensus sequence) is highly homologous to a sequence that is repeated twice in a major mammalian heterogeneous nuclear RNP protein, A1. The conservation of the repetitive RNP consensus sequence suggests an important function and a common evolutionary origin for messenger RNP and heterogeneous nuclear RNP proteins.

mRNA and heterogeneous nuclear RNA are found in eucaryotic cells in association with specific proteins to form RNA-protein complexes known as messenger ribonucleoprotein (mRNP) particles (mRNPs) and heterogeneous nuclear ribonucleoprotein particles (hnRNPs), respectively. The complexes are the structural units of these polynucleotides and are likely to be involved in mRNA biogenesis, metabolism, and function (11, 13, 32). Therefore, the proteins which form hnRNP and mRNP complexes are of considerable interest and have been extensively studied (1, 11-13, 24, 48). Despite a great deal of effort, little is known about the biochemical functions of RNP proteins. Until recently, it has been difficult to identify unambiguously the genuine RNP proteins since nonspecific RNA-protein interactions are likely to occur during cell fractionation. UV cross-linking of proteins to RNA in intact cells overcomes these difficulties and allows the identification of proteins which bind heterogeneous nuclear RNA and mRNA in vivo (11, 16, 51, 54). To learn more about these proteins and about the RNP complexes, we have immunized mice with purified UV-cross-linked RNP complexes and have produced antibodies to several of the major RNP proteins in vertebrate cells (1, 12, 13).

Much of the knowledge of RNP proteins has come from work with higher eucaryotes (12, 24, 48). However, little is known about the structure of mRNPs and hnRNPs in simple

eucaryotes such as the yeast *Saccharomyces cerevisiae*. We have undertaken the study of RNP proteins in yeast cells because, in addition to its intrinsic interest, yeast is an excellent organism for the study of RNP proteins. Large quantities of material can be readily prepared, strains with mutations in many aspects of RNA metabolism are available, and the genotype of yeast cells can be manipulated with relative ease, making it an ideal organism for determining the function of specific RNP proteins. Furthermore, because of phylogenetic divergence, we considered it likely that yeast RNP proteins would be very immunogenic, allowing production of antibodies to conserved RNP proteins derived from vertebrate cells. An example of such a case is the most prominent mRNP protein in vertebrates, the 72,000-molecular-weight (72K) mRNA poly(A)-binding protein. This protein was first described by Blobel in 1973 (7) and is one of the most conserved and extensively investigated eucaryotic RNP proteins. Although many important functions in mRNA metabolism have been attributed to this protein, little is known about its function, amino acid sequence, and metabolism. This is because specific probes, such as antibodies or gene sequences, have not been available from any species.

We report here the identification of the proteins that interact with poly(A)⁺ RNAs in yeast by UV cross-linking and the production of antibodies to the proteins. Using these antibodies, we identified the poly(A)-binding protein in yeast

* Corresponding author.

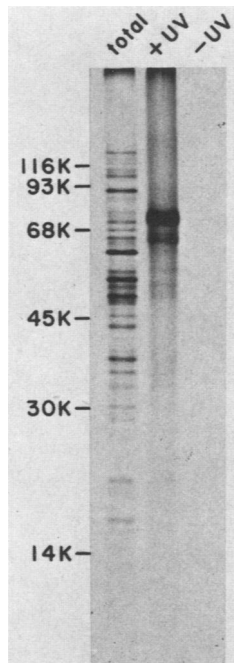


FIG. 1. [35 S]methionine labeling of UV-cross-linked yeast RNP particles. Log-phase yeast cultures were labeled for two generation times with [35 S]methionine, washed, and exposed to UV light for 15 min. The cells were broken by vortexing with glass beads under protein-denaturing conditions in the presence of protease and RNase inhibitors as described in Materials and Methods. Poly(A)⁺ RNA-protein complexes were selected by chromatography on oligo(dT)-cellulose and digested with RNase A and micrococcal nuclease before electrophoresis of the proteins in SDS-polyacrylamide gels. Lanes: total, total yeast lysate; +UV and -UV, proteins cross-linked to poly(A)⁺ RNA from cells with and without exposure to UV light, respectively. Molecular weights (in thousands) of protein standards are indicated to the left.

cells, isolated, expressed, and determined the sequence of its gene, and deduced the amino acid sequence of the protein.

MATERIALS AND METHODS

Cell growth and labeling. *S. cerevisiae* BJ926 (*MAT/a trp1/+ +his1 pep4-3/pep4-3 prc1-126/prc1-126 prb1-1122/prb1-1122 can1/can1*) was a gift from Lawrence Dumas. Cells were grown at 30°C in SD medium (0.67% yeast nitrogen base without amino acids, 2% dextrose) for labeling or in YPD medium (1% yeast extract, 2% Bacto-Peptone [Difco Laboratories], 2% dextrose) for large-scale culture. All experiments were done with mid- to late-log-phase cells. For labeling with [35 S]methionine, cells growing in SD medium were incubated with 10 μ Ci of [35 S]methionine per ml for 4 h.

UV cross-linking and RNA isolation. Cells were pelleted at $2,000 \times g$ and washed with cold phosphate-buffered saline (PBS). Yeast cells resuspended in PBS at a density of 5×10^9 cells per ml were irradiated with UV light for 15 min with a 15-W germicidal lamp as previously described (1, 11). Irradiated and nonirradiated (control) cells were lysed by vortexing with glass beads in lysis buffer (20 mM Tris chloride [pH 7.4], 1 mM EDTA, 50 mM LiCl, 1% sodium dodecyl sulfate [SDS], 1% 2-mercaptoethanol, 1 mg of heparin per ml, 10 mM vanadyl-adenosine, 0.2 TIU of aprotinin [Sigma Chemical Co.] per ml, 1 μ g each of

leupeptin and pepstatin A [Sigma] per ml, 10 mM sodium bisulfite, 0.2 mM benzamidine hydrochloride). UV-cross-linked poly(A)⁺ RNPs were isolated by oligo(dT)-cellulose (type 3; Collaborative Research, Inc.) chromatography as previously described (1, 11, 22).

Preparation of antisera. Preparation of antisera in BALB/c mice by repeated intraperitoneal injections was carried out as previously described (1, 12) with RNPs prepared from 10 liters of yeast culture per injection.

Gel electrophoresis and protein blotting. Proteins were analyzed on 12.5% SDS-polyacrylamide gels (11, 12), and protein blots were treated and processed as recently described (8, 11, 12). Two-dimensional gel electrophoresis was performed, with nonequilibrium pH gel electrophoresis in the first dimension, as described by O'Farrell et al. (35) for 1,600 V \cdot h, and SDS-polyacrylamide gel electrophoresis (PAGE) in the second dimension. Gels containing 35 S-labeled material were impregnated with PPO (2,5-diphenyloxazole) before being dried, and fluorography was performed with Kodak XAR-5 X-ray film (11, 31).

Screening of λ gt11 yeast genomic DNA library and purification of fusion protein. A yeast genomic DNA library constructed in the λ gt11 expression vector was a generous gift from Michael Snyder and Ronald Davis. The library produced by random mechanical shearing and blunt-end

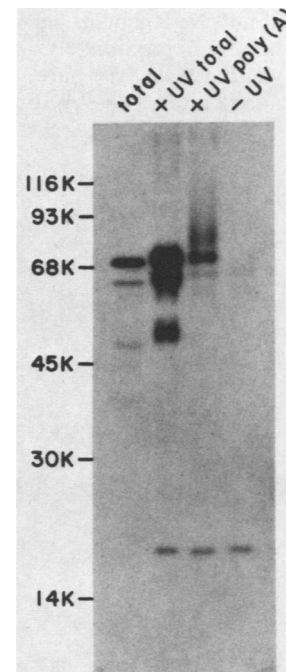


FIG. 2. Immunoblot analysis of UV-cross-linked RNPs with anti-yeast-RNP serum. UV-cross-linked poly(A)⁺ RNPs were isolated as described in the legend to Fig. 1, and the proteins were electrophoretically transferred to nitrocellulose paper after SDS-PAGE. Poly(A) segments of the RNPs were isolated by digestion of the total cross-linked RNPs with RNase A and RNase T₁ in high salt concentration, followed by selection of the undigested poly(A) on oligo(dT)-cellulose. The selected poly(A) segments were then completely digested with RNase A and micrococcal nuclease, and the protein bound to them was resolved by SDS-PAGE. Lanes: total, total yeast protein; +UV total, proteins cross-linked to total poly(A)⁺ RNAs; +UV poly(A), proteins cross-linked to poly(A) segments; -UV, proteins cross-linked to RNA in the absence of UV light. Molecular weights (in thousands) of protein standards are indicated to the left.

addition of *EcoRI* linkers (M. Snyder, personal communication) was plated on *Escherichia coli* Y1090. Immunological screening was carried out as recently described (34, 57) with the anti-yeast-RNP serum. Positive phage were purified and used to lysogenize Y1089 cells as described by Silhavy et al. (46) with the aid of the λ h Δ 80 selector bacteriophage. The lysogens were induced with 10 mM isopropyl- β -D-thiogalactopyranoside (Sigma) and temperature shift (42°C, 10 min) and were grown for an additional 2 h at 38°C. The cells were harvested, lysed by freeze-thawing in 5 ml of SM buffer (0.1 M NaCl, 0.016 M MgSO₄, 0.05 M Tris chloride [pH 7.5], 0.01% gelatin) per 250 ml of original culture, and treated with 10 μ g of DNase I (DPFF; Cooper Biomedical, Inc.) per ml–1 mM MgCl₂–28 mM 2-mercaptoethanol for 10 min on ice. A 2-ml amount of 50% sucrose in 0.1 M NaCl–0.01 M Tris chloride [pH 8.0]–1 mM EDTA was added along with 60 μ g of fresh lysozyme per ml and allowed to digest cells on ice for 30 min. After digestion, the cells were solubilized with 1 ml each of 10% Nonidet P-40 and 0.5 M EDTA and left on ice for 15 min. Then, 0.5 ml of 10% Zwittergent (Calbiochem-Behring) was added, and the mixture was sonicated for 30 s at a medium energy setting with a microtip sonicator (Heat Systems-Ultrasonics). The resulting suspension was layered on a 10-ml 40% sucrose cushion and centrifuged at 13,000 \times g for 30 min. The pellet was suspended in 1 to 5 ml of 8 M urea and recentrifuged for 15 min at 12,500 \times g. The supernatant contained the fusion protein (S. Michaeli and N. Agabian, personal communication). Typical yields were approximately 1 mg of protein per 50-ml culture and were at least 90% pure, as determined by Coomassie blue staining after SDS-PAGE.

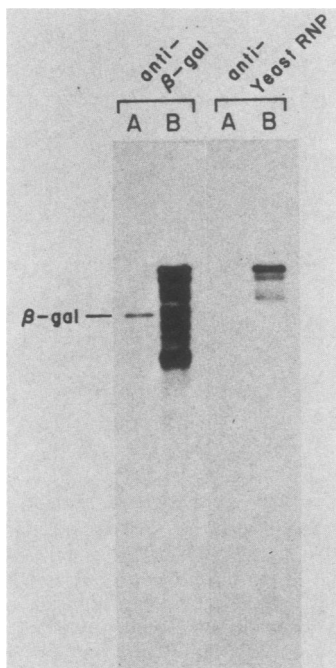


FIG. 3. Fusion protein produced by the lysogen λ YPA72.1 isolated from the yeast λ gt11 library. Total cell lysate of the induced lysogen was resolved by SDS-PAGE along with purified β -galactosidase and transferred to nitrocellulose. Identical immunoblots were probed with either the anti-yeast-RNP serum or a monoclonal antibody specific for β -galactosidase and were detected with ¹²⁵I-labeled goat anti-mouse IgG. Lanes marked A, β -galactosidase; lanes marked B, cell lysate of the induced lysogen.

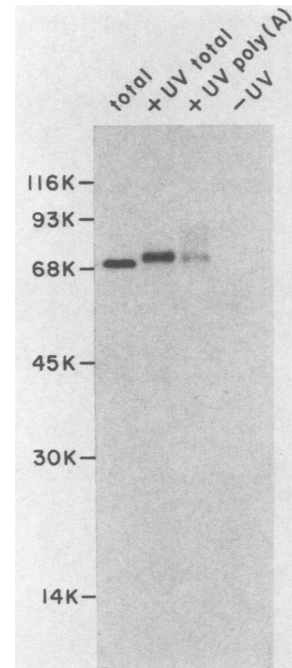


FIG. 4. Affinity purification of the antibody to the fusion protein and identification of the RNP protein gene in λ YPA72.1. An immunoblot of total cell lysate from the induced phage λ YPA72.1 was stained with amido black, and the area of the blot containing the fusion protein was cut out and used to adsorb antibodies specific for the fusion protein from the anti-yeast-RNP serum. The antibody was eluted from the nitrocellulose with glycine buffer and used to probe an immunoblot identical to that in Fig. 2. Lanes: total, total yeast protein; +UV total, proteins cross-linked to total poly(A)⁺ RNAs; +UV poly(A), proteins cross-linked to the poly(A) segment; –UV, proteins cross-linked to RNA in the absence of UV light. Molecular weights (in thousands) of protein standards are indicated to the left.

Affinity purification of antibody. The purified fusion protein preparation was resolved on a 7.5% SDS-polyacrylamide gel and transferred electrophoretically to nitrocellulose. A strip of the lane containing the protein was cut from the filter and stained with amido black to determine the position of the fusion protein. The region of the nitrocellulose corresponding to the fusion protein was cut out and blocked for 3 h in 5% nonfat dry milk in PBS. The nitrocellulose was then incubated with a 1:100 dilution of the anti-yeast-RNP serum in 3% bovine serum albumin for 3 h. After being washed, bound antibody was eluted in 0.2 M glycine [pH 2.5]–0.25 M NaCl for 3 min. The eluted antibody was neutralized with 2 M Tris base, and bovine serum albumin was added to a final concentration of 3%.

Poly(A) binding assay. The poly(A) binding assay is a slight modification of that described by Sachs and Kornberg (41). The fusion protein of λ YPA72.1 was bound to protein A-agarose beads (Boehringer-Mannheim Biochemicals) by incubation of 20 μ l of beads with a rabbit anti- β -galactosidase immunoglobulin G (IgG) fraction (Cooper Biomedical) in PBS, followed by incubation in 1 ml of PBS containing 20 to 50 μ g of the fusion protein for 1 h and washing to remove unbound protein. A 2-ng amount of poly(A) (Pharmacia Fine Chemicals), ³²P labeled with T4 polynucleotide kinase (International Biotechnologies Inc.) to a specific activity of 1.5×10^7 cpm/ μ g, was incubated in 100 μ l of buffer A (20 mM Tris chloride [pH 8.0], 100 mM NaCl, 1 mM EDTA) containing 10 μ g of unlabeled poly(A) or

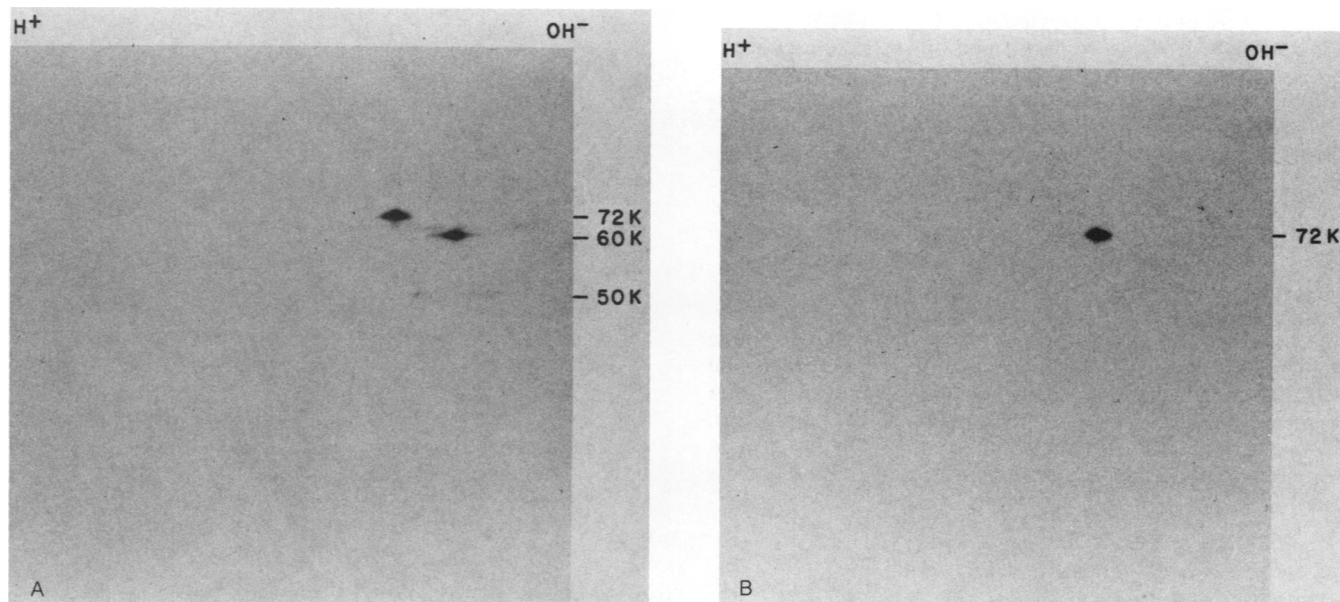


FIG. 5. Nonequilibrium pH gel electrophoresis of yeast RNP proteins. Total yeast lysate was resolved by electrophoresis through nonequilibrium pH gels with 3 to 10 ampholines. Identical gels, resolved in the second dimension by SDS-PAGE, were transferred to nitrocellulose and probed with either the anti-yeast-RNP serum or the antibodies affinity selected on the fusion protein from λ YPA72.1. (A) Immunoblot probed with the anti-yeast-RNP serum. (B) Immunoblot probed with the affinity-purified antibodies.

unlabeled poly(C) (Pharmacia) for 10 min at 23°C with the fusion protein-bound agarose beads. The beads were then washed four times in the same buffer without polynucleotide, and bound 32 P label was determined by Cerenkov radioactivity counting.

Immunofluorescence microscopy. Yeast cells were prepared for immunofluorescence microscopy as described by Kilmartin and Adams (26), with the following modifications. Early-log-phase yeast cells were harvested by centrifugation and washed in PBS. The cells were fixed by suspension in 1.2 M sorbitol–0.12 M K_2PO_4 –0.033 M citric acid (pH 5.9) containing 4% formaldehyde for 2 h at 23°C. The fixative was removed by two washings in the same buffer without formaldehyde. The fixed cells were digested with 0.5 mg of Zymolyase 5000 (Kirin Brewery) per ml in the sorbitol buffer without formaldehyde for 1 h at 37°C and washed by centrifugation. The cell pellet was quickly suspended in a small volume of methanol at -20°C , and a small drop was placed on a cover slip and allowed to evaporate. The yeast cells attached to the cover slip were incubated with a 1:200 dilution in 1% bovine serum albumin of the antiserum to the λ YPA72.1 fusion protein, followed by fluorescein isothiocyanate-conjugated goat anti-mouse $F(ab')_2$ (Cooper Biomedical). Photomicrographs were taken with a Zeiss model III photomicroscope equipped with a Planapo $\times 63$ oil objective.

Hybrid selection and in vitro translation. Poly(A)⁺ RNA was prepared from yeast cells by breaking the cells with glass beads in the presence of buffer-saturated phenol as described previously (21), followed by chromatography on oligo(dT)-cellulose (12, 22). Hybridization selection was performed by a modification of the technique of Ricciardi et al. (38). The plasmids pGEM-2 (Promega Biotech) and pY72.1 (pGEM-2 containing the 2.0-kilobase-pair [kbp] *EcoRI* fragment of λ YPA72.1) were linearized by digestion with *XbaI*, denatured by boiling and by treatment with 0.5 M NaOH, neutralized, and applied to nitrocellulose. Approximately 50 μg of yeast poly(A)⁺ RNA was hybridized to each

filter in 65% formamide–20 mM 1,4-piperazinediethanesulfonic acid (pH 6.4)–0.2% SDS–0.4 M NaCl–100 μg of yeast tRNA per ml for 4 h at 50°C. After an extensive washing at 65°C in 10 mM Tris chloride (pH 7.6)–0.15 M NaCl–1 mM EDTA–0.5% SDS, specifically bound RNAs were eluted with 10 mM methyl mercury hydroxide containing 100 μg of yeast tRNA per ml for 10 min at 22°C. Dithiothreitol was added to a final concentration of 5 mM, potassium acetate was added to a final concentration of 0.25 M, and the RNA was precipitated with 3 volumes of ethanol at -20°C overnight. Optimized in vitro translations in rabbit reticulocyte lysate (Promega Biotech) were performed for 60 min at 30°C in 25- μl reaction volumes and analyzed by SDS-PAGE and fluorography. Immunoprecipitations with the anti-fusion-protein antiserum were carried out in PLB (PBS containing 1 mM EDTA, 1% Triton X-100, 0.5% deoxycholic acid, 0.1% SDS, and 1% aprotinin) for 1 h at 4°C. Immune complexes were adsorbed on protein A-agarose (Boehringer-Mannheim) with rabbit anti-mouse $F(ab')_2$ and extensively washed with PLB. The adsorbed protein was eluted in SDS-PAGE sample buffer for electrophoresis.

RNA blot hybridization. Poly(A)⁺ RNA, prepared as described above, was resolved by electrophoresis in a 1.2% agarose gel in the presence of formaldehyde (34), and the nitrocellulose blot was prepared as described previously (49). The blot was probed with 32 P-labeled DNA produced by nick translation (39) of the DNA insert (2.0-kbp *EcoRI* fragment of λ YPA72.1) purified by agarose gel electrophoresis. Hybridization was carried out, essentially as described previously (34), for 24 h at 42°C in 50% formamide–5 \times SSC (1 \times SSC is 0.15 M NaCl plus 0.015 M sodium citrate [pH 7.0]–1 \times Denhardt solution–0.05 M sodium phosphate containing 250 μg of salmon sperm DNA per ml. The blot was washed in 2 \times SSC–0.1% SDS twice at 21°C and three times at 65°C before exposure to X-ray film for autoradiography.

DNA blot hybridization. Yeast high-molecular-weight genomic DNA was prepared by a rapid yeast DNA isolation

procedure as described previously (56). Southern blot hybridizations (34, 47) were performed on restriction endonuclease (International Biotechnologies Inc.)-digested yeast genomic DNA fractionated in an 0.8% agarose gel in Tris-borate buffer. The blot was probed with the ^{32}P -labeled DNA produced by nick translation (39) of the DNA insert (2.0-kbp *EcoRI* fragment of $\lambda\text{YPA72.1}$). Hybridization was at 42°C for 24 h in a solution of 50% formamide–5× SSC–5× Denhardt solution–0.1% SDS containing 200 g of salmon sperm DNA per ml. Blots were washed three times at 21°C in 2× SSC, followed by three washes at 42°C in 1× SSC, and exposed to X-ray film for autoradiography.

DNA sequence analysis. The two overlapping genomic clones λA_1 and λA_2 were subcloned into pGEM-1 and pGEM-2 as *EcoRI* inserts, and smaller overlapping DNA fragments were cloned into M13mp18 and M13mp19 (33). Sequence analysis was done by the dideoxy chain termination method (42). The sequencing strategy is outlined below (see Fig. 11).

RESULTS

Identification of yeast RNPs. Exposure of intact cells to UV light of sufficient intensity brings about covalent cross-linking of RNA to the proteins with which it is in contact *in vivo*. We have previously identified the proteins that are cross-linked by UV light to poly(A)⁺ RNA in HeLa cells and in several other vertebrate cell lines (1, 11–13). We now extend these studies to the yeast *S. cerevisiae* and examine the proteins that become cross-linked to poly(A)⁺ RNA. Log-phase yeast cell cultures were labeled with

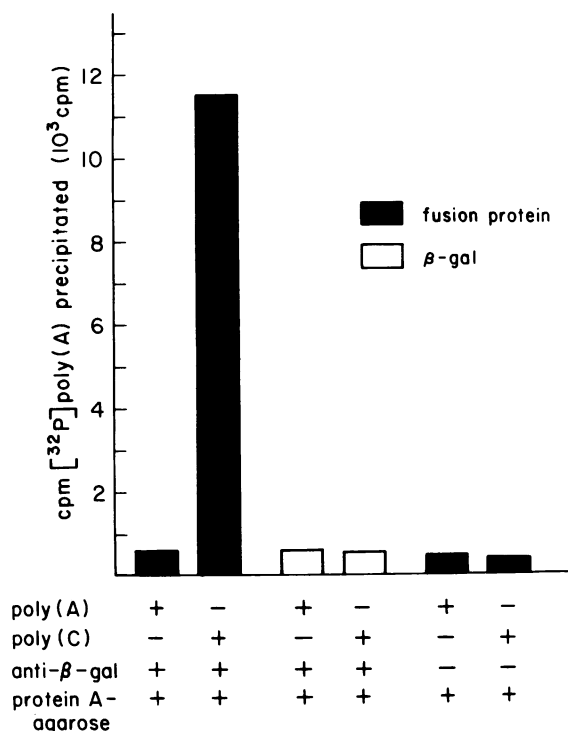


FIG. 6. Immunoprecipitation of poly(A)-binding activity in the fusion protein produced by $\lambda\text{YPA72.1}$. Anti- β -galactosidase antibody bound to protein A-agarose beads was used to adsorb either the fusion protein from $\lambda\text{YPA72.1}$ or the β -galactosidase from solution. The bound protein was tested for its ability to specifically bind ^{32}P -labeled poly(A), as described in Materials and Methods.

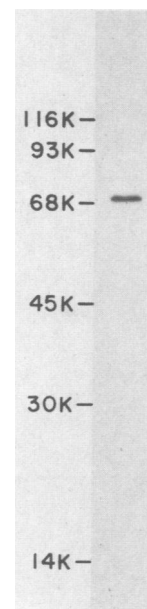


FIG. 7. Immunoblot analysis of antibodies produced to the fusion protein from $\lambda\text{YPA72.1}$. Antiserum produced in a BALB/c mouse to the fusion protein produced by $\lambda\text{YPA72.1}$ was used to probe an immunoblot of total yeast protein. Molecular weights (in thousands) of protein standards are indicated to the left.

^{35}S]methionine for two generation times. UV irradiation of the cells for 15 min resulted in covalent cross-linking of the yeast RNP complexes. The proteins which became cross-linked to poly(A)⁺ RNA in yeast cells are shown in Fig. 1. The predominant ^{35}S -labeled proteins had molecular weights of 72,000, 60,000, and 50,000. Several fainter bands of molecular weights 30,000 to 50,000 were also visible upon longer exposure. Silver staining of a similar gel suggested that the 72K protein was the most abundantly cross-linked RNP protein. The UV-cross-linked RNPs were isolated under conditions which minimized endogenous nuclease and protease digestion during cell lysis; high concentrations of SDS and 2-mercaptoethanol, as well as 10 mM vanadyl-adenosine (6) and several protease inhibitors, were included in the lysis buffer. Therefore, the proteins smaller than 72,000 molecular weight are probably distinct RNP proteins, not degradation products of the 72K protein. The detected proteins are strictly the products of UV cross-linking *in vivo* since no proteins were detected when RNA was isolated from cells that were not exposed to UV light (Fig. 1). Because of the high cytoplasmic-to-nuclear-volume ratio in yeast cells, it is likely that cytoplasmic proteins are detected more readily than nuclear proteins.

Production of antibodies to yeast RNP proteins. We have previously described a general strategy for the production of polyclonal and monoclonal antibodies to RNA-contacting proteins using purified cross-linked RNPs as antigens (1, 12). The anti-RNP sera can be used to identify RNP proteins that are not intensely labeled by ^{35}S]methionine and provide extremely useful reagents for the characterization of the proteins. Total poly(A)⁺ UV-cross-linked yeast RNPs were used to immunize BALB/c mice and produce polyclonal antisera to the yeast RNP proteins. When the serum of an immunized mouse was used to probe an immunoblot of total yeast protein, three polypeptides with apparent molecular weights of 72,000, 60,000, and 50,000 were detected (Fig. 2,

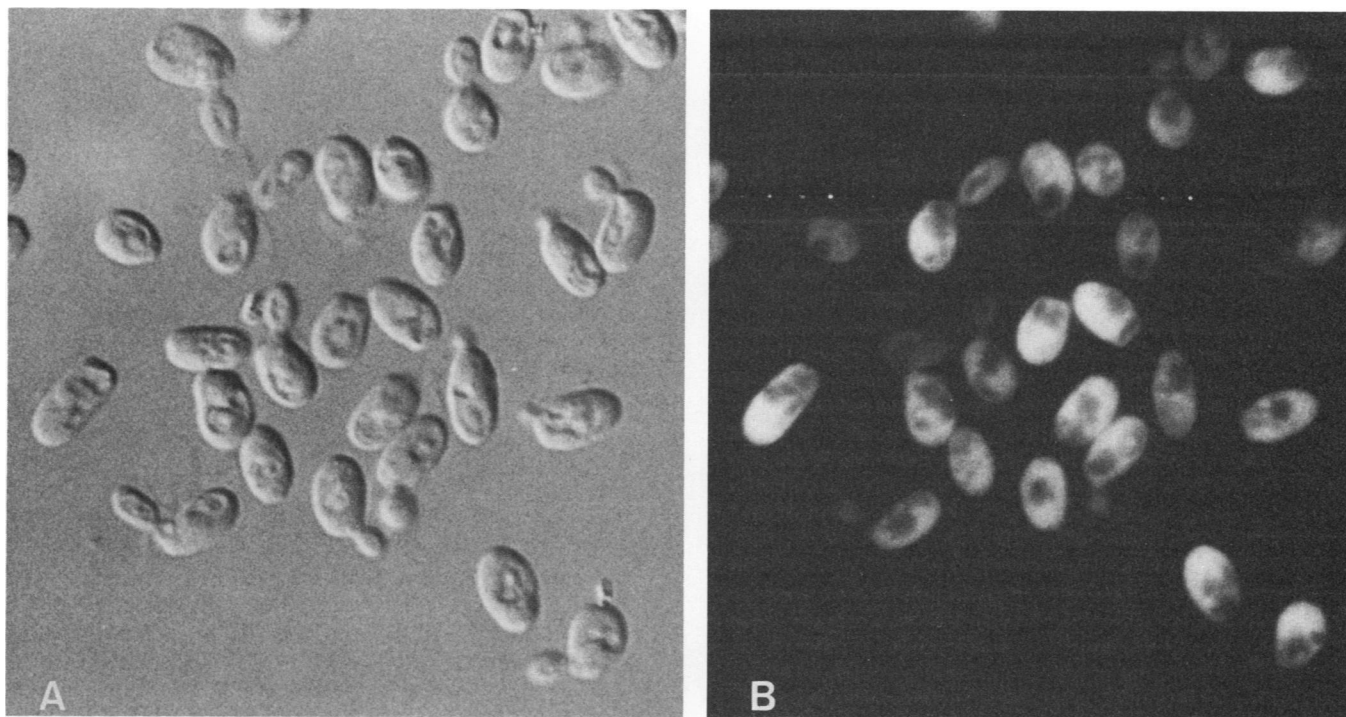


FIG. 8. Immunofluorescence of yeast cells with antiserum to the λ YPA72.1 fusion protein. Yeast cells were fixed and permeabilized as described in Materials and Methods and incubated with the same antiserum used in Fig. 7. Fluorescein isothiocyanate-labeled goat anti-mouse F(ab')₂ was used for detection. (A) Nomarski optics. (B) Fluorescence.

lane marked total). These three proteins were also visible by [³⁵S]methionine labeling (Fig. 1) and were also cross-linked to poly(A)⁺ RNA (Fig. 2, lane marked +UV total). The amount of the 50K protein which was cross-linked is variable and may reflect the metabolic state of the yeast cells at the time of cross-linking. The slight decrease in mobility of each cross-linked protein is characteristic of RNA-cross-linked proteins (12) and is probably due to the residual covalently bound nucleotide which cannot be removed by RNase digestion.

Intact poly(A) tails of mRNA and heterogeneous nuclear RNA can be isolated, and the proteins cross-linked to them can be analyzed after selective digestion of the total poly(A)⁺ RNPs with RNase A and RNase T₁ in 0.5 M NaCl, followed by reselection of the undigested poly(A) portion of the mRNA on oligo(dT)-cellulose. SDS-PAGE after complete digestion of the poly(A) with micrococcal nuclease and RNase A demonstrated that the 72K RNP protein was cross-linked to the poly(A) tail of yeast RNA [Fig. 2, lane marked +UV poly(A)]. This is similar to the poly(A)-binding protein of vertebrates which has a molecular weight of approximately 72,000 (4, 7, 13, 14, 18, 24, 30, 44, 45, 52). The band in Fig. 2 at 21,000 molecular weight and also the band at 65,000 molecular weight in lanes marked +UV poly(A) and -UV were due to the presence of antibodies to proteins in the RNase A and micrococcal nuclease preparations which were used to digest the RNPs and which were included in the immunization mixtures. Because these bands comigrated on the SDS gel at ca. 65,000 molecular weight with the UV-cross-linked 60K protein, we cannot rule out the possibility that a small amount of the 60K protein may also be associated with the poly(A) tail.

Isolation of yeast RNP genes by immunological screening of an expression vector library. The anti-yeast-RNP serum was

used for immunological screening of a yeast genomic DNA library constructed in the λ gt11 expression vector (57). Several clones were isolated, purified, and transferred to a lysogenic host. One of the phages, λ YPA72.1 bearing a 2.5-kbp insert, produced a β -galactosidase-yeast RNP fusion protein with an apparent molecular weight in SDS-PAGE of about 185,000 upon induction with isopropyl- β -D-thiogalactopyranoside. This fusion protein was recognized on an immunoblot by both the anti-yeast-RNP serum and an anti- β -galactosidase antibody (Fig. 3). The β -galactosidase portion of the fusion protein appeared to be partially degraded in the lysogen, although very little degradation of the yeast RNP protein was detected with the anti-yeast-RNP serum. Identification of the yeast RNP protein which is encoded by λ YPA72.1 was carried out by affinity purification of the antibodies specific for the fusion protein from the total anti-yeast-RNP serum. The region of an immunoblot containing the fusion protein was excised from the nitrocellulose sheet and used to adsorb antibodies from the total anti-yeast-RNP serum. The adsorbed antibodies were eluted in glycine buffer at low pH and, after neutralization, used to probe an immunoblot identical to that shown in Fig. 2. The affinity-purified antibody was specific for a 72K protein in a total yeast lysate (Fig. 4, lane marked total). This 72K protein was selectively cross-linked to total poly(A)⁺ RNPs and to the poly(A) portion of the mRNA (Fig. 4).

The possibility that more than one protein migrated at the 72,000-molecular-weight region on the one-dimensional SDS-polyacrylamide gels was tested by resolving total yeast proteins on two-dimensional gels by nonequilibrium pH gel electrophoresis (35) in the first dimension, followed by SDS-PAGE in the second. The proteins were then transferred to nitrocellulose and probed with either the total anti-yeast-RNP serum or the affinity-purified antibodies. The

total serum recognized all three RNP proteins, including a single 72K polypeptide with an apparent isoelectric point of about 8.0 (Fig. 5A). The affinity-purified antibody recognized an identical 72K protein with a similar isoelectric point (Fig. 5B). This isoelectric point is similar to that reported for the vertebrate mRNA poly(A)-binding protein (2). It is interesting that the other two yeast cell RNP proteins (60,000 and 50,000 molecular weight) identified here also had very basic isoelectric points (Fig. 5A).

The product of the cloned gene λ YPA72.1 is a specific poly(A)-binding protein. Up to this point, the possibility that the protein encoded by λ YPA72.1 is a general nucleic acid-binding protein could not be ruled out. To determine whether it is a specific-poly(A)-binding protein, we used the poly(A)-binding assay essentially as described by Sachs and Kornberg (41) to examine whether the fusion protein itself has specific-poly(A)-binding activity. The possibility that the fusion protein might have nucleic acid-binding activity was considered because its size suggests that it contains a large portion of the 72K protein sequence. Partial purification of the fusion protein, as described in Materials and Methods, yielded a preparation which was approximately 90% pure, as determined by Coomassie blue staining of an SDS gel. The

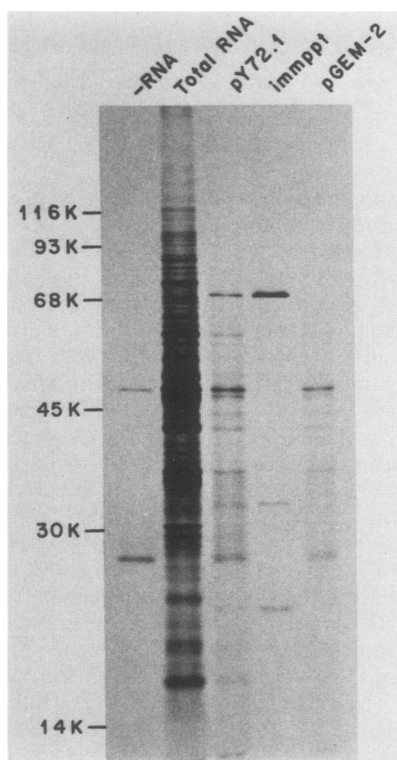


FIG. 9. Hybrid selection and translation of yeast mRNA with pY72.1. Yeast poly(A)⁺ RNA was hybrid selected and translated in a reticulocyte lysate as described in Materials and Methods. Immunoprecipitation was carried out on the translation products of mRNA selected on pY72.1 with the antiserum to the fusion protein of λ YPA72.1. Lanes: -RNA, background translation without addition of exogenous RNA; total RNA, translation products of the total poly(A)⁺ yeast RNA used for the hybrid selection; pY72.1, translation products of mRNA hybrid selected on plasmid pY72.1; immpppt, immunoprecipitation of translation products from lane pY72.1 with the antiserum to the fusion protein; pGEM-2, translation products of mRNA hybrid selected on pGEM-2. Molecular weights (in thousands) of protein standards are indicated to the left.

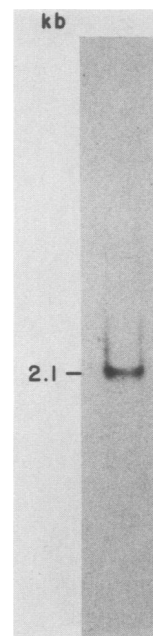


FIG. 10. RNA blot analysis of λ YPA72.1. Total yeast poly(A)⁺ RNA was resolved on a 1.2% formaldehyde agarose gel and transferred to nitrocellulose. The blot was probed with the nick-translated DNA 2.0-kbp *Eco*RI fragment of λ YPA72.1. The size marker is λ DNA digested with *Hind*III. kb, Kilobases.

fusion protein can be completely separated from the contaminants in this preparation by immunoadsorption on protein A-agarose beads with an anti- β -galactosidase antibody. The fusion protein bound to the beads was then tested for specific-poly(A)-binding activity by incubation with ³²P-labeled poly(A) and a 5,000-fold excess of unlabeled poly(A) or unlabeled poly(C). The beads were then washed extensively with the same buffer used for binding, and the amount of radioactivity bound was determined. Specific-poly(A) binding is defined as the amount of labeled poly(A) bound in the presence of poly(C) minus the amount bound in the presence of poly(A). The results of such an experiment indicated that the fusion protein had specific-poly(A)-binding activity (Fig. 6). In contrast, β -galactosidase had no poly(A)-binding activity. Therefore, the poly(A)-binding activity must be contained in the yeast-encoded 72K protein portion of the fusion protein of λ YPA72.1. A similar type of assay using fusion protein binding has been described recently by Johnson and Herskowitz (25) to demonstrate sequence-specific DNA binding of the yeast *MAT α 2* gene product.

Preparation of antibodies to the poly(A)-binding- β -galactosidase fusion protein. The partially purified fusion protein is a convenient and plentiful source of yeast poly(A)-binding protein free of contaminating yeast proteins. The fusion protein was used to immunize a BALB/c mouse and produce polyclonal antiserum to the poly(A)-binding protein. The serum contained antibodies which recognized a single 72,000-molecular-weight protein in total yeast lysate (Fig. 7), further confirming that the phage λ YPA72.1 isolated from the *lg*t11 library encodes the 72K yeast protein which is the yeast poly(A)-binding protein.

Immunofluorescence localization of the yeast poly(A)-binding protein. Vertebrate cells possess two different poly(A)-binding proteins, one with a molecular weight of

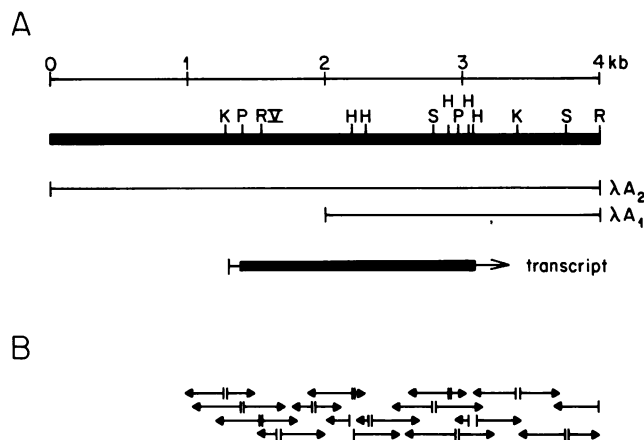


FIG. 11. Restriction map and sequencing strategy of λA_1 and λA_2 . (A) The restriction and transcription maps of the poly(A)-binding protein gene and flanking regions, as well as the positions of the λ gt11 inserts of λA_1 and λA_2 , are illustrated. The transcript was mapped with SP6 transcripts and by digestion of RNA-RNA hybrids with RNases A and T_1 ; therefore, the start and stop positions are approximate. Restriction sites are: H, *HindIII*; K, *KpnI*; P, *PvuII*; R, *EcoRI*; RV, *EcoRV*; and S, *SphI*. (B) DNA sequencing strategy. Sequencing was performed by the dideoxy chain termination procedure on fragments subcloned into M13mp18 and M13mp19. kb, Kilobases.

about 60,000, which is restricted to the poly(A) tails of heterogeneous nuclear RNA (41, 45), and a 72,000-molecular-weight mRNA poly(A)-binding protein in the cytoplasm (7, 13, 14, 17, 18, 24, 27, 44, 45, 52, 53). Indirect immunofluorescence on yeast cells with antibodies to the poly(A)-binding protein, produced by immunizing a mouse with the fusion protein of λ YPA72.1 (Fig. 7), shows that the 72K yeast poly(A)-binding protein is located primarily in the cytoplasm (Fig. 8). The dark areas in the cells in panel B are nuclei and vacuoles. Recently, Sachs and Kornberg (41) identified poly(A)-binding activities in two yeast proteins with approximate molecular weights of 72,000 (cytoplasmic) and 60,000 (nuclear). Taken together with what is known about the poly(A)-binding protein from higher organisms, the 72K poly(A)-binding protein product of λ YPA72.1 is an mRNA poly(A)-binding protein. However, we cannot rule out the possibility that some of the protein may also be found in the nucleus.

Hybrid selection and translation of the mRNA for the poly(A)-binding protein. The identity of the gene encoded by λ YPA72.1 was confirmed by hybrid selection and translation of the mRNA for the mRNA poly(A)-binding protein. The 2.5-kbp DNA insert was isolated from λ YPA72.1 by digestion with *EcoRI*, generating two fragments of 2.0 and 0.5 kbp which were gel purified and subcloned into the twin bacteriophage promoter plasmid pGEM-2. Yeast mRNA hybrid-selected on the plasmid containing the 2.0-kbp *EcoRI* fragment, designated pY72.1, when translated in vitro yielded a 72,000-molecular-weight protein (Fig. 9, lane marked pY72.1) which could be immunoprecipitated from the translation reaction with antiserum to the fusion protein which is specific for a 72,000-molecular-weight yeast protein (Fig. 7). In addition, two lower-molecular-weight proteins were also specifically immunoprecipitated, although their origins are unknown. No specific translation above background was seen when the plasmid containing the 0.5-kbp fragment was used for hybrid selection and translation,

indicating that all of the sequence for the fusion protein is encoded on the 2.0-kbp *EcoRI* fragment of λ YPA72.1.

RNA blot analysis of the mRNA poly(A)-binding protein mRNA. The 2.0-kbp *EcoRI* fragment was used to identify the homologous mRNA. Poly(A)⁺ RNA from yeast cells was fractionated by electrophoresis in a denaturing formaldehyde-agarose gel, blotted onto nitrocellulose paper, and hybridized with the nick-translated 2.0-kbp *EcoRI* fragment of λ YPA72.1. The probe hybridized at high stringency to one major band of about 2.1 kilobases in length (Fig. 10). One larger minor band was also detected.

DNA sequence. The complexity of the poly(A)-binding protein gene in yeast cells was investigated by Southern blot analysis (47), which indicated that a single-copy gene for the mRNA poly(A)-binding protein exists in yeast cells (data not shown). λ YPA72.1 was used to screen the genomic library by nucleic acid hybridization, and overlapping clones encompassing the entire gene were isolated. The restriction map of the gene which was subcloned as two phage recombinants, λA_1 (corresponding to λ YPA72.1, described above) and λA_2 , is shown in Fig. 11. The poly(A)-binding protein coding region and the sequencing strategy used are also indicated. The nucleotide sequence of the *KpnI* fragment, which includes the entire poly(A)-binding protein coding region, and the deduced amino acid sequence are shown in Fig. 12. The predicted size of the mRNA encoded by this transcription unit is 2.05 kilobases, in good agreement with the observed size of the mRNA (Fig. 10). Only one long open reading frame of 577 amino acids, coding for a protein of molecular weight 64,272, is found in this sequence. This is slightly lower than the apparent molecular weight in SDS-PAGE and could be the result of the high proline content. No intervening sequences are present (37). The designated initiator methionine is found in the commonly used consensus translation initiation sequence AXXATGG (28). Analysis of the codon usage shows a strong bias toward codons which are complementary to the anticodons of the major isoacceptor tRNA species (5) of *S. cerevisiae* (data not shown), suggesting a high level of expression for this protein. We noted that the sequence in the 5' untranslated region of the mRNA immediately upstream of the initiator codon is extremely adenosine rich and resembles a poly(A) segment to which the protein could bind.

From the amino acid sequence, it is apparent that the protein contains several distinct domains. Both the amino and carboxy termini of the protein are very acidic. The carboxy third of the molecule, separated from the rest of the protein by a segment of eight consecutive alanine residues, contains a segment (amino acids 426 to 489) of very high proline content and an overall positive charge due to the absence of negatively charged amino acids. Regions rich in proline and positively charged amino acids have been identified in other single-stranded-nucleic acid-binding proteins, such as the nucleocapsid protein of Semliki Forest virus (15) and the adenovirus type 5 single-stranded-DNA-binding protein (29). The amino-terminal half of the molecule has a very different structure. The sequence Arg-Val/Leu-Cys-Arg appears twice in this region, at amino acid positions 68 and 107. Perhaps the most striking feature of the entire protein sequence is a 13-amino-acid segment (underlined in Fig. 12) which is repeated three times in the middle section of the protein. The first and third segments are homologous over 12 of the 13 amino acids. The middle segment is homologous to the other two segments over eight of the amino acids of the repeat. It may also be significant that the spacing between the three elements is similar, the first and second elements


```

                20          40          60          80
GGTACCACCACCTAATATCTGGCTTTTTATTCAATAAAAACTCAAAAAAAAAAATCCAAAAAACTAAAAACCAATAAAAAATAAA

100          120          140          160
1  ATG GCT GAT ATT ACT GAT AAG ACA GCT GAA CAA TTG GAA AAC TTG AAT ATT CAA GAT GAC CAA AAG CAA GCC GCC
   Met Ala Asp Ile Thr Asp Lys Thr Ala Glu Gln Leu Glu Asn Leu Asn Ile Gln Asp Asp Gln Lys Gln Ala Ala 25

180          200          220
26  ACT GGT TCA GAA AGC CAA TCT GTT GAA AAC TCT TCT GCA TCA TTA TAT GTT GGT GAC TTA GAA CCT TCT GTT TCC
   Thr Gly Ser Glu Ser Gln Ser Val Glu Asn Ser Ser Ala Ser Leu Tyr Val Gly Asp Leu Glu Pro Ser Val Ser 50

240          260          280          300
51  GAA GCC CAC TTA TAT GAT ATC TTC TCT CCA ATC GGT TCA GTC TCC TCC ATT CGT GTC TGT CGT GAT GCC ATC ACT
   Glu Ala His Leu Tyr Asp Ile Phe Ser Pro Ile Gly Ser Val Ser Ser Ile Arg Val Cys Arg Asp Ala Ile Thr 75

320          340          360          380
76  AAG ACC TCT TTG GGC TAT GCT TAT GTT AAC TTT AAC GAC CAT GAA GCC GGC AGA AAA GCA ATT GAG CAA TTG AAC
   Lys Thr Ser Leu Gly Tyr Ala Tyr Val Asn Phe Asn Asp His Glu Ala Gly Arg Lys Ala Ile Glu Gln Leu Asn 100

400          420          440          460
101 TAC ACT CCA ATC AAG GGT AGA TTA TGC CGT ATT ATG TGG TCT CAA CGT GAC CCA TCA TTG AGA AAG AAG GGT TCT
   Tyr Thr Pro Ile Lys Gly Arg Leu Cys Arg Ile Met Trp Ser Gln Arg Asp Pro Ser Leu Arg Lys Lys Gly Ser 125

480          500          520
126 GGT AAC ATC TTT ATC AAG AAC TTG CAC CCT GAT ATT GAC AAC AAG GCT TTG TAT GAC ACT TTC TCT GTG TTT GGT
   Gly Asn Ile Phe Ile Lys Asn Leu His Pro Asp Ile Asp Asn Lys Ala Leu Tyr Asp Thr Phe Ser Val Phe Gly 150

540          560          580          600
151 GAC ATC TTG TCC AGC AAG ATT GCC ACC GAC GAA AAC GGA AAA TCC AAG GGT TTT GGG TTT GTT CAC TTC GAA GAA
   Asp Ile Leu Ser Ser Lys Ile Ala Thr Asp Glu Asn Gly Lys Ser Lys Gly Phe Gly Phe Val His Phe Glu Glu 175

620          640          660          680
176 GAA GGT GCT GCC AAG GAA GCT ATT GAT GCT TTG AAT GGT ATG CTG TTG AAC GGT CAA GAA ATT TAT GTT GCT CCT
   Glu Gly Ala Ala Lys Glu Ala Ile Asp Ala Leu Asn Gly Met Leu Leu Asn Gly Gln Glu Ile Tyr Val Ala Pro 200

700          720          740          760
201 CAC TTG TCC AGA AAG GAA CGT GAC TCT CAA TTG GAA GAG ACT AAG GCA CAT TAC ACT AAC CTT TAT GTG AAA AAC
   His Leu Ser Arg Lys Glu Arg Asp Ser Gln Leu Glu Leu Thr Lys Ala His Tyr Thr Asn Leu Tyr Val Lys Asn 225

780          800          820
226 ATC AAC TCC GAA ACT ACT GAC GAA CAA TTC CAA GAA TTG TTT GCC AAA TTT GGT CCA ATT GTT TCT GCC TCT TTG
   Ile Asn Ser Glu Thr Thr Asp Glu Gln Phe Gln Glu Leu Phe Ala Lys Phe Gly Pro Ile Val Ser Ala Ser Leu 250

840          860          880          900
251 GAA AAG GAT GCT GAT GGA AAA TTG AAG GGT TTC GGG TTT GTT AAC TAC GAA AAG CAT GAA GAC GCT GTG AAA GCT
   Glu Thr Asp Ala Asp Gly Lys Leu Lys Gly Phe Val Asn Tyr Glu Lys His Glu Asp Ala Lys Ala GAA GAA 275

920          940          960          980
276 GTT GAA GCT TTG AAT GAC TCT GAA CTA AAT GGA GAA AAG TTA TAC GTT GGT CGT GCC CAA AAG AAG AAT GAA CGT
   Val Glu Ala Leu Asn Asp Ser Glu Leu Asn Gly Glu Tyr Val Gly Arg Ala Gln Lys Lys Asn Glu Arg 300

1000          1020          1040          1060
301 ATG CAT GTC TTG AAG AAG CAA TAC GAA GCT TAC AGA TTG GAA AAA ATG GCC AAG TAC CAA GGT GTT AAT TTG TTT
   Met His Val Leu Lys Lys Gln Tyr Glu Ala Gln Tyr Arg Leu Glu Lys Met Ala Lys Met Ala Lys Tyr Gln Gly Val Asn Leu Phe 325

1080          1100          1120
326 GTT AAG AAC TTA GAT GAC AGC GTT GAT GAC GAA AAG TTG GAA GAA GAA TTT GCT CCA TAT GGT ACT ATC ACT TCT
   Val Lys Asn Leu Asp Asp Ser Val Asp Asp Glu Lys Leu Glu Glu Glu Phe Ala Pro Tyr Gly Thr Ile Thr Ser 350

1140          1160          1180          1200
351 GCA AAG GTT ATG AGA ACC GAA AAC GGT AAG TCT AAG GGT TTT GGT TGT TTC TCA ACT CCA GAG GAA GCT
   Ala Lys Val Met Arg Thr Glu Asn Gly Lys Ser Lys Gly Phe Gly Phe Val Cys Phe Ser Thr Pro Glu Glu Ala 375

1220          1240          1260          1280
376 ACT AAG GCC ATT ACA GAA AAG AAC CAA CAA ATT GTT GCT GGT AAG CCA TTA TAC GTT GCC ATT GCT CAA AGA AAA
   Thr Lys Ala Ile Thr Glu Lys Asn Gln Gln Ile Val Ala Gly Lys Pro Leu Tyr Val Ala Ile Ala Gln Arg Lys 400

1300          1320          1340          1360
401 GAC GTA AGA CGT TCT CAA TTG GCT CAA CAA ATC CAA GCC AGA AAT CAA ATG AGA TAC CAG CAA GCT ACT GCT GCC
   Asp Val Arg Arg Ser Gln Leu Ala Gln Gln Ile Gln Ala Arg Asn Gln Met Arg Tyr Gln Gln Ala Thr Ala Ala 425

1380          1400          1420
426 GCT GCC GCC GCC GCT GCC GGT ATG CCA GGT CAA TTC ATG CCT CCA ATG TTC TAT GGT GTT ATG CCA CCA AGA GGT
   Ala Ala Ala Ala Ala Ala Gly Met Pro Gly Gln Phe Met Pro Pro Met Phe Tyr Gly Val Met Pro Pro Arg Gly 450

1440          1460          1480          1500
451 GTT CCA TTC AAC GGT CCA AAC CCA CAA CAA ATG AAC CCA ATG GGC GGT ATG CCA AAG AAC GGC ATG CCA CCT CAA
   Val Pro Phe Asn Gly Pro Asn Pro Gln Gln Met Asn Pro Met Gly Gly Met Pro Lys Asn Gly Met Pro Pro Gln 475

1520          1540          1560          1580
476 TTT AGA AAT GGT CCG GTT TAC GGC GTC CCC CCA CAA GGT GGT TTC CCA AGA AAT GCC AAC GAT AAC AAC CAA TTT
   Phe Arg Asn Gly Pro Val Tyr Gly Val Pro Pro Gln Gly Gly Phe Pro Arg Asn Ala Asn Asp Asn Asn Gln Phe 500

1600          1620          1640          1660
501 TAT CAA CAA AAG CAA AGA CAA GCT TTG GGT GAA CAA TTA TAC AAG AAG GTT TCT GCT AAG ACT TCA AAT GAA GAA
   Tyr Gln Gln Lys Gln Arg Gln Ala Leu Gly Glu Gln Leu Tyr Lys Lys Val Ser Ala Lys Thr Ser Asn Glu Glu 525

1680          1700          1720
526 GCA GCT GGT AAA ATT ACT GGT ATG ATT TTG GAT TTG CCA CCT CAA GAG GTC TTC CCA TTG TTG GAA AGT GAT GAA
   Ala Ala Gly Lys Ile Thr Gly Met Ile Leu Asp Leu Pro Pro Gln Glu Val Phe Pro Leu Leu Glu Ser Asp Glu 550

1740          1760          1780          1800
551 TTG TTC GAA CAA CAC TAC AAA GAA GCT TCT GCT GCC TAT GAG TCT TTC AAA AAG GAG CAA GAA CAA CAA ACT GAG
   Leu Phe Glu Gln His Tyr Lys Glu Ala Ser Ala Ala Tyr Glu Ser Phe Lys Lys Glu Gln Glu Gln Gln Thr Glu 575

1820
576 CAA GCT TAA
   Gln Ala End 577

1840          1860          1880          1900          1920
TGCTCTATGTAATCACCTACTTCCCTACTCAACAACTATCATCTTTTTTTCTCCCTTCCCTTTTTTTTTTGTCTTAAATACTAGTTATTATCCTTTG
1940          1960          1980          2000          2020
TAACCTAAATACTCGCTCTATTGACAAAATGCAAACTTTTTTTTAACTGTCAGAGAAGTATCATTTTACTATTATTAGACTCTCATTACTGAGT
2040          2060          2080          2100          2120
TTATTAGTTGCGTTTTTTTGAGGTCTCATTATTAAACGGCAGAATGTAGTTTTTCGGGAAATAAAACAAAATAGCACGGAAAAAGACCAACAGGTTCCA
2140          2160
GGAAATACTAATTACAAAAAATTGACAAAATCAAGTGAGGTACC

```

FIG. 12. Nucleotide sequence of the poly(A)-binding protein gene and the deduced amino acid sequence. The amino acid repeat segments and the consensus poly(A) addition signal are underlined.

yeast	161	GLU	ASN	GLY	LYS	SER	LYS	GLY	PHE	GLY	PHE	VAL	His	PHE
mRNA														
poly(A)	255	Asp	GLY	LYS	Leu	LYS	GLY	PHE	GLY	PHE	VAL	Asn	TYR	
binding														
protein	357	GLU	ASN	GLY	LYS	SER	LYS	GLY	PHE	GLY	PHE	VAL	Cys	PHE
mammalian		52	Arg	SER	ARG	GLY	PHE	GLY	PHE	VAL	Thr	TYR		
A1 hnRNP														
protein	146	GLY	LYS	Lys	ARG	GLY	PHE	Ala	PHE	VAL	Thr	PHE		

FIG. 13. Sequence comparison of the yeast poly(A)-binding protein conserved repeated sequences and the homologous amino acid segments of mammalian A1 hnRNP proteins. The amino acid number is indicated as a superscript. The sequence of A1 is from the data of Cobianchi et al. (10) for the rat brain protein and is identical to the sequence of the UP1 calf thymus protein from the data of Williams et al. (55).

being separated by 84 amino acids and the second and third elements being separated by 92 amino acids. Extensive homology (67%) extends also beyond the underlined segments over large domains (92 amino acids each; 126 to 218, 219 to 311, and 322 to 414).

DISCUSSION

We identified the proteins that interact with poly(A)⁺ RNA in the yeast *S. cerevisiae* by UV cross-linking. As in higher eucaryotes, yeast poly(A)⁺ RNAs also exist in intact cells as RNA-protein complexes. The major proteins that become cross-linked to RNA in yeast have molecular weights of 72,000, 60,000, and 50,000. Several additional proteins with molecular weights of 30,000 to 45,000 were also visible after longer exposures of the autoradiograms. The 72K protein is cross-linked to the poly(A) segment of cytoplasmic RNA. Although the pattern of proteins that can be cross-linked to total cell poly(A)⁺ RNA in yeast cells is considerably simpler than that in vertebrate cells, at least one of these proteins, the 72K mRNA poly(A)-binding protein, appears to be very similar to that of higher organisms.

As demonstrated here and in previous reports (1, 12), UV-cross-linked RNPs can be efficiently used as antigens to generate specific antibodies to genuine RNP proteins (RNA-contacting proteins in vivo). Unfractionated antiserum containing specific antibodies to several of the major yeast RNP proteins was used to isolate recombinant phage expressing yeast RNP proteins from a yeast genomic DNA expression library, and the purified phage were further characterized to identify the specific RNP protein genes that they encode. One phage, λYPA72.1, was found to express the 72K poly(A)-binding protein. Antibodies prepared to the protein expressed from this gene in bacteria are specific for the poly(A)-binding protein.

Several previous reports indicate that the poly(A) tail of mRNA in vertebrate cells is associated with a 72,000- to 78,000-molecular-weight protein (7, 13, 14, 17, 18, 23, 24, 27, 44, 45, 52, 53), which can be cross-linked by UV light to the mRNA poly(A) tail in intact cells (13, 18). Of the proteins which bind poly(A)⁺ mRNA, the poly(A)-binding protein is the only one for which selective-sequence binding to the RNA has been identified. Kornberg and colleagues have characterized a 75,000-molecular-weight protein from mouse L cells which binds poly(A) and which can organize the poly(A) segment into a repeating RNP structure in vitro (3, 4). It is probable that this is the same protein which binds poly(A) in vivo. Recently, these workers (41) have also identified a cytoplasmic poly(A)-binding activity which can be eluted from an SDS-polyacrylamide gel of yeast whole-

cell extract in the 66,000-to-79,000-molecular weight region. Another poly(A)-binding activity elutes from gels of yeast nuclear extracts in the 50,000-to-59,000-molecular-weight region.

Although the mRNA poly(A)-binding protein was the first mRNA-binding protein to be described (7) and is quantitatively the most abundant of the mRNA-binding proteins, little is known about its function. It has been suggested that it is involved in various important aspects of mRNA metabolism, including nucleo-cytoplasmic transport (44), mRNA stability (58), and translation (1, 14, 18, 32, 43, 52, 53), and also that it may be related to a poly(A) polymerase (40). The 72K poly(A)-binding protein from vertebrates is poorly immunogenic in mice and rabbits, and previous attempts by a number of laboratories to produce antibodies to it have not been successful. We obtained, in addition to specific antibodies to the poly(A)-binding protein from yeast cells, antibodies which are specific for the poly(A)-binding protein from other organisms. This should facilitate the study of the metabolism and function of the protein. The identification of the yeast poly(A)-binding protein and the isolation and sequence of the gene and protein for this RNP protein will now enable us to study the function, metabolism, and evolution of this protein and of the RNP particles.

The amino acid sequence analysis of the poly(A)-binding protein revealed 92-amino-acid-long homologous repeating domains and a 13-amino-acid sequence that is repeated three times in the protein (underlined in Fig. 12). Routine computer search revealed no significant homology between this sequence and any other protein sequence that is in the National Biomedical Research Foundation protein data base. However, upon visual inspection of the recently published sequences of the UP1 calf thymus helix-destabilizing protein (55) and the rodent helix-destabilizing protein (10), we found that they both contain two regions highly homologous to the repeat sequence in the yeast poly(A)-binding protein that is defined by the consensus sequence (Fig. 13). Interestingly, it has recently been shown that antibodies to UP1 (which was first isolated by Herrick and Alberts [20] as a single-stranded-DNA-binding protein) react with a group of mammalian proteins in the 30- to 40-kilodalton range (50) which are the hnRNP proteins (36). The 24,000-dalton UP1 is derived from the hnRNP proteins by proteolysis which occurs during protein purification (36). Specifically, UP1 is the amino terminus nucleic acid-binding fragment of the A1 hnRNP protein (9; personal communication).

The amino acid sequence comparison of the major mRNP protein from yeast cells and one of the major hnRNP proteins of mammals identified a conserved and repeating RNP consensus sequence. We anticipate that this sequence may be common to other RNP proteins. The high degree of conservation and duplication of the sequence in such widely divergent organisms is striking. It suggests that these nuclear and cytoplasmic RNP proteins have a common evolutionary origin and most likely arose from a common ancestral gene. The conservation and internal duplication also imply an important, albeit yet unknown, function for this sequence. It is possible that the consensus sequences are involved in RNA binding. Sequential aromatic amino acids and positively charged amino acids may be important in the binding of protein to single-stranded nucleic acids (for reviews, see references 9 and 19) through sequential intercalation of the aromatic residues with the nucleotide bases and the interaction of the positively charged amino acids with the negatively charged phosphodiester backbone. The proline-rich region of the protein may also be involved in RNA binding,

as regions of similar general sequence character are found in other single-stranded-nucleic acid-binding proteins (15, 29). It is possible that the high affinity and binding specificity of the protein for the poly(A) tail are a consequence of an interaction involving these two distinct domains.

ACKNOWLEDGMENTS

We thank Lawrence Dumas for supplying the yeast, Michael Snyder and Ronald Davis for providing the λ gt11 library; Nina Agabian and Shulamit Michaeli for the fusion protein protocol; and Lawrence Dumas, Daniel Linzer, and David Miller for helpful discussions and comments on the manuscript.

This work was supported by grant GM-31888 from the National Institutes of Health.

LITERATURE CITED

- Adam, S. A., Y. D. Choi, and G. Dreyfuss. 1986. Interaction of mRNA with proteins in vesicular stomatitis virus-infected cells. *J. Virol.* **57**:614-622.
- Akhayat, O., A. Vincent, S. Goldenberg, A. Person, and K. Scherrer. 1983. The translation of the messenger for the poly(A)-binding protein associated with translated mRNA is suppressed. *FEBS Lett.* **162**:25-32.
- Baer, B. W., and R. D. Kornberg. 1980. Repeating structure of cytoplasmic poly(A)-ribonucleoprotein. *Proc. Natl. Acad. Sci. USA* **77**:1890-1892.
- Baer, B. W., and R. D. Kornberg. 1983. The protein responsible for the repeating structure of cytoplasmic poly(A) ribonucleoprotein. *J. Cell Biol.* **96**:717-721.
- Bennetzen, J. L., and B. D. Hall. 1982. The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase I. *J. Biol. Chem.* **257**:3018-3025.
- Berger, S., and C. S. Birkenmeier. 1979. Inhibition of intractable nucleases with ribonucleotide-vanadyl complexes: isolation of messenger ribonucleic acid from resting lymphocytes. *Biochemistry* **18**:5743-5749.
- Blobel, G. 1973. A protein of molecular weight 78,000 bound to the polyadenylate region of eukaryotic messenger RNA. *Proc. Natl. Acad. Sci. USA* **70**:924-928.
- Burnette, W. N. 1981. "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Anal. Biochem.* **112**:195-203.
- Chase, J. W., and K. R. Williams. 1986. Single-stranded DNA binding proteins required for DNA replication. *Annu. Rev. Biochem.* **55**:103-136.
- Cobianchi, F., D. N. SenGupta, B. Z. Zmudka, and S. H. Wilson. 1986. Structure of rodent helix-destabilizing protein revealed by cDNA cloning. *J. Biol. Chem.* **261**:3536-3543.
- Dreyfuss, G., S. A. Adam, and Y. D. Choi. 1984. Physical change in cytoplasmic messenger ribonucleoproteins in cells treated with inhibitors of mRNA transcription. *Mol. Cell. Biol.* **4**:415-423.
- Dreyfuss, G., Y. D. Choi, and S. A. Adam. 1984. Characterization of heterogeneous nuclear RNA-protein complexes in vivo with monoclonal antibodies. *Mol. Cell. Biol.* **4**:1104-1114.
- Dreyfuss, G., Y. D. Choi, and S. A. Adam. 1986. The ribonucleoprotein structures along the pathway of mRNA formation. In L. DeGroot (ed.), *Mechanism of action of thyroid hormones*. Academic Press, Inc., New York.
- Gaedigk, R., S. Oehler, K. Kohler, and B. Setyono. 1985. In vitro reconstitution of messenger ribonucleoprotein particles from globin messenger RNA and cytosol proteins. *FEBS Lett.* **179**:201-207.
- Garoff, H., A. M. Frischauf, K. Simons, H. Lehrach, and H. Delius. 1980. The capsid protein of Semliki Forest virus has clusters of basic amino acids and prolines in its amino-terminal region. *Proc. Natl. Acad. Sci. USA* **77**:6376-6380.
- Greenberg, J. R. 1980. Proteins crosslinked to messenger RNA by irradiating polyribosomes with UV light. *Nucleic Acids Res.* **8**:5685-5701.
- Greenberg, J. R. 1981. The polyribosomal mRNA-protein complex is a dynamic structure. *Proc. Natl. Acad. Sci. USA* **78**:2923-2926.
- Greenberg, J. R., and E. Carroll III. 1985. Reconstitution of functional mRNA-protein complexes in a rabbit reticulocyte cell-free translation system. *Mol. Cell. Biol.* **5**:342-351.
- Helene, C., and J. C. Maurizot. 1981. Interactions of oligopeptides with nucleic acids. *Crit. Rev. Biochem.* **10**:213-258.
- Herrick, G., and B. Alberts. 1971. Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. *J. Biol. Chem.* **251**:2124-2132.
- Hopper, A., F. Banks, and V. Evangelidis. 1978. A yeast mutant which accumulates precursor tRNAs. *Cell* **14**:211-219.
- Jain, S. K., M. G. Pluskal, and S. Sarkar. 1979. Thermal chromatography of eukaryotic messenger ribonucleoprotein particles on oligo(dT)-cellulose. *FEBS Lett.* **97**:84-90.
- Jeffrey, W. R. 1977. Characterization of polypeptides associated with messenger RNA and its polyadenylate segment in Ehrlich ascites messenger ribonucleic acid. *J. Biol. Chem.* **252**:3525-3532.
- Jeffrey, W. R. 1978. Composition and properties of messenger ribonucleoprotein fragments containing and lacking polyadenylate. *Biochim. Biophys. Acta* **521**:217-228.
- Johnson, A. D., and I. Herskowitz. 1985. A repressor (MAT α 2 product) and its operator control expression of a set of cell type specific genes in yeast. *Cell* **42**:237-247.
- Kilmartin, J. V., and A. E. M. Adams. 1984. Structural rearrangements of tubulin and actin during the cell cycle of the yeast *Saccharomyces*. *J. Cell Biol.* **98**:922-933.
- Kish, V. M., and T. Pederson. 1976. Poly(A)-rich ribonucleoprotein complexes of HeLa cell messenger RNA. *J. Biol. Chem.* **251**:5888-5894.
- Kozak, M. 1984. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin *in vivo*. *Nature (London)* **308**:241-246.
- Kruijer, W., F. M. A. van Schaik, and J. S. Sussenbach. 1981. Structure and organization of the gene for the DNA binding protein of adenovirus type 5. *Nucleic Acids Res.* **9**:4439-4457.
- Kumar, A., and T. Pederson. 1975. Comparison of proteins bound to heterogeneous nuclear RNA and messenger RNA in HeLa cells. *J. Mol. Biol.* **96**:353-365.
- Laskey, R. A., and A. D. Mills. 1975. Quantitative film detection of ^3H and ^{14}C in polyacrylamide gels by fluorography. *Eur. J. Biochem.* **56**:335-341.
- Liautard, J. B., B. Setyono, E. Spindler, and K. Kohler. 1976. Comparison of proteins bound to the different functional classes of messenger RNA. *Biochim. Biophys. Acta* **425**:373-383.
- Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* **101**:20-78.
- Nakagawa, T. Y., M. S. Swanson, B. J. Wold, and G. Dreyfuss. 1986. cDNA cloning of the nuclear RNP particle C proteins: a conserved gene family. *Proc. Natl. Acad. Sci. USA* **83**:2007-2011.
- O'Farrell, P. Z., H. M. Goodman, and P. H. O'Farrell. 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell* **12**:1133-1142.
- Pandolfo, M., O. Valentini, G. Biamonti, C. Morandi, and S. Riva. 1985. Single stranded DNA binding proteins derive from hnRNP proteins by proteolysis in mammalian cells. *Nucleic Acids Res.* **13**:6577-6590.
- Pikielny, G. W., J. L. Teem, and M. Rosbash. 1983. Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: implications for U1 RNA and metazoan mRNA splicing. *Cell* **34**:395-403.
- Ricciardi, R. P., J. S. Miller, and B. E. Roberts. 1979. Purification and mapping of specific mRNAs by hybridization selection and cell free translation. *Proc. Natl. Acad. Sci. USA* **76**:4927-4931.
- Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**:237-251.

40. **Rose, K. M., S. T. Jacob, and A. Kumar.** 1979. Poly(A) polymerase and poly(A) specific mRNA binding protein are antigenically related. *Nature (London)* **279**:260-262.
41. **Sachs, A. B., and R. D. Kornberg.** 1985. Nuclear polyadenylate-binding protein. *Mol. Cell. Biol.* **5**:1993-1996.
42. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
43. **Schmid, H.-P., M. Schonfelder, B. Setyono, and K. Kohler.** 1983. 76-kDa poly(A) protein is involved in the formation of 48S initiation complexes. *FEBS Lett.* **157**:105-110.
44. **Schwartz, H., and J. E. Darnell.** 1976. The association of protein with the polyadenylic acid of HeLa cell messenger RNA: evidence for a "transport" role of 75,000 molecular weight peptide. *J. Mol. Biol.* **104**:833-851.
45. **Setyono, B., and J. R. Greenberg.** 1981. Proteins associated with poly(A) and other regions of mRNA and hnRNA molecules as investigated by crosslinking. *Cell* **24**:775-783.
46. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
47. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.
48. **Sundquist, B., T. Persson, and U. Lindberg.** 1977. Characterization of mRNA-protein complexes from mammalian cells. *Nucleic Acids Res.* **4**:899-915.
49. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **77**:5201-5205.
50. **Valenti, O., G. Biamonti, M. Pandolfo, C. Morandi, and S. Riva.** 1985. Mammalian single-stranded DNA binding proteins and heterogeneous nuclear RNA proteins have common antigenic determinants. *Nucleic Acids Res.* **13**:337-346.
51. **van Eekelen, C. A., T. Riemen, and W. J. van Venrooij.** 1981. Specificity in the interaction of hnRNA and mRNA with proteins as revealed by *in vivo* crosslinking. *FEBS Lett.* **130**:223-226.
52. **van Venrooij, W. J., C. A. G. van Eekelen, R. T. P. Jansen, and J. M. G. Princen.** 1977. Specific poly-A-binding protein of 76,000 molecular weight in polyribosomes is not present on poly A of free cytoplasmic mRNP. *Nature (London)* **270**:189-191.
53. **Vincent, A., S. Goldenberg, and K. Scherrer.** 1981. Comparisons of the proteins associated with duck-globin mRNA and its polyadenylated segment in polyribosomal and repressed free messenger ribonucleoprotein complexes. *Eur. J. Biochem.* **114**:179-193.
54. **Wagenmakers, A. J. M., R. J. Reinders, and W. J. van Venrooij.** 1980. Crosslinking of mRNA to proteins by irradiation of intact cells with ultraviolet light. *Eur. J. Biochem.* **112**:323-330.
55. **Williams, K. R., K. L. Stone, M. B. LoPresti, B. M. Merrill, and S. R. Planck.** 1985. Amino acid sequence of the UP1 calf thymus helix-destabilizing protein and its homology to an analogous protein from mouse myeloma. *Proc. Natl. Acad. Sci. USA* **82**:5666-5670.
56. **Winston, F., F. Chumley, and G. R. Fink.** 1983. Eviction and transplacement of mutant genes in yeast. *Methods Enzymol.* **101**:211-228.
57. **Young, R. Y., and R. D. Davis.** 1983. Efficient isolation of genes using antibody probes. *Proc. Natl. Acad. Sci. USA* **80**:1194-1198.
58. **Zeevi, M., J. R. Nevins, and J. E. Darnell, Jr.** 1982. Newly formed mRNA lacking polyadenylic acid enters the cytoplasm and the polyribosomes but has a shorter half-life in the absence of polyadenylic acid. *Mol. Cell. Biol.* **2**:517-525.