Volume 14 Number 10 1986 Nucleic Acids Research

A cDNA clone of the hnRNP C proteins and its homology with the single-stranded DNA binding protein UP2

Debomoy K.Lahiri and John O.Thomas

Department of Biochemistry, New York University School of Medicine, NY 10016, USA

Received 27 January 1986; Accepted 22 April 1986

ABSTRACT

A cDNA clone which expresses a protein that cross-reacts immunologically with the human C1 and C2 hnRNP core proteins has been isolated. The clone was selected by a sensitive immunochemical assay employing an avidin-biotin complex for detection, and identified as a clone for the hnRNP C proteins by a highly sensitive antibody select assay that is described here. The clone contains 677 nucleotides, and, as shown by northern blotting, is derived from a 1.5 Kb poly(A)$^+$ mRNA. There are regions of strong homology between the human and mouse genes, weak homology is seen with chicken DNA, and very little, if any, homology can be detected with Drosophila, Artemia, sea urchin, or yeast DNAs. Two peptides (a total of 24 amino acids) of the calf thymus single-stranded DNA binding protein UP2 show perfect homology with the deduced amino acid sequence of the clone, suggesting that UP2 is related to the hnRNP C proteins. There is also a region that has a sequence very similar to two regions of the single-stranded DNA binding protein UP1 that contain proposed DNA binding sites.

INTRODUCTION

As it is being synthesized, hnRNA combines with a set of proteins to form heterogeneous nuclear ribonucleoprotein (hnRNP) complexes (1,2). This is an early event in the metabolism of nuclear RNA, occurring on nascent transcripts (3); all subsequent steps in the processing and transport of RNA take place with RNA-protein complexes as substrates, not free RNA. This may be of considerable importance since the conformation of the RNA is probably altered by the proteins with which it is associated. For example, the major hnRNP protein component of the brine shrimp Artemia is a helix destabilizing protein that unwinds most of the secondary structure of the RNA, and then compacts the unwound RNA into a structure that has a "beads-on-a-string" appearance (4,5). There is also evidence to suggest that the distribution of hnRNP proteins on the RNA may be sequence dependent (3,6). Moreover, it is likely that these hnRNP complexes play a role in the splicing of pre-mRNA as suggested by studies in which the inhibition of an in

vitro splicing system by monoclonal antibodies to the hnRNP C proteins were investigated (7). While the proteins of the hnRNP complex may function in both the packaging of hnRNA and in its metabolism, very little is known about them.

Nuclear pre-mRNAs can be isolated as nucleoprotein complexes that, in humans, contain several major proteins along with a large number of minor components (1,2). In interphase cells, the hnRNP complexes are restricted to the nucleus, but during mitosis the major hnRNP proteins, in association with high molecular weight RNA, are distributed throughout the cell (8). The major protein components, the core hnRNP proteins, have molecular weights between 30,000 and 45,000. When electrophoresed, these proteins appear as three doublets, and are referred to as the A1, A2, B1, B2, C1, and C2 core hnRNP proteins in order of increasing molecular weight (1). Each of these six protein bands can be further separated into several components on two dimensional gels (2). The proteins are immunologically cross-reactive (8-10), suggesting that they are closely related, and peptide mapping suggests that they are derived from a smaller number of precursor proteins (11). The amino acid compositions of the proteins are similar, characterized by a high content of glycine (about 20%), very little cystine, blocked NH_2 termini, and the modified amino acid dimethylarginine (2). Although the exact molecular weight distribution of the hnRNP proteins differs between species, these general features are well conserved, extending to crustaceans as evidenced by the properties of HD40, the major hnRNP protein of Artemia (12), and to plants, as evidenced by the properties of the major hnRNP proteins of wheat (13). There is also a great deal of immunological cross-reactivity between the hnRNP proteins of different species (8-10,14), supporting the evolutionarily conserved nature of these proteins.

Recently, it has become apparent that there is a considerable amount of homology between the hnRNP proteins and several single-stranded DNA binding proteins that have been isolated from mammalian cells. The properties of several of these proteins, including UP1 and UP2 from calf thymus, and HDP-1 from mouse myeloma have been investigated (15-18). Immunological studies have indicated that these proteins and hnRNP proteins are antigenically related (17), and an analysis of amino acid sequences shows that there is a clear relationship between the hnRNP A1 protein and UP1, and between the hnRNP A2 protein and HDP-1 (19). Trypsin and trypsin-like enzymes have been shown to cleave hnRNP proteins to yield single stranded

DNA binding proteins of 24,000-28,000 molecular weight, suggesting a product precursor relationship between some of these proteins, particularly the hnRNP A1 protein and UP1 (18,19).

To facilitate the analysis of the hnRNP complexes, we have isolated a partial cDNA clone of the human hnRNP C proteins. In this report we describe this clone, and present sequencing data that show that the single stranded DNA binding protein UP2 is closely related to the hnRNP C proteins.

EXPERIMENTAL PROCEDURES

The chicken anti-bovine core hnRNP antibodies that were used have been described (8). Although bovine hnRNP core proteins were used for immunization, the antibodies recognize the entire set of hnRNP core proteins, both human and bovine.

Immunodetection.

Immunodetection by a double antibody method was done as follows. Free sites on nitrocellulose paper containing the adsorbed antigen were blocked with 3% gelatin in TBS (20 mM Tris, pH 7.4, 0.5 M NaCl) for 1 hr, and the paper was washed in TBST (TBS with 0.05% Tween-20). The paper was incubated with anti-hnRNP antiserum at a dilution of 1:400 in TBSG (TBS with 1% gelatin) for 1 hr with gentle shaking, then washed three times with TBST. Horseradish peroxidase conjugated rabbit anti-chicken IgG (Cappel Labs) at a dilution of 1:800 in TBSG was then applied for 1 hr followed by a wash with TBST. Color was generated by reacting with color substrate (4 ml of 0.3% Bio-Rad color development reagent containing 4-chloro-1-naphthol in methanol added to 20 ml TBS with 12 µl of 30% H_2O_2 added immediately before use). Immunodetection using an avidin-biotin complex (ABC) was done by replacing the peroxidase conjugated second antibody used in the above procedure with biotinylated rabbit anti-chicken IgG prepared by the method of Hsu *et al*. (20). Prior to the addition of color substrate, the paper was incubated with an avidin-biotinylated horseradish peroxidase complex (Vectastain ABC kit, Vector Labs, Burlingame, CA) for 1 hr at room temperature and washed as above.

Isolation of a cDNA clone by immunoscreening.

A λgt11 cDNA expression library, made from human hepatoma Li-7 poly $(A)^+$ RNA, was obtained from Dr. Jeffrey DeWet. The library contains an average insert of 0.78 Kb, and has been described (21). To screen this library, *E. coli* Y1090 (22) was infected with phage at a low m.o.i. and plated at a density of 50,000 plaques per 150 mm plate according to the

method of Young and Davis (22). Plaques were transfered to dry nitrocellulose filters (Schleicher and Schuell) that had been soaked in 10 mM isopropyl ß-D-thiogalactopyranoside (IPTG). The presence of positive plaques was detected by immunoassay as described above, and purified by at least two additional rounds of screening. The inserts were then subcloned into pBR322 for restriction analysis and preparation of probes. Labeled probes were prepared by nick translation (26).

Antibody selection.

Lysogens of E. coli Y1089 (22) were made with λDL4 and with λgt11. The lysogens were grown at 30° to 0.4-0.5 A_{610}, induced at 42° for 20 min, and, following the addition of IPTG to 10 mM, they were grown for 2.5 hr at 37°. Cells were harvested quickly at room temperature, and the pellet resuspended in 1/10 volume of 10 mM Tris HCl, 10 mM NaCl, 0.1 mM PMSF, pH 7.5 (TNP). The cells were lysed by three cycles of rapid freeze/thawing, and the lysate was clarified by centrifugation at 5,000g for 10 min at 4°. The ß-galactosidase or ß-galactosidase fusion protein was partially purified (all steps at 4°) by ammonium sulfate precipitation. To 10 ml of lysate, solid ammonium sulfate was added to 20% saturation (1.14 g). After stirring for 30 min, the precipitate was collected by centrifugation at 10,000g for 15 min. The supernatant was adjusted to 30% saturation by the addition of 0.59 g solid ammonium sulfate. After stirring for 30 min, the precipitate was collected by centrifugation at 10,000g for 15 min. The pellet was resuspended in 1 ml of TNP and dialyzed extensively against TNP. This partially purified fusion protein was adsorbed onto the surface of a 24 mm diameter nitrocellulose filter (1 hr at room temperature). A siliconized glass scintilation vial makes a convenient vessel. The filters were then incubated with 1% gelatin for 1 hr to block any remaining protein binding sites. Serum, diluted 1:10 in TBSG was added (2 ml) and incubated with gentle shaking for 1 hr at room temperature. The unbound serum was removed, and the filter was given four 5 min washes with 5 ml of TBST. The selected antibodies were eluted by incubating the filter for 5 min with 0.5 ml of 0.25 M glycine-HCl, pH 2.5, 0.5 M NaCl, removing the filter, and immediately adjusting the pH by adding 0.5 ml 1 M Tris, pH 8.0. The eluted antibody was used to probe western blots of hnRNP complexes. These complexes were prepared by sedimenting a pH 8.0 HeLa cell nuclear extract (23) through a sucrose gradient, and collecting the 40S peak, which contains the major portion of the hnRNP core proteins complexed with nuclear RNA (1).

Blotting Procedures.

For western blots, proteins separated on Laemmli gels (29) were electrophoretically transferred to nitrocellulose (24). Antigens were detected by ABC immunoassay. For northern blots, poly(A)$^+$ RNA (2.5 µg) was denatured and subjected to electrophoresis in 1.5% agarose-formaldehyde gels (25). The RNA was then transferred to nitrocellulose, and hybridized as described (26). For Southern blots, 10 µg of DNA was digested with the indicated enzyme, electrophoresed on a 1% agarose gel, transferred to nitrocellulose, and hybridized as described (26). Probes were prepared from pBR322 containing the subcloned cDNA insert. The filters were washed three times with 2 X SSC, 0.1% SDS at room temperature followed by three washes with either 0.2 X SSC, 0.1% SDS at 68° (high stringency) or with 2 X SSC, 0.1% SDS at 50° (low stringency).

Sequencing.

The DNA insert, cleaved by Eco RI, was subcloned into the Eco RI sites of M13mp8 and M13mp9 (27). Two fragments, generated by cleavage of the Eco RI insert by Bgl II at a site 327 nucleotides from one end were also subcloned into the Eco RI/Bam HI sites of M13mp8 and M13mp9 (27). Sequencing was done by the dideoxy chain termination method (28). The nucleic acid sequences were analyzed using MicroGenie (Beckman) software.

RESULTS

A human cDNA expression library constructed in λgt11 was screened with antibodies that cross-react with the major 30,000-45,000 molecular weight human hnRNP core proteins. These antibodies, which have been described (8), were raised in chickens that were immunized with the complete set of the 30,000-45,000 molecular weight hnRNP proteins isolated from bovine brain. During the course of the screening, several different methods for detecting the presence of an expressed antigen were investigated. The two methods that provided the greatest sensitivity and best signal to noise ratio were an avidin-biotin complex (ABC) assay, and an indirect second antibody assay. A comparison of these two assays using hnRNP proteins as a standard antigen shows that the ABC assay is at least four times more sensitive, and gives about the same amount of background color (Fig. 1A,B). The appearance of positive plaques in the second round of screening is shown in Fig. 1C. It can be seen that an ABC immunochemical detection system provides a sensitive method for screening λgt11 expression libraries.

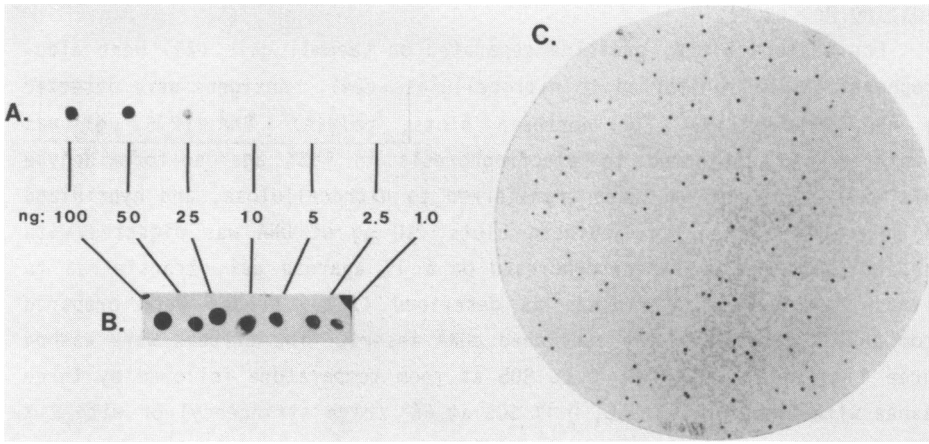

FIG. 1. Immunodetection of antigens and plaques by double antibody (A) and ABC (B,C) immunoassays. Filters containing the indicated amounts of antigen (A,B) or plaques (c) of a secondary screening were developed as described in Experimental Procedures.

A screening of more than 200,000 recombinants by the ABC immunochemical assay yielded several putative hnRNP clones. After several rounds of purification and immunological screening (the appearance of the filter after the secondary screening is shown in Fig. 1C), one clone, λDL4 was selected for the studies that are reported here. In order to further characterize the fusion protein, lysogens of λgtll and λDL4 were prepared. After induction by IPTG, cell lysates of these lysogens were analyzed by SDS polyacrylamide electrophoresis. As shown in Fig. 2A, the synthesis of both ß-galactosidase and the fusion protein is dependent on induction with IPTG. Somewhat more protein is synthesized by cells in which the phage have been induced by heating to 42° and growing at 37° than in the cells grown at 30°. A fusion protein with a molecular weight of about 140,000 is synthesized by λDL4. This corresponds to a fragment of about 24,000 molecular weight fused to the ß-galactosidase of λgtll, and is the size expected from a 700 base pair cDNA insert. This fusion protein cross-reacts very strongly with anti-hnRNP antibodies but ß-galactosidase does not, as shown by immunoblotting experiments (Fig. 2B).

Based on the results of a sensitive antibody select analysis, we have identified the 24,000 molecular weight polypeptide insert as a fragment of the C1 and/or C2 hnRNP core proteins. As shown in Fig. 3B, antibodies that

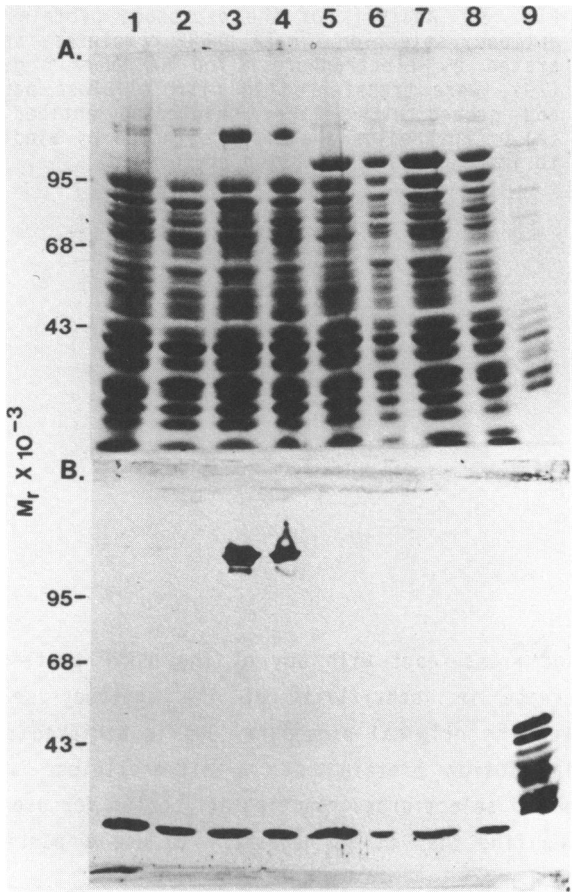

FIG. 2. Fusion proteins produced by λDL4 and λgtll lysogens. Lysates of λDL4 (lanes 1-4) and λgtll (lanes 5-8) were analyzed by electrophoresis on 8% Laemmli gels (29) (A) and by western blotting (B). The lysogens were induced at 42° and grown for 2 hr at 37° (lanes 2,3,6,7) or grown at 30° with no temperature shift (lanes 1,4,5,8) in the presence of 10 mM IPTG (lanes 3,4,7, 8) or in its absence (lanes 1,2,5,6). A HeLa cell pH 8 nuclear extract, which contains the hnRNP complexes, is in lane 9.

have been selected by the fusion protein are highly enriched in antibodies that cross-react with these two proteins as compared to unselected serum. When hnRNP proteins are used to select antibodies in place of the fusion protein, the selected antibodies cross-react the same as unselected serum (not shown). Antibodies selected with ß-galactosidase obtained and purified by the same procedure as the fusion protein but from bacteria infected

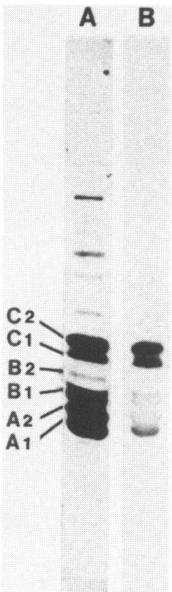

FIG. 3. Analysis of the expressed protein by antibody selection. HeLa hnRNP complexes, separated by electrophoresis on 8% Laemmli gels (29), were transferred to nitrocellulose paper and probed with either unselected antibodies (A) or antibodies that were selected by binding to immobilized λDL4 fusion protein (B).

with λgtll do not cross-react with any of the hnRNP proteins in western blots. The increase in sensitivity of the antibody select analysis employed here over the original procedure (30) is achieved by adsorbing a partially purified fusion protein onto a nitrocellulose solid support. This is then used to select cross-reacting antibodies for use as probes for western blots. We find that it is necessary to use a partially purified

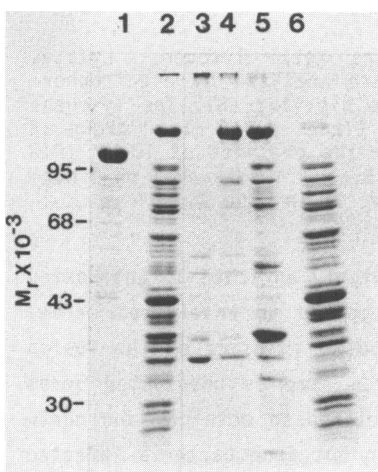

FIG. 4. Purification of ß-galactosidase fusion proteins for selection of antibodies. A lysate of an induced λDL4 lysogen was fractionated by ammonium sulfate precipitation, and the fractions were analyzed by electrophoresis on an 8% Laemmli gel (29). Lane 1, commercial ß-galactosidase; 2, crude lysate; 3-6, ammonium sulfate precipitates at 0-20%, 20-30%, 30-40%, and 40-60% saturation respectively.

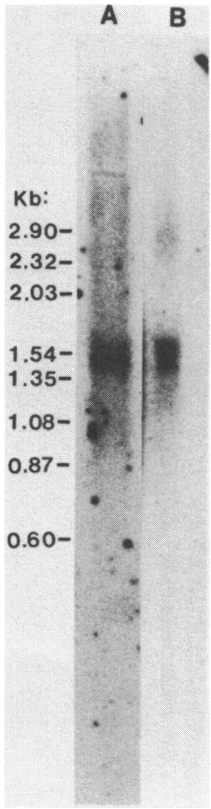

FIG. 5. Northern analysis of human (HeLa cell) (A), and mouse liver (B) poly(A)+ mRNA probed at high stringency. The two blots are at different exposures. When exposed under the same conditions, the band in (B) is about 10 fold weaker than the band in (A).

fusion protein since nitrocellulose has a limited protein binding capacity. Conceivably, a different matrix such as a cyanogen bromide activated gel could be used, but nitrocellulose provides a very convenient, easy to use, and in our hands, more reliable support. The fusion protein can be easily enriched to about 80% purity by ammonium sulfate precipiation; both ß-galactosidase and the fusion protein are found at maximum purity in the 20-30% saturation fraction (Fig. 4, Lane 4). By using a partially purified fusion protein, a several fold increase in sensitivity of antibody selection is obtained.

Northern blots of human (HeLa cell) and mouse liver poly(A)$^+$ RNA probed with the cloned cDNA both show hybridizing RNAs of about 1.5 Kb (Fig. 5). RNA of this size is somewhat larger than the minimum size required to code for a protein of 42,000 molecular weight. The difference in size may reflect the presence of 5' and 3' untranslated regions of the

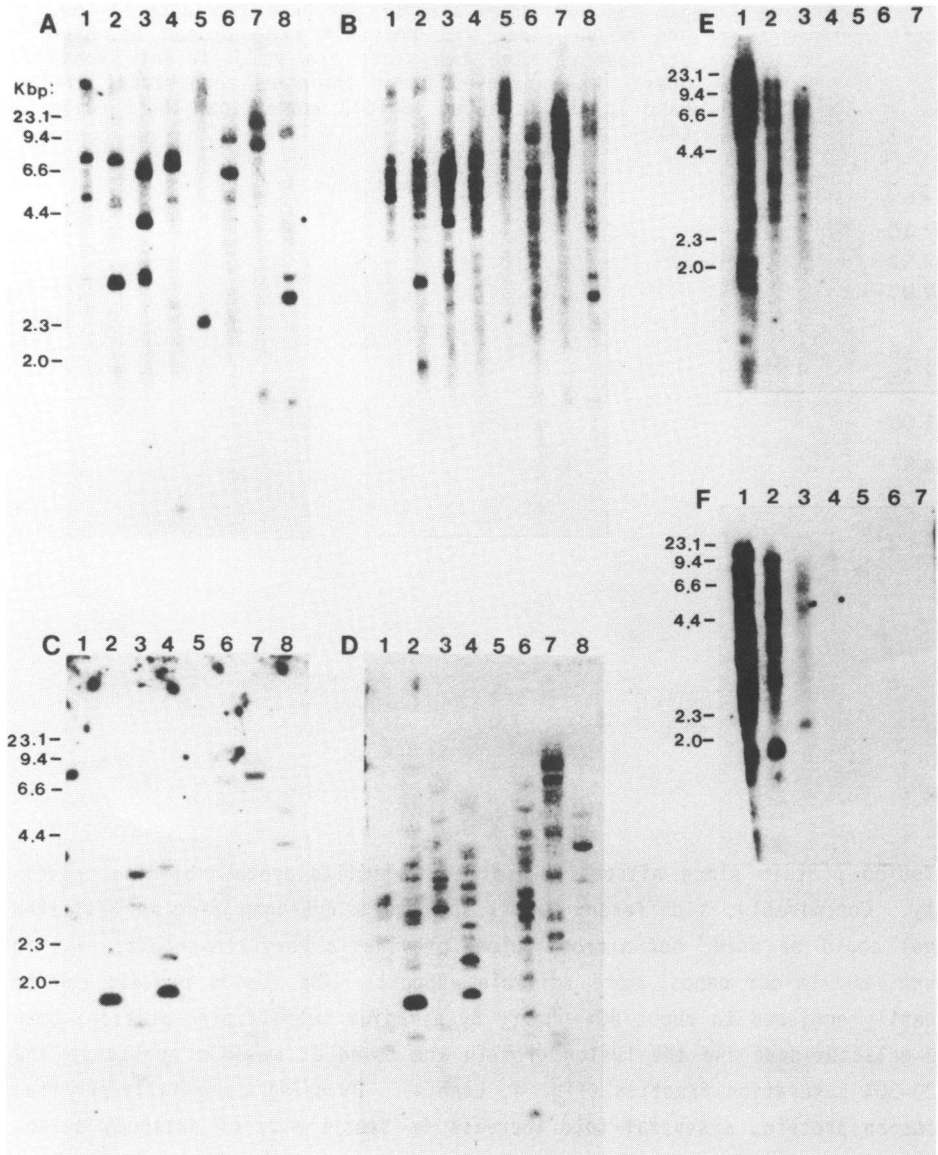

FIG. 6. Southern blots of human DNA (A,B), and mouse DNA (C,D) cut with Bam HI, Bgl II, Eco RI, Hind III, Pst I, Sal I, Sac I, or Kpn I (lanes 1-8 respectively) were probed at high (A,C) or low (B,D) stringency. DNAs from human, mouse, chicken, Artemia, sea urchin, Drosophila, and yeast (lanes 1-7 respectively) were cut with Bam HI (E), or Bgl II (F) and probed at low stringency.

mRNA. RNAs from both humans and mice hybridize with the probe, and the hybridizing RNAs from the two sources are about the same size. The mouse RNA however, gives a band that is about ten fold less intense than the human RNA. This could be a reflection of differences in abundancies between the mouse liver and HeLa cell messages or, more likely, a less than perfect homology between the two RNAs.

Sequence homologies are also apparent at the DNA level as shown by Southern blots (Fig. 6). While there is a considerable amount of homology between the mouse and human hnRNP C protein genes, no, or very little, homology with chicken, Drosophila, sea urchin or yeast DNA is detected when probed at a high stringency (not shown). When Southern blots are probed at a low stringency, several additional bands of weaker homology can be seen in the mouse and human DNAs, and some weak homology with chicken DNA can be detected. At the RNA level, new bands are not seen at the lower stringency (not shown), suggesting that either the DNA of the lower stringency bands is not expressed, or yields RNAs of about 1.5 Kb. The intensities of the bands, when compared to cDNA standards that were loaded and run in parallel, correspond to a low copy number. While these data are not accurate enough to imply that the genes are present in a single copy, they do suggest that the copy number is low.

The nucleotide sequence of the cloned DNA is shown in Fig. 7. Sequencing was done on Eco RI fragments subcloned in both directions into M13mp8 and mp9, and on Eco RI/Bgl II fragments subcloned into the Eco RI/Bam HI sites of M13mp8 and mp9. The sequences of these fragments give a complete overlap of the Eco RI/Bam HI fragment extending from nucleotide 1 to 327 with an Eco RI fragment, and a small overlap between the 328-678 Eco RI/Bam HI fragment and an Eco RI fragment in the opposite orientation. The deduced amino acid sequence has one open reading frame as determined by the lack of termination codons. When this sequence is compared with the amino acid sequences of peptides from single-stranded DNA binding proteins, we find, as discussed below, that there is a complete homology with two peptides from the calf thymus protein UP2. There is also a region that has a sequence very similar to two regions of the calf thymus protein UP1 (Fig. 8).

DISCUSSION

We have isolated and characterized a cDNA clone which expresses a protein that cross-reacts immunologically with both the C1 and C2 core hnRNP proteins but not with any of the other core hnRNP proteins. An analysis of

```
          10                  20                  30                  40                  50                  60
CCCG GAG GAG CAG TTC GGC GGG GAC GGG GCG GCG GCA CGG GCA ACG GCG GCG GTA GGC GGC
 Pro Glu Glu Gln Phe Gly Gly Asp Gly Ala Ala Ala Arg Ala Thr Ala Ala Val Gly Gly

          70                  80                  90                  100                 110                 120
TCG GCG GGC GAG CAG GAG GGA GCC ATG GTG GCG GCG ACA CAG GGG GCA GCG GCG GCG GCG
Ser Ala Gly Glu Gln Glu Gly Ala Met Val Ala Ala Thr Gln Gly Ala Ala Ala Ala Ala

          130                 140                 150                 160                 170                 180
GGA AGC GGA GCG GGA CCG GGG GCG GAA CCG CGT CTG GAG GCA CCG AAG GGG CAG CGC CAG
Gly Ser Gly Ala Gly Pro Gly Ala Glu Pro Arg Leu Glu Ala Pro Lys Gly Gln Arg Gln

          190                 200                 210                 220                 230                 240
TCG GAG GGG GCG AAG ATT GAC GCC AGT AAG AAC GAG GAG GAT GAA GGC CAT TCA AAC TCC
Ser Glu Gly Ala Lys Ile Asp Ala Ser Lys Asn Glu Glu Asp Glu Gly His Ser Asn Ser

          250                 260                 270                 280                 290                 300
TCC CCA CGA CAC TCT GAA GCA GCG ACG GCA CAG CGG GAA GAA TGG AAA ATG TTT ATA AGA
Ser Pro Arg His Ser Glu Ala Ala Thr Ala Gln Arg Glu Glu Trp Lys Met Phe Ile Arg

          310                 320                 330                 340                 350                 360
GCC TTA GCT GGG ACA CTA CAA AGA AAG ATC TCA GAG CTA CTT TCC AAT TGT AAG TGT AGC
Ala Leu Ala Gly Thr Leu Gln Arg Lys Ile Ser Glu Leu Leu Ser Asn Cys Lys Cys Ser

          370                 380                 390                 400                 410                 420
TGC ACT CTG AGT TAC ATC CTA TCA CAG GCA TCA AGG TTT GCT TTG CTA TTT AAA GAA TCG
Cys Thr Leu Ser Tyr Ile Leu Ser Gln Ala Ser Arg Phe Ala Leu Leu Phe Lys Glu Ser

          430                 440                 450                 460                 470                 480
AAT GTA GAT AAG GTC ATG ATC AAA AAG AAC ATA AAT TCA TGG AAG GTG ATT GAT CCT AAA
Asn Val Asp Lys Val Met Ile Lys Lys Asn Ile Asn Ser Trp Lys Val Ile Asp Pro Lys

          490                 500                 510                 520                 530                 540
AGG GCC AAA GCC ATG AAA ACA AAA GAG CCG GTT AAA AAA ATT TTT GTT GGT GGC CTT TCT
Arg Ala Lys Ala Met Lys Thr Lys Glu Pro Val Lys Lys Ile Phe Val Gly Gly Leu Ser

          550                 560                 570                 580                 590                 600
CCA GAT ACA CCT GAA GAG AAA ATA AGG GAG TAC TTT GGT GGT TTT GGT GAG GTG GAA TCC
Pro Asp Thr Pro Glu Glu Lys Ile Arg Glu Tyr Phe Gly Gly Phe Gly Glu Val Glu Ser

          610                 620                 630                 640                 650                 660
ATA GAG CTC CCC ATG GAC AAC AAG ACC AAT AAG AGG CGT GGG TTC TGC TTT ATT ACC TTT
Ile Glu Leu Pro Met Asp Asn Lys Thr Asn Lys Arg Arg Gly Phe Cys Phe Ile Thr Phe

          670
AAC CAA GAA GAA CCG G
Asn Gln Glu Glu Pro
```

FIG. 7. The nucleotide sequence of the cloned cDNA, and the amino acid sequence deduced from it. The regions that show perfect homology with two peptides from UP2 are boxed, and the region that shows homology with UP1 (as shown in Fig. 8) is underlined.

the DNA sequence of this clone shows that it also shares homology with the single-stranded DNA binding protein UP2, and has a region that is similar to UP1.

The clone was isolated from a human hepatoma cDNA expression library constructed in λgt11 that was screened with antibodies that cross-react

```
DL4:      VKKIFVGGLSPDTPEEK IREYFGGFGEVESIELPMDNKTNKR RGFCFITF
            :.:.:::: .: .:  .:  :                   :: ::: :.:
UP1:    14-KLFIGGLSFETTDES LRSHF-33              51-KRSRGFGFVTY-61
            ::::::::. :: ::  .:.::                  :  ::: :.::
UP1: 103-VKKIFVGGIKEDT EEHHLRDYF-124             143-KK RGFAFVTF-152
```

FIG. 8. Homology with the UP1 amino acid sequence. Amino acids indicated by (:) are identical with those of DL4; those indicated by (.) represent Ile and Val, Ile and Leu or Glu and Asp matches. The region that is homologous with the proposed UP1 DNA binding sites (31) is underlined. The sequence corresponds to nucleotides 512-661 of the DL4 sequence shown in Fig. 7.

with all of the core hnRNP proteins. During the course of the screening, two commonly used immunochemical methods for detecting antigens were investigated: a double antibody assay and an avidin-biotin complex (ABC) assay, both employing horseradish peroxidase for the development of a color reaction. We find that the ABC assay gives at least a four-fold increase in sensitivity over the double antibody method with very little increase in background. Although the ABC method requires additional steps, the increase in sensitivity that was obtained made this the method of choice for use with the antigen and antibodies that were investigated in this study.

The identity of the clone was determined by a highly sensitive antibody select procedure. We have found that for an antibody select procedure to be successful it is necessary to have a sufficient amount of selected antibody at a high enough concentration to insure a strong signal when it is used for the development of western blots. By adsorbing a purified fusion protein onto nitrocellulose filters (which have a limited protein binding capacity) rather than crude extracts, the amount and concentration of selected antibody that can be obtained from each filter is increased several fold. This results in a corresponding increase in the sensitivity of the antibody select procedure over the original procedure (30). In this report we describe a simple scheme for the enrichment of ß-galactosidase fusion proteins from bacterial lysates.

Based on peptide mapping experiments, it has been suggested (11) that the hnRNP proteins probably belong to several closely related families. The fusion protein that is expressed by λDL4 selects an antibody that cross-reacts equally well with both the C1 and C2 proteins but not with any of the other core hnRNP proteins. This suggests that the C1 and C2 core hnRNP proteins are very closely related, perhaps being derived from a common precursor. The lack of reactivity of the λDL4 fusion protein selected antibodies with the A and B hnRNP proteins (Fig. 3B) suggests

that the A and B hnRNP proteins do not share this particular antigenic relationship.

By SDS-polyacrylamide gel electrophoresis, the C1 protein is about 41,000 molecular weight, and the C2 protein is about 43,000. An mRNA of at least 1.2 Kb is required to encode these proteins or their precursor. On northern blots (Fig. 5), poly(A)$^+$ mRNA of about 1.5 Kb is detected by hybridization with the cloned DNA, suggesting the lack of a large untranslated region. The mRNA that encodes HD40, the major hnRNP protein of Artemia (which has a molecular weight of 40,000) is also about 1.5 Kb in length (31). The DL4 probe detects both human and mouse mRNAs of 1.5 Kb, even at high stringency, suggesting that portions of the mRNAs that code for the C proteins are fairly well conserved between the two species. The intensity of the 1.5 Kb band is about ten-fold weaker in the mouse RNA than in the human RNA suggesting that while there are regions of good homology between the two RNAs, these regions only comprise portions of the total RNA. An alternative explanation, that the amount of RNA coding for the hnRNP C proteins differs greatly between mouse liver and HeLa cells, seems less likely since the concentration of hnRNP C protein appears to be about the same in both sources.

From Southern blots (Fig. 6) of genomic DNA, we calculate a copy number close to one for the hnRNP C protein gene, indicating that the gene is present in low copy number. Blots done at high (0.2 X SSC at 68°) and low (2 X SSC at 50°) stringency indicate that there is a considerable amount of homology between the human and mouse hnRNP C protein genes, and regions of weak homology with the chicken genes. We see no, or very little, homology with sea urchin, Drosophila, yeast, or Artemia DNA. In Southern blots probed with a cDNA clone of the Artemia protein HD40, homology was seen between the HD40 probe and yeast, pea, chicken, mouse, and human DNA, with human DNA giving 8-10 defined bands ranging in size from 0.8 to 15 Kb (31). The fact that we see no homology with Artemia DNA suggests that while there may be homology between HD40 and some of the other hnRNP proteins, this homology may not extend to the C proteins.

The strong homology between the human and mouse DNAs and RNAs that encode the hnRNP C proteins is not particularly surprising in light of the great deal of immunological cross-reactivity between different mammalian species (8-10). It has also been observed by a number of workers that it is very difficult to induce the synthesis of antibodies against the mam-

malian hnRNP core proteins in mammals. Antibodies against the mammalian hnRNP proteins react only very weakly with HD40 (7).

Recently, it has been suggested that at least some of the eukaryotic single stranded DNA binding proteins, a number of which have been described, may be related to the core hnRNP proteins. Three peptides from UP2, which is isolated from calf thymus, have been sequenced (32). A comparison of the amino acid sequences of these peptides with the amino acid sequence predicted from the nucleotide sequence of DL4 shows a perfect match with two of the three peptides (Fig. 7). There is no match with the third peptide, which has the sequence Lys-Tyr-His-Asn-Val-Gly-Leu-Ser-Lys. Since DL4 is a partial clone, smaller than UP2, it is possible that this peptide lies outside of the region covered by DL4. This strong homology not only suggests a relation between the hnRNP C proteins and UP2, but also indicates that the amino acid sequence of these proteins has been well conserved between humans and cattle. In a similar note, very good homologies have been shown between amino acid sequences of peptides from the human hnRNP A1 protein (M_r=32,000) and the 22,000 molecular weight calf thymus single-stranded DNA binding protein UP1, and between peptides from the human hnRNP A2 protein (M_r=34,000) and the single-stranded DNA binding protein UP1-B from calf liver and HDP-1 from mouse myeloma (M_r=24,000-27,000) (19). Based on these homologies, immunological evidence (18), and proteolytic digestion studies in which it was observed that the A1 and A2 proteins give rise to 24,000-28,000 molecular weight peptides when digested with trypsin or with proteases present in cellular extracts (18,19), it has been proposed that some of the helix-destabilizing proteins are derived from hnRNP proteins by proteolysis. The exact relation between UP2 and the hnRNP C proteins is not clear, but the similarity in molecular weights suggests that UP2 may be the same as the hnRNP C1 protein. Alternatively, UP2 may be derived from the hnRNP C proteins by cleavage of a small fragment. It should be noted however, that the digestion of the hnRNP C1 protein with trypsin gives rise to a relatively stable 28,000 molecular weight fragment; a 39,500 molecular weight fragment is not observed (19).

Merrill et al (32) have observed that there is a partial homology between one of the UP2 peptides and two regions of UP1. These two regions, it is proposed, compose two domains of UP1 that share internal homologies. By comparing the amino acid sequence of UP1 with the sequence derived from DL4, we see that there is a large region of DL4 that is homologous with two regions of UP1: amino acids 14-61 and amino acids 103-152 (Fig. 8). This

region includes one of the UP2 peptides that have been sequenced. There is also a considerable homology between this region and the mouse myeloma protein HDP-1 (which is related to the hnRNP A2 protein (19)), but this is not particularly surprising since UP1 and HDP-1 have very similar amino acid sequences (33). The rest of DL4 shows very little homology with UP1. It has been suggested that UP1 might contain two structurally related DNA binding sites: one containing residues 51-61, and the other residues 143-152 (32). As shown in Fig. 8, DL4 contains a region that is very similar in composition to these sites. Of the eight amino acids in this region, six are the same in all three sequences, and the two that differ might be expected to have little effect on the structure of the protein: an Ile in DL4 which is Val in the two UP1 sequences, and a Cys which is Gly or Ala in UP1. Just as striking, however, is the homology between DL4 and UP1 residues 14-34 and 103-125. Since it appears to be highly conserved, this region may also play a central role in the function of these proteins. Although the λDL4 fusion protein appears to contain the proposed DNA (and RNA) binding site, it does not bind to mRNA as assayed by nitrocellulose filter binding (not shown). The proposed binding site is, however, near the junction with ß-galactosidase, and may not be in the proper conformation. A comparison of the nucleotide sequence of DL4 with that of UP1 (34) shows very little similarity at the nucleotide level between the two, even in the regions of conserved amino acid sequence.

One striking feature of the hnRNP proteins is their high content of glycine: about 20% for the C proteins. From DL4, we can deduce the sequence of a 225 amino acid region. This region contains about 10% glycine, a value closer to the 7.6 mole% average value for proteins in the PIR Protein Sequence Database. Consequently, the remainder of the hnRNP C protein must have a very high glycine content of about 30%. This is similar to the hnRNP A1 and A2 proteins which have an NH_2 terminal domain with a normal glycine content and a carboxy terminal domain that is very rich in glycine (19).

ACKNOWLEDGMENTS

We wish to thank Drs A. Kumar, W. Szer, N.J. Cowan, and G.N. Godson for helpful discussions, and K.R. Williams and S.H. Wilson for providing amino acid and nucleotide sequences prior to publication. This work was supported by N.I.H.grant GM28747.

REFERENCES

1. Beyer, A.L., Christensen, M.E., Walker, B.W. and LeStourgeon, W.M. (1977) Cell 11, 127-138.
2. Wilk, H.E., Werr, H., Friedrich, D., Klitz, H.H. and Schafer, K.P. (1985) Eur. J. Biochem. 146, 71-81.
3. Osheim, Y.N., Miller, O.L.,Jr. and Beyer, A.L. (1985) Cell 43, 143-151.
4. Thomas, J.O., Glowacka, S.K. and Szer, W. (1983) J. Mol. Biol. 171, 439-455.
5. Nowak, L., Marvil, D.K., Thomas, J.O., Boublik, M. and Szer, W. (1980) J Biol. Chem. 255, 6473-6478.
6. Patton, J.R., Ross, D.A. and Chae, C.-B. (1985) Mol. Cell. Biol. 5, 1220-1228.
7. Choi, Y.D., Grabowski, P.J., Sharp, P.A. and Dreyfuss, G. (1986) Science, 231, 1534-1539.
8. Lahiri, D.K. and Thomas, J.O., (1985) J. Biol. Chem. 260, 598-603.
9. Jones. R.E., Okamura, C.S. and Martin, T.E. (1980) J. Cell Biol. 86, 235-243.
10. Leser, G.P., Escara-Wilke, J. and Martin, T.E. (1984) J. Biol. Chem. 259, 1827-1833.
11. Fuchs, J.-P., Judes, C. and Jacob, M. (1980) Biochemistry 19, 1087-1094.
12. Marvil, D.K., Nowak, L. and Szer, W., (1980) J. Biol. Chem. 255, 6466-6472.
13. Raziuddin, Thomas, J.O. and Szer, W. (1982) Nucl. Acids Res. 10, 7777-7789.
14. Choi, Y.D. and Dreyfuss, G. (1984) J. Cell. Biol. 99, 1997-2004.
15. Herrick, G., Delius, H. and Alberts, B. (1976) J. Biol. Chem. 251, 2142-2146.
16. Plank, S.R. and Wilson, S.H. (1980) J. Biol. Chem. 255, 11547-11556.
17. Valentini, O., Biamonti, G., Pandolfo, M., Morandi, C. and Riva, S. (1985) Nucl. Acids Res. 13, 337-346.
18. Pandolfo, M., Valentini, O., Biamonti, G. and Riva, S. (1985) Nucl. Acids Res. 13, 6577-6590.
19. Kumar, A., Williams, K.R. and Szer, W. (1986) J. Biol. Chem. in press.
20. Hsu, S.M., Raine, L. and Fanger, H. (1981) J. Histochem. Cytochem. 29, 577-580.
21. DeWet, J.R., Fukushima, H., Dewji, N.N., Wilcox, E., O'Brien, J.S. and Helinski, D.R. (1984) DNA 3, 437-447.
22. Young, R.A. and Davis, R.W. (1983), Proc. Natl. Acad. Sci. USA 80, 1194-1198.
23. Samarina, O.P., Lukanidin, E.M., Molnar, J. and Georgiev, G.P. (1968) J. Mol. Biol. 33, 251-263.
24. Towbin, J., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350-4354.
25. Boedtker, H. (1971) Biochim. Biophys. Acta 240, 448-453.
26. Perbal, B. (1984) A Practical Guide to Molecular Cloning, John Wiley & Sons.
27. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
28. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
29. Laemmli, U.K. (1970) Nature 227 680-685.
30. Hall, R., Hyde, J.E., Goman, M., Simmons, D.L., Hope, I.A., Mackay, M., Scaife, J., Merkli, B., Richle, R. and Stocker, J. (1984) Nature 311, 379-382.
31. Cruz-Alvarez, M., Szer, W. and Pellicer, A. (1985) Nucl. Acids Res. 13, 3917-3930.

32. Merrill, B.M., LoPresti, M.B., Stone, K.L. and Williams, K.R. (1986) J. Biol. Chem. 261, 878-883.
33. Williams, K.R., Stone, K.L., LoPresti, M.B. and Plank, S.R. (1985) Proc. Natl. Acad. Sci. USA 82, 5666-5670.
34. Cobianchi, F., SenGupta, D., Zmudzka, B.Z. and Wilson, S.H. (1986) J. Biol. Chem. 261, 3536-3543.