

A Nuclear Localization Domain in the hnRNP A1 Protein

Haruhiko Siomi and Gideon Dreyfuss

Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6148

Abstract. The heterogeneous nuclear RNP (hnRNP) A1 protein is one of the major pre-mRNA/mRNA binding proteins in eukaryotic cells and one of the most abundant proteins in the nucleus. It is localized to the nucleoplasm and it also shuttles between the nucleus and the cytoplasm. The amino acid sequence of A1 contains two RNP motif RNA-binding domains (RBDs) at the amino terminus and a glycine-rich domain at the carboxyl terminus. This configuration, designated 2× RBD-Gly, is representative of perhaps the largest family of hnRNP proteins. Unlike most nuclear proteins characterized so far, A1 (and most 2× RBD-Gly proteins) does not contain a recognizable nuclear localization signal (NLS). We have found that

a segment of ca. 40 amino acids near the carboxyl end of the protein (designated M9) is necessary and sufficient for nuclear localization; attaching this segment to the bacterial protein β -galactosidase or to pyruvate kinase completely localized these otherwise cytoplasmic proteins to the nucleus. The RBDs and another RNA binding motif found in the glycine-rich domain, the RGG box, are not required for A1 nuclear localization. M9 is a novel type of nuclear localization domain as it does not contain sequences similar to classical basic-type NLS. Interestingly, sequences similar to M9 are found in other nuclear RNA-binding proteins including hnRNP A2.

THE heterogeneous nuclear RNP (hnRNP)¹ A1 protein is an abundant nuclear protein and is one of the major pre-mRNA binding proteins (Piñol-Roma et al., 1988; reviewed in Cobianchi et al., 1990; Dreyfuss et al., 1993). A1 binds with high affinity to sequences that resemble pre-mRNA 3' and 5' splice sites (Swanson and Dreyfuss, 1988; Burd and Dreyfuss, 1994a). A1 has an effect on alternative splicing in vitro and in vivo as the amount of A1 relative to that of the splicing factor SF2/ASF determines the utilization of alternative 5' splice sites (Mayeda and Krainer, 1992; Fu et al., 1992; Cáceres et al., 1994; Yang et al., 1994), although it does not seem to be essential for constitutive splicing (Ben-David et al., 1992). One of the most intriguing properties of A1 is its cellular localization and transport. A1 is a nuclear protein but it is not confined to the nucleus, rather it shuttles between the nucleus and the cytoplasm (Piñol-Roma and Dreyfuss, 1992). Furthermore, its nuclear

localization requires RNA polymerase II (polII) transcription since in the absence of polII transcription A1 that is exported from the nucleus remains in the cytoplasm (Piñol-Roma and Dreyfuss, 1991; Piñol-Roma and Dreyfuss, 1992; Piñol-Roma and Dreyfuss, 1993). By immunofluorescence microscopy, A1 is detected only in the nucleus, presumably because the return of A1 to the nucleus is normally very rapid. While in the cytoplasm A1 can bind mRNA, and therefore it is likely to have functions in the cytoplasm (e.g., in mRNA localization and stability), and it may also play an important role in the transport of mRNA from the nucleus to the cytoplasm. Importantly, recent findings indicate that this pattern of localization and transport is not unique to A1, but is found for many of the other pre-mRNA binding proteins (Piñol-Roma, S., and G. Dreyfuss, unpublished results). The amino acid sequence of A1 from mammals and amphibians has been determined (Williams et al., 1985; Cobianchi et al., 1988; Buvoli et al., 1988; Kay et al., 1990; Good et al., 1993) and it contains two RNP motif RNA-binding domains (RBDs; Dreyfuss et al., 1988; Bandziulis et al., 1989; Kenan et al., 1991; Burd and Dreyfuss, 1994b) at the amino terminus and a glycine-rich domain at the carboxyl terminus. This structure, designated 2× RBD-Gly, is representative of perhaps the largest family of hnRNP proteins (Matunis et al., 1992; Dreyfuss et al., 1993; Burd and Dreyfuss, 1994b). A1 also contains another motif found in many RNA-binding proteins, an arginine- and glycine-rich

Address all correspondence to G. Dreyfuss, Howard Hughes Medical Institute, University of Pennsylvania School of Medicine, Clinical Research Building, Room 328, 422 Curie Blvd., Philadelphia, PA 19104-6148. Ph: (215) 898-0398. Fax: (215) 573-2000.

1. *Abbreviations used in this paper:* Gly, glycine; HA, hemagglutinin; hnRNP, heterogeneous nuclear RNP; HPABP, human poly(A)-tail mRNA binding protein; NLS, nuclear localization signal; PCR, polymerase chain reaction; PK, pyruvate kinase; polII, polymerase II; RBD, RNA-binding domains.

domain that contains a cluster of the tripeptide repeat and termed the RGG box (Kiledjian and Dreyfuss, 1992). This motif has been demonstrated to have RNA binding activity (Kilejian and Dreyfuss, 1992). Unlike most nuclear proteins characterized so far, the A and B group hnRNP proteins do not contain recognizable nuclear localization signals (NLSs) which generally resemble either the single basic domain SV-40 large T-antigen NLS (Kalderon et al., 1984; for review see Nigg, 1990) or the double basic (bipartite) domain nucleoplasmin NLS (Robbins et al., 1991; reviewed in Dingwall and Laskey, 1991). However, other hnRNP proteins, such as the C group and U proteins, contain candidate NLSs. It is likely that the C group and U proteins are imported to the nucleus by mechanisms similar to those used by other nuclear proteins that contain similar basic NLSs. On the other hand, the mechanism and pathway of the nuclear import of the A and B group proteins may be different from those used by other nuclear proteins that contain classical types of NLSs. We therefore set out to determine the sequence within A1 that localizes it to the nucleus. Here we report that a sequence of the carboxyl terminus of A1 is necessary and sufficient for the nuclear localization of this protein and it can completely localize a large heterologous bacterial protein, β -galactosidase, the cytoplasmic enzyme pyruvate kinase, and the mRNA poly(A)-binding protein to the nucleus. Interestingly, similar sequences to the nuclear localization domain of A1 are found in several other nuclear RNA-binding proteins including A2, and the sequence is highly conserved in vertebrates as *Xenopus laevis* A1/A2 proteins contain a similar domain.

Materials and Methods

Plasmid Constructions

The cDNA encoding human hnRNP A1 (pBSO1; Buvoli et al., 1990) was a kind gift from Dr. Silvano Riva (CNR, Pavia, Italy). cDNAs encoding human hnRNP C1 and mRNA poly(A)-binding protein (HPABP) were described previously (Swanson et al., 1987; Görlach et al., 1994a). Epitope tagging of hnRNP A1, hnRNP C1, and HPABP was performed as described by Kolodziej and Young (1991) using the eukaryotic expression vector pHHSI (Siomi et al., 1993) which was derived from pcDNA I (Invitrogen San Diego, CA) by insertion of a 72-bp BstXI fragment encoding a 9-amino acid epitope from the influenza virus hemagglutinin (HA) protein with an initiator methionine followed by Sal I and EcoRI restriction sites. The HA epitope for the 12CA5 anti-HA monoclonal antibody (Niman et al., 1983) was placed at the N-termini of all constructs.

HA-tagged hnRNP A1 with the entire protein-coding region (wild-type) was constructed using the polymerase chain reaction (PCR). A 34-base synthetic single-stranded DNA PCR primer (5'-A1) that included 9 bases to create a SalI site and 25 bases complementary to positions +1 through +25 was synthesized. The 5'-A1 primer, in conjunction with a primer (3'-A1) that included 9 bases to create NotI site and 25 bases complementary to the sequence encoding the carboxyl terminal end of hnRNP A1, was used to prime a PCR using plasmid pBSO1 as template DNA. The amplified fragment was digested with SalI and NotI and ligated into SalI-NotI-digested pHHSI, resulting in plasmid p16FA1. The NH₂-terminal deletion mutants Nd97, Nd112, Nd149, Nd155, and Nd186 were also generated by PCR using primers that included 9 bases to create a SalI site and 25 bases complementary to the positions indicated and a primer 3'-A1. The COOH-terminal deletion mutants of hnRNP A1 were constructed as follows: plasmid p16FA1 was digested at its unique NotI site, and then exonuclease III and S1 nuclease digestion (Guo and Wu, 1983; Siomi et al., 1990) were performed to create deletions in the 3' portion of the hnRNP A1 coding region using double-stranded nested deletion kit (Pharmacia, Piscataway, NJ). After nuclease digestion, a synthetic HpaI linker (TTAAGTAACTTAA; Pharmacia) was ligated at the terminus. Deletion fragments were then digested with HpaI

and SalI, and inserted into pHHSI that had been digested with NotI and treated with Klenow fragment to create blunt end and then digested with SalI. To determine exact nucleotide endpoints, clones were sequenced using T3 primer.

HA-tagged wild-type hnRNP C1 and its NH₂-terminal deletion (Nd87) were generated by PCR as described above for hnRNP A1 constructs. Synthetic 5' PCR primers that included a SalI site and 25 bases complementary to sequences encoding the NH₂-terminal end of hnRNP C1 and the NH₂-terminal end of the auxiliary domain, in conjunction with a 3' primer that included a NotI site and 25 bases complementary to the sequence encoding the carboxyl terminal end of hnRNP C1, were used to prime PCRs using plasmid pHC12 (Swanson et al., 1987) as template DNA. The amplified fragments were digested with SalI and NotI and ligated into SalI-NotI-digested pHHSI, resulting in plasmids p16FC1 and p16Nd87, respectively. The COOH-terminal deletion mutants of hnRNP C1 were constructed using plasmid p16FC1 as described for hnRNP A1 deletions.

HA-tagged wild-type HPABP (Görlach et al., 1994a) was constructed using PCR-generated fragment as described above. PABP4 \times RBDs was also PCR-generated using the 5' primer that included a SalI site and 25 bases complementary to positions +1 through +25 and the 3' primer that included NotI site, a termination codon and 25 bases complementary to the sequence encoding the end of the four RNA-binding domains. PABP-Gly was also generated by PCR in several steps involving an overlap extension protocol (Horton et al., 1990).

The β -galactosidase expression vector (p16 β -Gal) was constructed by inserting a HindIII-BamHI fragment from pCH110 (Hall et al., 1983; Clontech, Palo Alto, CA) into pcDNA1 that has been digested with HindIII and BamHI. β -Gal A1 and β -Gal M1 were PCR-generated by annealing a primer carrying, in series, a EcoRI site and 20 bases corresponding to the human hnRNP A1 cDNA (pBSO1) and a 3' primer (3'-A1). The amplified fragments were digested with EcoRI and NotI and ligated into EcoRI-NotI-digested p16 β -GalA1. β -Gal M2 to M4 were constructed as follows; plasmid p16 β -GalA1 was digested at its unique EcoRI site, and then treated with nucleases as described for hnRNP A1 deletions. After nuclease digestion, a synthetic EcoRI linker (GGAATTCC) was ligated at the terminus. Deletion fragments were then digested with EcoRI and NotI, and inserted into p16 β -Gal that has been digested with EcoRI and NotI. To determine the exact nucleotide endpoints, clones were sequenced using a primer of 20 nucleotides spanning the region 50-70 nucleotides upstream of the EcoRI site of the lacZ gene. β -Gal M5 to M8 were constructed as described for hnRNP A1 COOH-terminal deletion mutants.

Myc-tagged pyruvate kinase (PK)-A1 fusions were constructed as follows: synthetic 5' PCR primer that included an EcoRI site which corresponds to codon 12 of chicken muscle pyruvate kinase, in conjunction with a 3' primer that included a NotI site and 25 bases complementary to the sequence encoding the carboxyl terminal end of PK, were used to prime PCR using the PK cDNA-containing plasmid, RLPK (a gift from Dr. William Lee, University of Pennsylvania, Philadelphia, PA; Dang and Lee, 1988) as template DNA. The amplified fragment was digested with EcoRI and NotI and ligated into EcoRI-NotI-digested pcDNA3 (Invitrogen). The resulting plasmid containing PK cDNA under the control of the CMV early promoter was cut at the BstXI and EcoRI sites and ligated to a double-stranded oligonucleotide encoding the peptide sequence MEQKLISEEDL to create an epitope tag at the amino terminus. This sequence is the epitope for the 9E10 anti-myc monoclonal antibody (Evan et al., 1985). To fuse portions of A1 to the 3' end of the PK cDNA, we generated various KpnI-NotI fragments of A1 by PCR as described above. These KpnI-NotI fragments were subcloned into the KpnI and NotI sites of the myc-tagged pcDNA3-PK. The KpnI site corresponds to codon 443 of PK (Roberts et al., 1987; Dang and Lee, 1988).

Cell Culture and Transfection

Monkey kidney COS-1 cells were grown in Dulbecco's modified Eagle's medium (DME; GIBCO BRL, Gaithersburg, MD) supplemented with 10% FCS (GIBCO BRL). Cells grown on glass coverslips in 30 mm dishes were transfected with 5 μ g DNA (all plasmids were purified by Cs-EtBr density gradient centrifugation) by the calcium phosphate co-precipitation technique with the following modifications: a total of 5 μ g of plasmid DNA was added to 250 ml of Hepes-buffered saline (140 mM NaCl, 1.5 mM Na₂HPO₄, 25 mM Hepes, pH 7.05) and precipitated by adding 16.5 μ l of 2 M CaCl₂ with constant stirring. After incubating the cells with the co-precipitate for 4-5 h, cultures were washed twice with 2 ml DME (without FCS), and then were incubated with fresh DME and FCS. Cells were fixed for immunofluorescence assays 48 h after transfection.

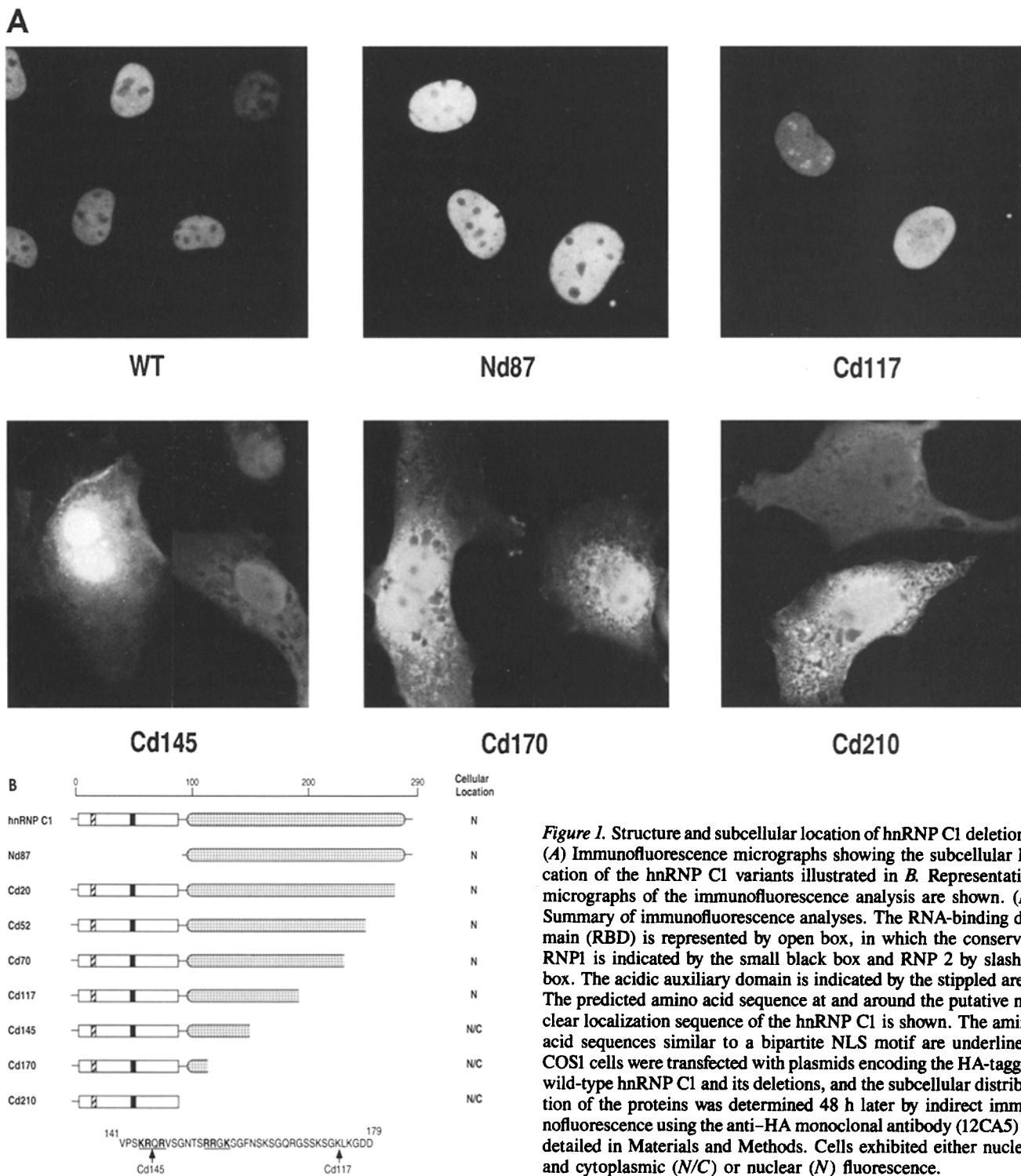


Figure 1. Structure and subcellular location of hnRNP C1 deletions. **(A)** Immunofluorescence micrographs showing the subcellular location of the hnRNP C1 variants illustrated in **B**. Representative micrographs of the immunofluorescence analysis are shown. **(B)** Summary of immunofluorescence analyses. The RNA-binding domain (RBD) is represented by open box, in which the conserved RNP1 is indicated by the small black box and RNP 2 by slashed box. The acidic auxiliary domain is indicated by the stippled area. The predicted amino acid sequence at and around the putative nuclear localization sequence of the hnRNP C1 is shown. The amino acid sequences similar to a bipartite NLS motif are underlined. COS1 cells were transfected with plasmids encoding the HA-tagged wild-type hnRNP C1 and its deletions, and the subcellular distribution of the proteins was determined 48 h later by indirect immunofluorescence using the anti-HA monoclonal antibody (12CA5) as detailed in Materials and Methods. Cells exhibited either nuclear and cytoplasmic (N/C) or nuclear (N) fluorescence.

Immunofluorescence Microscopy

Indirect immunofluorescence analysis was carried out as described for human HeLa cells (Choi and Dreyfuss, 1984) with the following modifications: cells were fixed with 2% formaldehyde (Polysciences, Warrington, PA) in PBS for 30 min at room temperature followed by 3 min in acetone at -20°C . Anti-HA tag staining was performed by incubation for 1 h at room temperature with 1:1,500 12CA5 anti-HA mAb (ascites fluid diluted in PBS with 3% BSA). Cells were then washed three times with PBS and incubated for 1 h at room temperature with 1:1,500 fluorescein-conjugated

anti-mouse antibody (Cappel Laboratories, Cochranville, PA) in PBS containing 3% BSA. After washing three times, coverslips were inverted and mounted on glass microscope slides. Anti- β -galactosidase staining was using mouse monoclonal antibody to *Escherichia coli* β -galactosidase (Boehringer Mannheim Biochemicals, Indianapolis, IN; 1:1,500 in PBS with 3% BSA) as the primary antibody. Immunofluorescent visualization of the PK-A1 fusion proteins was achieved in an identical manner to that of the 12CA5 antibody using a monoclonal anti-myc 9E10 antibody (1:1,500 dilution in PBS with 3% BSA).

A

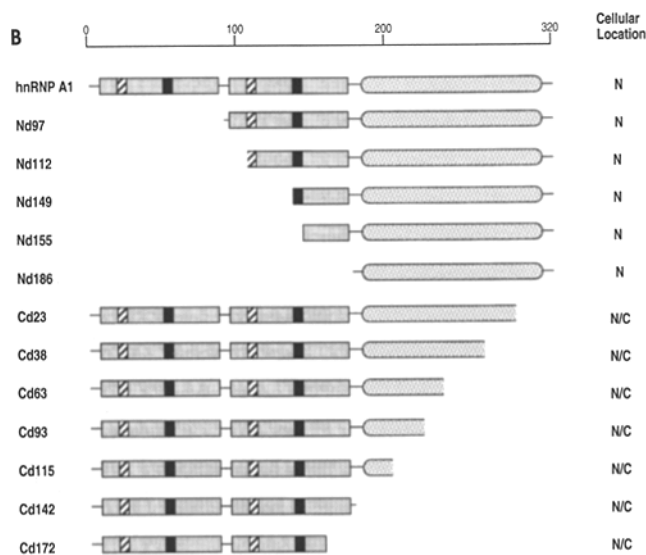
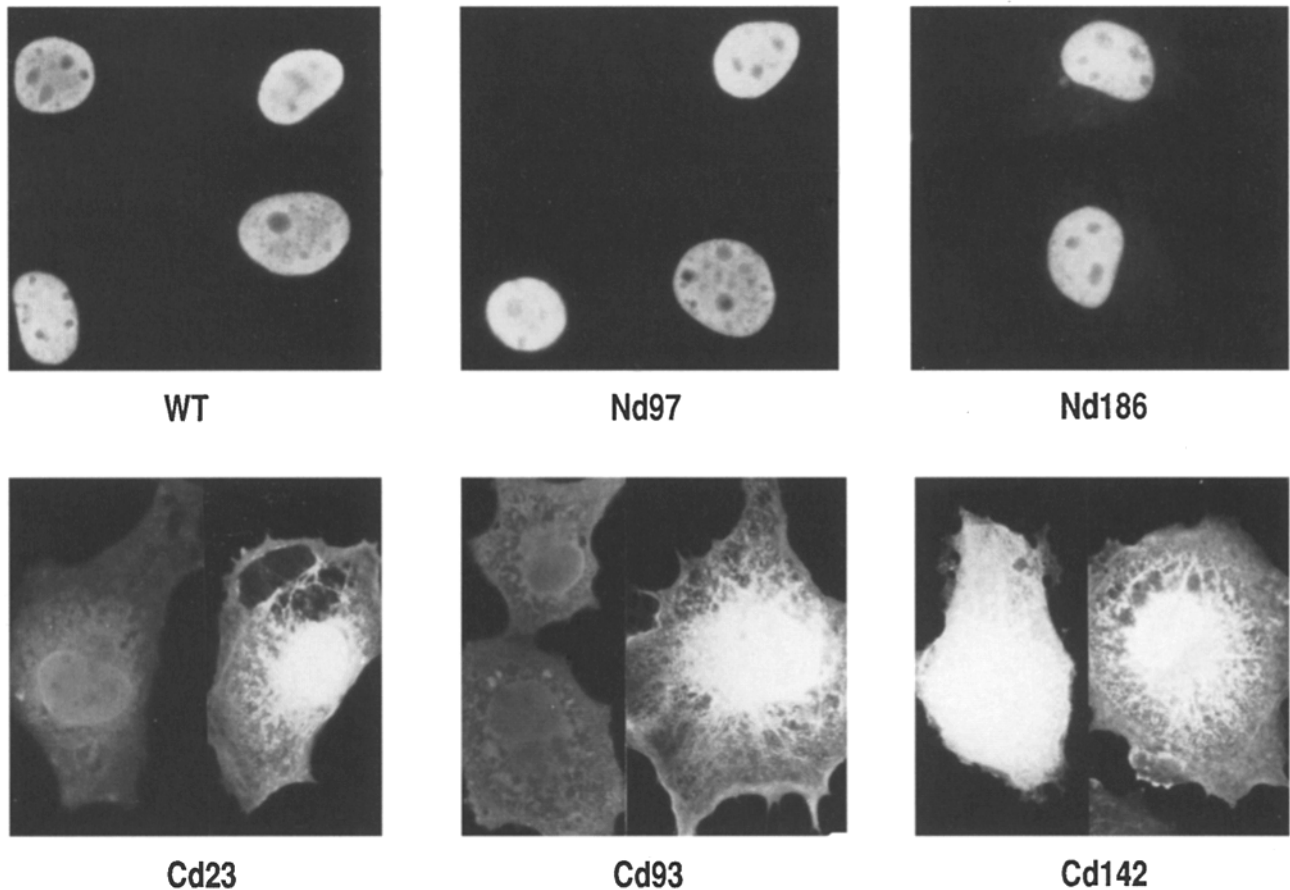


Figure 2. Structure and subcellular location of hnRNP A1 deletions. (A) Photomicrographs showing the subcellular location of the hnRNP A1 deletion mutants illustrated in B. (B) Structure of hnRNP A1 and its deletions. The two RNA-binding domains are indicated by shaded boxes and the glycine-rich auxiliary domain is indicated by the stippled area. Plasmids encoding HA-tagged hnRNP A1 deletions were transfected into COS1 cells, and the subcellular distribution of the proteins was determined 48 h later as described in the legend to Fig. 1. Cells exhibited either nuclear (N) or nuclear and cytoplasmic (N/C) fluorescence.

Gel Electrophoresis and Immunoblotting

COS-1 cells were transfected with 10 µg plasmid DNA per 30 mm petri dish as described above. Cells were harvested 48 h after transfection by scraping in SDS sample buffer. Samples were separated on a 12% SDS-polyacrylamide gel and electrotransferred to nitrocellulose membrane as described previously (Siomi et al., 1993). The membrane was blocked with 5% dried milk powder in PBS for 1 h, 9E10 anti-myc tag antibody (1:1,500 ascites fluid) was added and incubated for additional 1 h at room temperature. After five washes in PBS with 0.05% NP-40, the blot was incubated with

peroxidase-conjugated goat anti-mouse IgG (Jackson Immuno Research Laboratories, 1:10,000) for 1 h at room temperature, washed five times in PBS/ 0.05% NP-40, and developed with an enhanced chemiluminescence Western blotting detection system (Amersham International, Buckinghamshire, UK).

Results

To determine the sequence in A1 that localizes it to the nu-

A

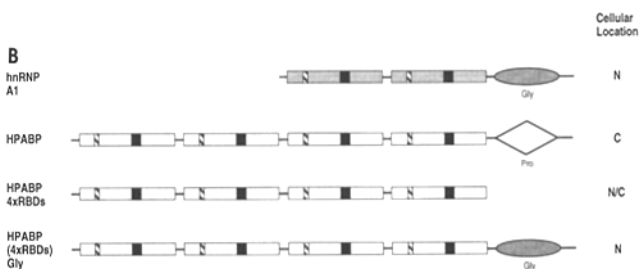
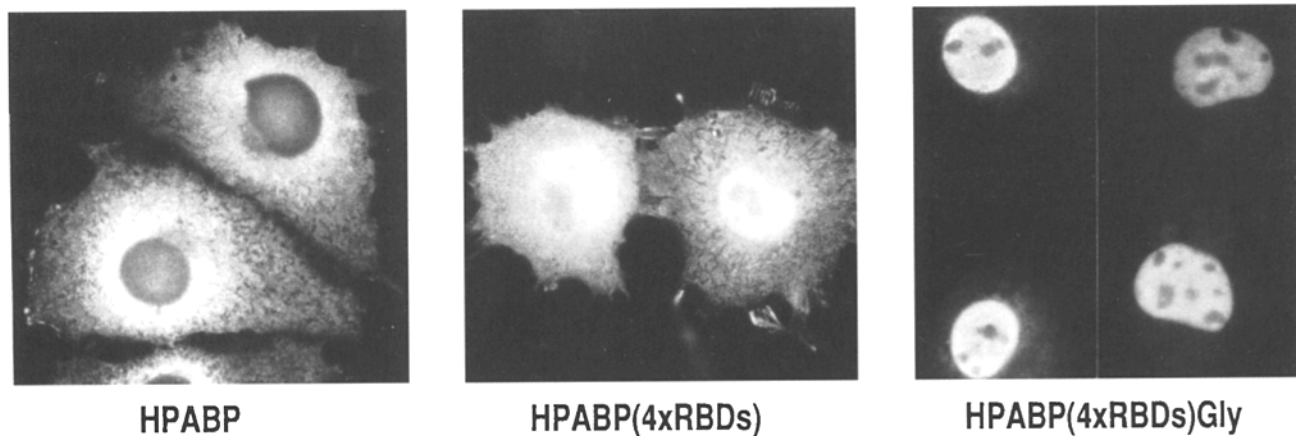


Figure 3. Structure and subcellular location of HPABP fusion proteins. (A) Subcellular location of HPABP-hnRNP A1 Gly fusion protein. COS1 cells were transfected with the plasmids encoding deleted or fusion proteins illustrated in panel B. Forty eight hours later the cells were fixed, and the HA-tagged proteins were visualized by immunofluorescence microscopy as described in the legend to Figure 1. (B) Structure of an HPABP-hnRNP A1 Gly domain fusion protein. The four open boxes represent the four RNA-binding domains present in the HPABP (Görlach et al., 1994a). The open diamond represents the proline-rich auxiliary domain of HPABP. The symbols C and N/C refer to cytoplasmic or nuclear and cytoplasmic localization, respectively.

cleus, parts of A1 were cloned as fusion proteins with the influenza HA peptide (Kolodziej and Young, 1991) at the amino terminus to provide an epitope tag for detection of the expressed polypeptides in transfected cells by immunofluorescence microscopy. All the constructs were confirmed by DNA sequencing and expression of proteins of the expected size, amount, and immunologic reactivity were confirmed by immunoblotting. For reference, we carried out similar experiments on the hnRNP C1 protein which contains a candidate bipartite basic NLS (amino acids 144 to 157). As expected, the entire C1 protein fused to HA localized entirely to the nucleus (Fig. 1). Amino terminal deletion of the RNA-binding domain of hnRNP C1 (Görlach et al., 1994b) showed that the carboxyl terminal auxiliary domain by itself, Nd87, localized to the nucleus. Thus the RNA-binding activity of C1 is not required for its nuclear localization. Carboxyl terminal deletions indicated that the sequence between amino acids 145 and 175 are necessary for nuclear localization, in good agreement with NLS deduced from the protein sequence.

The Carboxyl Terminal Region of hnRNP A1 Contains the Determinant for Nuclear Localization

HnRNP A1 is distinguished from most other nuclear proteins by the absence of a short, highly basic nuclear localization sequence and by its transcription-dependent nuclear localization (reviewed in Piñol-Roma and Dreyfuss, 1993). To identify the regions of A1 important for nuclear localization, we constructed a series of amino terminal and carboxyl terminal deletions. The primary sequence of wild-type A1 is shown schematically in Fig. 2 B and Fig. 6. The major char-

acteristics of this protein are two RNA-binding domains within the amino terminus, a stretch rich in glycine and arginine residues (RGG box: Kiledjian and Dreyfuss, 1992) in the central region, and a glycine-rich (Gly) domain at the carboxyl terminus. Like hnRNP C1, the entire hnRNP A1 protein tagged with HA also completely localized to the nucleus (Fig. 2). The carboxyl domain by itself (Nd97 to Nd186) also showed exclusively nuclear staining but the two amino terminal RBDs (Cd142), which efficiently bind RNA (data not shown; Riva et al., 1986; Cobianchi et al., 1990), and all other carboxyl terminal deletions did not localize to the nucleus (Fig. 2). We therefore, concluded that the nuclear localizing sequence of A1 resides in the carboxyl part (Gly domain) of the protein.

Further evidence that the RBDs of A1 have no role in the localization of A1 to the nucleus and that the Gly domain of A1 is sufficient for nuclear localization was obtained by constructing a human poly(A)-tail mRNA binding protein (HPABP; Görlach et al., 1994a) that contains the A1 Gly domain instead of the normal, proline-rich carboxyl domain (Adam et al., 1986; Sachs et al., 1986), (Fig. 3). HA-tagged PABP is localized to the cytoplasm and appears to be excluded from the nucleus (Fig. 3). This distribution is identical to that of endogenous PABP (Görlach et al., 1994a). Deletion of the proline-rich carboxyl domain of PABP (4xRBDs) causes a mixed distribution such that a proportion of the cells show nuclear fluorescence (this may reflect the distribution of poly(A) RNAs), suggesting that the proline-rich auxiliary domain may be responsible for cytoplasmic confinement of PABP. The fusion protein between the four RBDs of PABP and the Gly domain of A1 is localized to the nucleus in contrast to the wild-type and proline-rich car-

A

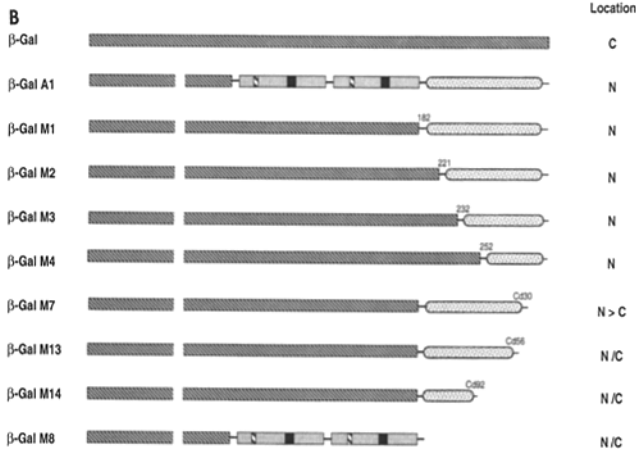
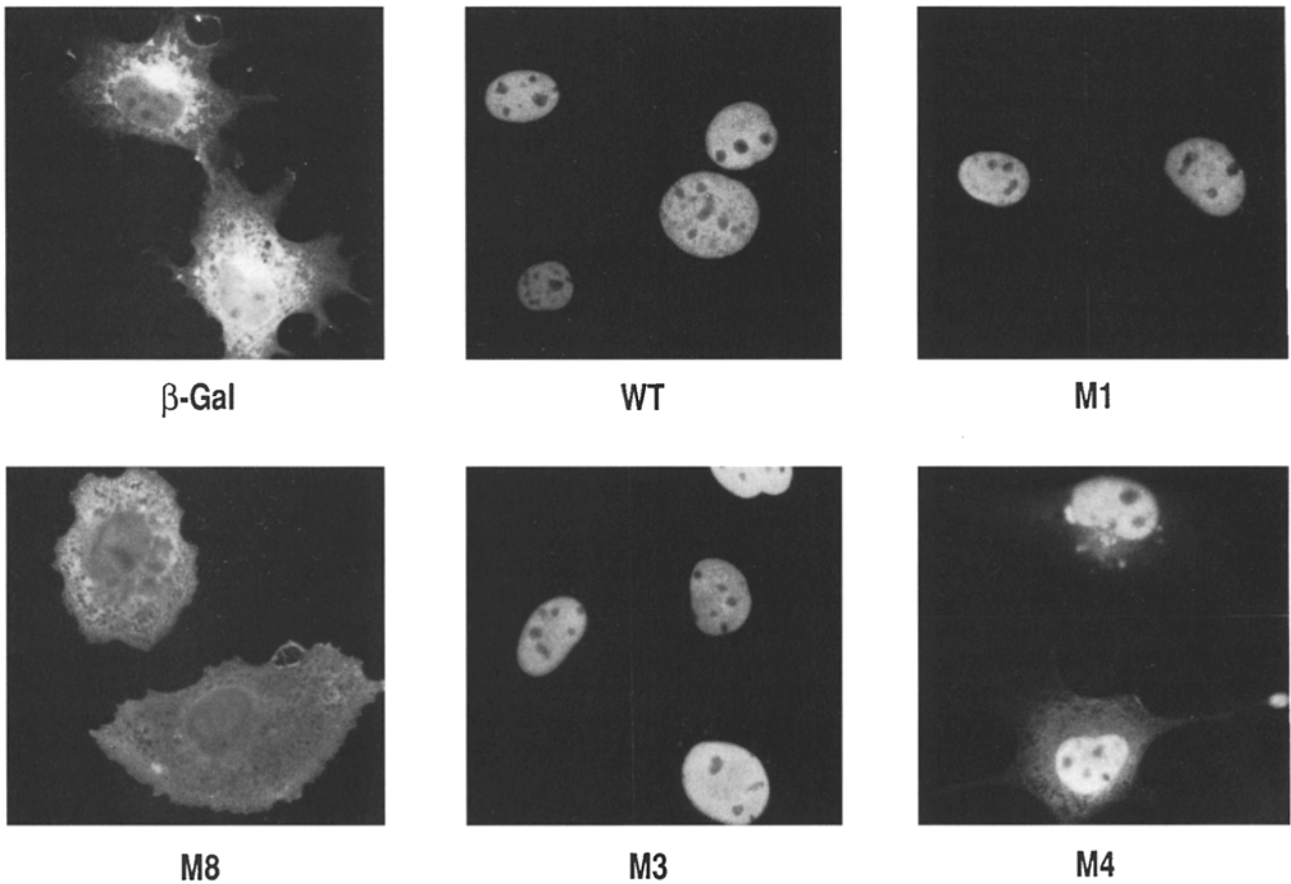


Figure 4. Structure and subcellular distribution of β -galactosidase-hnRNP A1 fusion derivatives. (A) Subcellular location of β -galactosidase-hnRNP A1 fusion proteins. COS1 cells were transfected with the plasmids encoding the fusion proteins illustrated in B. 48 h later the cells were fixed and then examined by indirect immunofluorescence using anti- β -galactosidase antibody. WT refers to β -gal-A1. (B) Structure of β -galactosidase-hnRNP A1 chimeric proteins. The chimeric proteins contain a constant portion of β -galactosidase at the amino terminus and the entire A1 or its deleted sequences at the carboxyl terminus. β -Galactosidase is represented by the hatched box. Numbers at the fusion points between β -galactosidase and hnRNP A1 portions indicate amino terminal amino acid positions of hnRNP A1 moieties. The β -galactosidase portion in the fusion proteins consists of 1004 residues. Assessment of the fluorescence staining was as described in the legends to Figs. 1-4. The symbol N>C refers to predominantly nuclear staining accompanied by some cytoplasmic staining.

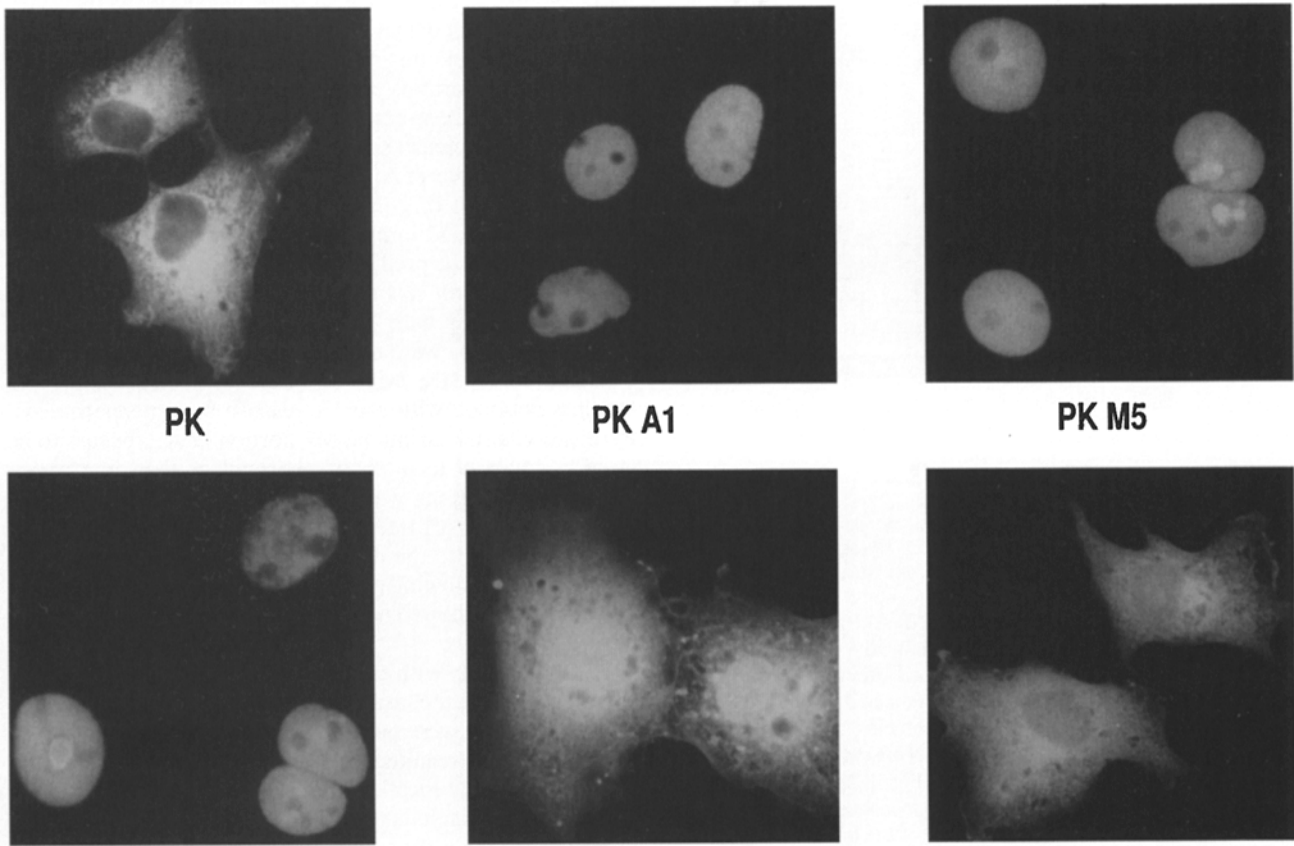
boxyl domain-deleted PABP (4 \times RBDs) (Fig. 3). This experiment also demonstrated the capacity of the A1 Gly domain to localize to the nucleus other RBDs than those found in A1, and a heterologous protein larger than A1.

A Domain in the Carboxyl Region of A1 Localizes Heterologous Proteins to the Nucleus

To further delineate the sequence responsible for the nuclear localization of A1 and to obviate the concern of diffusion of small proteins into the nucleus, we fused the A1 glycine-rich carboxyl domain and parts of it to β -galactosidase (M, 116,000) and carried out transfection experiments and de-

tected the expressed proteins with anti- β -galactosidase antibodies. Bacterial β -galactosidase is a well-characterized reporter for nuclear localization (Bürglin and De Robertis, 1987; Siomi et al., 1988; see also Silver, 1991 for review). β -gal-A1 completely localized to the nucleus whereas β -galactosidase by itself was mostly cytoplasmic (Fig. 4). These data suggest that nuclear import of A1 is an active transport because the size of the fusion protein is well beyond the reported diffusion limit for the nuclear pores (Dworetzky et al., 1988; Peters, 1986). The Gly domain and amino terminal deleted portions of this domain up to amino acid residue 252 (Fig. 4), act as targeting signals capable of directing

A



PK

PK A1

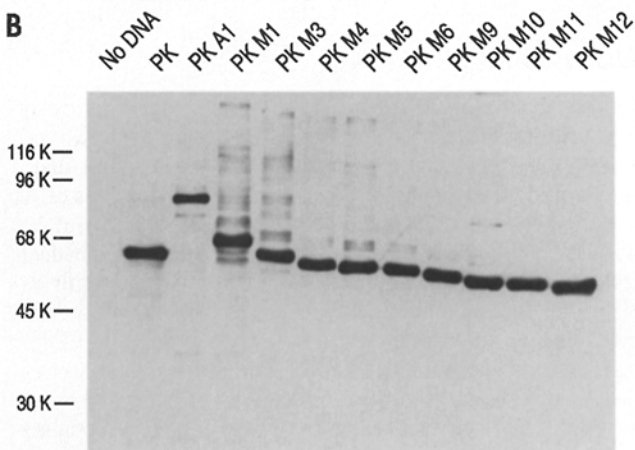
PK M5

PK M9

PK M11

PK M12

B



C

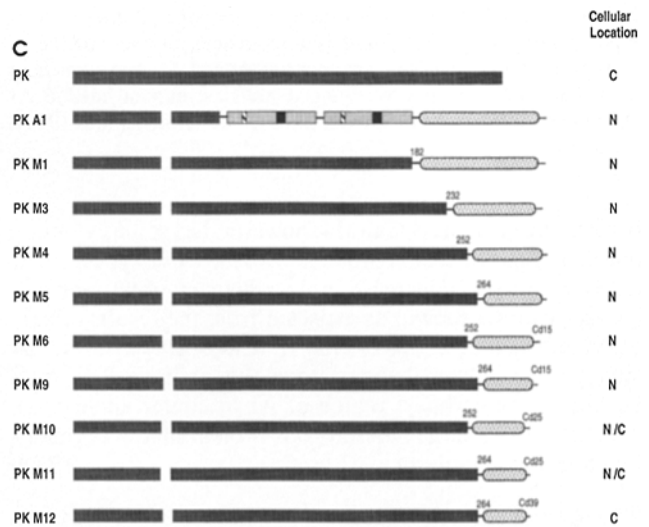


Figure 5. Structure, immunoblotting analysis and subcellular location of pyruvate kinase-hnRNP A1 fusion derivatives. **(A)** Subcellular location of pyruvate kinase-A1 fusion proteins. *Myc*-tagged pyruvate kinase-A1 fusion constructs were transfected into COS-1 cells. The subcellular localization of each protein was determined using anti-*myc* monoclonal antibody. **(B)** Immunoblotting analysis of pyruvate kinase-A1 fusion proteins. Lysates from transiently transfected COS-1 cells were analyzed by Western blotting using the anti-*myc* epitope antibody as a probe. Molecular weight standards are indicated on the left in kD. **(C)** Structure of pyruvate kinase-A1 fusion proteins. A series of fusion constructs similar to those described in Fig. 4 were constructed using chicken muscle pyruvate kinase in place of β -galactosidase. Note that the constructs were made using the *Kpn*I site in PK which truncates the protein at amino acid 443, hence, PK-M series of constructs are slightly smaller than WT PK. The *myc* epitope was placed at the NH₂ termini of all constructs.

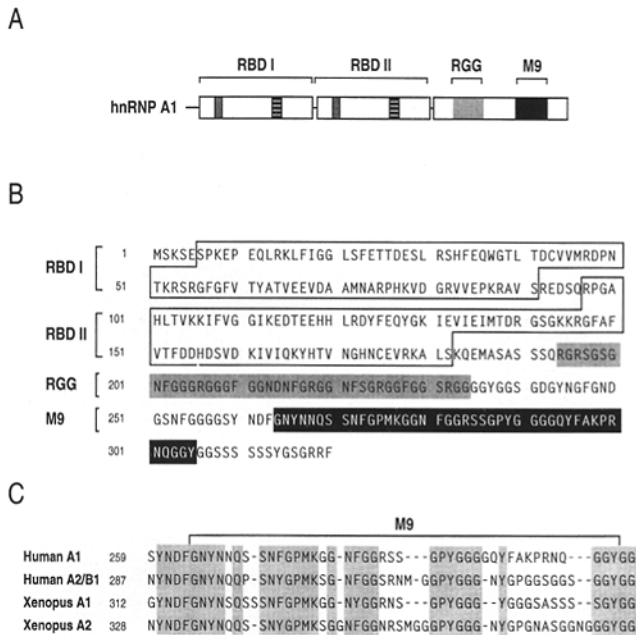


Figure 6. Functional domains of the hnRNP A1. (A) The structural organization of the hnRNP A1 is schematically represented with functional domains that have been mapped previously and in the present work. Functional domains are indicated by RBD (RNA-binding domain), RGG (RGG box; amino acids 194-234) and nuclear localization domain (M9). The RNA-binding domains are represented as described in the legend to Fig. 1. M9 is the smallest segment of hnRNP A1 which completely localizes pyruvate kinase to the nucleus. (B) Amino acid sequence of the hnRNP A1 protein and location of the functional domains. (C) An hnRNP A1-like motif occurs in the Gly domains of other nuclear RNA-binding proteins. Sequence alignment of the nuclear localization domain of hnRNP A1 (M9) with the corresponding domains of other nuclear RNA-binding proteins. Conserved sequences in the hnRNP A groups from mammals and *Xenopus* are shaded. Protein sequences were obtained from the following references: human hnRNP A1 (Cobianchi et al., 1988); human hnRNP A2 (Burd et al., 1989); *Xenopus* hnRNP A1 and A2 (Good et al., 1993).

β -gal to the nucleus. β -Gal-M4, however, has some cytoplasmic staining. The fusion proteins with the carboxyl-terminal deletions caused a decrease in the efficiency of nuclear targeting (data not shown) as expected from the results of carboxyl terminal deletions of HA-A1. The β -galactosidase fusion proteins clearly demonstrate the ability of a small portion of the carboxyl region of A1 to confer nuclear accumulation of a large heterologous protein that is normally found in the cytoplasm.

The Region of A1 Which Is Necessary and Sufficient for Nuclear Import Comprises a 40-Amino Acid Domain Near the Carboxyl Region of the Protein

We wished to define further the sequences within the carboxyl portion of A1 that are necessary for its nuclear localization. The last 60 amino acids of A1 have been shown above to be sufficient to localize *E. coli* β -galactosidase to the nucleus. However, the results of Kalderon et al. (1984) have questioned the suitability of this protein as a reporter for nuclear localization sequences; despite its large size and

bacterial origin, *E. coli* β -galactosidase may be able to partially localize to the nucleus of some eukaryotic cells. Thus, this bacterial protein may contain a partially functional NLS.

In order to avoid this possibility, we constructed gene fusions between portions of A1 and chicken muscle pyruvate kinase cDNA, whose encoded protein is stable and excluded from the cell nucleus (Kalderon et al., 1984; Richardson et al., 1986). Portions of A1 were fused to the 3' end of the *myc*-tagged PK cDNA (Fig. 5). Each gene fusion was expressed in COS-1 cells, and immunoblotting analysis demonstrated that a protein of the predicted sized was produced, and little, if any, degradation was apparent (Fig. 5).

Fusion proteins with NH₂-terminal deletions up to residue 252 (PK-M4) were nuclear in localization as judged by immunofluorescence with anti-*myc* antibody, supporting the data obtained with β -galactosidase fusion experiments. Pyruvate kinase bearing the M4 portion of A1 appears to be more completely localized to the nucleus than is β -galactosidase, for reasons we do not understand. Further deletions resulted in a protein (PK-M5) which was also completely localized to the nucleus. It should be noted that the fusion proteins showing nuclear localization, except PK-wild type A1 fusion, showed nonhomogeneous nuclear accumulation in some cells.

Fusion proteins with carboxyl terminal deletions demonstrated that while the last 15 amino acids of the protein (PK-M6 and PK-M9) were not essential for nuclear localization, further deletions resulted in A1 fusion proteins that were no longer completely localized to the nucleus. From these experiments, we conclude that the amino acids comprising M9 (boxed sequence in Fig. 6) are necessary and sufficient for complete nuclear localization of a heterologous protein that is otherwise cytoplasmic. It is possible that a few additional amino acids at the amino and carboxyl ends of M9 can also be removed without loss of nuclear localization, as we have not carried out further single amino acid deletions.

Discussion

By transfection of DNAs encoding epitope-tagged cytoplasmic proteins fused to portions of A1 and using an immunofluorescence microscopy assay we have found that a segment of 40 amino acids near the carboxyl terminus of A1 is necessary and sufficient for nuclear localization. hnRNP A1 and C1 differ in the requirements for their nuclear localization; A1 requires ongoing RNA polII transcription for nuclear localization whereas C1 does not. This suggests that a mechanistic difference exists concerning the subcellular trafficking of these hnRNP proteins (Piñol-Roma and Dreyfuss, 1993). Nuclear entry of hnRNP C1 is mediated by the well characterized NLS pathway whereas nuclear localization of hnRNP A1 occurs by a novel mechanism.

The assay for nuclear localization of a series of hnRNP A1 deletions and fusions to β -galactosidase and pyruvate kinase determined that neither the two RBDs nor an RGG box, the two types of RNA-binding elements which were previously found to mediate the RNA-binding activity of this protein, are required for nuclear localization. The results from the fusion of carboxyl terminal hnRNP A1 fragments to pyruvate kinase demonstrate that the sequences spanning the region from amino acids 264 to 305 are responsible for nuclear localization of hnRNP A1. This segment of A1 (termed M9,

Fig. 6) does not contain any short basic stretches similar to the basic type NLS previously identified in large T of SV-40 or a bipartite NLS previously found in nucleoplasmin (for reviews see Nigg, 1990; Forbes, 1992) and therefore is a novel type of nuclear localization sequence. The results of a search for proteins containing regions of similarity to the nuclear localization domain of hnRNP A1 are presented in Fig. 6. Similar amino acid sequences are found in the hnRNP A/B groups from mammals and amphibians.

Thus far, two other proteins whose nuclear localization is mediated by unusually large sequences rather than short basic NLSs have been described, namely the *trans*-acting protein Tax of human T-cell leukemia virus type I (Smith and Greene, 1992) and the snRNP protein U1A (Kambach and Mattaj, 1992). Neither of these have significant sequence similarity to the NLS of hnRNP A1. The U1A NLS is very large (more than 100 amino acids) but does contain several basic amino acid clusters which bear a resemblance to the SV-40 T antigen class of NLSs. In the case of Tax, the amino terminal 48 residues comprise a functional NLS which contains a zinc finger-like cysteine-rich sequence.

Specific protein-protein interactions have been shown to contribute to the nuclear localization of several proteins including adenovirus encoded DNA polymerase (Zhao and Padmanabhan, 1988) and the pancreas-specific transcription factor PTF1 (Sommer et al., 1991). It is possible that hnRNP A1 could be transported to the nucleus by complex formation through the Gly domain with a protein that contains a classical NLS (piggyback transport). Alternatively, hnRNP A1 protein may be targeted to the nucleus by a mechanism independent of that which targets classical NLS-bearing proteins. The import of m₃GpppN-containing snRNPs is kinetically non-competitive with the SV-40 T-antigen-like import pathway (Fischer and Lüthmann, 1990; reviewed in Goldfarb and Michaud, 1991), suggesting that multiple nuclear targeting mechanisms exist. It will be of interest to test whether free m₃GpppG dinucleotide and/or T-antigen NLS peptide compete nuclear import of hnRNP A1. Whether the components of the classical NLS pathway for nuclear import are used by hnRNP A1 or not, it is clear that some component(s) of the A1 import pathway is unique, since A1 import is dependent on RNA polII transcription. Furthermore, it is possible that A1 import is subject to different regulatory mechanisms than classical basic NLS-bearing protein import.

It has been shown that hnRNP A1 shuttles between the nucleus and cytoplasm and can be cross-linked to mRNA in the cytoplasm (Piñol-Roma and Dreyfuss, 1992, 1993), suggesting that a mechanism exists that regulates the nuclear reaccumulation of hnRNP A1 by, for example, monitoring the release of the mRNAs in the cytoplasm. Identifying the components of the transport machinery with which the nuclear localization domain of hnRNP A1 interacts will be important for determining the mechanism of transcription-dependent localization of RNA-binding proteins and of the nucleo-cytoplasmic shuttling of these proteins.

We are grateful to Erika Matunis and Michael Matunis for helpful discussions on protein sequences and invaluable help with immunofluorescence microscopy, to Vartkas Apkarian for oligonucleotide synthesis and to Gina Daly and Susan Kelchner for help with the manuscript. We are grateful to Dr. Bill Lee for the PK plasmid. We thank members of our laboratory, espe-

cially Serafin Piñol-Roma, Sara Nakielny, and Matthew Michael for many discussions and for helpful comments on the manuscript.

This work was supported by a grant from the National Institutes of Health. G. Dreyfuss is an Investigator of the Howard Hughes Medical Institute.

Received for publication 20 July 1994 and in revised form 6 February 1995.

References

- Bandziulis, R. J., M. S. Swanson, and G. Dreyfuss. 1989. RNA-binding proteins as developmental regulators. *Genes & Dev.* 3:431-437.
- Ben-David, Y., M.-R. Bani, B. Chabot, A. De Koven, and A. Bernstein. 1992. Retrioviral insertions downstream of the heterogeneous ribonucleoprotein A1 gene in erythroleukemia cells: evidence that A1 is not essential for cell growth. *Mol. Cell. Biol.* 12:4449-4455.
- Burd, C. G., and G. Dreyfuss. 1994b. Conserved structures and diversity of functions of RNA-binding proteins. *Science (Wash. DC)*. 265:615-621.
- Burd, C. G., M. S. Swanson, M. Görlach, and G. Dreyfuss. 1989. Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts. *Proc. Natl. Acad. Sci. USA*. 86:9788-9792.
- Burd, C. G., and G. Dreyfuss. 1994a. RNA binding specificity of hnRNP A1: hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO (Eur. Mol. Biol. Organ.) J.* 13:1197-1204.
- Bürglin, T. R., and E. M. De Robertis. 1987. The nuclear migration signal of *Xenopus laevis* nucleoplasmin. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:2617-2625.
- Buvoli, M., G. Biamonti, P. Tsoulfas, M. T. Bassi, A. Ghetti, S. Riva, and C. Morandi. 1988. cDNA cloning of human hnRNP protein A1 reveals the existence of multiple RNA isoforms. *Nucleic Acids Res.* 16:3751-3770.
- Buvoli, M. G., F. Cobianchi, M. G. Bestagno, A. Mangiarotti, M. T. Bassi, G. Biamonti, and S. Riva. 1990. Alternative splicing in the human gene for the core protein A1 generates another hnRNP protein. *EMBO (Eur. Mol. Biol. Organ.) J.* 9:1229-1235.
- Cáceres, J. F., S. Stamm, D. M. Helfman, and A. R. Krainer. 1994. Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science (Wash. DC)*. 265:1706-1709.
- Choi, Y. D., and G. Dreyfuss. 1984. Isolation of the heterogeneous nuclear RNA-ribonucleoprotein complex (hnRNP): A unique supramolecular assembly. *Proc. Natl. Acad. Sci. USA*. 81:7471-7475.
- Cobianchi, F., R. L. Karpel, K. R. Williams, V. Notario, and S. H. Wilson. 1988. Mammalian heterogeneous nuclear ribonucleoprotein complex protein A1. *J. Biol. Chem.* 263:1063-1071.
- Cobianchi, F., G. Biamonti, M. T. Bassi, M. Buvoli, and S. Riva. 1990. The protein A1 of mammalian hnRNP: a new insight into protein complexity and gene family structure. In *The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assemblies*. Volume 2. P. R. Strauss and S. H. Wilson, editors. The Telford Press, Caldwell, N.J. 561-577.
- Dang, C. V., and W. M. F. Lee. 1988. Identification of the human c-myc protein nuclear translocation signal. *Mol. Cell. Biol.* 8:4048-4054.
- Dingwall, C., and R. A. Laskey. 1991. Nuclear targeting sequences—a consensus? *Trends Biochem. Sci.* 16:478-481.
- Dworetzky, S. I., R. E. Lanford, and C. M. Feldherr. 1988. The effects of variations in the number and sequence of targeting signals on nuclear uptake. *J. Cell Biol.* 107:1279-1287.
- Dreyfuss, G., M. S. Swanson, and S. Piñol-Roma. 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.* 13:86-91.
- Dreyfuss, G., M. J. Matunis, S. Piñol-Roma, and C. G. Burd. 1993. hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* 62:289-321.
- Evan, G. I., G. K. Lewis, G. Ramsay, and J. M. Bishop. 1985. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.* 5:3610-3616.
- Fischer, U., and R. Lüthmann. 1990. An essential signaling role for the m₃G cap in the transport of U1 snRNP to the nucleus. *Science (Wash. DC)*. 249:786-790.
- Forbes, D. J. 1992. Structure and function of the nuclear pore complex. *Annu. Rev. Cell Biol.* 8:495-527.
- Fu, X.-D., A. Mayeda, T. Maniatis, and A. R. Krainer. 1992. General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5' and 3' splice site selection. *Proc. Natl. Acad. Sci. USA*. 89:11224-11228.
- Goldfarb, D., and N. Michaud. 1991. Pathways for the nuclear transport of proteins and RNA. *Trends Cell Biol.* 1:20-24.
- Good, P. J., M. L. Rebertus, and I. B. Dawid. 1993. Three new members of the RNP protein family in *Xenopus*. *Nucleic Acids Res.* 21:999-1006.
- Görlach, M., C. G. Burd, and G. Dreyfuss. 1994a. The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. *Exp. Cell Res.* 211:400-407.
- Görlach, M., C. G. Burd, and G. Dreyfuss. 1994b. The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J. Biol. Chem.* 269:23074-23078.

- Guo, L.-H., and R. Wu. 1983. Use for DNA sequence analysis and in specific deletions of nucleotides. *Methods Enzymol.* 100:60-96.
- Hall, C. V., P. E. Jacob, G. M. Ringold, and F. Lee. 1983. Expression and regulation of *Escherichia coli lacZ* gene fusions in mammalian cells. *J. Mol. Appl. Genet.* 2:101-109.
- Horton, R. M., Z. Cai, S. N. Ho, and L. R. Pease. 1990. Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques.* 8:528-535.
- Kalderon, D., B. L. Roberts, W. D. Richardson, and A. E. Smith. 1984. A short amino acid sequence able to specify nuclear location. *Cell.* 39:499-509.
- Kambach, C., and I. W. Mattaj. 1992. Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1 snRNA. *J. Cell Biol.* 118:11-21.
- Kay, B. K., R. K. Sawhney, and S. H. Wilson. 1990. Potential for two isoforms of the A1 ribonucleoprotein in *Xenopus laevis*. *Proc. Natl. Acad. Sci. USA.* 87:1367-1371.
- Kenan, D. J., C. C. Query, and J. D. Keene. 1991. RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.* 16:214-220.
- Kiledjian, M., and G. Dreyfuss. 1992. Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO (Eur. Mol. Biol. Organ.) J.* 11:2655-2664.
- Kolodziej, P. A., and R. A. Young. 1991. Epitope tagging and protein surveillance. *Methods Enzymol.* 194:508-519.
- Mayeda, A., and A. R. Krainer. 1992. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell.* 68:365-375.
- Matunis, E. L., M. J. Matunis, and G. Dreyfuss. 1992. Characterization of the major hnRNP proteins from *Drosophila melanogaster*. *J. Cell Biol.* 116:257-269.
- Nigg, E. A. 1990. Mechanisms of signal transduction to the cell nucleus. *Adv. Cancer Res.* 55:271-310.
- Niman, H. L., R. A. Houghten, L. E. Walker, R. A. Reisfeld, I. A. Wilson, J. M. Hogle, and R. A. Lerner. 1983. Generation of protein-reactive antibodies by short peptides is an event of high frequency: implications for the structural basis of immune recognition. *Proc. Natl. Acad. Sci. USA.* 80:4949-4953.
- Peters, R. 1986. Fluorescence microphotolysis to measure nucleo-cytoplasmic transport and intracellular mobility. *Biochim. Biophys. Acta.* 864:305-359.
- Piñol-Roma, S., and G. Dreyfuss. 1991. Transcription-dependent and transcription independent nuclear transport of hnRNP proteins. *Science (Wash. DC).* 253:312-314.
- Piñol-Roma, S., and G. Dreyfuss. 1992. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature (Lond.).* 355:730-732.
- Piñol-Roma, S., and G. Dreyfuss. 1993. hnRNP proteins: localization and transport between the nucleus and the cytoplasm. *Trends Cell Biol.* 3:151-155.
- Piñol-Roma, S., Y. D. Choi, M. J. Matunis, and G. Dreyfuss. 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes & Dev.* 2:215-227.
- Riva, S., C. Morandi, P. Tsoulfas, M. Pandolfo, G. Biamonti, B. Merrill, K. R. Williams, G. Multaup, K. Beyreuther, H. Werr, B. Henrich, and K. P. Schafer. 1986. Mammalian single-stranded DNA binding protein UP1 is derived from the hnRNP core protein A. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:2267-2273.
- Robbins, J., S. M. Dilworth, R. A. Laskey, and C. Dingwall. 1991. Two interdependent basic domains in nucleoplasmic nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell.* 64:615-623.
- Roberts, B. L., W. D. Richardson, and A. E. Smith. 1987. The effect of protein context on nuclear location signal function. *Cell.* 50:465-475.
- Silver, P. A. 1991. How proteins enter the nucleus. *Cell.* 64:489-497.
- Siomi, H., H. Shida, S. H. Nam, T. Nosaka, M. Maki, and M. Hatanaka. 1988. Sequence requirements for nucleolar localization of human T cell leukemia virus type I pX protein, which regulates viral RNA processing. *Cell.* 55:197-209.
- Siomi, H., H. Shida, M. Maki, and M. Hatanaka. 1990. Effects of a highly basic region of human immunodeficiency virus Tat protein on nucleolar localization. *J. Virol.* 64:1803-1807.
- Siomi, H., M. C. Siomi, R. L. Nussbaum, and G. Dreyfuss. 1993. The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell.* 74:291-298.
- Smith, M. R., and W. C. Greene. 1992. Characterization of a novel nuclear localization signal in the HTLV-I Tax transactivator protein. *Virology.* 187:316-320.
- Sommer, L., O. Hagenbüchle, P. K. Wellauer, and M. Strubin. 1991. Nuclear targeting of the transcription factor PTF1 is mediated by a protein subunit that does not bind to the PTF1 cognate sequence. *Cell.* 67:987-994.
- Swanson, M. S., and G. Dreyfuss. 1988. RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:3519-3529.
- Swanson, M. S., T. Y. Nakagawa, K. Levan, and G. Dreyfuss. 1987. Primary structure of the human ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA and pre-rRNA-binding proteins. *Mol. Cell. Biol.* 7:1731-1739.
- Williams, K. R., K. L. Stone, M. B. LoPresiti, 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.
- Yang, W., M.-R. Bani, S.-J. Ju, S. Rowan, Y. Ben-David, and B. Chabot. 1994. The A1 and A1^B proteins of heterogeneous nuclear ribonucleoproteins modulate 5' splice site selection in vivo. *Proc. Natl. Acad. Sci. USA.* 81:6924-6928.
- Zhao, L.-J., and R. Padmanabhan. 1988. Nuclear transport of adenovirus DNA polymerase is facilitated by interaction with preterminal protein. *Cell.* 55:1005-1015.