The U1 Small Nuclear Ribonucleoprotein (snRNP) 70K Protein Is Transported Independently of U1 snRNP Particles via a Nuclear Localization Signal in the RNA-Binding Domain

JOELLE M.-J. ROMAC,[†] DAVID H. GRAFF, AND JACK D. KEENE*

Department of Microbiology, Duke University Medical Center, Durham, North Carolina 27710

Received 29 November 1993/Returned for modification 5 January 1994/Accepted 21 April 1994

Expression of the recombinant human U1-70K protein in COS cells resulted in its rapid transport to the nucleus, even when binding to U1 RNA was debilitated. Deletion analysis of the U1-70K protein revealed the existence of two segments of the protein which were independently capable of nuclear localization. One nuclear localization signal (NLS) was mapped within the U1 RNA-binding domain and consists of two typically separated but interdependent elements. The major element of this NLS resides in structural loop 5 between the β 4 strand and the α 2 helix of the folded RNA recognition motif. The C-terminal half of the U1-70K protein which was capable of nuclear entry contains two arginine-rich regions, which suggests the existence of a second NLS. Site-directed mutagenesis of the RNA recognition motif NLS demonstrated that the U1-70K protein can be transported independently of U1 RNA and that its association with the U1 small nuclear ribonucleoprotein particle can occur in the nucleus.

U1 small nuclear ribonucleoprotein (snRNP) particles are located in the nuclei of eukaryotic cells and perform a transactivating role in splicing of pre-mRNAs (for reviews, see references 11 and 38). U1 snRNP particles contain U1 RNA and a set of nine proteins termed the Sm complex, which is also present in the other abundant U snRNPs (21, 23). In addition, the U1-70K, U1-A, and U1-C proteins are unique to the U1 snRNP particle (3, 29).

Assembly of the U1 snRNP particle is thought to occur in the cytoplasm (25). It is believed that cytoplasmic assembly and nuclear transport require both hypermethylation of the 5' cap structure of U1 RNA (12) and a binding site for the Sm proteins on U1 RNA (8, 12). However, more recent experiments suggested that the Sm core domain can bind to a transport receptor and can be transported to the nucleus independently of the m³G cap (9). Also, pulse-chase experiments suggest that the Sm proteins associate with U1 RNA in the cytoplasm (7). Consistent with this model of U1 snRNP transport were the findings that the U1-70K and U1-A proteins are dispensable for nuclear import of U1 snRNP particles (12). Furthermore, microinjection experiments in *Xenopus* oocytes using U1-A protein indicated that this protein contains sequences necessary for nuclear localization within a stretch of 110 residues between amino acids 94 to 204 (17). These amino acids belong to the spacer region separating the two RNA recognition motifs (RRMs). The U1-C protein has also been reported to migrate to the nuclei of oocytes on its own and to associate with preassembled U1 snRNP (15).

The U1-70K protein possesses a single defined U1 RNA-binding domain near the N terminus and two Arg-rich domains in the C-terminal portion (30). Similar Arg-rich sequences are also present in known splicing factors including tra, tra-2 (30), su (w^a) (4), and in a unique set of proteins termed SR (Ser-Arg-rich) proteins (41).

The U1 RNA-binding domain of the U1-70K protein contains a characteristic RRM (30). RRM is generally com-

posed of 80 amino acids (1, 18), and a model of the three-dimensional structure has been deduced for the first RRM of U1-A by crystallography (27) and nuclear magnetic resonance (14). It is unlikely that all of these residues are required for direct binding with the cognate RNA. Thus, certain amino acids in the RRM are necessary for correct folding, while others are necessary for direct RNA contact, as well as other functions. One example is the U2-B' protein, in which structural loop 3 of its RRM is involved in protein-protein interactions with the U2-A' protein, thus augmenting the affinity of RNA binding (2, 36). In this study, we demonstrate that the U1-70K RRM contains elements of a nuclear localization signal (NLS) in loop 5 which presumably interact with cellular components involved in protein shuttling (37). Using transient expression in mammalian cells, we show that the expressed protein is readily transported to the nucleus and that this transport does not require U1 RNA binding or presumably U1 snRNP association.

MATERIALS AND METHODS

Cloning procedures. g10 was added to the U1-70K constructs through subcloning of these cDNAs into pET-3A (35). g10-U1-70K constructs were derived from U1-70K clones previously described (30) or by restriction digestion. cDNA constructs of the g10-U1-70K clone were subcloned into the pBC/CMV/IL2 vector (5) by removal of the interleukin-2 (IL-2) cDNA. Point mutants, with the exception of the octamer mutant, were obtained by PCR-directed oligonucleotide mutagenesis, followed by standard cloning procedures. Cloned PCR fragments were verified by sequencing with Sequenase version 2.0 (U.S. Biochemical Corp.). The RRM octamer point mutant in which a phenylalanine was replaced by a tyrosine was subcloned into the pBC/CMV/g10 from the pGEM-(92-216)Y148 plasmid (31a).

Transfection. COS cells were maintained in Iscove medium (Gibco) supplemented with 10% fetal bovine serum (Flow/ICN). Plasmid DNA was transfected into COS cells by the DEAE-dextran method of transfection (6). DNA, at a concentration of 2.5 μ g/ml in phosphate-buffered saline containing

* Corresponding author.

[†] Present address: Department of Medicine, Duke University Medical Center, Durham, NC 27710.

FIG. 1. COS cells transfected with epitope (g10)-tagged U1-70K protein (a and b) or epitope (g10)-tagged U1-A protein (c and d). Cells were fixed by the PLP fixation method, made permeable by treatment with 0.1% Triton X-100, incubated with the anti-g10 serum (1/100), and stained with anti-rabbit immunoglobulin G serum conjugated with rhodamine (1/100). Cells were observed by phase-contrast (a and c) or by epifluorescence (b and d) microscopy with a 25 $\times$ objective.

500 μ g of DEAE-dextran (Pharmacia) per ml, was added to 5×10^4 cells per cm^2 for 30 min at 37°C. Medium with 100 μ M chloroquine (Sigma) was added, and the cells were incubated for 2.5 h. The medium was aspirated, and the cells were incubated with 10% dimethyl sulfoxide in Iscove medium for 2 min. Dimethyl sulfoxide-containing medium was removed, fresh medium was added, and the cells were incubated at 37°C. The cells were harvested after 48 h for immunofluorescence or 72 h for immunoprecipitation.

Immunofluorescence. Cells were fixed by PLP (10 mM periodate, 75 mM lysine, 2% paraformaldehyde) (26) for 15 min and then made permeable by treatment with 0.1% Triton X-100 for 15 min. Subsequently, the cells were incubated with the anti-g10 serum (1/100) for 30 min at room temperature followed by incubation with a mouse anti-rabbit immunoglobulin G conjugated with rhodamine (Jackson ImmunoResearch) at a dilution of 1/100 for another 30 min at room temperature. The cells were then washed and mounted in 60% glycerol with 2.5% *n*-propyl-gallate (Sigma). They were observed by phase-contrast or epifluorescence microscopy with a 25 $\times$ (Leitz) or 40 $\times$ objective (Zeiss). Photographs were recorded on Tmax 400 film (Kodak).

Mobility shift RNA binding assay. cDNA clones of the g10-U1-70K protein in the pET vector (35) were transcribed in vitro with T7 polymerase (U.S. Biochemical Corp.), and the mRNAs were then translated in vitro in reticulocyte lysate (Promega) in the presence of [35 S]methionine (ICN) at 37°C for 1 h. The proteins were dialyzed against KNET3 buffer (13) for 1 h.

The cDNA of the U1 RNA stem-loop 1 (SL1) fragment was transcribed in vitro as previously described (31); 4 μ g of RNA was added to 35 S-labelled proteins in KNET3 buffer (13) and incubated for 20 min at 37°C in the presence of nonspecific competitor RNAs (30). The complex was resolved by electrophoresis in 90 mM Tris-borate buffer (pH 8.3)–5% polyacrylamide native gel at 10 V/cm at 4°C (30). The gel was then rinsed in water and soaked in 25 mM Tris-base–0.1 N NaOH for 10 min. Then, it was soaked twice for 10 min in 0.5 M sodium salicylate (24), dried, and photographed by fluorography with hyperfilm-MP (Amersham).

Immunoprecipitation. The cells were lysed in STM buffer (10 mM Tris [pH 7.2], 100 mM NaCl, 1 mM MgCl_2) by sonication and clarified by centrifugation, and the cell extract was incubated with the anti-g10 serum for 30 min on ice. The immune complex was purified with protein A-Sepharose (Sigma) and washed with NET2+ buffer (2). The complex was released from the protein A-beads with 0.1 M glycine-HCl buffer, pH 3. RNA was extracted with phenol, precipitated with ethanol, and resuspended in water. RNA was hybridized overnight with a U1 cDNA probe, spanning the nucleotides 1 to 71 and containing additional nucleotides from the pGEM vector (Promega) at the 5' end so that the probe could be differentiated from the protected fragment. The S1 nuclease protection assay was performed as follows. The RNA-DNA complexes were hybridized at 55°C overnight and then digested for 1 h at 37°C. Protected fragments were expected to be 71 bp long. RNA-DNA hybrids were electrophoresed in a 5% polyacrylamide–8.3 M urea gel in 50 mM Tris-borate buffer

FIG. 2. Localization of U1-70K protein fragments containing the g10 epitope tag following transfection into COS cells. Tagged U1-70K fragments were detected by indirect immunofluorescence as described in the legend to Fig. 1. (A) Schematic of human U1-70K and relevant deletion constructs. The U1 RNA-binding domain (stippled box), RRM motif (black box), and the Arg-rich regions (white boxes) are indicated. AA, amino acids; WT, wild type; N, nucleus; C, cytoplasm. (B) Immunofluorescence of transfected COS cells. Phase-contrast (left panels) and immunofluorescence (right panels) staining of the same fields. Cells were observed on a Zeiss Axioskop microscope with a 40 $\times$ objective. wt, wild type.

at 10 W. The gel was fixed in 10% glacial acetic acid–10% methanol, dried, and autoradiographed with X-AR film (Kodak).

RESULTS

Overexpressed U1-70K and U1-A proteins accumulate in the nucleus. The constructs used for overexpression of U1-70K or U1-A protein in COS cells were derived from human cDNA clones described previously (24, 30), and each was fused to an N-terminal epitope of 12 amino acids termed g10 (39). Figure 1 shows immunofluorescence of cells transfected with g10-U1-70K (Fig. 1b) or g10-U1-A (Fig. 1d). At 48 h posttransfection, the cells were fixed, reacted with the rabbit anti-g10 serum, and stained with a mouse anti-rabbit antibody conjugated with rhodamine. The corresponding fields of cells were also photographed with phase-contrast microscopy (Fig. 1a and c). Both wild-type g10-U1-70K and g10-U1-A accumulated in interphase nuclei but with different staining patterns (compare Fig. 1b with Fig. 1d). The g10-U1-A was diffused throughout the nucleoplasm (Fig. 1d). g10-U1-70K was present in two nuclear locales. The protein showed faint diffuse staining throughout the nucleoplasm in addition to being concentrated into large discrete granules (Fig. 1b).

The brightly staining granules seen in the U1-70K transfectants correlated with dense nuclear structures that were apparent even with phase-contrast microscopy (Fig. 1a, arrow). The granules varied in size and number among individual cells. Having demonstrated that the U1-70K protein accumulated in the nucleus, the next logical step was to determine the elements responsible for its accumulation. Therefore, cellular

location of the U1-70K deletion fragments was determined by indirect-immunofluorescence microscopy, while assembly of these deletion constructs into U1 snRNP particles was assessed by coimmunoprecipitation with U1 RNA.

Analysis of U1-70K localization using deletion fragments. COS cells were transfected with various cDNA deletion fragments, and the cellular location of the overexpressed protein was analyzed by indirect immunofluorescence at 48 h posttransfection. These fragments also carried the g10 epitope at their amino termini. A diagram of the various mutants (Fig. 2A) and examples of cell staining experiments (Fig. 2B) are shown. Immunofluorescence demonstrated that some fragments of U1-70K (d1, d2, d3, d6, d7, d9, d10, d11, and d12) were distributed exclusively in the nuclei of transfected cells, while other U1-70K protein fragments (d4, d5, d8, and d13) were distributed throughout the nuclei and cytoplasm of the cells. Comparison of d3 with d4 indicated that amino acids 183 to 199 were critical for nuclear accumulation. d11, corresponding to a U1-70K fragment 92-199 (fragment containing amino acids 92 to 199), was the smallest fragment that still retained the capability to accumulate in the nucleus. This fragment encompasses the RRM but is shorter than the U1-70K RNA-binding domain defined by *in vitro* studies (30). Two other mutants, d7 and d12, lacking most of the RNA-binding domain remained strictly nuclear. Further deletions of these fragments at the amino termini (compare d7 with d8) or at the carboxy termini (compare d12 with d13) rendered the protein fragments cytoplasmic as well as nuclear (Fig. 2B). From this deletion analysis, we were able to delineate two regions of U1-70K that are sufficient for accumulation in the nucleus; the

FIG. 2—Continued.

first region is located between residues 92 and 199; the other requires C-terminal amino acids 183 to 437.

The carboxy terminus of U1-70K protein does not associate with U1 snRNP particles in vivo. To determine the ability of the deletion peptides to associate with U1 snRNP particles, we carried out immunoprecipitation experiments using extracts from cells transfected with various epitope-tagged fragments of U1-70K protein. An aliquot of each extract was used for detection of the expressed protein by Western blotting (immunoblotting) with anti-g10 antibody (Fig. 3B). As noted previously (30), the full-length U1-70K protein and the C-terminal half of the protein migrate aberrantly in sodium dodecyl sulfate-polyacrylamide gels (Fig. 3B, lanes 1, 4, 5, and 7). In addition to immunoblotting, expressed U1-70K fragments co-

immunoprecipitated U1 RNA, as detected by S1 nuclease protection (Fig. 3A). The 90-nucleotide-long probe yielded a 71-nucleotide protected fragment when hybridized with U1 RNA. As expected, all U1-70K protein fragments encompassing the U1 RNA-binding domain were associated with U1 snRNPs (U1-70K [Fig. 3A, lane 1], d1 [lane 2], d2 [lane 3], and d9 [lane 6]). Consistent with in vitro results (30), protein fragments lacking part (d7 [lane 4] and d12 [lane 7]) or all (d8 [lane 5]) of the U1 RNA-binding domain did not associate detectably with U1 snRNP particles.

Interestingly, both d7 and d12 were located in the nucleus following transfection (Fig. 2B). Since these mutants did not associate in U1 snRNP particles, it is most likely that U1-70K protein contains amino acids necessary for nuclear transport in

FIG. 3. Expression and U1 RNA binding of U1-70K protein fragments transfected into COS cells. (A) Immunoprecipitated U1 RNA from COS cells transfected with either wild-type U1-70K (lane 1), d1 (lane 2), d2 (lane 3), d7 (lane 4), d8 (lane 5), d9 (lane 6), or d12 (lane 7). Extracts were immunoprecipitated with the anti-g10 antibody, and RNA was phenol extracted and subjected to S1 nuclease protection analysis. The RNA-DNA protected fragments were resolved in 7 M urea-5% polyacrylamide gel and visualized by autoradiography. The protected fragments are indicated by an arrow. Lane M contains molecular size markers (indicated on the left in base pairs). (B) Aliquots of the cell extracts were subjected to Western blot analysis using the anti-g10 antibody to demonstrate expression of each fragment in COS cells. The numbers on the left correspond to molecular mass markers in kilodaltons.

the carboxy-terminal half of the molecule. In our assay, the d11 protein fragment could not be detected by Western blotting; therefore, no conclusion could be derived as to its association with U1 snRNPs *in vivo*. However, on the basis of *in vitro* binding experiments (30), the protein fragment represented by d11 will not bind U1 RNA. Therefore, we designed experiments to determine whether the d11 protein fragment, which encompasses the RRM, harbors a separate NLS.

Loop 5 of the U1-70K RNA-binding domain contains basic amino acids required for nuclear localization. Our strategy was based on the following rationale. From deletion analysis of U1-70K (Fig. 2), it is evident that amino acids 183 to 199 are necessary for nuclear accumulation. In general, NLSs (10) contain basic residues, and as described below, the three-dimensional model of the RRM of the U1-70K protein can accommodate an NLS within structural loop 5. As depicted in Fig. 4A, the U1-70K RRM, modeled on the U1-A RNA-binding domain (24, 27), is composed of four antiparallel β strands forming a surface for interaction with RNA. Two α helices are located behind the β -sheet surface, with the first helix connecting β 1 with β 2 and the second helix connecting β 3 with β 4. Loops 1, 3, and 5 connect these structures (Fig. 4A). A number of basic residues are situated in loop 5 of U1-70K (amino acids 167 to 173) and in the region found to be necessary for nuclear targeting (amino acids 183 to 199). Therefore, specific amino acid substitutions were placed in loop 5 and in residues 190 and 191. The initial construct that was used for oligonucleotide-directed mutagenesis was a cDNA encoding the structurally defined domain 92-199 (d11) which lacks detectable binding affinity for U1 RNA *in vitro* (30) yet was still able to accumulate in the nucleus (Fig. 3B). M0 has a tyrosine substituted for Phe-148 in the RNP 1 octamer. This mutation, when present in the 92-216 fragment, abolished binding to U1 RNA *in vitro* (31a). Residue 148 is on

the surface of the molecule and exposed to solvent (15). Thus, the mutation is not expected to grossly affect the folding of the RNA-binding domain. Since M0 (92-199 fragment, Y-148) still accumulated in the nucleus, it was used as a control for nuclear staining in the indirect immunofluorescence experiment described below.

Residues G-166, K-167, K-168, and R-173 were mutated either individually or in conjunction with one another. The M1 to M5 mutants contained single (P-173 [M3]; Q-173 [M4]), double (F-167 and P-168 [M1]), or triple (D-166, F-167, and P-168 [M2]; F-167, P-168, and P-173 [M5]) mutations in the RNA-binding domain loop 5 (see Fig. 7A). It is unlikely that the structure of the RNA-binding domain was compromised in this experiment, because the substitutions in loop 5 retained U1 RNA binding activity (see Fig. 5 and 7). The substituted amino acids F-167, P-168, and P-173 of U1-70K were chosen for their locations on the surface, modeled from the corresponding positions deduced from amino acids in the alignment of U1-A protein (18). Arginine residues at positions 190 and 191 in a carboxy-terminal region of the RRM that are still part of the functional RNA-binding domain (30) were substituted with threonines (M6). Examples of immunofluorescence of cells transfected with these constructs show clearly that the M0 peptide was strongly nuclear (Fig. 4B). M1, M2, M3, M4, M5 (not shown) and M6 mutants did not accumulate in the nucleus and instead were distributed throughout the cell. These results clearly demonstrate that three basic residues of loop 5 and two basic residues outside of the RRM, but still within the RNA-binding domain, are required for transport of U1-70K protein to the nucleus. Thus, it is apparent that loop 5 of the U1-70K RNA-binding domain is part of a NLS, is accessible to solvent, and is in a plane different from that of the four β strands that interact with the RNA molecule (see Fig. 7A). It is possible that other basic residues near loop 5 are involved in nuclear targeting. These findings are compatible with the existence of a bipartite or split NLS in U1-70K (37) (see below).

Arginine residues in NLS 1 are involved in U1 RNA binding. To address whether amino acids of the U1-70K loop 5 can participate in U1 RNA binding, we examined the ability of loop 5 mutants to bind to SL1 (nucleotides 16 to 48) of U1 RNA *in vitro* (31). cDNA d10 constructs encoding amino acids 92 to 216 but harboring the point mutations described above, were subcloned into pET vectors containing a T7 promoter site (35). *In vitro*-translated proteins were expressed from *in vitro*-synthesized transcripts and labelled with [35 S]methionine by using rabbit reticulocyte lysates. RNA representing SL1 of U1 RNA, the binding site of U1-70K, was either omitted from the binding (Fig. 5, odd-numbered lanes) or added (Fig. 5, even-numbered lanes). Labelled protein was detected on non-denaturing polyacrylamide gels. The U1-70K peptide, fragment 92-216, was used as a positive control, and the expected mobility shift was evident when SL1 RNA was added (compare lanes 1 and 2). As expected from previous RNA binding studies (30, 31a), no change in the migration of M0 (U1-70K, octamer mutant) was detected upon the addition of RNA (lane 16). All of the loop 5 mutants of U1-70K retained the ability to bind U1 SL1 RNA as shown (M1 [lane 4], M2 [lane 6], M3 [lane 8], M4 [lane 10], and M5 [lane 12]). Thus, we conclude that the basic residues in loop 5 involved in nuclear localization are not critical for U1 RNA binding. Also, these results demonstrate that these amino acid substitutions did not significantly alter the overall RRM fold required for RNA binding. However, the changes at positions 190 and 191 (M6) of arginines to threonines did affect RNA binding in this assay (lane 14).

Nuclear transport of U1-70K protein is independent of U1

FIG. 4. Mutational analysis of NLS in the U1 RNA-binding domain of the U1-70K protein. (A) Schematic of the point mutants in the RNA-binding domain (92-199) used for intracellular localization by indirect immunofluorescence. The RRM (black box) and RNA-binding domain (stippled box) (18, 30) are shown. (B) Immunofluorescence of COS cells transfected with either M0 (a), M1 (b), M2 (c), M3 (d), M4 (e), or M6 (f) constructs and reacted with anti-g10 serum as described in the legend of Fig. 1. M0 mutant is the 92-216 construct. Nuclear staining similar to that of M0 was observed with the 92-199 construct (not shown).

snRNP particle transport. As shown above, detailed analyses of the deletion and point mutants of U1-70K showed that two protein fragments could accumulate independently of one another in the nucleus. The first fragment, d11 (92-199 fragment) contained a NLS, as demonstrated above. The other fragment representing the entire C-terminus d12 (183-437 fragment) could not be truncated further without losing its ability to accumulate in the nucleus. The two fragments of U1-70K (d13 and d8) were both cytoplasmic. Therefore, we designed a construct representing U1-70K residues 92 to 216 with triple mutations in loop 5 in order to determine definitively whether U1-70K could still be transported to the nucleus via a cytoplasmic association with nascent U1 snRNP particles. As shown in Fig. 5, lane 12, the U1-70K construct with 92-216 fragment and F-167, P-168, and P-173 retained U1 RNA binding activity. By indirect immunofluorescence, as depicted in Fig. 6, this protein fragment did not accumulate in the nucleus, however. Instead, these transfected COS cells displayed cytoplasmic and nuclear staining (Fig. 6b). Overall,

these data suggest that the transport of U1-70K protein to the nucleus is not dependent upon its association with the U1 snRNP particle as mediated by U1 RNA binding.

DISCUSSION

In this study, we examined the localization and U1 snRNP association of the U1-70K protein by overexpressing the protein and its mutants in COS cells. Indirect immunofluorescence microscopy and coimmunoprecipitation of the protein with U1 RNA strongly suggest that the U1-70K protein enters the nucleus independently of the U1 snRNP particle. In addition, indirect immunofluorescence staining of COS cells expressing U1-70K protein fragments indicated the presence of two regions of the protein that are able to accumulate in the nucleus: one contains amino acids 92 to 199, and the other contains amino acids 183 to 437. However, shorter fragments such as fragments 161-216 (not shown), d13 (183-260), and d8 (280-437) failed to accumulate in the nucleus.

FIG. 5. RRM loop 5 of U1-70K is not required for U1 RNA binding. Mobility shift of ^{35}S -labelled U1-70K protein (18, 30), demonstrating that basic residues R-190 and R-191 in the RRM are required for U1 RNA binding, while the RRM loop 5 is not required. Experiments used the wild-type RNA-binding domain 92-216 (18, 30) or point mutants in the RNA-binding domain. Lanes 1 and 2, wild-type residues 92 to 216; lanes 3 and 4, mutant M1; lanes 5 and 6, M2; lanes 7 and 8, M3; lanes 9 and 10, M4; lanes 11 and 12, M5; lanes 13 and 14, M6; lanes 15 and 16, M0. Unlabelled RNAs used for binding were in vitro-transcribed U1 RNA SL1 (even-numbered lanes) or no RNA added (odd-numbered lanes) prior to electrophoresis in nondenaturing gels. B, bound; U, unbound.

Evidence that U1-70K does not require U1 RNA binding to enter the nucleus has been obtained by two independent experiments. First, the d7 and d12 fragments both accumulated in the nucleus yet could not be detected in association with U1 snRNP by immunoprecipitation. Second, the U1-70K 92-216 fragment with triple mutations in its NLS retained a functional U1 RNA-binding domain yet failed to accumulate in the nucleus.

Detailed analysis of the 92-199 fragment has shown the presence of a segmented or interdependent NLS, similar in nature to known NLSs. NLSs are present on many karyophilic proteins and generally contain a short sequence of basic residues (10). NLSs have been found at various locations within regions of primary structure, although they appear to consistently reside on the surface of a folded protein (34). Some proteins contain only one NLS (16), while others can harbor several NLSs that can function independently of one another (32). Examples of interdependent or segmented NLSs have also been described (19, 28, 33).

Interestingly, the NLS that we have characterized on the d11 fragment is segmented and is situated within the defined RNA-binding domain which encompasses an RRM. A general model for the RRM tertiary structure has been deduced from crystallographic and nuclear magnetic resonance data of the U1-A RNA-binding domain and can be extended to the U1-70K protein. This model indicates that part of the NLS of the U1-70K protein resides within a structural loop that connects an α helix 2 and a β strand 4 (Fig. 7A). The presence of an NLS in the RNA-binding domain of U1-70K indicates that this domain can harbor residues involved in functions other than RNA binding. For example, protein-protein interactions of the NLS with putative NLS-binding proteins are responsible for shuttling karyophilic proteins to the nuclear pore complex (37). Thus, the size of the RRM (80 amino acids) is large enough to allow multiple macromolecular interactions. This is the second example of a region defined within an RRM which is involved in protein-protein interactions. Association

FIG. 6. Nuclear transport of U1-70K protein is independent of U1 snRNP particle transport. The RNA-binding domain containing the NLS loop 5 mutant (g10-U1-70K with 92-216 fragment and F-169, P-170, and P-173) was overexpressed in COS cells, and its subcellular location was analyzed by indirect immunofluorescence. (a) phase-contrast microscopy; (b) indirect-immunofluorescence microscopy.

between U2-A' and loop 3 of the first RRM of U2-B' has been shown to increase U2 RNA binding affinity of U2-B' by at least 2 orders of magnitude (2, 36). While the interaction of the U2 RNA with the U2-B' protein occurs concomitantly with the interaction of the U2A' protein with the U2-B' protein, it is possible that this differs in the case of U1-70K. Our mutagenesis studies indicate that certain arginine residues belonging to the NLS are also involved in U1 RNA binding. This opens the possibility that association of U1-70K protein with the putative NLS-binding proteins might exclude the interaction of U1-70K with U1 RNA and possibly with U1 snRNPs in the cytoplasm.

Since RRMs are present in a large number of proteins, some of which are nuclear, loop 5 sequences of known nuclear proteins can be compared to determine whether similar residues reside in this region of other RNA binding domains. As shown in Fig. 7B, it appears that none of the known RRM family proteins have strongly similar loop 5 sequences, suggesting that the NLS in loop 5 is unique to the U1-70K protein (Fig. 7B). Only loop 5 of the first RRM of Sex lethal has three basic residues, two being at the same positions as those of U1-70K. Whether the Sex lethal loop 5 can function as an NLS is not known.

The C-terminal half of the U1-70K protein (183-437 fragment) which contains numerous highly charged repeated motifs also has the ability to accumulate in the nucleus. Li and Bingham (22) had previously demonstrated that similar Arg-rich regions in *su(w^a)* are involved in nuclear localization. Our attempts to dissect this region of U1-70K showed that shorter deletion fragments remained in the cytoplasm (d8 [280-437] and d13 [183-260]). Since d8 and d13 failed to accumulate in the nucleus, this suggested two possibilities. Either the sequence between the residues 260 and 280 contains a single

B

U1 snRNP 70K	human	165	D	G	K	K	I	D	G	R	R	173
U1 snRNP 70K	frog		D	G	K	K	I	D	G	R	R	
U1 snRNP 70K	fly		D	G	K	K	I	D	G	R	R	
U1 snRNP A #1	human		Q	G	F	P	F	Y	D	K	P	
U1 snRNP A #2	human		Q	G	F	K	I	T	Q	N	N	- A
U1 snRNP B'' #1	human		Q	G	F	P	F	Y	G	K	P	
U1 snRNP B'' #2	human		Q	G	F	K	I	T	P	S	N	- A
hnRNP A1 #1	bovine		P	H	K	V	D	G	R	V	V	- A
hnRNP A1 #2	bovine		Y	H	T	V	M	G	M	M	G	- E
hnRNP A2/B1 #1	human		P	H	S	I	D	G	R	V	V	- E
hnRNP A2/B1 #2	human		Y	H	T	I	N	G	H	N	A	- E
hnRNP C1C2	human		D	G	R	M	I	A	G	Q	V	
hnRNP L #1	human		A	D	N	Q	I	Y	I	A	G	H P
hnRNP L #2	human		N	G	A	D	I	Y	S	G	C	T
hnRNP L #3	human		N	N	N	F	M	F	G	Q	G	
pPTB #1	human		T	S	V	T	P	V	L	R	G	Q P
pPTB #2	human		D	G	Q	N	I	Y	N	A	C	T
pPTB #3	human		N	G	H	K	L	H	G	K	P	
nucleolin #1	murine		T	G	L	K	V	P	G	N	E	
nucleolin #2	murine		Q	G	A	E	I	D	G	R	S	
nucleolin #3	murine		N	K	N	E	I	E	G	R	T	
nucleolin #4	murine		E	D	G	E	I	D	G	N	G	
ASF/SF2	human		R	G	Y	D	Y	D	G	Y	R	
Sxl #1	fly		N	G	I	T	V	R	N	K	R	
Sxl #2	fly		N	N	V	I	P	E	G	G	S	Q P
tra-2	fly		S	G	I	E	V	D	G	R	R	
U2AF ⁶⁵ #1	human		G	I	I	F	Q	G	Q	S		
U2AF ⁶⁵ #2	human		N	G	M	Q	L	G	D			
U2AF ⁶⁵ #3	human		T	G	R	K	F	A	N			
SC35	human		D	G	R	E	L	R	V	Q	M	

FIG. 7. Presence of a nuclear localization element within loop 5 of the RNA-binding domain of U1-70K protein in comparison with other RRM family members. (A) Model of tertiary structure of the U1-70K RNA-binding domain based upon nuclear magnetic resonance analysis of U1-A protein (14) showing the location of the NLS sequences in loop 5. (B) Sequence comparison of the structural loop 5 (for a review, see reference in 18) in several nuclear RRM-containing proteins. The basic amino acid residues in loop 5 are shown in boldface type.

NLS or the NLS is complex and interspersed throughout this region. In the first case, examination of the residues present between 260 and 280 shows two conserved repeats with the common motif babaRbRRRSRSRab. This motif is also present in the d8 and d13 mutants but neither was able to enter the nucleus. With respect to the second possibility, the U1-A protein has been shown to contain a large NLS, spanning 110 amino acids which could not be dissected further (17). These findings with U1-70K and U1-A proteins are both consistent with the presence of a multicomponent NLS which may be dispersed over a larger region. Thus, we favor the second possibility, but new methods will be required to define this function precisely.

Given that this region is known to interact with other proteins, it is also possible that U1-70K protein is being shuttled to the nucleus as an RNP complex. Kohtz et al. (20) have recently demonstrated that the arginine-rich regions of the U1-70K protein can interact in vitro with the Arg- and Ser-rich domain of the SR splicing factors. These and other studies have suggested a role for these SR proteins in splice site selection and alternative splicing (20, 40).

In conclusion, our experiments demonstrate the presence of a segmented NLS in the U1 RNA-binding domain of the U1-70K protein. Some of the residues belonging to the NLS reside in loop 5 of the U1-70K RRM. Further experiments indicated that the nuclear transport of U1-70K is independent of U1 snRNP associated. These results open the possibility that U1-70K may function independently of U1 snRNP in the nucleus.

ACKNOWLEDGMENTS

We thank Bryan Cullen, George Leser, Terence Martin, and Laurence Rimsky for advice; Daniel Kenan for anti-g10 antibody and Ameeta Kelekar and Dragana Antic for comments on the manuscript.

This work was supported by research grant CA60083 from the NIH to J.D.K. J.M.-J.R. was supported by Neurobehavioral Training grant 5T32MH5177 and Viral Oncology Training grant 5T32CA09111. D.H.G. was a Howard Hughes Medical Student Fellow.

REFERENCES

- Bandziulis, R. J., M. S. Swanson, and G. Dreyfuss. 1989. RNA-binding proteins as developmental regulators. *Genes Dev.* 3:431-437.
- Bentley, R. C., and J. D. Keene. 1991. Recognition of U1 and U2 small nuclear RNAs can be altered by a 5-amino-acid segment in the U2 small nuclear ribonucleoprotein particle (snRNP) B' protein and through interactions with U2 snRNP-A' protein. *Mol. Cell. Biol.* 11:1829-1839.
- Billings, P. B., and S. O. Hoch. 1984. Characterization of U small nuclear RNA-associated proteins. *J. Biol. Chem.* 259:12850-12856.
- Chou, T.-B., Z. Zachar, and P. M. Bingham. 1987. Developmental expression of a regulatory gene is programmed at the level of splicing. *EMBO J.* 6:4095-4104.
- Cullen, B. R. 1986. Trans-activation of human immunodeficiency virus occurs via a bimodal mechanism. *Cell* 46:973-982.
- Cullen, B. R. 1987. Use of eukaryotic expression technology in the functional analysis of cloned genes. *Methods Enzymol.* 152:684-704.
- Feeney, R. J., R. A. Sauterer, J. L. Feeney, and G. W. Zieve. 1989. Cytoplasmic assembly and nuclear accumulation of mature small nuclear ribonucleoprotein particles. *J. Biol. Chem.* 264:5776-5783.
- Fischer, U., and R. Luhrmann. 1990. An essential signaling role for the m³G cap in the transport of U1 snRNP to the nucleus. *Science* 249:786-790.
- Fischer, U., V. Sumpter, M. Sekine, T. Satoh, and R. Luhrmann. 1993. Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m³G cap. *EMBO J.* 12:573-583.
- Garcia-Bustos, J., J. Heitman, and M. N. Hall. 1991. Nuclear protein localization. *Biochim. Biophys. Acta* 1071:83-101.
- Green, M. R. 1986. Pre-mRNA splicing. *Annu. Rev. Genet.* 20:671-708.
- Hamm, J., E. Darzynkiewicz, S. M. Tahara, and I. W. Mattaj. 1990. The trimethylguanosine cap structure of U1 snRNA is a component of a bipartite nuclear targeting signal. *Cell* 62:569-577.
- Harper, D., L. D. Fresco, and J. D. Keene. 1992. RNA binding specificity of a *Drosophila* snRNP protein that shares sequence homology with mammalian U1-A and U2-B' proteins. *Nucleic Acids Res.* 20:3645-3650.
- Hoffman, D. W., C. C. Query, B. L. Golden, S. W. White, and J. D. Keene. 1991. RNA-binding domain of the A protein component of

- the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc. Natl. Acad. Sci. USA* **88**:2495–2499.
15. Jantsch, M. F., and J. G. Gall. 1992. Assembly and localization of the U1-specific snRNP C protein in the amphibian oocyte. *J. Cell Biol.* **119**:1037–1046.
 16. Kalderon, D., W. D. Richardson, A. F. Markhan, and A. E. Smith. 1984. Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature (London)* **311**:33–38.
 17. 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.
 18. Kenan, D. J., C. C. Query, and J. D. Keene. 1991. RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.* **16**:214–220.
 19. Kleinschmidt, J. A., and A. Seiter. 1988. Identification of domains involved in nuclear uptake and histone binding of protein N1 of *Xenopus laevis*. *EMBO J.* **7**:1605–1614.
 20. Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, and J. L. Manley. 1994. Protein-protein interaction and 5' splice site recognition in mammalian mRNA precursors. *Nature (London)* **368**:119–124.
 21. Lehmeier, T., K. Foulaki, and R. Luhrmann. 1990. Evidence for three distinct D proteins, which react differentially with anti-Sm autoantibodies, in the cases of the major snRNPs U1, U2, U4/U6 and U5. *Nucleic Acids Res.* **18**:6475–6484.
 22. Li, H., and P. M. Bingham. 1991. Arginine/serine-rich domains of the *su(w^h)* and *tra* RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. *Cell* **67**:335–342.
 23. Luhrmann, R. 1988. SnRNP proteins, p. 71–99. *In* M. L. Birnstiel (ed.), *Structure and function of major and minor small nuclear ribonucleoprotein particles*. Springer-Verlag, New York.
 24. Lutz-Freyermuth, C., C. C. Query, and J. D. Keene. 1990. Quantitative determination that one of the two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem loop II of U1 RNA. *Proc. Natl. Acad. Sci. USA* **87**:6393–6397.
 25. Mattaj, I. W. 1988. U snRNP assembly and transport, p. 110–114. *In* M. L. Birnstiel (ed.), *Structure and function of major and minor small nuclear ribonucleoprotein particles*. Springer-Verlag, New York.
 26. McLean, I. W., and P. K. Nakane. 1974. Periodate-lysine-paraformaldehyde fixative: a new fixative for immunoelectron microscopy. *J. Histochem. Cytochem.* **22**:1077–1083.
 27. Nagai, K., C. Oubridge, T. H. Jessen, J. Li, and P. R. Evans. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature (London)* **348**:515–520.
 28. Nath, S. T., and D. P. Nayak. 1990. Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). *Mol. Cell. Biol.* **10**:4139–4145.
 29. Petterson, I., M. Hinterberger, T. Mimori, E. Gottlieb, and J. A. Steitz. 1984. The structure of mammalian small nuclear ribonucleoproteins. *J. Biol. Chem.* **259**:5907–5914.
 30. Query, C. C., R. C. Bentley, and J. D. Keene. 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**:89–101.
 31. Query, C. C., R. C. Bentley, and J. D. Keene. 1989. A specific 31-nucleotide domain of U1 RNA directly interacts with the 70K small nuclear ribonucleoprotein component. *Mol. Cell. Biol.* **9**:4872–4881.
 - 31a. Query, C. C., and J. D. Keene. Unpublished data.
 32. Richardson, W. D., B. L. Roberts, and A. E. Smith. 1986. Nuclear location signals in polyoma virus large-T. *Cell* **44**:77–85.
 33. Robbins, J., S. M. Dilworth, R. A. Laskey, and C. Dingwall. 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**:615–623.
 34. 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.
 35. Rosenberg, A. H., B. N. Lade, D.-S. Chui, S. W. Lin, J. J. Dunn, and F. W. Studier. 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* **56**:125–135.
 36. Scherly, D., W. Boelens, N. A. Dathan, W. J. van Venrooij, and I. W. Mattaj. 1990. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B' and their cognate RNAs. *Nature (London)* **345**:502–506.
 37. Silver, P. A. 1991. How proteins enter the nucleus. *Cell* **64**:489–497.
 38. Steitz, J. A., D. L. Black, V. Gerke, K. A. Parker, A. Kramer, D. Frendewey, and W. Keller. 1988. Functions of abundant U-snRNPs, p. 115–154. *In* M. L. Birnstiel (ed.), *Structure and function of major and minor small nuclear ribonucleoprotein particles*. Springer-Verlag, New York.
 39. Tsai, D. S., D. J. Kenan, and J. D. Keene. 1992. In vitro selection of an RNA epitope immunologically cross reactive with a peptide. *Proc. Natl. Acad. Sci. USA* **89**:8864–8868.
 40. Wu, J. Y., and T. Maniatis. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**:1061–1070.
 41. Zahler, A. M., W. S. Lane, J. A. Stolk, and M. B. Roth. 1992. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* **6**:837–847.