A Novel Set of Nuclear Localization Signals Determine Distributions of the αCP RNA-Binding Proteins

Alexander N. Chkheidze and Stephen A. Liebhaber*

Departments of Genetics and Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104

Received 10 June 2003/Returned for modification 22 July 2003/Accepted 29 August 2003

αCPs comprise a subfamily of KH-domain-containing RNA-binding proteins with specificity for C-rich pyrimidine tracts. These proteins play pivotal roles in a broad spectrum of posttranscriptional events. The five major αCP isoforms are encoded by four dispersed loci. Each isoform contains three repeats of the RNA-binding KH domain (KH1, KH2, and KH3) but lacks other identifiable motifs. To explore the complexity of their respective functions, we examined the subcellular localization of each αCP isoform. Immunofluorescence studies revealed three distinct distributions: αCP1 and αCP2 are predominantly nuclear with specific enrichment of αCP1 in nuclear speckles, αCP3 and αCP4 are restricted to the cytoplasm, and αCP2-ΔL, an αCP2 splice variant, is present at significant levels in both the nucleus and the cytoplasm. We mapped nuclear localization signals (NLSs) for αCP isoforms. αCP2 contains two functionally independent NLS. Both NLSs appear to be novel and were mapped to a 9-amino-acid segment between KH2 and KH3 (NLS I) and to a 12-amino-acid segment within KH3 (NLS II). NLS I is conserved in αCP1, whereas NLS II is inactivated by two amino acid substitutions. Neither NLS is present in αCP3 or αCP4. Consistent with mapping studies, deletion of NLS I from αCP1 blocks its nuclear accumulation, whereas NLS I and NLS II must both be inactivated to block nuclear accumulation of αCP2. These data demonstrate an unexpected complexity in the compartmentalization of αCP isoforms and identify two novel NLS that play roles in their respective distributions. This complexity of αCP distribution is likely to contribute to the diverse functions mediated by this group of abundant RNA-binding proteins.

Posttranscriptional controls play a major role in the regulation of eukaryotic gene expression (24, 65). These controls (i) can increase the complexity of nuclear RNAs via alternative splicing and editing, (ii) can modulate information flow from the nucleus to cytoplasm, and (iii) can alter levels and sites of protein synthesis via controls over mRNA stability, translation efficiency, and subcellular localization (4, 56, 68). RNA-binding proteins that mediate these controls can be categorized based on the presence of one or more conserved RNA-binding motifs (for reviews, see references 6 and 37). The sequence specificity of these proteins, the identities of their RNA targets, and the respective mechanisms of action are therefore of significant interest.

Studies from our laboratory and others have focused on the structures and actions of a subfamily of RNA-binding proteins, the αCPs (31, 39). These proteins, also referred to as PCBP1 (17) and hnRNP Es (34, 58), contain a triplication of the KH domain (43, 69). The 70-amino-acid KH domain comprises a triple-β-sheet platform supporting three α-helical segments (35, 36, 50, 51). Cocystal structures reveal that the KH domain can interact in a highly specific manner with four to five contiguous bases in a target RNA (5, 27). Two KH domain subtypes have been identified: the type 1 KH domain (e.g., KH3 of hnRNP K) has a C-terminal β extension, and the type 2 KH domain (e.g., ribosomal protein S3) contains an N-terminal αβ extension (21). The KH domains in the αCPs are type 1 (40). KH domains are often represented in proteins in multiple copies. Since each KH domain has the potential to independently interact with a target RNA sequence, the complexity and specificity of RNA interaction for these proteins can be quite high (66, 74; our unpublished data).

Our laboratory has focused on the role of αCPs in mRNA stabilization. These studies have defined a cytosine (C)-rich cis-acting stability element within the 3′ untranslated region (UTR) of the human (h) α-globin mRNA that serves as a binding site for αCP both in vitro and in vivo (28, 32, 83). Studies in our laboratory and others have identified closely related αCP binding sites in the 3′ UTRs of additional highly stable mRNAs (7, 10, 25, 61, 72, 86). In the case of the α-globin mRNA, the RNP “α-complex” in the 3′ UTR appears to represent a 1:1 interaction of αCP with the C-rich motif (8), although there is evidence that additional proteins may also bind at this site (30, 63, 84, 85). These data suggest that the α-complex may represent a widely distributed and general determinant of mRNA stabilization (25).

The αCPs involved in α-complex assembly can represent several isoforms (8, 31, 39). Whether these various αCP isoforms are performing unique functions or are redundant in their actions is not known. The functions mediated by αCP RNP complexes are in fact quite diverse (for review, see reference 41). As mentioned above, assembly of the 3′ UTR α-complex in α-globin mRNA mediates mRNA stabilization (32, 82, 83). This stabilization may reflect direct steric protection of the 3′ UTR from endonucleolytic attack (79), and/or it may reflect protection of the poly(A) tail from rate-limiting decay via interactions in cis between the bound αCP and the poly(A)-binding protein (30, 49, 78). αCPs also mediate translational controls. An array of αCP binding sites within the 3′ UTR of the 15-lipoxygenase mRNA has been linked to devel-
opmentally regulated translational repression during erythroid maturation (56–58). In contrast, association of αCP with the 5' UTR of the polio viral RNA serves as an enhancer of internal ribosome entry site-mediated translation (2, 3). αCP binding within the 3' UTR has also been implicated in the activation of maternal mRNA translation in early embryonic development in *Xenopus* via control of cytoplasmic polyadenylation (59).

Additional systems are reported to involve αCP binding in the control of various aspects of mRNA expression (63, 84, 85; reviewed in reference 41). Thus, the targets and actions of the αCPs are quite diverse and may reflect the actions of one or more of the defined αCP isoforms.

αCP isoforms are encoded by four unlinked loci in the human and mouse genomes: *PCBP1*, *PCBP2*, *PCBP3*, and *PCBP4* (38, 39, 75) (see also Fig. 1, left). Each locus has been mapped, sequenced, and characterized for mRNA structure (38, 39). A total of five major αCP isoforms have been identified in human or mouse tissues: αCP1, αCP2, αCP3, αCP4, and a major αCP2 splice variant, αCP2-KL, that differs from αCP2 by the exclusion of a 31-amino-acid segment in the region between the KH2 and KH3 encoded by a single exon (exon 8a) (17, 38, 39). These proteins are broadly expressed in human and mouse tissues and demonstrate polyC-binding specificity (34, 38, 39; unpublished observations). αCP1 and αCP2 share the highest level of amino acid sequence similarity at 89% (75), αCP3 is more divergent, and αCP4 is the most distantly related (52% divergence from αCP2 [39]). Each protein contains three similarly spaced KH domains; two KH repeats are located in the N terminus followed by a nonconserved region of variable length, and the third KH domain is located at the C terminus. Posttranslational modifications may regulate the binding of αCPs to RNA. For example, phosphorylation of αCP1 and αCP2 results in a marked decrease in RNA-binding activity (34). An additional major determinant of αCP isoform function may relate to subcellular localization. Although the defined roles of αCP in mRNA stability and translational control suggest a cytoplasmic localization, prior attempts to sublocalize αCPs did not resolve the question of whether αCP is in fact nuclear or cytoplasmic (17, 18).

Protein transport from cytoplasm to nucleus is mediated by multiple families of soluble factors. It is generally accepted that nuclear proteins carry a nuclear localization signal (NLS) (11, 45, 53, 67). NLSs vary considerably from relatively short stretches of residues to large protein domains with relaxed sequence conservation. Although different in structure, NLSs all seem to play the same role: recognition by soluble factors that mediate transport through the nuclear pore complexes (19, 20, 42, 54, 55, 77).

In the present study each of the five major isoforms was individually assessed for subcellular localization, and the signals underlying their nuclear versus cytoplasmic distribution were defined. These data demonstrate an unexpected complexity in αCP isoform compartmentalization based on a novel set of nuclear localization motifs.

**MATERIALS AND METHODS**

**Plasmids.** A pcDNA1 (Clontech)-based eukaryotic expression vector that encodes pyruvate kinase (Pk) with a c-myc epitope tag (12) at its amino terminus and a polylinker at its carboxyl terminus was used for expression studies (a gift from G. Dreyfuss, University of Pennsylvania) (45). The *myc*-Pk-NLS plasmid (a gift from M. Malim, Guy’s Hospital), containing a 17-amino-acid bipartite basic NLS of hrNRP K protein (45) at the carboxy terminus of Pk has been described elsewhere (13) and was used as a positive control in expression studies. In the cases of αCP2-KL, αCP3, and αCP4A, PCR-amplified fragments corresponding to the coding region of the respective genes or subdomains were inserted into a plasmid vectors as Xhol-XbaI fragments. Given the internal XbaI site for αCP1 and αCP2 and compatibility between XbaI and SpeI ends, αCP1/2 full-length coding regions were cloned as *XhoI*-SpeI fragments into the same parental vector, pBluescript II (a gift from S. Smale, University of California at Los Angeles) (22), pGibD-αCP1 and pGibD-αCP2 (gifts from M. Kulikjed, Rutgers University) (30), and pET-αCP1 and pET-αCP4A were used as PCR templates to obtain a series of myc-tagged Pk-αCP fusion plasmids: myc-Pk-αCP2-KL, myc-Pk-αCP1, myc-Pk-αCP2, myc-Pk-αCP3, myc-Pk-αCP4A, respectively. Deletion mutants of myc-Pk-αCP1, myc-Pk-αCP2, and myc-Pk-αCP2-KL were generated by targeted PCR. The corresponding primers contained 5' XhoI site and a 3' XbaI site, and the amplified fragments after restriction nuclease digestion were inserted into myc-Pk. Deletion mutants contained the following αCP amino acid sequences. For the N-terminal deletion series (Nd), mutants contained sequences from amino acids 17 to 331, 36 to 331, 330 to 331, 257 to 331, and 306 to 331 of αCP-KL (see Fig. 4A and B) and 284 to 356 of αCP1 (see Fig. 7). For the C-terminal deletion series (Cd), mutants contained sequences from amino acids 1 to 17, 1 to 63, 1 to 101, 1 to 150, and 1 to 306 of αCP-KL (see Fig. 5), 1 to 150 of αCP1 (see Fig. 7), and 1 to 306 of αCP-2 (see Fig. 6C and D). For the internal segment series (Is), mutants contained sequences from amino acids 150 to 256 and 257 to 306 of αCP-2-KL (see Fig. 5); 151 to 290, 151 to 228, 292 to 290, 270 to 290, 290 to 290, and 270 to 279 of αCP-2 (see Fig. 6A and B); 291 to 339, 291 to 334, 318 to 339, 332 to 339, and 328 to 339 of αCP-2 (see Fig. 6C and D); and 151 to 283 of αCP-1 (see Fig. 7).

Codons encoding T at position 331 and D at position 332 of αCP2 in Pk-αCP2 were mutated by using a QuickChange kit (Stratagene) to S and G, respectively, myc-Pk-αCP1/ΔNLSI, myc-Pk-αCP2/ΔNLSI, and myc-Pk-αCP2/ΔNLSI/mutNLSI were constructed by spliced-overlap extension by using myc-Pk-αCP1, myc-Pk-αCP2, and myc-Pk-αCP2/ΔNLSI as templates. Two segments of a corresponding template gene were PCR amplified independently and then fused in a subsequent reaction. The amplified fragments were inserted as *XhoI*-SpeI fragments into the myc-Pk vector.

**Cell culture and transfection.** HeLa cells were cultured in Dulbecco modified Eagle medium (Gibco-BRL) supplemented with 10% fetal bovine serum, 100 μg of streptomycin per ml, and 100 U of penicillin per ml. Cells were seeded into Falcon 2 chamber tissue culture glass slides (Becton Dickinson Labware, Franklin Lakes, N.J.) at a density of 1 × 10^5 to 2 × 10^5 in 20 h prior to transfection. Transfection of cells was performed by calcium phosphate transfection (5 Prime–3 Prime, Inc., Boulder, Colo.). Medium containing the DNA mixture was removed 18 to 20 h after transfection and replaced with fresh medium. The time when the DNA was added was considered *T₀* and cells were fixed for immunofluorescence at 24 to 48 h.

**Microscopy.** Indirect immunofluorescence microscopy was carried out as described previously (9). After transfection, the cells were washed in 1× phosphate-buffered saline (PBS) and fixed in 1× PBS plus 3% paraformaldehyde (EM Science, Gibbstown, N.J.) for 30 min at room temperature. After fixation the cells were washed in PBS, permeabilized by the addition of 1× PBS–10% goat serum–1% Triton X-100 at room temperature for 10 min, and washed in 1× PBS. The cells were then blocked in antibody blocking-inubcation buffer (ABB; 1× PBS, 10% goat serum, 0.1% Tween) for 60 min at room temperature prior to primary antibody incubation. Monoclonal anti-αCP antibody 9E10 (Santa Cruz Biotechnology, Inc., Santa Cruz, Calif.) was then diluted 1:1,000 in ABB and applied to the cells at room temperature for 60 min. The cells were then washed and incubated with fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse secondary antibody (Kirkegaard & Perry Laboratories, Gaithersburg, Md.) diluted at 1:100 in ABB for 20 min. The cells were then washed and glass slides were mounted with mounting medium (ProLong Antifade kit; Molecular Probes, Eugene, Ore.) and sealed with nail polish.

For detection of endogenous αCPs and for colocalization experiments, cells were fixed in 1× PBS plus 2% paraformaldehyde for 15 min at room temperature and incubated serially with anti-αCP antibodies (F11, FF2, and FF3) (8) and with a mouse monoclonal anti-SC35 antibody (1:150; Pharmingen, San Diego, Calif.), followed by incubation with FITC-conjugated goat anti-rabbit and Texas red (TX)-conjugated anti-mouse antibodies (1:100; Kirkegaard & Perry) in ABB for 60 min. Samples were examined by using a Leitz DMR microscope (Leica GmbH, Wetzlar, Germany), and images were captured by using a Hamamatsu Orca Hg4 cooled charge coupled device (CCD) camera equipped with Slideview software (Hamamatsu, Bridgewater, N.J.). Image analysis was performed by using Adobe Photoshop software (Adobe Systems, Inc., San Jose, Calif.). Optical sections were obtained by using a con-
focal laser scanning microscope (Leica). Fluoview was operated at excitation wavelengths of 488 nm (FITC) and 568 nm (TRITC [tetramethyl rhodamine isothiocyanate]) from an argon-krypton laser. Fluorescent signals of both fluorochromes were recorded simultaneously by two detectors.

**RESULTS**

**αCP proteins are present in both nucleus and cytoplasm.** Initial studies of αCP subcellular localization focused on the most abundantly expressed isoforms: αCP1, αCP2, and αCP2-KL (Fig. 1). HeLa cells were stained with antisera specific to αCP1 and to αCP2 (antisera FF1 and FF2, respectively) and a third antibody raised against an epitope common to αCP2 and αCP2-KL but absent from αCP1 (antisera FF3) (8). The specificities of these antisera were confirmed by Western blotting (8). In previous studies we demonstrated that the first two antisera, when used against mouse and human cell extracts, recognized a single band (αCP-1 and αCP-2, respectively), whereas the third antisera revealed a band comigrating with αCP-2 and an additional, smaller band. The size of this additional band was consistent with that of αCP2-KL. The first two antisera (FF1 and FF2) revealed exclusive nuclear localization of αCP1 and αCP2. In contrast, the third antisera (FF3) revealed dual nuclear and cytoplasmic staining. Since full-length αCP2 is exclusively nuclear (FF2 staining), this dual-staining pattern with the FF3 antibody indicates that at least a portion of the αCP2-KL is cytoplasmic (see below for further analysis). Direct analyses of endogenous αCP3 and αCP4 isoforms were similarly attempted. However, in these two cases, the isoform-specific antisera failed to give unambiguous signals. The distributions of these proteins were addressed by detection of epitope-tagged fusion proteins (see below). We conclude from these data that the major isoforms of αCP have distinct subcellular distributions.

**αCP1 is enriched in nuclear speckles.** The intracellular localization of endogenous αCP1 and αCP2 were further characterized by confocal microscopy (Fig. 2). Staining with the αCP1-specific antibody revealed that the nuclear protein was
FIG. 3. Expression and subcellular localization of αCP isoforms as fusion proteins in HeLa cells. (A) Schematic drawings of the Pk-αCP isoform fusion proteins. The chicken muscle Pk (black ellipse) linked to an N-terminal myc epitope tag (open box) was fused in a continuous ORF corresponding to each of the indicated αCP isoforms (shaded boxes). The bipartite basic NLS of hnRNP K is denoted by the stippled oval. (B) Western analysis of fusion proteins expressed in HeLa cells. Expression vectors containing each of the indicated proteins (see panel A for schematic details) were transfected into HeLa cells, and extracts were harvested at 36 h. Proteins were analyzed by Western analysis with an anti-myc antibody (FITC). Cells in the field were visualized by phase-contrast imaging, and the nuclei were identified by DAPI staining.

Subcellular localization of epitope-tagged αCP isoforms reveals three distribution patterns. Localization of αCP1 and αCP2 to the nucleus was unexpected since these proteins lacked identifiable nuclear import signals (53). In addition, prior studies had identified poly(C)-binding activity in cytosolic extracts (31, 80). Thus, it was likely that the cytosolic poly(C)-binding activity might reflect αCP2-KL, αCP3, and/or αCP4 content. To explore these possibilities, full-length open reading frames (ORFs) encoding each of the five major αCP isoforms were isolated and fused in frame to the ORF encoding the 55-kDa Pk and an N-terminal myc epitope tag (Fig. 3A). The high molecular weight of the myc-Pk-αCPs fusion proteins prevents nuclear entry by passive diffusion; their nuclear sub-localization should accurately reflect active transport. Control constructs included the myc-Pk tag alone and the myc-Pk tag fused to a known NLS. The resultant cDNAs were expressed in transfected HeLa cells. Cell extracts were analyzed by immunoblotting to confirm size, amount, and immunoreactivity of each expressed fusion protein (Fig. 3B). Localization of each fusion protein was then determined by immunostaining with a myc epitope monoclonal antibody (Fig. 3C). myc-Pk was appropriately restricted to the cytoplasm and was efficiently targeted to the nucleus when fused to the 18-amino-acid bipartite basic NLS from hnRNP K (Pk-NLS; Fig. 3C) (64). The recombinant myc-Pk-tagged αCP1 and αCP2 proteins were both confined to the nucleus. This was consistent with the analysis of the native proteins (Fig. 1 and 2). Analysis of the epitope-tagged αCP2-KL demonstrated that this isoform is present in both the nucleus and the cytoplasm. In contrast, the αCP3 and αCP4 fusion proteins were both confined to the cytoplasm. These data confirmed and extended the analysis of endogenous proteins and demonstrated three distinct patterns of nuclear/cytoplasmic compartmentalization: αCP1 and αCP2 are predominantly nuclear, αCP2-KL is both nuclear and cytoplasmic, and αCP3 and αCP4 are cytoplasmic.

Mapping NLSs. The αCP1, αCP2, and αCP2-KL isoforms are all present in the nucleus. However, inspection of the primary sequences failed to reveal matches with the bipartite-
basic or simian virus 40 large T-type NLS (11), the A1 M9 domain (45, 67), or other NLSs reported at lower frequencies (53). The sequence determinant(s) responsible for the nuclear localization therefore appeared to represent novel structures and were mapped by a functional assay (12, 13, 45). 

**FIG. 4.** Mapping nuclear import signals in αCP2-KL (A) N-terminal deletion set. The three KH domains are denoted by shaded boxes; numbers above the protein body indicate its size and the positions of KH domains in amino acids. The N and C terminus of each expressed protein is indicated along with its name. The positions of the KH domains are shown for reference. Other symbols are as defined in Fig. 3. (B) Subcellular localization of the αCP-KL Nd proteins. Representative micrographs of the immunofluorescence analyses with anti-myc are shown. The identity of each expressed protein is indicated below the frame. (C) C-terminal deletion set. Details are as described for panel A. (D) Subcellular localization of the αCP-KL Cd proteins. Details are as described for panel B.

The combined results of the N- and C-terminal truncation sets could not be reconciled on the basis of a single NLS. The most parsimonious model comprised two separate NLSs: one located in the region between KH2 and a second located within KH3 itself. This model of two distinct NLSs was tested.

**FIG. 5.** αCP-2KL contains two independent NLSs. Schematics of two internal segments of the αCP2-KL fused with myc-Pk are shown. Immunofluorescence micrographs show the subcellular localization of the indicated αCP2-KL fusion proteins expressed in transfected HeLa cells.
[Is(150-256)] and the region encompassing KH3 through the C terminus of the αCP2-KL protein [Is(257-306)] were individually tested for NLS function (Fig. 5). The segments from amino acids 150 to 256 and amino acids 257 to 306 each individually resulted in nuclear localization of the fused myc-Pk reporter. These data confirmed that αCP2-KL contains at least two separable and independent NLSs: NLS I, located in the region between KH2 and KH3, and NLS II, located within the third KH domain.

Since neither NLS I nor NLS II contained a recognizable nuclear localization motif, we mapped the region between KH2 and KH3 at a higher resolution. To maximize the information from these studies, the mapping was done on αCP2 that contains the additional 31-amino-acid intra-KH2/KH3 segment lacking in αCP2-KL. A nested set of six subfragments of this region was generated (Fig. 6A). Each fragment was fused to the myc-Pk ORF and expressed in HeLa cells. As shown in Fig. 6B, a 10-amino-acid fragment located just N terminal to KH3 domain (i.e., from amino acids 270 to 279) was sufficient for NLS function. In a similar fashion, a set of six nested fragments representing the C terminus of the αCP2 KH3 domain were tested for NLS function (Fig. 6C). The expression patterns of these fusion proteins were sufficient to localize NLS II activity to a 12-amino-acid fragment at the C terminus of the KH3 domain (Fig. 6D). Thus, two short peptides derived from αCP2 could independently target a fusion protein to the nucleus.

Selecting conservation of NLS I in αCP1. The sequences corresponding to the two NLS in αCP2 were compared to the other αCP isoforms. The sequence of NLS I in αCP1 is fully conserved in αCP1 (Fig. 7A). In contrast, the region corresponding to the αCP2 NLS II contained two divergent residues in αCP1, and even greater divergence was noted for the cytoplasmic αCP3 and αCP4 (Fig. 7B). The functional conservation of NLS motifs in αCP1 was tested. αCP1 was divided into three segments (Fig. 8A). Each segment was fused to myc-Pk, and the distribution of the expressed protein was determined in transfected cells (Fig. 8B). The fragment encompassing the region between KH2 and KH3, Is(151-283), directed recombinant protein transport to the nucleus. In contrast, the fragment encompassing KH3 and the C terminus, Nd(284-356), lacked NLS function. To clarify the functional impact of these two amino acid substitutions on NLS II function, the αCP1 peptide corresponding to NLS II was directly tested by fusing it to the myc-Pk reporter. The expressed fusion protein was limited to the cytoplasm (Fig. 6C and D). Similarly, the corresponding regions of αCP3, αCP4, and hnRNP K (Fig. 7B) were all incapable of directing myc-Pk reporter to nucleus (data not shown). These data demonstrate that αCP1 carries a single NLS (NLS I) and that αCP2 and αCP2-KL contain two inde-
NLS DETERMINE DISTRIBUTION OF αCP RNA-BINDING PROTEINS

FIG. 7. Sequence alignments of major αCP isoforms at NLS I and NLS II. (A) Alignment of the sequences corresponding to NLS I. This alignment shows perfect conservation between αCP1 and αCP2. There is no region presented in αCP3 or αCP4 that has significant homology to this sequence. (B) Alignments of the sequences corresponding to NLS II. These alignments reveal divergence among αCP2, the other αCP isoforms, and hnRNP K. Conserved amino acids are shaded.

FIG. 8. NLS activity in αCP1 is restricted to the region between KH2 and KH3. (A) Division of αCP1 into three segments for NLS mapping. The termini of each segment are noted. (B) Immunofluorescence analysis of the αCP1 subregions. myc-Pk-tagged αCP1 subregions were expressed in HeLa cells, and the subcellular distribution of each protein was determined. Representative micrographs of the immunofluorescence analysis are shown.

DISCUSSION

αCPs comprise a highly abundant subset of RNA-binding proteins. The four dispersed PCBPs encode five distinct αCP isoforms. These proteins play prominent roles in post-transcriptional control of cellular and viral mRNAs. To what extent αCP isoforms overlap in functions and/or mediate specific controls is not known. As a step in this direction, we have defined the intracellular localizations of five major αCP isoforms. In the process of these studies, we have identified a novel set of NLSs that contribute their respective distributions.

αCP1 and αCP2 localize predominantly to the nucleus. This conclusion is based on immunofluorescence analysis of endogenous proteins with isomeric-specific antisera, as well as epitope-tagged fusion proteins expressed in cells. The nuclear localization of these two αCP isoforms can be compared to two prior data sets. In one study, mCBP (mouse homolog of αCP2) appeared to be predominantly nuclear (17), whereas a second study reported cytoplasmic localization of PCBP1 (αCP1) and PCBP2 (αCP2) (18). Our present data agree with the first study. To try to clarify the discrepancy with the second cited study, we repeated analyses by using the fixation technique, cells, and staining protocols specified in that study, including the use of the same antibody preparations (gifts from R. Andino). Despite these efforts, we were unable to resolve the conflicting data since αCP1 and αCP2 were consistently detected only in the nuclear compartment.

Although both αCP1 and αCP2 are localized to the nucleus, their respective distributions are not identical. αCP1 is selectively concentrated in nuclear speckles, whereas αCP2 is more diffusely distributed (Fig. 2). Speckles, or interchromatin granule clusters (IGCs) (reviewed in references 33 and 70), represent sites at which splicing factors are concentrated prior to assembly on newly synthesized transcripts (26, 29, 48). Recently, Mintz et al. biochemically purified IGCs. Consistent with the present study, αCP1 was specifically identified as one of 75 IGC-associated proteins (47). Association of other αCP isoforms with IGCs was not observed. On the other hand,
Funke et al. presented evidence for in vivo interaction between αCP2-KL and splicing factor 9G8 (17). We were not able to detect convincing data that support a selective enrichment of αCP1 proteins. Plasmids encoding myc-Pk-αCP1 and myc-Pk-αCP1/ΔNLSI proteins were transfected into HeLa cells, and the subcellular distribution of the proteins was determined. Representative micrographs of the immunofluorescence analysis are shown. (C) Schematic of wild-type αCP2 and the mutants lacking one or both of NLS I and NLS II. A schematic of wild-type myc-Pk-αCP2 and three derivative mutants in which NLS I, NLS II, or NLS I plus NLS II were inactivated (myc-Pk-αCP2/ΔNLSI, myc-Pk-αCP2/mutNLSII, and myc-Pk-αCP2/ΔNLSI/mutNLSII, respectively) is shown. The two cross-hatched boxes represent NLS I and NLS II. The position of deleted NLS I is indicated by the gap, and the boxes with ovals depict the mutated amino acid sequences. (D) Immunofluorescence micrographs showing the subcellular localization of the wild-type αCP2 and the derivative NLS mutants. Plasmids encoding myc-Pk-tagged αCP2, myc-Pk-αCP2/ΔNLSI, myc-Pk-αCP2/mutNLSII, and myc-Pk-αCP2/ΔNLSI/mutNLSII were transfected into HeLa cells, and the subcellular distribution of the proteins was determined. Representative micrographs of the immunofluorescence analysis are shown.

The structural basis for the distinct subcellular compartmentalization of the various αCP isoforms is undoubtedly complex. The present study focuses on determinants of αCP nuclear localization. Two functional NLSs were identified in αCP2, and one of these is conserved in αCP1. Remarkably, the newly identified NLS I and NLS II have no apparent similarity to each other and bear little or no resemblance to NLSs described in the literature. Specifically, they lack similarity to either the bipartite-basic type or the simian virus 40 large T-type NLSs or to shutting signals found in hnRNP A1 (45, 67), hnRNP K (46), HuR (16, 62), and human immunodeficiency virus type 1 Rev (44, 71). Protein basic local alignment search tool (BLAST) (1) searches optimized for searching for small sequences (i.e., 1, “search for short nearly exact matches,” matrix PAM30; gap penalties: existence, 9; extension, 1) failed to identify other proteins with significant primary sequence homology. Although it is unlikely that these two NLSs are unique to the αCPs, they do appear to be uncommon. The presence of two independent NLSs in one protein is not unprecedented. As an example, hnRNP K contains a classical NLS at the N-terminal end and KNS (i.e., hnRNP K nuclear shuttling domain) between the KH2 and KH3 RNA-binding domains (46). It is interesting that hnRNP K does not contain the two NLSs in the closely related αCP1 and αCP2 (Fig. 7).
intracellular localization of the endogenous or epitope-tagged αCP1 or αCP2 (data not shown). In addition, the identified NLS I and NLS II sequences do not match to the previously described importin-α and importin-β recognition signals (23, 60, 76). It remains to be determined whether the apparently novel NLSs of αCP1 and αCP2 represent noncanonical binding sites of the currently known transport receptors, or whether they interact with novel factors of the intracellular transport.

The distribution of αCP2-KL in both nuclear and cytoplasmic compartments is of particular interest because this protein comprises the major αCP isoform in certain cell types (8, 38, 39). The structure of the nuclear αCP2 differs from that of αCP2-KL by the inclusion of a 31-amino-acid alternatively spliced segment located between the KH2 and KH3 RNA-binding domains (see Fig. 1). It is formally possible that this 31-amino-acid segment encodes an additional NLS. However, this appears unlikely since this isolated segment is unable to localize to a fused myc-Pk to the nucleus (Fig. 6B). A more plausible model would be that this segment encodes a nuclear retention signal that anchors αCP2 in the nucleus. The presence of a nuclear retention signal in hnRNP C proteins establishes a precedent for this model (52).

Visual examination of the αCP amino acid sequences reveals a candidate N-terminal Leu-rich motif between that may serve a direct nuclear export function. Although we observed that its a 31-amino-acid alternatively spliced segment located between the KH2 and KH3 RNA-binding domains (see Fig. 1). It is formally possible that this 31-amino-acid segment encodes an additional NLS. However, this appears unlikely since this isolated segment is unable to localize to a fused myc-Pk to the nucleus (Fig. 6B). A more plausible model would be that this segment encodes a nuclear retention signal that anchors αCP2 in the nucleus. The presence of a nuclear retention signal in hnRNP C proteins establishes a precedent for this model (52).

The colocalization of NLS II with the KH3 RNA-binding domain of αCP2 suggests a possible interplay between RNA-binding domain and subcellular transport. αCP2 may initially bind to target mRNAs in the nucleus during transcription and/or processing. An initial association of αCPs with target mRNAs in the nucleus is supported by the recent finding that expression of RNA decoys to αCP selectively in the nucleus results in a block in both nuclear (splicing) and cytoplasmic (translation and stabilization) activities (40; Eastmond et al., unpublished). Binding of αCP to a target mRNA in the nucleus may block NLS II which is located within an RNA-binding (KH) domain. This might facilitate redistribution of the αCP-RNP complex to the cytoplasm because both NLS I and NLS II are needed for efficient nuclear localization (Fig. 9). Once in the cytoplasm, the cytoplasmic function(s) of the αCP2-RNP complex can be achieved; the αCP2 may then dissociate from the cytoplasmic mRNA. The newly exposed NLS II, in conjunction with NLS I, could then mediate efficient reimportation of αCP2-KL to the nucleus. A similar model of RNP cycling via reversible NLS function has been suggested for HIV Rev (23). The proposed reversible RNA association of αCPs with target mRNAs may be further facilitated by posttranscriptional modifications (34). αCP may thus bind to a target mRNA in the nucleus, facilitate mRNP export, have an impact on cytoplasmic control over mRNA stability and/or translation, and then return to the nucleus. Given the present delineation of αCP localization and underlying NLS motifs, these models can now be further explored.

ACKNOWLEDGMENTS

We thank R. Andino, G. Dreyfuss, and M. Malim for generously sharing reagents used in this study. We thank Julia Morales for early contributions to this project, Brian Calvi for advice on confocal microscopy, and Lili Wan for critical review of the manuscript. This work was supported by NIH grant RO1 HL65449 and by the generosity of the Doris Duke Foundation.

REFERENCES

2. Andino, R., N. Boddeker, D. Silvera, and A. Gamarnik. 1999. Intracellular localization of the endogenous or epitope-tagged αCP1 or αCP2 (data not shown). In addition, the identified NLS I and NLS II sequences do not match to the previously described importin-α and importin-β recognition signals (23, 60, 76). It remains to be determined whether the apparently novel NLSs of αCP1 and αCP2 represent noncanonical binding sites of the currently known transport receptors, or whether they interact with novel factors of the intracellular transport.


