Molecular Biology of the Cell
Vol. 4, 1239–1250, December 1993

Determination of the Functional Domains Involved in Nucleolar Targeting of Nucleolin

Laurent Créancier,* Hervé Prats,† Catherine Zanibellato,† Francois Amalric,* and Béatrix Bugler*

*Laboratoire de Biologie Moleculaire Eucaryote, Institut de Biologie Cellulaire et Genetique, Centre National de la Recherche Scientifique, 31062 Toulouse Cedex, France; and †Laboratoire d'Endocrinologie Experimentale, Institut Louis Bugnard, Institut National de la Santé et de la Recherche Medicale, 31054 Toulouse Cedex, France

Submitted August 23, 1993; Accepted October 7, 1993

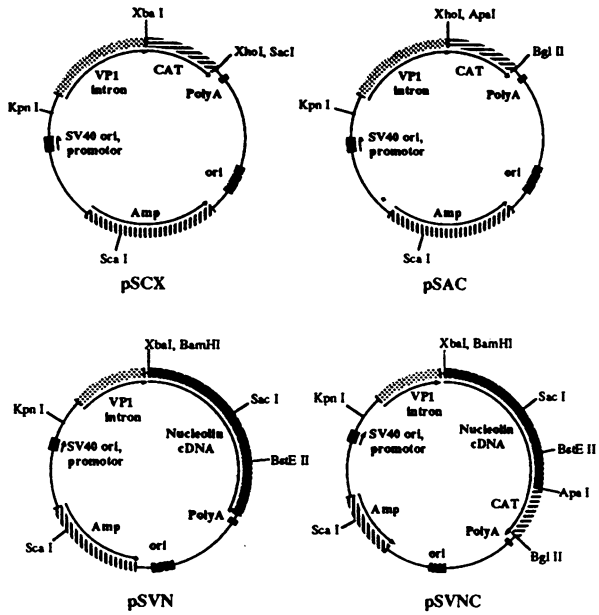
Nucleolin (713 aa), a major nucleolar protein, presents two structural domains: a N-terminus implicated in interaction with chromatin and a C-terminus containing four RNA-binding domains (RRMs) and a glycine/arginine-rich domain mainly involved in pre-rRNA packaging. Furthermore, nucleolin was shown to shuttle between cytoplasm and nucleolus. To get an insight on the nature of nuclear and nucleolar localization signals, a set of nucleolin deletion mutants in fusion with the prokaryotic chloramphenicol acetyltransferase (CAT) were constructed, and the resulting chimeric proteins were recognized by anti-CAT antibodies. First, a nuclear location signal bipartite and composed of two short basic stretches separated by eleven residues was characterized. Deletion of either motifs renders the protein cytoplasmic. Second, by deleting one or more domains implicated in nucleolin association either with DNA, RNA, or proteins, we demonstrated that nucleolar accumulation requires, in addition to the nuclear localization sequence, at least two of the five RRMs in presence or absence of N-terminus. However, in presence of only one RRM the N-terminus allowed a partial targeting of the chimeric protein to the nucleolus.

INTRODUCTION

In addition to the pre-rRNA and a subset of ribosomal proteins, the nucleolus contains proteins and RNAs that are involved in ribosome biogenesis but are not part of mature cytoplasmic ribosomes. These components can be roughly divided into factors involved in rDNA transcription (polymerases, transcription factors, and transactivators), in the maturation of preribosomes, and in the shuttling between cytoplasm and nucleolus. Immunolocalization studies permitted assignment of certain of these proteins to the various nucleolar subcompartments. The mammalian protein B23 (Spector *et al.*, 1984), its amphibian counterpart protein No38 (Peculis and Gall, 1992), and ribocharin (Hügle *et al.*, 1985) were found associated with the precursor of the large ribosomal subunit in the granular component. Fibrillarin (Ochs *et al.*, 1985; Lapeyre *et al.*, 1990) and nucleolin (Bugler *et al.*, 1982) were predominantly recovered in the dense fibrillar component. With the last characterized nucleolar protein Nopp 140 (Meier and Blobel, 1992), B23/No38 and nucleolin have been shown to shuttle between nucleolus and cytoplasm (Borer *et al.*, 1989), but little is known about the mechanisms by which these nucleolar proteins are targeted to the nucleolus.

Nucleolin/C23 is a multifunctional protein (for review see Jordan, 1987) that has been implicated in the control of the transcription of rRNA genes (Olson *et al.*, 1983; Bouche *et al.*, 1984) and in the packaging of nascent ribosomal RNA (Herrera and Olson, 1986; Bugler *et al.*, 1987). The nucleolin structure revealed a modular organization in functional domains: the N-terminal region contains alternating basic/acidic stretches associated with nucleolar chromatin in interphase (Erard *et al.*, 1988) and nucleolar organizer regions of metaphasic chromosomes (Gas *et al.*, 1985). The carboxy two-thirds consists of four RNA-binding domains (RRMs) and a glycine/arginine-rich (GAR) domain where arginines can be dimethylated (GDMA) (Lischwe *et al.*, 1985). The GAR domain interacts with RNA in an efficient but nonspecific manner and unfolds RNA secondary structure. However, it is essential for efficient specific binding

A



B

Mutant name	Matrix	Primer	Cloning plasmid	Cloning sites
pCGDM	pSVN	1 2	pSCX	XhoI SacI
pKSN	pSNH	3 4	pKS	HindIII ApaI
pSVN1			pSVN pKSN	ScaI BstEII
pSVNC1	pSNH	3 5	pSVNC	ApaI XbaI
pSVNC2	pSVN	6 4	pSAC	XhoI ApaI
pSVNC3	pSVN'	7 8	pSVNC pSVNC2	XbaI XhoI
pSVNC4	pSVN'	7 9	pSVNC	XbaI ApaI
pSVNC41			pSVNC4 pCGDM	XbaI ScaI
pSVNC5	pSVN'	7 10	pSVNC	XbaI ApaI
pSVNC51			pSVNC5 pCGDM	XbaI ScaI
pSVNC6			pSVNC61 pSVNC1	BstEII BglII
pSVNC61	pSVNC	11 12	pSVNC	XbaI BglII
pSVNC7	pSVNC4	11 12	pSVNC	ApaI BglII
pSVNC71			pSVNC7 pCGDM	XbaI ScaI
pSVNC8	pSVNC5	11 12	pSVNC	XbaI BglII
pSVNC81			pSVNC8 pCGDM	XbaI ScaI
pSVNC82	pSVNC pSVNC8	1 12 11 13	pSVNC8	XbaI BglII
pSVNC9	pSVNC1	14 12	pSVNC91	XhoI BglII
pSVNC91	pSVNC51 pSVNC	7 15 14 12	pSVNC8	XbaI BglII
pSVNC10			pSVNC101 pSVNC9	XhoI BglII
pSVNC101	pSVNC8	11 15	pSVNC91	XbaI XhoI

C

1 : 5' TTTCTCGAGGTGAAGGTGGCTTTGGT 3'
2 : 5' TTTGAGCTCATTATTCAAACTTCGTCT 3'
3 : 5' CATACGATTTAGGTGACACTATAG 3'
4 : 5' AAAGGGCCCTCAAACTTCGTCTTCTTTC 3'
5 : 5' AAAGGGCCCTGTTCACCCTTAGGTTTGGC 3'
6 : 5' AAACTCGAGAAATGACCAAACAGAAAG 3'
7 : 5' TAATACGACTCACTATAGGGAGA 3'
8 : 5' TTTCTCGAGGTGCTGGTTTAACAGG 3'
9 : 5' TTTGGGCCCACTTTCTGTTTCTTTGCT 3'
10 : 5' TTTGGGCCCTTACTATCTCTTCCTTTTGGT 3'
11 : 5' TTTTCTAGAGCCATGGGCCCTGTTAAACCAGCACCT 3'
12 : 5' TTTGAGCTCAGATCTCATTACGCCCCGCCCTGCCA 3'
13 : 5' AAAGTCGACCCTCGAGTTTTACTATCTCTTCCTTTTGG 3'
14 : 5' TTTCTCGAGGACCCAGGGGGTCACCTAATG 3'
15 : 5' TTTCTCGAGCCACTTTCTGTTTCTTTGCT 3'

D

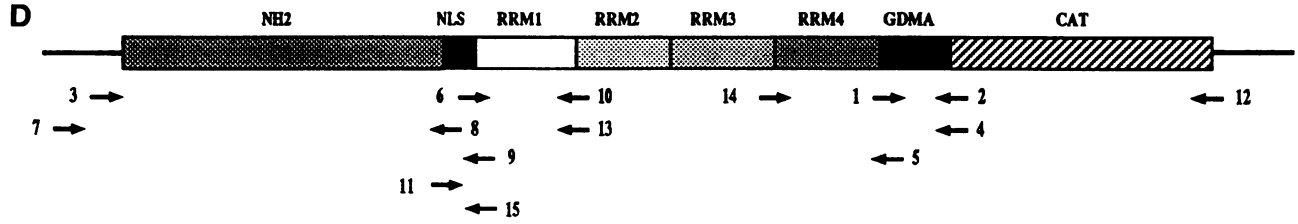


Figure 1. Vectors and cloning strategies. (A) Structure of the four basic vectors used for cloning serial nucleolin-CAT expression plasmids listed in B. The plasmids pSCX and pSAC derived from pSVL contain the CAT protein coding sequence that can be fused in frame with the nucleolin mutant or wild-type cDNA in the CAT C-terminus of pSCX or in the CAT N-terminus of pSAC. The pSVL-derived plasmid pSVN

of nucleolin to pre-rRNA that occurs through the four RRMs (Ghisolfi *et al.*, 1992a,b). Given the number and variety of characterized functional domains, we became interested in identifying the sequences or domains of nucleolin responsible for its nuclear and nucleolar localization. To this end, we constructed a series of deletion mutants within the nucleolin coding region in frame with the chloramphenicol acetyltransferase (CAT) coding sequence. Using a CAT specific antiserum, we were able to distinguish the mutant recombinant proteins from the endogenous cellular protein after transient transfection into L929 mouse cells. The results establish that a nuclear localization sequence (NLS) is strictly necessary for nucleolar targeting, implying that the protein has to be transported as other nuclear proteins across the nuclear membrane. Nucleolar accumulation requires in addition at least two RRMs. However the NH2 terminus also seems to play a role by partially compensating the lack of any one RNA binding region.

MATERIALS AND METHODS

Vectors and Cloning Strategies

To construct the nucleolin-CAT eukaryotic expression plasmids, we used four basic pSVL-derived vectors (Pharmacia, Piscataway, NJ) represented in Figure 1A: pSCX and pSAC, which contain the CAT gene (Bugler *et al.*, 1991). The whole mouse nucleolin cDNA between *Xba* I and *Eco*RI sites (Kleenow traited) (Lapeyre *et al.*, 1987) was inserted in the *Xba* I/*Sma* I sites of pSVL vector to construct pSVN. To obtain pSVNC, *Apa* I site was created by usual polymerase chain reaction (PCR) in place of the nucleolin stop codon TAA. Then the *Xba* I/*Apa* I nucleolin fragment was inserted between the hybrid vector *Apa* I/*Sca* I fragment of pSAC and *Sca* I/*Xba* I fragment of pSCX to get pSVNC.

The deleted nucleolin fragments were inserted between *Xho* I and *Apa* I sites in frame with the CAT N-terminus in pSAC and between *Xho* I and *Sac* I in frame with the CAT C-terminus in pSCX to get all the nucleolin-CAT mutant plasmids as described in Figure 1B. A set of these recombinant plasmids was constructed after amplification of the nucleolin cDNA fragments by PCR. For example, the plasmid pSVNC2 was obtained after amplification of the nucleolin fragment comprised between the oligonucleotides 4 and 6 (Figure 1, C and D) followed by insertion between *Xho* I and *Apa* I sites of the plasmid pSAC. When transfected in cells, this plasmid allows expression of the NH2-deleted nucleolin-CAT protein NC2, schematically represented in Figure 3A. The other set of recombinant plasmids (pSVN1, pSVNC41, 51, 6, 71, 81, and 10) were obtained by standard techniques. For example, the *Xba* I/*Sca* I nucleolin fragment of pCGDM was inserted between the *Xba* I/*Sca* I sites of pSVNC4 to allow expression of the nucleolin-CAT protein NC41.

Cell Culture and Transfection

Cos-7 and L929 cells were cultured in Dulbecco's minimal essential medium (DMEM) (GIBCO, Grand Island, NY) containing 5% fetal calf serum in 5% CO2. Cos-7 cells were transfected as described (Bugler *et al.*, 1991). L929 were grown on 12-mm^2 glass coverslips in 25-mm^2 plastic Petri dishes for immunofluorescence to 40–60% confluence for transfection. Cells were washed with phosphate-buffered saline (PBS) and incubated with 2–5 μg plasmid DNA in 1.5 ml PBS containing 0.4 mg/ml cellulose diethylaminoethyl-cellulose dextran (Sigma, St. Louis, MO) and 80 μM of Chloroquine (Sigma) for 20 min at 37°C. After 1 min dimethyl sulfoxide choc (10% in DMEM) and three washes, cells were cultured for 48 h before microscopy. At least four separate experiments with different plasmid preparations were done for each construct.

Immunofluorescence

Transfected cells coverslips were prepared for immunofluorescence microscopy as follows: coverslips were washed with PBS and fixed with 3% paraformaldehyde in PBS for 7 min at 20°C. After washing, cells were permeabilized with 0.05% Saponine (Fluka, Buchs, Switzerland) in PBS for 20 min and incubated in 1/100 dilution of goat preimmune serum (Nordic, Tilburg, Netherlands) for 30 min then incubated for 1 h at 37°C with 1/1000 dilution of serum against CAT protein (Tebu, Paris, France). Coverslips were washed three times and incubated with 1/400 dilution Texas red anti rabbit antibody (Nordic) for 40 min then washed and mounted in an antifading solution (Cityfluor, London, UK). They were viewed under an Axiophot (Zeiss, Thornwood, NY) and photographed with 400 ASA HP5 film (Ilford, Basildon, UK).

Western Blot Analysis

Proteins from total lysates of transfected Cos-7 cells were separated on 8%, 10%, or 12% sodium dodecyl sulfate (SDS) polyacrylamide gels and transferred to nitrocellulose in buffer containing 48 mM tris(hydroxymethyl)aminomethane (Tris) 39 mM glycine, 0.05% SDS, and 25% methanol for 2 h at 4°C. The filters were incubated for 1 h in blocking buffer (10 mM Tris/HCl pH 8, 150 mM NaCl, 0.2% Tween 20 [TBST], 3% nonfat milk) and then overnight in the blocking buffer containing 1/10 000 dilution of rabbit anti-CAT antibody. After three washes in TBST, the filters were probed with horseradish peroxidase (HRP)-labeled second antibody (HRP-linked whole antibody from donkey, Amersham, Arlington Heights, IL), and signals were visualized using electrochemiluminescence Western blotting protocol (Amersham).

contains the whole mouse nucleolin cDNA and pSVNC comprises the fused nucleolin-CAT coding sequence where the CAT AUG initiation codon was eliminated. Only the restriction sites used for subcloning are mentioned on the maps. (B) Cloning strategy. The recombinant plasmids (mutant name) coding for all the deleted nucleolin-CAT chimeric proteins have been constructed either after amplification by usual PCR of the cloned fragment or by standard techniques in the designated cloning plasmids and cloning sites. In the case of standard cloning, the upper cloning plasmid always represents the cloning vector, and the second one is used to excise the fragment to insert. When the deleted derivates were constructed by PCR for each construct, "matrix" designates the template used for amplification with the two corresponding oligonucleotides detailed in C (first primer = N-terminal primer, second primer = C-terminal primer). The three first mutant plasmids pCGDM, pKSN, and pSVN1 correspond to intermediate constructs for subsequent cloning. The plasmid pSNH contains the whole mouse nucleolin cDNA inserted in *Bam*HI/*Eco*RI sites of pSP64 (Promega, Madison, WI). All other plasmids encode the different nucleolin-CAT fusion proteins after transfection in eukaryotic cells. pKS was supplied by Stratagene (La Jolla, CA). (C) Sequences of all the synthetic oligonucleotides numbered as in B and D. (D) Approximate positions of the 15 synthetic primers all along the nucleolin-CAT expression plasmids. The different functional domains of nucleolin are represented by shaded area, and arrows indicate the direction of polymerization from each primer listed in C. NH2, N-terminal domain including acidic and basic stretches; RRM1-2-3-4, RNA recognition motifs; CAT, CAT protein fused with each recombinant nucleolin.

RESULTS

Expression of Nucleolin-CAT Fusion Proteins in Transfected Cells

To discriminate between endogenous nucleolin and recombinant proteins expressed from the transfected genes, we constructed a set of mutants in frame with the CAT prokaryotic gene. Vectors and cloning strategies are presented in Figure 1. After transfection in L929 cells the complete nucleolin in fusion with CAT was detected by indirect immunofluorescence in the nucleolus by anti-CAT antibodies with a same distribution pattern as the endogenous protein detected by a serum raised against nucleolin even if in a very few percentage of cells a cytoplasmic location was also observed (Figure 2, A and B).

Characterization of recombinant proteins was carried out after transfection in Cos-7 cells that allowed production of large amounts of protein. As shown by Western blotting after SDS gel electrophoresis, the nucleolin-CAT protein was recognized both by nucleolin and CAT anti-sera, and, as expected, the fusion protein migrated slower than nucleolin (Figure 2C). Strikingly, it can be seen that simian endogenous nucleolin unlike hamster nucleolin is not recognized by this anti mouse nucleolin antibody suggesting the main antigenic epitopes differ in the two species (Figure 2C, lane 1).

As shown by its sequence analysis (Lapeyre *et al.*, 1987), nucleolin can be divided into four structural domains: a basic-acidic N-terminal region, four RRMs in the central part, a GDMA-GAR region in the carboxy terminus, and a putative NLS between the acido-basic stretches and the RRM (Figure 2D). To determine the role of each domain in the localization of nucleolin, constructions were done to delete the GAR domain in NC1, the putative NLS in NC3, one or all the RRM in NC41, NC51, or the NH2 terminal domain in NC61. All other constructs were derived from multiple deletions (Figure 3A).

As we can see in Figure 3, B–D, all proteins containing the N-terminal region (NC and the derivates NC1, 3, 4, 41, 5, 51, 9, 91) migrate with apparent molecular

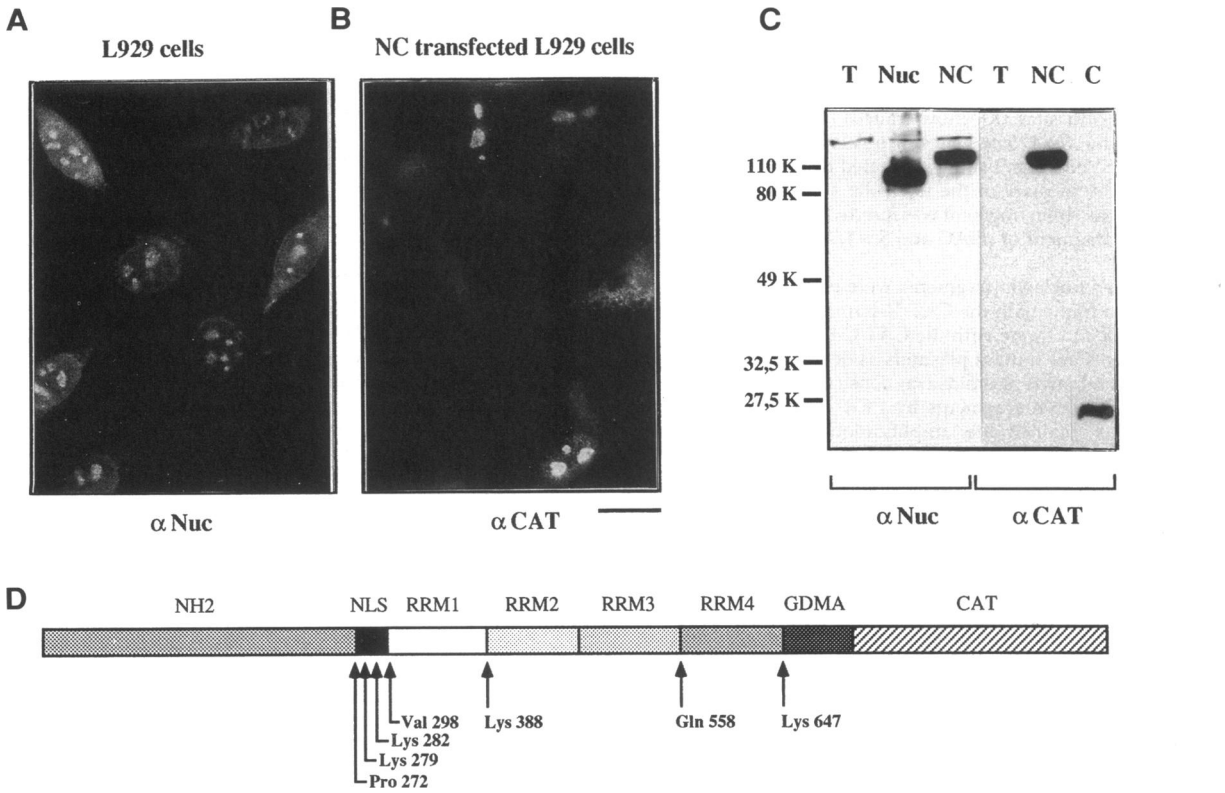


Figure 2. Nucleolar expression of the endogenous nucleolin in L929 cells and whole nucleolin-CAT fusion protein in L929 transfected cells. (A) Immunofluorescence image after staining with anti nucleolin and Texas Red-labeled goat anti rabbit IgG in control cells. (B) Same staining with anti CAT antibody in nucleolin-CAT expression plasmid (pSVNC)-transfected cells. Bar, 10 μm. (C) Western blots of total proteins from nontransfected cells (T), nucleolin (Nuc), nucleolin-CAT (NC), and CAT (C)-expressing cells. The first three lanes on the left correspond to protein extracts probed with anti-nucleolin and the three lanes on the right to extracts probed with anti-CAT antibody. (D) Deletion boundaries along the nucleolin coding sequence. The different functional domains defined in Figure 1D were deleted according to the cloning strategy to produce the mutant proteins represented in Figure 3A. The residues and their positions in the nucleolin coding sequence delimit the precise boundaries of each deletion shown in Figure 3A.

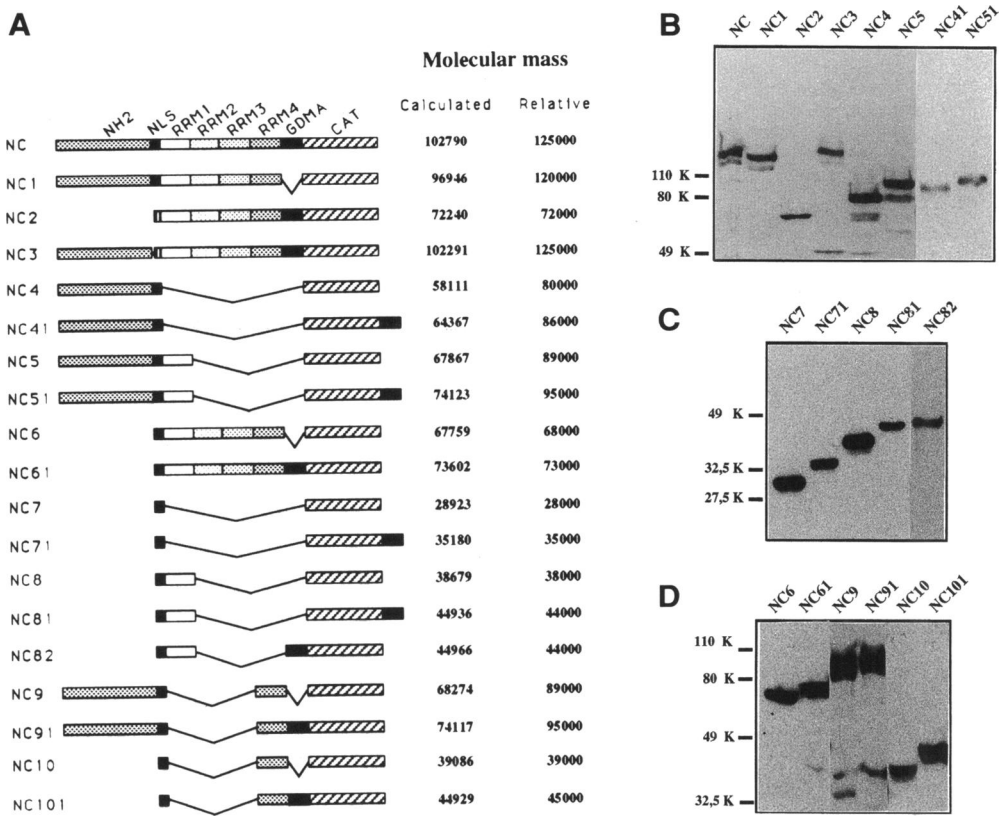


Figure 3. Western blot analysis of nucleolin-CAT fusion proteins. (A) Schematic representation of nucleolin-CAT fusion proteins in functional domains described in Figure 1D. To allow alignment of all derivates, internal deletions are represented by broken lines. The precise limits of each deletion are defined in the Figure 2D. The molecular mass calculated from the predicted protein cDNA and the mass relative deduced from migration in SDS- are indicated for each fusion protein. (B–D) Western blot analysis of protein extracts from Cos-7–transfected cells expressing the nucleolin-CAT fusion proteins represented in A and probed with anti-CAT antibody. Molecular weight standards are indicated on the left in kDa.

weights larger than the expected ones. This could result from posttranslational modifications, particularly phosphorylation (Caizergues-Ferrer *et al.*, 1987; Belenguer *et al.*, 1990) or from the acid-rich composition of the domain (Dingwall *et al.*, 1987; Meier and Blobel, 1992). An additional band at ~10 000 under the major one that is sometimes resolved into a doublet is routinely observed. It is because of the maturation products of nucleolin deleted from its N-terminal region (Bugler *et al.*, 1982; Bourbon *et al.*, 1983). All the N-terminal–deleted mutants NC2 (Figure 3B), NC7, 71, 8, 81, 82 (Figure 3C) and NC6, 61, 10, 101 (Figure 3D) have the expected apparent molecular weights.

To test the incidence of the GAR position inside the recombinant proteins that is normally excluded from the globular structure of native nucleolin (Erard *et al.*, 1990), we joined it to the C terminus of the CAT protein in NC41, NC51, NC71, and NC81 to avoid its buring within the folded protein (Figure 3A). These mutants were never detected by anti-CAT antibodies as well as their respective counterparts NC4, NC5, NC7, and NC8 in four separate transfection experiments and use of different plasmid preparations (Figure 3, B and C). Change in chimeric protein conformation masking CAT-epitopes could explain these differences in the immunodetection assays, because anti-nucleolin antibodies revealed a similar expression pattern between GAR-included proteins and GAR-deleted proteins.

Subcellular Localization of Nucleolin-CAT Fusion Proteins in L929-transfected Cells: Characterization of the Nucleolin NLS

For analysis of the nuclear translocation and nucleolar accumulation processes, we used L929 cells in place of Cos-7 cells to avoid overexpression of the proteins that could result in abnormal localization. All fusion proteins listed in Figure 3A were transfected and detected by indirect immunofluorescence as described in Materials and Methods using anti-CAT antibodies. This polyclonal serum did not react with any cellular component because untransfected cells were unlabeled (Figures 4 to 6).

The chimeric wild-type nucleolin (NC) is predominantly nucleolar, whereas the mutant lacking the N-terminal region and about one-half of the putative NLS until lysine 282 is cytoplasmic (NC2). Mutant NC3 lacking only four residues included in the putative NLS (Lys 279 until Lys 282) were cytoplasmic as the CAT protein (Figure 4). On the other hand, the sequence comprised between residues 272 and 298 joined to the CAT protein is able to translocate NC7 to the nucleus.

The sequence residues 272 to 298 are sufficient to direct a cytoplasmic protein such as CAT to the nucleus and thus can be defined as the mammalian nucleolin NLS that is conserved between human, mouse, and rat. It must be noted that the NC7-transfected cells showed

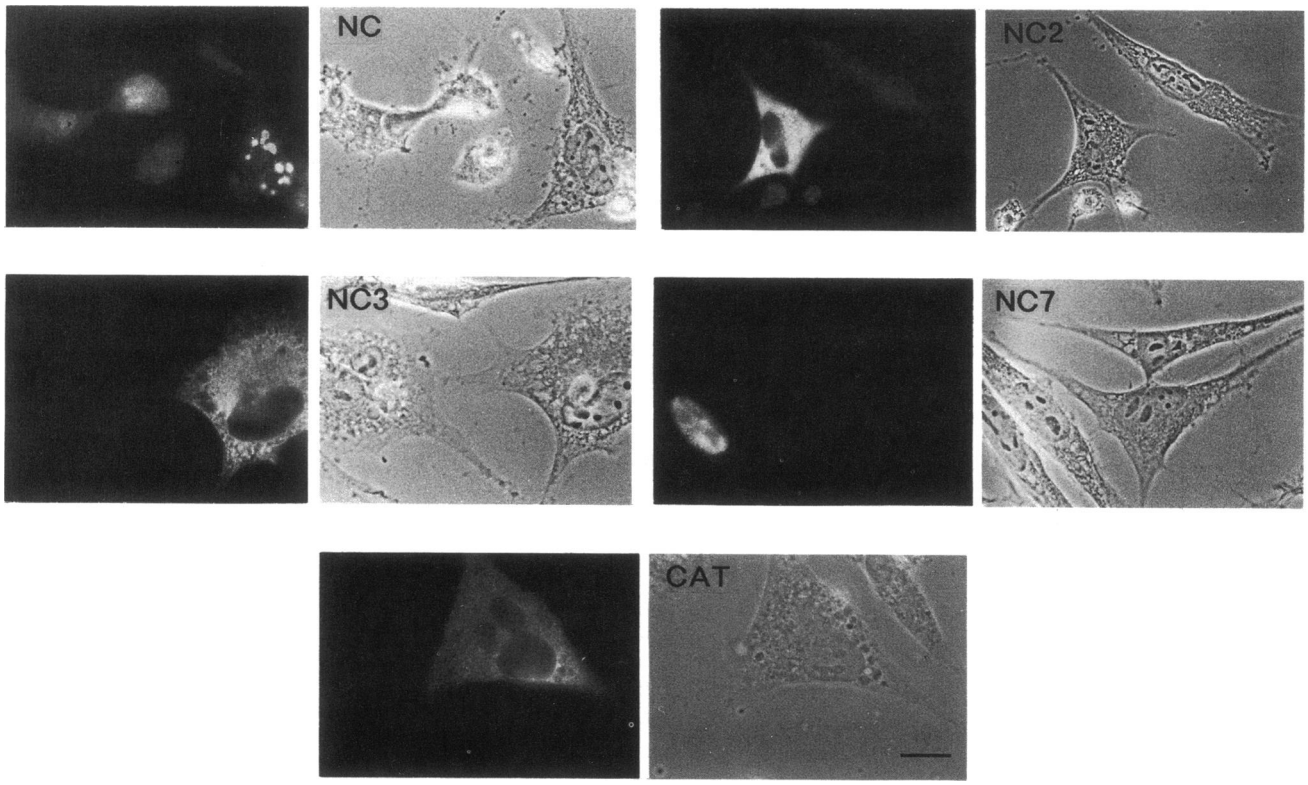


Figure 4. Location of CAT fusion nucleolin in L929-transfected cells. Rabbit anti serum to CAT protein and fluorescent Texas red anti rabbit antibody were used for the indirect immunofluorescent staining assay. Phase contrast microscopy is shown on the right of the corresponding fluorescent staining. NC, full-length nucleolin fused with CAT; NC2, N-terminal–deleted fusion nucleolin (see also text and Figure 3A); NC3, deleted part of NLS fusion nucleolin; NC7, nucleolin NLS fused with CAT reporter protein. Bar, 10 μm.

only nucleoplasmic fluorescent staining with dark unstained nucleolar structures.

Domains Involved in Nucleolar Accumulation

Because the characterized NLS was insufficient to target proteins to the nucleolus, other nucleolin domains were successively added to chimeric proteins. To see whether the RRM participating in RNA binding played a role in nucleolar accumulation, we deleted either all of them (NC41: deletion from Val 298 to Lys 647), the last three (NC51: deletion from Lys 388 to Lys 647) or the first three (NC91: deletion from Val 298 to Gln 558) (Figures 2 and 3). As we can see in Figure 5, only proteins with RRM1 (NC51) or RRM4 (NC91) were exclusively nucleolar, whereas the complete deletion of RNA binding domains (NC41) prevented nucleolar accumulation in ~50–70% of transfected cells. However, when the GAR domain was removed of proteins comprising only one RRM (NC5 and NC9), the pattern of localization looked almost like NC41 protein that include the sole GAR motif as RNA-binding domain. The GAR motif added to the NLS did not localize NC71 in the nucleolus (Figure 6). On the other hand, the GAR domain was not essential when the four RNA binding domains were present (Figure 6, NC1) even if the N-terminal domain was deleted up to the proline 272 (NC6). Thus the GAR domain appears to act in synergy with RRM1 or RRM4 to target protein to the nucleolus.

An unexpected result was routinely observed with NC61 (4 RRM and GAR). Although in 50% of cells the expected nucleolar location was observed, in the remaining transfected cells the signal was found in cytoplasm without any nucleolar staining. One hypothesis is that the assignment of this recombinant protein could be particularly sensitive to physiological parameters like the different phases of the cell cycle.

Additional experiments suggested that the N-terminus domain could play a subsidiary role in nucleolin location. Whereas the GAR domain with the NLS only accumulate NC71 in the nucleoplasm, addition of the N-terminal domain localizes the resulting protein NC41 both in the nucleoplasm and the nucleolus (Figure 5). Furthermore, two other constructs showed good agreement with this potential role: the fusion protein NC8 including only RRM1 is exclusively nucleoplasmic (Figure 6) and NC5 including this domain and the N-terminus is partially recovered in the nucleolus (Figure 5). Addition of GAR domain relocalizes the proteins NC81

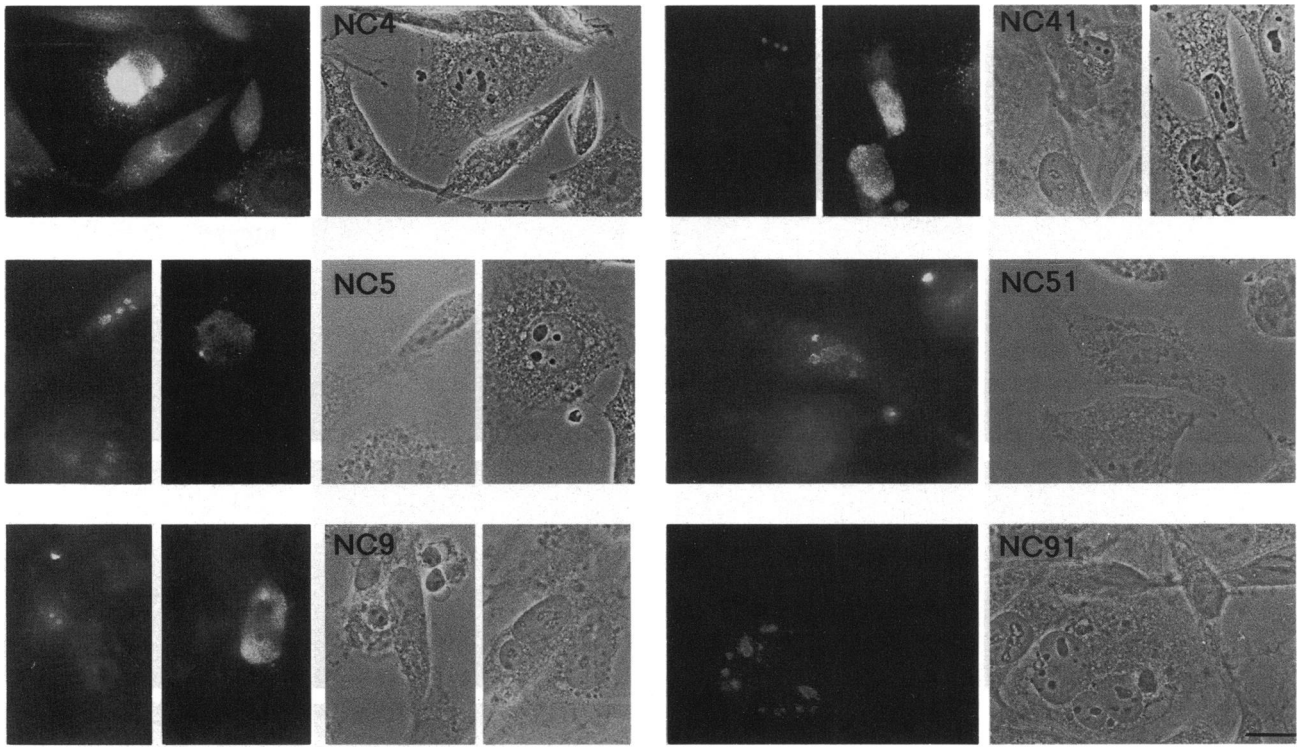


Figure 5. Effect of RRM deletion and role of GAR motif in L929-transfected cells. NC4 and NC41, deletion of all RRM domains (see also text and Figure 3A); NC5 and NC51, deletion of RRMs 2–4; NC9 and NC91, deletion of RRMs 1–3. NC4, NC5, and NC9 correspond respectively to the same RRM deletions as NC41, NC51, and NC91 but deleted of GAR motif. Bar, 10 μm.

and NC82 into the nucleolus independently of its position in the molecule (Figure 6). The same observation can be made for NC10 and NC101 with RRM4 in place of RRM1.

As summarized in Table 1, different domains of nucleolin can accumulate the protein in the nucleolus with a graded efficiency. Four RNA-binding domains are as efficient as only one domain plus the GAR motif (NC6 versus NC81, NC82, and NC101, Figure 6); the N-terminal acido-basic rich region can play a subsidiary role increasing nucleolar accumulation in proteins that comprise only one type of RNA-binding domain, either RRM or GAR motif (NC41, NC5, and NC9) (Figure 5).

DISCUSSION

The Nuclear Targeting Sequence of Nucleolin Is Bipartite

Nucleolin (Lapeyre *et al.*, 1987) is organized into three domains schematically represented in Figure 2D. The N-terminal domain comprises four acidic stretches containing almost exclusively glutamate and aspartate residues and a putative NLS between lysine 279 and valine 298. Although the lysine cluster at the amino terminal end of the nucleolin NLS, between lysine 279 and 282, resembles the SV40 large T antigen nuclear targeting sequence, Dang and Lee (1989) found that it was not sufficient to direct the pyruvate kinase to the nucleus. Similarly, this cluster deletion alone (NC2 and NC3, Figure 4) blocks the CAT-nucleolin fusion protein transport to the nuclear compartment. However, the Dang's construct differs from ours by the absence of the last six carboxy terminal NLS residues including three lysine (Table 2). When the full sequence (proline 272 to valine 298) is fused to CAT, efficient nuclear targeting is observed (NC7, Figure 4).

Although further work is needed to define the precise boundaries and the role of the individual amino acids in nuclear localization of nucleolin, the data suggest that two spatially distinct sequences are required as is the case for nucleoplasmin (Robbins *et al.*, 1991). This bipartite sequence comprises two interdependent clusters of basic amino acids separated by 10 intervening spacer residues for the nucleoplasmin and 11 for nucleolin (Table 2). Both basic stretches are necessary for nuclear targeting but not sufficient alone to address a reporter protein to the nucleus. A bipartite motif is also present in the steroid receptor protein family (Picard *et al.*, 1987, 1990), the nucleosomal protein N1 (Kleinschmidt and Seiter, 1988), the SWI 5 yeast protein (Nasmyth *et al.*, 1990), the polymerase basic protein 1 of influenza virus (Nath and Nayak, 1990), the antion-

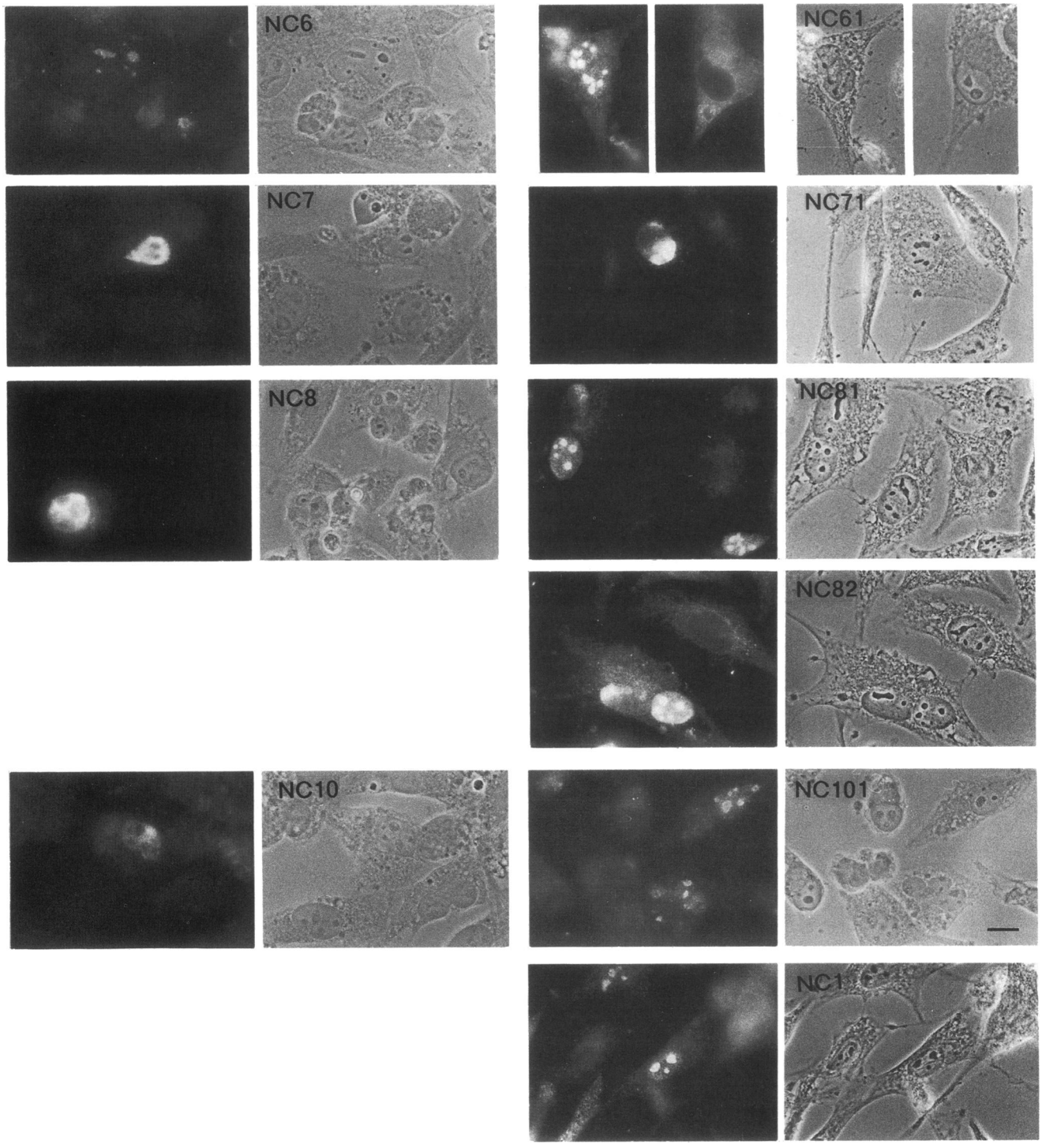
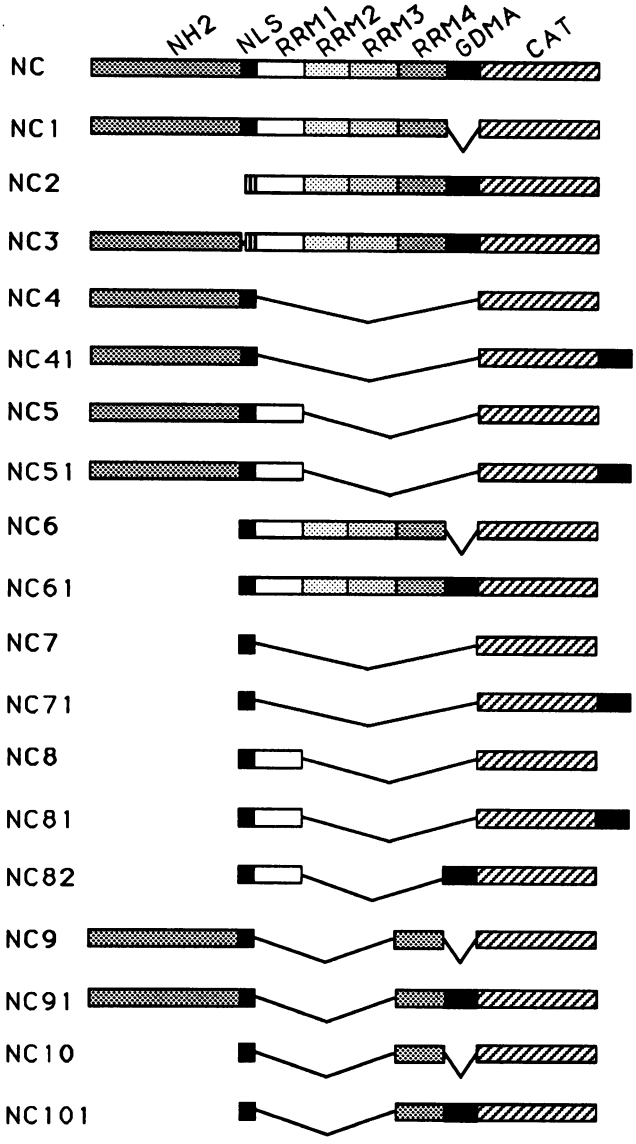


Figure 6. Effect of multiple deletions on location of nucleolin in L929-transfected cells. NC6 and NC61, deletion of the N-terminal domain; NC7 and NC71, nucleolin NLS in fusion with the CAT protein; NC8, NC81, and NC82, deletion of the N-terminal domain and the last three RRMs; NC10 and NC101, deletion of the N-terminal domain and the first three RRMs. NC6, NC7, NC8, and NC10 correspond respectively to the same deletions as NC61, NC71, NC81/NC82, and NC101 but deleted of the GAR motif. The NC81 GAR is fusioned in the C-terminus of CAT, whereas the NC82 GAR is fusioned in the N-terminus. The nucleolar NC1 protein is only deleted of its GAR motif.

Table 1. Localization of chimeric nucleolin-CAT proteins

Proteins	Cytoplasm	Nucleoplasm	Nucleolus
NC			+
NC1			+
NC2	+		
NC3	+		
NC4		+	
NC41		+	+
NC5		+	+
NC51			+
NC6			+
NC61	+		+
NC7		+	
NC71		+	
NC8		+	
NC81			+
NC82			+
NC9		+	+
NC91			+
NC10		+	
NC101			+



This table summarizes the ability of the different nucleolin fusion proteins to accumulate in the cytoplasm, nucleoplasm, or nucleolus.

cogene p53 (Addison *et al.*, 1990), and the *Xenopus* nucleolar No38 (Peculis and Gall, 1992). Interestingly, this same motif is present in almost one-half the nuclear proteins in the *Swiss protein data base* but in <5% of nonnuclear proteins. The mechanism by which two distinct regions are required in nuclear translocation is not clear. The two basic sequences could act cooperatively for nuclear transport by binding to a nuclear receptor or one of them provide the proper conformation for presenting the binding receptor signal. Therefore the deletion of one region may alter the conformation of the protein thereby masking a distant signal.

Nucleolar Localization of Nucleolin Is Determined Essentially by its RNA-Binding Domains

The central domain of nucleolin is composed of four closely related sequence elements. Each constitutes a RRM that is shared by a large family of RNA-binding proteins (for reviews see Keene and Query, 1991; Kenan *et al.*, 1991).

Table 2. Sequence comparison of nuclear targeting sequences in nucleolin and nucleoplasmin

Nucleolin	WT	^{279}**KRKK**EMTKQKEVPEA**KKQK**V^{298}	Nx	(This work)
	DM	EMTKQKEVPEA**KKQK**V	Cyto (100%)	
		KRKKEMTKQKEVP	Cyto (80%)	(Dang and Lee, 1989)
Nucleoplasmin	WT	^{155}**KR**PAATKKAGQA-**KKKK**L^{171}	Nx	
	PM	**KR**.....-**KNK**NL	Cyto	(Robbins *et al.*, 1991)
		N**R**.....-**KKKK**L	Nx = Cyto	
		KS.....-**KKKK**L	Nx = Cyto	

To allow alignment of identical or similar residues between both NLS, lacking residues in the nucleoplasmin sequences are represented by dashes. Conserved residues are represented by points. Basic amino-acids involved in nuclear localization are in bold-face letters. WT = wild type, DM = deletion mutants, PM = point mutants, Cyto = cytoplasm, Nx = nucleus.

As shown in Figure 5, deletion of the four RRMs maintains a proportion of nucleolin in the nucleoplasm with some nucleolar staining (NC41), whereas the presence of only one RRM (1 or 4) is sufficient to direct the protein in the nucleolus (NC51 and NC91). These RRMs specifically recognize a site in the 5′ external transcribed spacer region of the ribosomal gene even RNA affinity between RRM1 and RRM4 is different (Ghisolfi *et al.*, 1992b). Moreover removal of the GAR C-terminal domain from these nucleolin mutants results in a 10 times drop in RNA affinity. In this study NC5 and NC9 GAR-deleted proteins (corresponding to low RNA affinity mutants) are nucleoplasmic and nucleolar, whereas their counterparts NC51 and NC91 (GAR included) proteins are strictly nucleolar. Thus the affinity for pre-rRNA is correlated with a nucleolar location.

The GAR motif acts independently of its position in the molecule (compare NC81 and NC82, Figure 6) as was suggested by the structural organization of other members of the same nucleolar family that possess similar motifs. It is central for SSB1 (Jong *et al.*, 1987), C-terminal for nucleolin, NSR1 (Lee *et al.*, 1991), NOP3 (Russell and Tollervey, 1992), N-terminal for fibrillarin (Lapeyre *et al.*, 1990), and at both extremities for GAR1 (Girard *et al.*, 1992). GAR1-deleted of the two GAR domain is still able to reach the nucleolus and to assume a normal pre-rRNA processing in the yeast complementation test (Girard *et al.*, unpublished data).

To summarize, association of both types of RNA-binding domains (RRM and GAR) in presence of NLS, play essential roles in the nucleolar localization of nucleolin and are sufficient to bring the CAT reporter protein into the nucleolus (NC6, Figure 6 and Table 1).

The NH2 Terminal Domain Can Modulate Subnuclear Location

When the N-terminal domain is removed, the nucleolar localization is more or less affected depending on the deletion mutant. Deletion of the glycine-rich region leads to nucleoplasmic localization of deleted N-terminal NC8 and NC10 proteins, whereas the presence of the NH2 terminus permits partial recovery of the proteins in the nucleolus in 50% of transfected cells (compare NC8, Figure 6; NC5, Figure 5; NC10, Figure 6; and NC9, Figure 5). These results suggest a possible involvement of this domain in nucleolin localization by association with nucleolar chromatin at the level of the nontranscribed A/T rich rDNA spacer (Olson *et al.*, 1983; Kharrat *et al.*, 1991).

Structure/Location Comparison Between Nucleolin and Other Nucleolar Proteins

Another abundant nucleolar protein of 38 kDa related to the histone-binding nucleoplasmin, named B23 in mammalian cells and No38 in *Xenopus*, is thought to be involved in packaging of preribosomal particles (Chan *et al.*, 1986; Schmidt-Zachmann *et al.*, 1987). Nucleolin and B23/No38 share several properties: two acidic stretches with a serine in a good CKII phosphorylation context followed by a consensus TPXKK cdc2 kinase site that constitute the central part of the molecule. If this portion is removed, the deleted protein is not detected in the nucleolus (Peculis and Gall, 1992). However although this is needed for nucleolar accumulation, it is not alone sufficient to target a reporter protein to the nucleolus. These results suggest that the N-terminus domain of No38 is also necessary in nucleolar targeting or that conformational change in the deleted protein prevents its nucleolar localization as seems to be the case for the heat shock HSp70 protein (Milarski and Morimoto, 1989).

The nucleolar targeting sequences were determined in the viral proteins *tat* (Ruben *et al.*, 1989), *rev* (Kubota *et al.*, 1989; Cochrane *et al.*, 1990), and HTLV-1 pX (Siomi *et al.*, 1988). They show little homology but consist of or contain long basic sequences resembling previously characterized NLS. In eukaryotic proteins it seems that nucleolar accumulation is not conferred by extended NLS as in the viral proteins but seems more likely to implicate different functional domains in a properly folded configuration. The capacity to concentrate in the nucleolus could reflect the ability of nucleolar

proteins to interact with some specific components, nucleic acids, and/or proteins involved in ribosome biogenesis. Consequently, the more likely monomolecular targeting of one viral protein and the multiple interactions of one cellular protein to different nucleolar components must be driven by different proccesses.

The transcription factor Upper Binding Factor (UBF) represents, with nucleolin, a good example of requirement of functional domains for nucleolar localization. It is also transferred to the nucleus by its NLS and is sequestered in the nucleolus by its specific binding to the rDNA promoter via High Mobility Group (HMG) boxes and the putative SL1 factor binding site via its acidic tail (Maeda *et al.*, 1992). Deletion of HMG box 1 suppresses UBF transcriptional activity but also results in the loss of nucleolar targeting. Moreover SL1 strongly enhances the binding of UBF to the rDNA promotor and its transcriptional activity, but deletion of the putative binding SL1 site prevents UBF nucleolar location. Therefore, domains necessary for transcription initiation are also required for nucleolar targeting.

The mammalian nucleolar phosphoproteins Nopp 140 (Meier and Blobel, 1992), B23, nucleolin (Borer *et al.*, 1989), and the yeast NSR1 (Lee and Meleze, 1989) could play a role in the nucleocytoplasmic transport of ribosomal components. These nucleolar proteins are clustered in several acidic repeats revealing some potentially or effective consensus sites for CKII phosphorylation interspaced with basic amino acid stretches overlapping p34cdc2 phosphorylation sites. Whether these features are directly implicated either in nucleocytoplasmic shuttling and/or in its regulation and/or in nucleolar activities will require further efforts. It is however striking that nucleolar proteins are modular with small numbers of functional units: basic repeats, acidic/serine blocks, specific RNA recognition motifs, and GAR domains. Only nucleolin displays all of these domains. The fact that two of them in each of these proteins are sufficient to target the native peptide into the nucleolus is in good agreement with our results where the presence of two types of RNA-binding domains or only one plus the NH2 terminal DNA/protein-binding region allows nucleolar accumulation.

ACKNOWLEDGMENTS

This work was supported by grants from the Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, and the Association pour la Recherche sur le Cancer.

Note added in proof. When this paper was submitted, Schmidt-Zachmann and Nigg published a study of nucleolar targeting domains in nucleolin. They reach similar conclusions about the bipartite NLS and the lack of linear nucleolar targeting sequences. But they observed a strict requirement of RNA recognition motifs and GAR domain to address the protein into the nucleolus. Different hypotheses could explain this discrepancy with our results: the recipient cells, human versus mouse cells, and the nucleolin origins, chicken versus hamster, are different. Sequences alignment shows quite sensitive differences between both NH2 terminus, which could explain the lack of role of this part of the chicken nucleolin in human cells. Moreover, the reporter proteins were joined in fusion by the N-terminal part of the nucleolin and could also prevent appropriate folding of the chimeric molecule compared with the COOH terminal fusion we used in our constructs. Finally, they used N1 and SV40-NLS in place of nucleolin NLS, whose role in nucleolar location has to be determined. In any case it is striking that the two essential kinds of RNA binding domains are not sufficient to target two reporter proteins in the nucleolus.

REFERENCES

Addison, C., Jenkins, J., and Sturzbecher, H.W. (1990). The p53 nuclear localization signal is structurally linked to a p34cdc2 kinase motif. Oncogene *5*, 423–426.

Belenguer, P., Caizergues-Ferrer, M., Labbe, J.-C., Doree, M., and Amalric, F. (1990). Mitosis-specific phosphorylation of nucleolin by p34cdc2 protein kinase. Mol. Cell. Biol. *10*, 3607–3618.

Borer, R.A., Lehner, C.F., Eppenberger, H.M., and Nigg, E.A.O. (1989). Major nucleolar proteins shuttle between nucleus and cytoplasm. Cell *56*, 379–390.

Bouche, G., Caizergues-Ferrer, M., Bugler, B., and Amalric, F. (1984). Interrelations between the maturation of the 100 kDa nucleolar protein and pre-rRNA synthesis in CHO cells. Nucleic Acids Res. *12*, 3025–3035.

Bourbon, H.M., Caizergues-Ferrer, M., Amalric, F., and Zalta, J.P. (1983). Maturation of a 100kDa protein associated with preribosomes in Chinese hamster ovary cells. Mol. Biol. Rep. *9*, 39–47.

Bugler, B., Amalric, F., and Prats, H. (1991). Alternative initiation of translation determines cytoplasmic or nuclear localization of bFGF. Mol. Cell. Biol. *10*, 573–577.

Bugler, B., Bourbon, H.M., Lapeyre, B., Wallace, M., Chang, J.H., Amalric, F., and Olson, M. (1987). RNA binding fragments from nucleolin contain the ribonucleoprotein consensus sequence. J. Biol. Chem. *262*, 10922–10925.

Bugler, B., Caizergues-Ferrer, M., Bouche, G., Bourbon, H.M., and Amalric, F. (1982). Detection and localization of a class of proteins immunologically related to a 100kDa nucleolar protein. Eur. J. Biochem. *128*, 475–480.

Caizergues-Ferrer, M., Belenguer, P., Lapeyre, B., Amalric, F., Wallace, M., and Olson, M. (1987). Phosphorylation of nucleolin by a nucleolar type NII protein kinase. Biochemistry *26*, 7876–7883.

Chan, P.K., Chan, W.Y., Yung, B.Y.M., Cook, R.G., Aldrich, M.B., Ku, D., Goldknopf, I.L., and Busch, H. (1986). Amino acid sequence of a specific antigenic peptide of protein B23. J. Biol. Chem. *261*, 14335–14341.

Cochrane, A.W., Perkins, A., and Rosen, C.A. (1990). Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization and function. J. Virol. *64*, 881–885.

Dang, C.V., and Lee, W.M.F. (1989). Nuclear and nucleolar targeting sequences of c-erb-A, c-myb, N-myc, p53, HSP70, and HIV tat proteins. J. Biol. Chem. *264*, 18019–18023.

Dingwall, C., Dilworth, S.M., Black, S.M., Kearsey, S.E., Cox, L.S., and Laskey, R.A. (1987). Nucleoplasmin cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences homologous to putative nuclear localization signals. EMBO J. *6*, 69–74.

Erard, M., Belenguer, P., Caizergues-Ferrer, M., Pantaloni, A., and Amalric, F. (1988). A major nucleolar protein, nucleolin, induces chromatin decondensation by binding to histone H1. Eur. J. Biochem. *175*, 525–530.

Erard, M., Lakhdar-Ghazal, M., and Amalric, F. (1990). Repeat peptide motifs which contain β turns and modulate DNA condensation in chromatin. Eur. J. Biochem. *191*, 19–26.

Gas, N., Escande, M.L., and Stevens, B.J. (1985). Immunolocalization of the 100 kDa nucleolar protein during the mitotic cycle in CHO cells. Biol. Cell *53*, 209–218.

Ghisolfi, L., Joseph, G., Amalric, F., and Erard, M. (1992a). The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. J. Biol. Chem. *267*, 2955–2959.

Ghisolfi, L., Kharrat, A., Joseph, G., Amalric, F., and Erard, M. (1992b). Concerted activities of the RNA recognition and the glycine-rich C-terminal domains of nucleolin are required for efficient complex formation with preribosomal RNA. Eur. J. Biochem. *209*, 541–548.

Girard, J.-P., Lehtonen, H., Caizergues-Ferrer, M., Amalric, F., Tollervey, D., and Lapeyre, B. (1992). GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast. EMBO J. *11*, 673–682.

Herrera, A.H., and Olson, M.O.J. (1986). Association of protein C23 with rapidly labeled nucleolar RNA. Biochemistry *25*, 6258–6264.

Hügle, B., Hazan, U., Scheer, W., and Franke, W. (1985). Ribocharine: a nuclear Mr 40000 protein specific to precursor particles of the large ribosomal subunit. Cell *41*, 615–627.

Jong, A.Y., Clark, M.W., Gilbert, M., Oehm, A., and Campbell, J.L. (1987). *Saccharomyces cerevisiae* SSB1 protein and its relationship to nuclear RNA-binding proteins. Mol. Cell. Biol. *7*, 2947–2955.

Jordan, G. (1987). At the heart of the nucleolus. Nature *329*, 489–490.

Keene, J., and Query, C. (1991). Nuclear-RNA binding proteins. In: Progress in Nucleic Acid Research and Molecular Biology, eds. W. E. Cohn and K. Moldave, San Diego, CA: Academic Press *41*, 179–202.

Kenan, D., Query, C., and Keene, J. (1991). RNA recognition: towards identifying determinants of specificity. Trends Biochem. Sci. *16*, 214–220.

Kharrat, A., Derancourt, J., Doree, M., Amalric, F., and Erard, M. (1991). Synergistic effect of histone H1 and nucleolin on chromatin condensation in mitosis: role of a phosphorylated heteromer. Biochemistry *30*, 10329–10336.

Kleinschmidt, J.A., and Seiter, A. (1988). Identification of domains involved in nuclear uptake and histone binding of protein N1 of *Xenopus laevis*. EMBO J. *7*, 1605–1614.

Kubota, S., Siomi, H., Satoh, T., Endo, S.I., Maki, M., and Hatanaka, M. (1989). Functional similarity of HIV-1 rev and HTLV-1 rex proteins: identification of a new nucleolar-targeting signal in rev protein. Biochem. Biophys. Res. Commun. *162*, 963–970.

Lapeyre, B., Bourbon, H.M., and Amalric, F. (1987). Nucleolin, the major nucleolar protein of growing eukaryotic cells: an unusual protein structure revealed by the nucleotide sequence. Proc. Natl. Acad. Sci. USA *84*, 1472–1476.

Lapeyre, B., Mariottini, P., Mathieu, C., Ferrer, P., Amaldi, F., Amalric, F., and Caizergues-Ferrer, M. (1990). Molecular cloning of *Xenopus* fibrillarin, a conserved U3 small nuclear ribonucleoprotein recognized by antisera from humans with autoimmune disease. Mol. Cell. Biol. *10*, 430–434.

Lee, W., and Melese, T. (1989). Identification and characterization of a nuclear localization sequence-binding protein in yeast. Proc. Nat. Acad. Sci. USA *86*, 8808–8812.

Lee, W.-C., Xue, Z., and Melese, T. (1991). The NSR1 gene encodes a protein that specifically binds nuclear localization sequences and has two RNA recognition motifs. J. Cell Biol. *113*, 1–12.

Lischwe, M.A., Cook, R.G., Soo Ahn, Y., Yeoman, L.C., and Busch, H. (1985). Clustering of glycine and N,N-dimethylarginine in the nucleolar protein C23. Biochemistry *24*, 6025–6028.

Maeda, Y., Hisatake, K., Kondo, T., Hanada, K., Song, C.Z., Nishimura, T., and Muramatsu, M. (1992). Mouse rRNA gene transcription factor mUBF requires both HMG-box1 and an acidic tail for nucleolar accumulation: molecular analysis of the nucleolar targeting mechanism. EMBO J. *11*, 3695–3704.

Meier, U.T., and Blobel, G. (1992). Nopp140 shuttles on tracks between nucleolus and cytoplasm. Cell *70*, 127–138.

Milarski, K.L., and Morimoto, R.I. (1989). Mutational analysis on the human HSP70 protein: distinct domains for nucleolar localization and adenosine triphosphate binding. J. Cell Biol. *109*, 1947–1962.

Nasmyth, K., Adolf, G., Lydall, D., and Seddon, A. (1990). The identification of a second cell cycle control on the HO promoter in yeast: cell cycle regulation of SW15 nuclear entry. Cell *62*, 631–647.

Nath, S.T., and Nayak, D.P. (1990). Function of two discrete regions is required for nuclear localization of polymerase basic protein1 of A/WSN/33 influenza virus (H1 N1). Mol. Cell. Biol. *10*, 4139–4145.

Ochs, R.L., Lischwe, M.A., Sphon, W.H., and Busch, H. (1985). Fibrillarin: a new protein of the nucleolus identified by autoimmune sera. Biol. Cell *54*, 123–134.

Olson, M.O.J., Rivers, Z.M., Thompson, B.A., Kao, W.Y., and Case, S.T. (1983). Interaction of nucleolar phosphoprotein C23 with cloned segments of rat ribosomal deoxyribonucleic acid. Biochemistry *22*, 3345–3351.

Peculis, B.A., and Gall, J.G. (1992). Localization of the nucleolar protein No38 in amphibian oocytes. J. Cell Biol. *116*, 1–14.

Picard, D., Kumar, V., Chambon, P., and Yamamoto, K. (1990). Signal transduction of steroïd hormones: nuclear localization is differentially regulated in estrogen and glucocorticoid receptors. Cell Regul. *1*, 291–299.

Picard, D., and Yamamoto, K. (1987). Two signals mediate hormone dependent nuclear localization of the glucocorticoid receptor. EMBO J. *6*, 3333–3340.

Robbins, J., Dilworth, S.M., Laskey, R.A., and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmic nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell *64*, 615–623.

Ruben, S., Perkins, A., Purcell, R., Joung, K., Sia, R., Burghoff, R., Haseltine, W.A., and Rosen, C.A. (1989). Structural and functional characterization of human immunodeficiency virus tat protein. J. Virol. *63*, 1–8.

Russell, I.D., and Tollervey, D. (1992). NOP3 is an essential yeast protein which is required for pre-rRNA processing. J. Cell Biol. *119*, 737–747.

Schmidt-Zachmann, M.S., Hügle-Dörr, B., and Franke, W.W. (1987). A constitutive nucleolar protein identified as a member of the nucleoplasmin family. EMBO J. *6*, 1881–1890.

Siomi, H., Shida, H., Nam, S.H., Nosaka, T., Maki, M., and Hatanaka, M. (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.

Spector, D.L., Ochs, R.L., and Busch, H. (1984). Silver staining, immunofluorescence and immunoelectron microscopic localization of nucleolar phosphoproteins B23 and C23. Chromosoma *90*, 139–148.