BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **252,** 97–102 (1998)
ARTICLE NO. RC989606

NOLP: Identification of a Novel Human Nucleolar Protein and Determination of Sequence Requirements for Its Nucleolar Localization

Nobuhide Ueki,*,[1] Maiko Kondo,* Naohiko Seki,† Kazuhiro Yano,* Tamaki Oda,* Yasuhiko Masuho,* and Masa-aki Muramatsu*,‡

*Biological Technology Laboratory, Helix Research Institute, Inc., 1532-3 Yana, Kisarazu-shi, Chiba 292-0812, Japan; †Kazusa DNA Research Institute, 1532-3 Yana, Kisarazu-shi, Chiba 292-0812, Japan; and ‡Department of Biological Cybernetics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-0034, Japan

Received September 30, 1998

This study reports cDNA isolation and partial characterization of a novel human nucleolar protein isolated by "nuclear transportation trap" described previously. The cDNA encodes a putative polypeptide of 524 amino acids with a short *Escherichia coli* DNA helicase homologous region, an acid-rich domain, three potential base-rich nuclear localization signals (NLSs), a serine-rich domain, and a deduced coiled-coil domain. The protein has no known prominent similarities with any other protein in the protein databases. Tissue distribution analysis demonstrated a predominant expression in brain and testis. To determine the sequence requirements for nucleolar targeting, a set of deletion constructs with a fluorescent tag were transiently expressed in COS-7 cells. We revealed that a region of 30 amino acids (position 342–371), which overlaps the first and second NLS, is sufficient for nucleolar localization. Furthermore, the adjacent region of 30 amino acids (position 372–401), which contains the third NLS, is sufficient for nuclear localization. These results suggest that this novel nucleolar protein has at least two distinct domains for directing to different subnuclear destinations.

Nuclear proteins have an important role in nuclear events such as transcription, mRNA processing, replication, and chromosomal organization. Many nuclear proteins have intrinsic signals that mediate active transport across nuclear pore complexes (NPCs). The primary sequences for selective import into the nucleus, known as nuclear localization signals (NLSs), and their cognate receptors, such as the family of karyopherins (Kaps), have been identified (for reviews, see 1–6). The best known NLSs are the classical NLSs, which consist of a short stretch of basic lysine/arginine (K/R)-rich residues (potential consensus: monopartite, $[K/R]_2$-X-[K/R]; bipartite, $[K/R]_2$-$X_{10\text{-}12}$-$[K/R]_3$) and are recognized by karyopherin or importin α/β heterodimer. Recently, the numbers of distinct NLSs and their cognate Kaps have rapidly increased, suggesting that protein transport into the nucleus is mediated by various distinct pathways (1–6).

In contrast, little is known regarding the mechanisms and the topogenic sequences that determine subnuclear localization. The nucleolus is the main site of ribosome biosynthesis (7–9), as well as the locus where several viral core proteins transiently accumulate during viral replication (10–14). In addition, it was recently proposed that the nucleolus is a critical site of cellular aging (15). A number of sequence requirements for nucleolar localization were reported for proteins such as ribosomal proteins L31 and L5 (16, 17), viral proteins Tat and Rev of the human immunodeficiency virus (10–12), non-ribosomal protein nucleolin/C23 (18, 19), NO38/B23/nucleophosmin (19–21), and transcription factor mUBF (22). None of the consensus motifs, however, has been identified and the mechanisms of nucleolar localization are still obscure.

In the present study, we isolated a full-length cDNA encoding a novel nucleolar protein, termed NOLP (Nucleolar Localized Protein), a portion of which was previously screened using a yeast nuclear transportation trap (NTT) system to identify nuclear targeted proteins

[1] To whom correspondence should be addressed at Pharmaceuticals Discovery Laboratory, Mitsubishi Chemical Corporation, Kamoshida-cho 1000, Aoba-ku, Yokohama 227-8502, Japan. Fax: +81-45-963-3992. E-mail: goose@rc.m-kagaku.co.jp.

Abbreviations used: NLS, nuclear localization signal; EST, expressed sequence tag; EGFP, enhanced green fluorescent protein.

0006-291X/98 $25.00

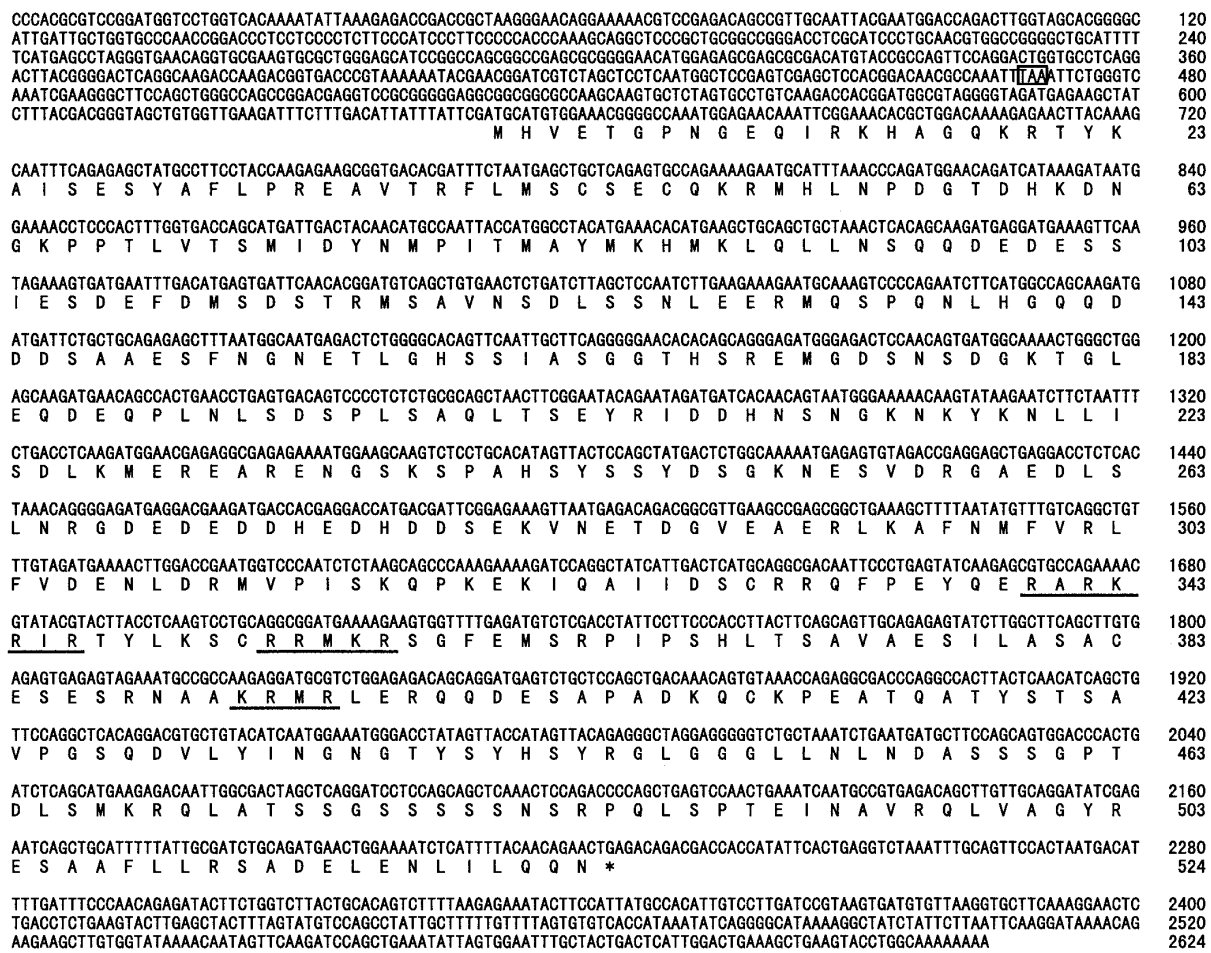

FIG. 1. Nucleotide and predicted amino acid sequence of human NOLP. An in-frame termination codon (position 468) preceding the presumed initiation methionine is boxed. Three potential nuclear localization signals are underlined. The nucleotide sequence data shown here has been deposited in the GenBank/EMBL/DDBJ nucleotide sequence databases under GenBank Accession No. AB017800.

(Ueki *et al.*, manuscript submitted). Its tissue-specific expression and the presence of two distinct domains for nucleolar localization and for nuclear localization was also revealed.

MATERIALS AND METHODS

Cloning of NOLP cDNA. A combination of conventional polymerase chain reaction (PCR) and rapid amplification of cDNA ends (RACE) was used to clone the full-length NOLP cDNA from the SUPERSCRIPT human fetal brain cDNA Library (Life Technologies, Inc.).

Tissue distribution analysis. Primers (NU174 [5′-TTT GAA TTC CAA TCT TGA AGA AAG AAT GCA AAG TCC-3′], NU162 [5′-TTT GTC GAC GGC ACG CTC TTG ATA CTC AGG-3′]) were used for reverse transcription (RT)-PCR. The cDNA templates were synthesized from each human tissue poly(A)+ RNA (Clontech) with excess Superscript II reverse transcriptase (Life Technologies, Inc.) and random hexamer primers. Thirty cycles of PCR amplification were performed (20 s at 95°C and 1 min at 62°C) in 10-μl reactions. PCR products were resolved on a 2.5% Nusive GTG agarose gel (FMC, Rockland, ME) with a 1-kb ladder DNA marker (Life Technologies, Inc.).

Plasmid construction. A series of EGFP-fusion expression plasmids were constructed by cloning E*co*RI/N*ot*I- or E*co*RI/S*al*I-digested DNA fragments into the corresponding sites of pEGFP-C1′ that had been created by inserting a linker (oligonucleotides [5′-GAT CTG GAA TTC ATA TGG CCA TGG CGG CCG CTG CA-3′]/[5′-GCG GCC GCC ATG GCC ATA TGA ATT CCA -3′]) between the *Bgl*II/*Pst*I sites of pEGFP-C1 (CLONTECH). The DNA fragments encoding 222–524 were excised with *Eco*RI/*Not*I from the partial cDNA previously isolated. The others were amplified from the full-length NOLP cDNA using PCR, followed by *Eco*RI/*Sal*I digestion. Primer pairs used were NU181/NU204 for 1–524, NU205/NU204 for 12–524, NU131/NU204 for 282–524, NU131/NU166 for 282–461, NU131/NU164 for 282–401, NU131/NU163 for 282–371, NU131/NU162 for 282–341, NU133/NU164 for 342–401, NU133/NU163 for 342–371, and NU134/NU164 for 372–401: NU181 [5′-TTT GAA TTC TCA TGC ATG TGG AAA CGG GGC CAA ATG GAG-3′], NU204 [5′-TTT GTC GAC TGA ATA TGG TGG TCG TCT GTC TCA GTT CTG-3′], NU205 [5′-TTT GAA TTC GGA AAC ACG CTG GAC-3′], NU131 [5′-TTT GAA TTC TTA AAG TTA ATG AGA CAG ACG GCG TTG AAG C-3′], NU166 [5′-TTT GTC GAC TCC ACT GCT GGA AGC ATC ATT CAG-3′], NU164 [5′-TTT GTC GAC ATC CTG CTG TCT CTC CAG ACG-3′], NU163 [5′-TTT GTC GAC AGT AAG GTG GGA AGG AAT AGG-3′], NU162 [5′-TTT GTC GAC GGC ACG CTC TTG ATA CTC AGG-3′]. A plasmid pEGFP-′lacZ was constructed by cloning a mod-

FIG. 2. Sequence alignment of the *Tra*I homologous region. *E. coli Tra*I/relaxase (1677–1715), human NOLP (7–54), and a part of the Drosophila hypothetical protein translated from genomic DNA (Accession No. AC004360) are shown. Identical and similar amino acid residues are represented by * and /, respectively. Highly conserved residues in the core consensus are boxed.

ified β-galactosidase fragment (the ′lacZ contains a synthetic *Bgl*II linker at the 5′-end and its 3′-end was truncated at the *Eco*RI site) into the *Bgl*II/*Eco*RI sites of pEGFP-C1′ to express an in-frame EGFP-′lacZ fusion. The PCR products encoding 342–371 and 372–401 described above were ligated into the corresponding sites of the pEGFP-′lacZ to express in-frame fusion constructs of EGFP-′lacZ. All constructs were verified by DNA sequencing.

Fluorescence studies in COS-7 cells. COS-7 cells were maintained with Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum, penicillin G (100 U/ml), and streptomycin (100 mg/ml). Plasmids to express EGFP fusion proteins were introduced into COS-7 cells with Lipofectamine Plus Reagent (Life Technologies, Inc.) according to the manufacturer's instructions. The fluorescence was visualized using fluorescence microscopy with an Axiovert 135 (Carl Zeiss, Tokyo, Japan) microscope, and the images were collected from a ZVS-3C75DE (Carl Zeiss) device through a DIGITAL IMAGE FILE DF-30M (Fujifilm, Tokyo, Japan).

RESULTS AND DISCUSSION

Isolation of full-length NOLP cDNA. A partial NOLP cDNA previously isolated (accession number AB015339) was used for cloning the full-length cDNA from a human fetal brain cDNA library. The cloned cDNA of 2624 bp reveals that the presumed initiation methionine is preceded in the cDNA by an in-frame termination codon, and thus a single open reading frame (ORF) of 1572 bp is predicted (Fig. 1). The ORF encodes a deduced polypeptide of 524 amino acids (aa) with a calculated molecular mass of 58 kDa. The protein has an acidic aspartic acid/glutamic acid-rich region (79%, aa 268–281), three putative base-rich NLSs (NLS1, RARKRIR, aa 340–346; NLS2, RRMKR, aa 353–357; NLS3, KRMR, aa 392–395), a serine-rich region (42%, aa 458–498), and a potential coiled-coil domain (23) (aa 492–524). No known RNA-binding motifs, such as the ribonucleoprotein (RNP) motif (24), were detected. Although highly acid-rich residues and base-rich residues are frequently found in several nucleolar proteins, a search of the amino acid sequences in the protein databases (NCBI Protein Database and Swiss Protein Database) revealed no significant similarities with any known protein. A local homology (50% sequence identity in 36 aa overlap) was discovered between the N-terminal short residues (aa 7–40) of NOLP and the C-terminal portion (aa 1677–1715) of the *E. coli* DNA helicase, *Tra*I/relaxase (Fig. 2). TraI/relaxase has an essential role in the initiation and termination of conjugative DNA transfer (25). Whereas its critical regions for catalytic activities are proposed to reside in the N-terminus (26), the role of the C-terminal region is unknown. In addition, a highly similar (71% identity in 41 aa overlap) hypothetically translated sequence was found from a Drosophila genomic DNA (accession number AC004360), suggesting that this motif might be well-conserved throughout evolution (Fig. 2).

A search of the GenBank/EMBL/DDBJ databases using the BLAST program (27) revealed no overall similarity with known genes, but a number of highly similar human EST sequences, most of which are identical to a unique gene (Hs. 6414) in the human UniGene database. According to the database, it is mapped to chromosome 18q12 (between markers D18S457 and D18S1124), and its identical EST clones have been derived from brain, eye, testis, and a small cell lung cancer (SCLC) cell line.

Tissue distribution of NOLP. To investigate the expression profiles of NOLP, the level of mRNA expression in 12 tissues was examined using RT-PCR. A PCR product of the expected size (664 bp) was detected exclusively in fetal brain, brain, and testis (Fig. 3). Northern blot analysis revealed a single transcript of approximately 3.5 kb in brain and testis (data not shown). These results are partially consistent with the information from the database searches. Notably, no transcript signals were detectable in normal lung using either RT-PCR or Northern blot analysis, whereas several identical EST clones (Accession Nos. AA121747, AA129951, AA626487, and AA127451) were derived from an SCLC cell line (NCI-H69). This observation raises the possibility that NOLP is involved in neuroendocrine tumorigenesis, and thus could be a potential oncogene or cancer-associated marker.

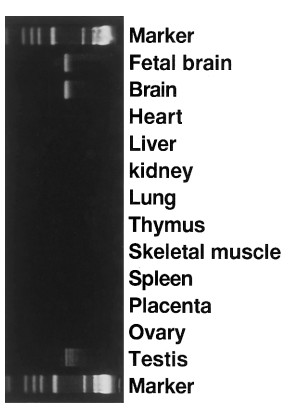

FIG. 3. Tissue distribution analysis using RT-PCR. The 12 tissues examined are indicated above each lane.

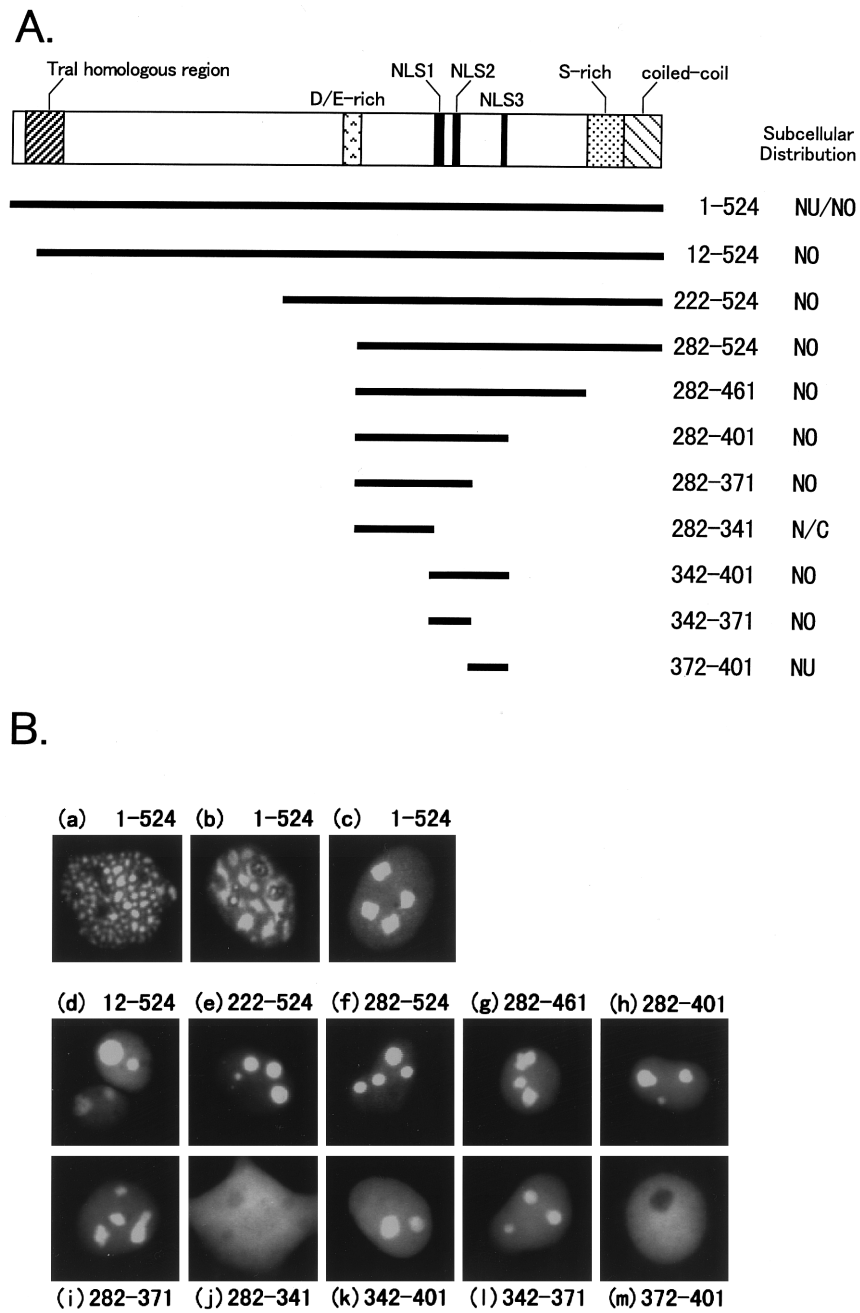

FIG. 4. Domain requirements for nucleolar localization of NOLP protein. (A) Schematic description of deletion constructs. A schematic drawing of the predicted domains of NOLP protein is given in the top panel: D/E-rich, aspartic acid/glutamic acid-rich residues; NLS1, NLS2, NLS3, potential nuclear localization signals; S-rich, serine-rich residues. The subcellular distribution patterns of each construct as determined in B are summarized on the right: NU/NO, some are nuclear and some are nucleolar; NO, constitutive nucleolar; N/C, nonspecific (throughout the cell); NU, constitutive nuclear. (B) Subcellular localization of EGFP-hybrid proteins after expression in COS-7 cells. Transfected cells were examined after 24 h.

Determination of sequence requirements for nucleolar and nuclear localization. To initiate our studies on the structural elements of NOLP, we wanted to define the sequence requirements that confer a specific subcellular distribution. To do so, a series of EGFP-tagged deletion constructs of NOLP were made and transiently expressed in COS-7 cells. The subcellular distribution was then examined (Fig. 4). Whereas the originally isolated portion (222–524) conferred constitutive nucleolar localization (Fig. 4, e), the intact NOLP (1–524) had three major distribution patterns, such as speckled with nucleolar exclusion, patchy, and nucleolar (Fig. 4, a, b, c), along with increased deformation of the nuclei. It is currently unclear, however,

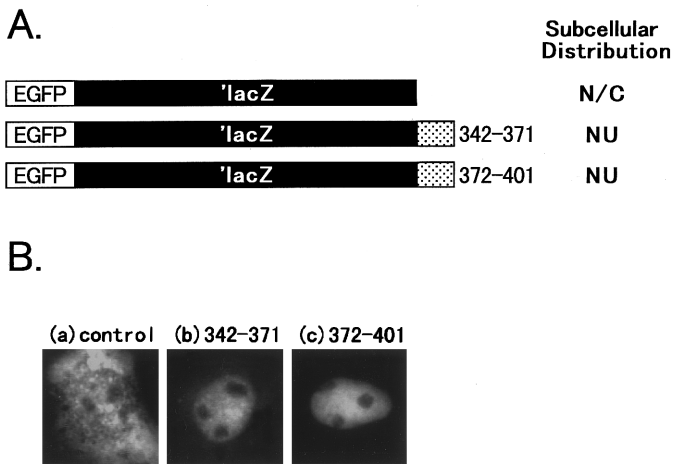

FIG. 5. Construction and analysis of EGFP-β-galactosidase hybrid proteins. (A) Schematic description of fusion protein constructs. The subcellular distribution patterns of each constructs as determined in B. are summarized on the right: N/C, nonspecific (throughout the cell); NU, constitutive nuclear. (B) Subcellular localization of EGFP hybrid proteins after expression in COS-7 cells. (a) EGFP-′lacZ (control), (b) EGFP-′lacZ-[342–371], (c) EGFP-′lacZ-[372–401]. Transfected cells were examined after 24 h.

why these diversities in nuclear structure should occur. In contrast, residues 12–524, which lack the N-terminal 11 residues, did not present such distribution patterns, showing constitutive nucleolar localization (Fig. 4, d). These results suggests that the distal N-terminal 11 residues, a part of which overlap with the TraI homologous region, are necessary for proper localization and function of NOLP.

As for constitutive nucleolar targeting, further studies revealed that neither deletions of the N-terminal moiety (aa 1–281) that includes acid-rich residues, nor the C-terminal moiety (aa 372–524) affected nucleolar localization (Fig. 4B, f, i). Further deletion of aa 342–371 completely abolished the exclusive nucleolar accumulation (Fig. 4B, j), presenting the same pattern as EGFP itself (data not shown), and conversely, aa 342–371 alone could still induce nucleolar localization (Fig. 4B, k). Therefore, we conclude that the 30-amino acid residue (aa 342–371), where the NLS1 and the NLS2 exist, is sufficient for nucleolar localization. Furthermore, the adjacent 30 amino acid residues (aa 372–401), where the NLS3 resides, serves as an active NLS (Fig. 4B, j).

To examine whether aa 342–371 could direct a known reporter protein to the nucleolus, and also whether aa 372–401 could direct it to the nucleus, we fused each peptide to the C-terminus of EGFP-β-galactosidase (EGFP-′lacZ) hybrid protein. Whereas aa 372–401 was sufficient for nuclear localization (Fig. 5B, c), aa 342–371 directed the chimera protein to the nucleus, but not to the nucleolus, showing an apparent exclusion from the nucleolus (Fig. 5B, b). This might be due to a folding artifact that inhibits proper nucleolar localization. The reason for this is currently unclear, however the inability of transferring a nucleolar localization signal onto a heterologous reporter protein was previously noted for other nucleolar proteins (17, 28, 29).

In the present study, we isolated a novel nucleolar protein, examined its tissue distribution, and identified two distinct domains for nucleolar localization and nuclear localization. Because of its tissue-specific distribution, it might be involved in a process of determining cell specificity. Although the biological function(s) of NOLP require further study, it would be worthwhile, as a first step, to identify functional elements and dissect their functions. The nucleolar transport mechanisms of NOLP remain to be elucidated as well.

ACKNOWLEDGMENT

We greatly thank all of the staff at HRI for supporting our work.

REFERENCES

1. Görlich, D., and Mattaj, I. W. (1996) *Science* **271,** 1513–1518.
2. Nigg, E. A. (1997) *Nature* **386,** 779–787.
3. Ullman, K. S., Powers, M. A., and Forbes, D. J. (1997) *Cell* **90,** 967–970.
4. Pemberton, L. F., Blobel, G., and Rosenblum, J. S. (1988) *Curr. Opin. Cell Biol.* **10,** 392–399.
5. Wozniak, R. W., Rout, M. P., and Aitchison, J. D. (1988) *Trends Cell Biol.* **8,** 184–188.
6. Ohno, M., Fornerod, M., and Mattaj, I. W. (1998) *Cell* **92,** 327–336.
7. Hadjiolov, A. A. (1985) *in* The Nucleolus and Ribosome Biogenesis, Springer Verlag, New York.
8. Scheer, U., and Weisenberger, D. (1994) *Curr. Opin. Cell Biol.* **6,** 354–359.
9. Pederson, T. (1988) *Nucleic Acids Res.* **26,** 3871–3877.
10. Dang, C. V., and Lee, W. M. (1989) *J. Biol. Chem.* **264,** 18019–18023.
11. Cochrane, A. W., Perkins, A., and Rosen, C. A. (1990) *J. Virol.* **64,** 881–885.
12. Hatanaka, M. (1991) *in* Genetic Structure and Regulation of HIV (Haseltine, W. A., and Wong-Staal, F., Eds.), pp. 264–287, Raven Press, New York.
13. Siomi, H., Shida, H., Nam, S. H., Nosaka, T., Maki, M., and Hatanaka, M. (1988) *Cell* **55,** 197–209.
14. Mears, W. E., Lam, V., and Rice, S. A. (1995) *J. Virol.* **69,** 935–947.
15. Johnson, F. B., Marciniak, R. A., and Guarente, L. (1998) *Curr. Opin. Cell Biol.* **10,** 332–338.
16. Quaye, I. K., Toku, S., and Tanaka, T. (1996) *Eur. J. Cell Biol.* **69,** 151–155.
17. Michael, W. M., and Dreyfuss, G. (1996) *J. Biol. Chem.* **271,** 11571–11574.
18. Lapeyre, B., Bourbon, H., and Amalric, F. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1472–1476.
19. Busch, H. (1984) *in* Chromosomal Nonhistone Proteins (Hnilica, L. S., Ed.), Vol. 4, pp. 233–286, CRC Press, Boca Raton, FL.
20. Schmidt-Zachmann, M. S., Hugle-Dorr, B., and Franke, W. W. (1987) *EMBO J.* **6,** 1881–1890.
21. Chan, W. Y., Liu, Q. R., Borjigin, J., Busch, H., Rennert, O. M., Tease, L. A., and Chan, P. K. (1989) *Biochemistry* **28,** 1033–1039.

22. Maeda, Y., Hisatake, K., Kondo, T., Hanada, K., Song, C. Z., Nishimura, T., and Muramatsu, M. (1992) *EMBO J.* **11,** 3695–3704.
23. Lupas, A. (1996) *Methods Enzymol.* **266,** 513–525.
24. Burd, C. G., and Dreyfuss, G. (1994) *Science* **265,** 615–621.
25. Byrd, D. R., and Matson, S. W. (1997) *Mol. Microbiol.* **25,** 1011–1022.
26. Balzer, D., Pansegrau, W., and Lanka, E. (1994) *J. Bacteriol.* **176,** 4285–4295.
27. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
28. Peculis, B. A., and Gall, J. G. (1992) *J. Cell Biol.* **116,** 1–14.
29. Schmidt-Zachmann, M. S., and Nigg, E. A. (1993) *J. Cell Sci.* **105,** 799–806.