A Novel Heterogeneous Nuclear RNP Protein with A Unique Distribution on Nascent Transcripts

Serafín Piñol-Roma, Maurice S. Swanson, Joseph G. Gall,* and Gideon Dreyfuss

Department of Biochemistry, Molecular Biology, and Cell Biology, Northwestern University, Evanston, Illinois 60208; and *Department of Embryology, Carnegie Institution, Baltimore, Maryland 21210

Abstract. Immediately after the initiation of transcription in eukaryotes, nascent RNA polymerase II transcripts are bound by nuclear proteins resulting in the formation of heterogeneous nuclear ribonucleoprotein (hnRNP) complexes. hnRNP complexes from HeLa cell nuclei contain >20 major proteins in the molecular mass range of 34,000–120,000 D. Among these are the previously described A, B, and C groups of proteins (34,000–43,000 D) and several larger, and as yet uncharacterized, proteins. Here we describe the isolation and characterization of a novel hnRNP protein termed the L protein (64–68 kD by mobility in SDS-polyacrylamide gels). Although L is a bona fide component of hnRNP complexes, it also appears to be a different type of hnRNP protein from those previously characterized. A considerable amount of L is found outside hnRNP complexes, and monoclonal antibodies to the L protein also strongly stain unidentified discrete nonnucleolar structures, in addition to nucleoplasm, in HeLa cell nuclei. Interestingly, the same antibodies stain the majority of nonnucleolar nascent transcripts from the loops of lampbrush chromosomes in the newt, but the most intense staining is localized to the landmark giant loops. The L protein is the first protein of giant loops identified so far, and antibodies to it thus provide a useful tool with which to study these unique RNAs. In addition, isolation and sequencing of cDNA clones for the L protein from human cells predicts a glycine- and proline-rich protein of 60,187 D, which contains two 80 amino acid segments only distantly related to the RNP consensus sequence-type RNA-binding domain. The L protein, therefore, is a new type of hnRNP protein.

HETEROGENEOUS nuclear ribonucleoprotein (hnRNP)[1] complexes are the structures that contain hnRNAs and their associated proteins. These complexes, rather than naked RNA, provide the substrate for the processing events that pre-mRNAs undergo before becoming functional, translatable mRNAs in the cytoplasm (for reviews see Dreyfuss, 1986; Chung and Wooley, 1986). It is thus essential to understand the composition and structure of these complexes to gain further insights into the posttranscriptional pathway of gene expression. To accomplish this, we have produced antibodies to several hnRNP proteins and used them as specific probes for investigating the structure and the function of hnRNP complexes. Antibodies to hnRNP proteins have proven to be invaluable tools to study the intracellular localization of specific hnRNP proteins and their mode of association with hnRNP complexes (Leser et al., 1984; Dreyfuss et al., 1984*b*; Choi and Dreyfuss, 1984*a*,*b*), to study the involvement of specific proteins in pre-mRNA processing events such as splicing (Choi et al., 1986; Sierakowska et al., 1986), to probe their mode of interaction with specific RNA sequences (Swanson and Dreyfuss, 1988*a*,*b*), and to obtain cDNA clones for these proteins (Nakagawa et al., 1986; Swanson et al., 1987) with the subsequent structural information afforded by the deduced amino acid sequences.

Moreover, immunopurification with monoclonal antibodies to hnRNP proteins has proven to be an effective way of purifying hnRNP complexes away from other cellular structures in nucleoplasm (Choi and Dreyfuss, 1984*a*; Piñol-Roma et al., 1988), and has allowed a detailed analysis of their protein constituents. Two-dimensional gel electrophoresis of such complexes revealed that >20 different major polypeptides in the molecular mass range of 34,000 to 120,000 D are associated with the hnRNA in HeLa cells. While several of the components previously identified in sucrose gradient–isolated 30–40S hnRNP particles (Martin et al., 1974; Beyer et al., 1977; Karn et al., 1977; Jacob et al., 1981; Wilk et al., 1985) are abundant in immunopurified complexes, numerous additional components are also apparent. Among these is a prominent set of proteins of ~68 kD and neutral isoelectric point which is comprised of two distinct proteins designated L and M (Piñol-Roma et al., 1988). Proteins of similar molecular masses are prominent among those induced to cross-link to hnRNA by ultraviolet light in

1. *Abbreviations used in this paper*: CS-RBD, ribonucleoprotein consensus type RNA binding domain; GV, germinal vesicle; hnRNP, heterogeneous nuclear ribonucleoprotein; RNP-CS, ribonucleoprotein consensus sequence; ss, single stranded.

vivo (Mayrand et al., 1981; Dreyfuss et al., 1984*b*) indicating they are bound to hnRNA in the cell.

To facilitate the production of monoclonal antibodies to these novel hnRNP proteins, we have used the ability of most hnRNP proteins to bind single-stranded (ss)DNA in a heparin- and salt-resistant manner (Pandolfo et al., 1987; Piñol-Roma et al., 1988). Here we characterize one of these abundant higher molecular mass hnRNP proteins, the L protein. In comparison to other hnRNP proteins characterized so far, L is unique in its distribution on nascent transcripts and in its amino acid sequence. Of particular interest is the finding of an association of L with a distinct set of nascent transcripts—those of the landmark giant loops of amphibian lampbrush chromosomes. L is the first protein found to localize to these nuclear structures, and it is likely that the antibodies to L that have been produced will make it possible to learn more about the giant loops.

Materials and Methods

Cell Culture, Labeling, and Cell Fractionation

HeLa S3 and the HeLa monolayer-adapted clone JW36 cells were cultured in monolayer to subconfluent densities in DME, supplemented with penicillin and streptomycin, and containing 10% calf serum at 37°C. Cells were labeled with [^{35}S]methionine at 20 μCi/ml for 20 h in DME containing one-tenth the normal methionine level and 5% calf serum. The nucleoplasmic fraction was prepared essentially according to Pederson (1974), as previously detailed (Choi and Dreyfuss, 1984*a*).

RNase Digestion

Digestions of the nucleoplasmic fraction were carried out with micrococcal nuclease (Pharmacia LKB Biotechnology, Piscataway, NJ) at the indicated concentrations for 10 min at 30°C, in the presence of 1 mM $CaCl_2$. The reactions were stopped by adding EGTA to a final concentration of 5 mM.

Preparation of Monoclonal Antibodies

The monoclonal antibodies 4F4, to the C proteins, and 4B10, to the A1 protein, were prepared as described previously (Choi and Dreyfuss, 1984*b*; Piñol-Roma et al., 1988). The anti-L protein monoclonal antibody 4D11 was obtained by immunization of a BALB/c mouse with hnRNP proteins purified by affinity chromatography on ssDNA agarose (Piñol-Roma et al., 1988). The antigen was composed of proteins eluting from the column at 2 M NaCl after a heparin wash at 1 mg/ml in 100 mM NaCl. Hybridoma production and screening were as previously detailed (Choi and Dreyfuss, 1984*b*).

Immunopurification of Proteins and hnRNP Complexes

The hnRNP complex was immunopurified from the nucleoplasm as described previously (Choi and Dreyfuss, 1984*a*) for 10 min at 4°C with the anti-C proteins monoclonal antibody 4F4 or anti-L monoclonal antibody 4D11 bound to protein A-agarose. Rabbit anti-mouse IgG antiserum was used with the 4D11 antibody, since 4D11 does not bind protein A directly. The same secondary antiserum was included with all the SP2/0 nonimmune controls. Ascites fluid from a BALB/c mouse that was inoculated intraperitoneally with the parent myeloma line SP2/0 was used for the nonimmune control immunopurifications with each experiment. Antibody specificities were confirmed by immunoblotting and by immunopurification in the presence of the ionic detergent Empigen BB at 1%, 1 mM EDTA, and 0.1 mM DTT as described (Choi and Dreyfuss, 1984*b*).

Gel Electrophoresis and Immunoblotting

Protein samples were subjected to electrophoresis on an SDS-containing discontinuous PAGE system (SDS-PAGE) (Dreyfuss et al., 1984*a*). The separating gel had a final acrylamide concentration of 12.5%. After electrophoresis of [^{35}S]methionine-labeled proteins, the gel was stained with Coomassie Blue and impregnated with 2,5-diphenyloxazole for fluorography (Laskey and Mills, 1975). Two-dimensional NEPHGE was carried out by the procedure of O'Farrell et al. (1977). The first dimension was separated by using pH 3–10 ampholine gradients for 4 h at 400 V, and the second dimension was by SDS-PAGE as described above. Immunoblotting procedures were as described previously (Choi and Dreyfuss, 1984*b*).

Sucrose Gradient Sedimentation

Sucrose gradients (10–30% [wt/vol]) were sedimented in a rotor (model SW41; Beckman Instruments Inc., Palo Alto, CA) at 38,000 rpm for 5 h at 4°C. 22 fractions (0.6 ml each) were collected from the bottom, and proteins in each fraction were precipitated with trichloroacetic acid added to a final concentration of 10% for analysis by SDS-PAGE and immunoblotting as described above. Sucrose solutions were made up of 10 mM Tris-HCl (pH 7.4) buffer containing 100 mM NaCl and 2.5 mM $MgCl_2$. The 28S ribosomal marker was from phenol-extracted cytoplasmic fraction sedimented similarly in the presence of 10 mM EDTA.

Immunofluorescence Microscopy of Human Cells

Immunofluorescence microscopy was essentially as previously described (Dreyfuss et al., 1984*b*). Monolayer-adapted HeLa cells (clone JW36), cultured on glass coverslips, were fixed with 2% formaldehyde in PBS for 30 min at room temperature, followed by permeabilization with acetone at −20°C for 3 min. Ascites fluid dilutions were 1:1,000 for both 4D11 and the anti-A1 monoclonal antibody 4B10 (Piñol-Roma et al., 1988). Detection of the mouse antibodies was with fluorescein isothiocyanate–conjugated goat anti-mouse F(ab')2 (Cappel Laboratories, Malvern, PA) used at a 1:50 dilution in 1% BSA in PBS. Immunofluorescence was carried out with a Zeiss Photomicroscope III.

Lampbrush Chromosome Preparation and Immunofluorescent Staining

Lampbrush chromosomes from the newt *Notophthalmus viridescens* were prepared as previously described (Gall et al., 1981) with the addition of 1 mM $MgCl_2$ to both the isolation and spreading solutions. After centrifugation to attach the chromosomes to the slide, preparations were fixed for 30 min in 4% HCHO in amphibian Ringer (Lacroix et al., 1985). They were rinsed in PBS and carried through the immunofluorescence procedure as previously described (Roth and Gall, 1987) using rhodamine-labeled goat anti-mouse IgG (1:200) as the second antibody.

RNA–Protein Cross-linking in Intact Cells

Photochemical RNA-protein cross-linking by UV light irradiation of cells on culture dishes and isolation and analysis of RNPs was carried out as previously described (Dreyfuss et al., 1984*a*,*b*).

Isolation of cDNA Clones and Affinity Purification of Antibodies

Mouse antisera (1:250 dilution) were used to directly screen a λgt11 HeLa cell cDNA library (Nakagawa et al., 1986). Positive plaques were purified, plated at high density (5×10^4 phage/100-mm plate), and filter replicated onto 82-mm nitrocellulose filters (Schleicher & Schuell, Inc., Keene, NH). These filters were used to affinity purify antibodies from the total mouse antisera using the technique of Snyder et al. (1987). Those antibodies affinity selected by the filter-bound fusion proteins were then used to screen Western blots of total HeLa cell proteins to identify L protein clones. A full-length clone, pHCL3, was subsequently isolated using the original L protein clone, pHCL1, as a hybridization probe.

RNA Blot Analysis, Hybrid Selection, and In Vitro Translation

Poly(A)$^+$ RNA was prepared from HeLa S3 cells as previously described (Nakagawa et al., 1986), resolved by electrophoresis on formaldehyde-containing 1.4% agarose gels (Lehrach et al., 1977), and the fractionated RNA blotted onto nitrocellulose (Maniatis et al., 1982). Both pHCL1 and pHCL3, and various subfragments, were used as hybridization probes, and were prepared by nick translation with [^{32}P]dCTP (Rigby et al., 1977). Hybridization selection and in vitro translations were performed as previously described (Nakagawa et al., 1986; Swanson et al., 1987) using either pHC12, a clone for the human C proteins, or pHCL1.

DNA Sequence Analysis

Overlapping restriction fragments of pHCL1 and pHCL3 were subcloned into M13mp18 and M13mp19 and sequenced by the dideoxy chain termination method (Sanger et al., 1977) using deaza-GTP and the Klenow fragment of DNA Polymerase I as described previously (Swanson et al., 1987). Sequence analysis was performed with the University of Wisconsin Genetics Computer Group (UWGCG) Sequence Analysis Programs. The universal sequence database searching programs FASTA and TFASTA (Pearson and Lipman, 1988) were used to search six databases for sequence similarities.

Results

L Protein Is a Component of Immunopurified hnRNP Complexes

The protein composition of hnRNP complexes isolated from nucleoplasm of human HeLa cells with monoclonal antibodies to the hnRNP A1 or C proteins has been described previously (Choi and Dreyfuss, 1984*a*; Piñol-Roma et al., 1988). These complexes are composed of ~20 major proteins in the molecular mass range of 34,000–120,000 D as resolved by two-dimensional gel electrophoresis, and these are designated alphabetically as shown in Fig. 1 *left* (see Piñol-Roma et al., 1988 for detailed discussion). A major component of hnRNP complexes obtained by immunopurification is a group of proteins of ~64–68 kD, referred to as L and M. Proteins of similar molecular mass are some of the major proteins that become cross-linked to RNA by UV light in vivo (Dreyfuss et al., 1984*b*), and therefore it was of interest to produce specific probes for these proteins to facilitate their study. Antibodies to the L protein were generated by immunizing mice with fractions containing HeLa hnRNP proteins partially purified by affinity chromatography on ssDNA–agarose, a procedure that enables the large scale purification of most hnRNP proteins (Piñol-Roma et al., 1988). Reactivity towards several of these proteins was observed, and a mouse that showed good response against the L protein was used for production of monoclonal antibodies. The specificity of the monoclonal antibody produced by one of the stable hybridomas, 4D11, for the L protein is shown by two-dimensional gel electrophoresis of material immunoprecipitated from total HeLa cell proteins in the presence of the ionic detergent Empigen BB (Choi and Dreyfuss, 1984*b*) (Fig. 1 *right*). A comparison of the proteins immunopurified in Fig. 1 *right* and *left* also demonstrates that 4D11 only recognizes a subset of the proteins in the region of L and M, hence the distinction made among proteins in this region. Several isoelectric forms, probably the result of posttranslational modifications, are seen for both L and M. A series of monoclonal antibodies specific for M (Adam, S., S. Piñol-Roma, and G. Dreyfuss, unpublished data), as well as polyclonal antisera against L, confirm the immunological relatedness of the proteins within each group, and the lack of immunological cross-reactivity between the two groups.

Further evidence that L is an authentic component of hnRNP complexes was obtained by using 4D11 in immunopurification experiments starting with HeLa nucleoplasm with or without addition of ionic detergent. In the presence

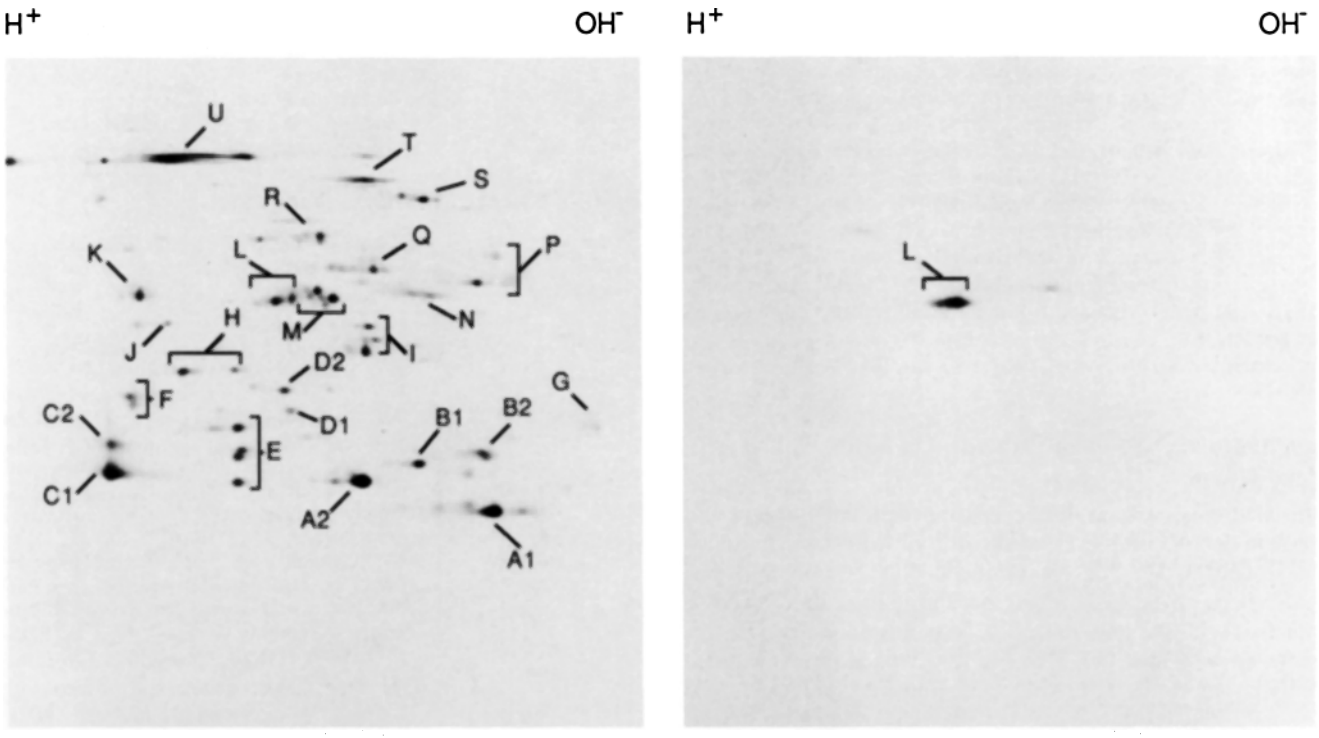

Figure 1. Two-dimensional gel electrophoresis of immunopurified hnRNP complexes and immunopurified L protein. (*Left*) hnRNP complexes were immunopurified from [^{35}S]methionine-labeled HeLa nucleoplasm with the anti-C proteins monoclonal antibody 4F4 as described in the text. The immunopurified complexes were resolved by two-dimensional gel electrophoresis, with NEPHGE in the first dimension, and SDS-PAGE in the second dimension. The proteins were visualized by autoradiography. (*Right*) L protein was immunopurified from [^{35}S]methionine-labeled HeLa cells with the 4D11 monoclonal antibody, in the presence of the ionic detergent Empigen BB in order to dissociate protein–RNA and protein–protein interactions. The immunopurified protein was then resolved by two-dimensional gel electrophoresis simultaneously with the sample shown on the left to allow for direct comparison of electrophoretic behavior of the individual proteins.

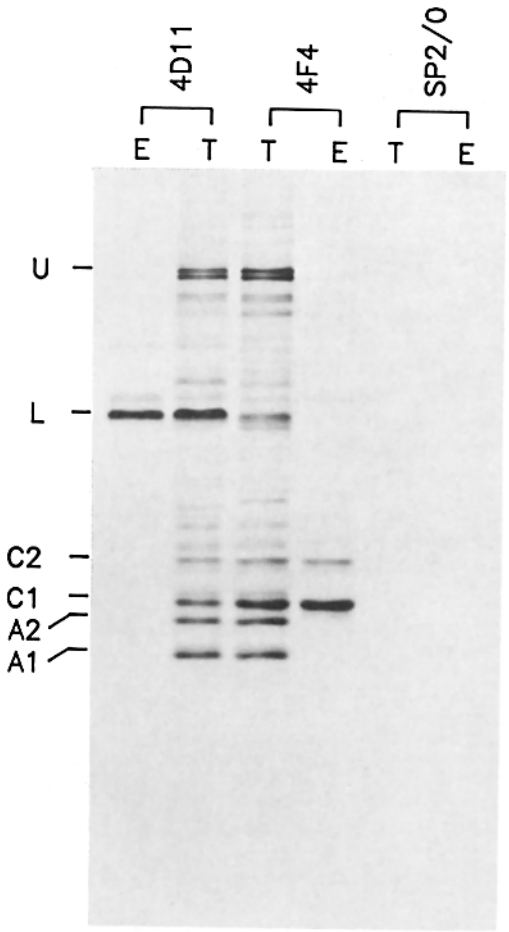

Figure 2. Immunopurification of hnRNP complexes with antibodies to the hnRNP L protein (*4D11*) or the hnRNP C proteins (*4F4*). The hnRNP complexes were immunopurified from the nucleoplasm of [^{35}S]methionine-labeled HeLa cells in the presence of the nonionic detergent Triton X-100 at 0.5% (lanes *T*). Antigen specificity of the antibodies was demonstrated by immunoprecipitation in the presence of the ionic detergent Empigen BB at 1% (lanes *E*). Control immunoprecipitations were carried out with ascites fluid of a BALB/c mouse inoculated intraperitoneally with the parent myeloma line SP2/0. Gel positions of some of the hnRNP proteins are indicated on the left.

of the ionic detergent Empigen BB, the 4D11 antibody isolates only the L protein (Fig. 2, lane *4D11*, *E*). However, if this detergent, which dissociates most protein–protein and protein–RNA complexes while still allowing efficient antibody–antigen binding, is omitted, 4D11 immunopurifies hnRNP complexes of similar protein composition to those purified with the anti-hnRNP C proteins antibody 4F4 (cf. lanes *4D11*, *T* and *4F4*, *T* in Fig. 2). This indicates a stable association of the L protein with hnRNP complexes. The amount of L immunopurified with 4D11 relative to that of other hnRNP proteins is much higher than that found in complexes isolated with antibodies to other hnRNP proteins. This is most likely due to the fact that there are L proteins outside hnRNP complexes that are also directly bound by the antibody, although it is possible that some L dissociates from the particles during the isolation procedure. Nevertheless, the ability of 4D11 to immunopurify a subset of nuclear proteins, which is virtually identical to that immunopurified under the same conditions with anti-C protein antibodies, indicates that L is part of the same hnRNP complexes.

L Protein Is Cross-linked by UV Light to Poly(A)-containing RNA in Intact Cells

To examine whether L is indeed in contact with hnRNA in the living cell, we induced covalent cross-linking of RNA to proteins that are associated with it in intact cells by irradiation with UV light. This procedure allows selection of RNA with the covalently bound proteins, under conditions that eliminate adventitious association of proteins with the RNA during the fractionation procedures (van Eekelen et al., 1981; Mayrand et al., 1981; Dreyfuss et al., 1984*a*,*b*). HeLa cells grown in monolayer culture were irradiated with UV light, and poly(A)-containing RNA was isolated by affinity chromatography on oligo(dT)-cellulose, after boiling with SDS and β-mercaptoethanol. The proteins that are cross-linked to the selected RNA were released by digestion of the RNA with ribonucleases, and resolved by electrophoresis on SDS-containing polyacrylamide gels. An immunoblot of such a gel probed with the anti-L antibody 4D11 is shown in Fig. 3. It demonstrates that the L protein is bound to

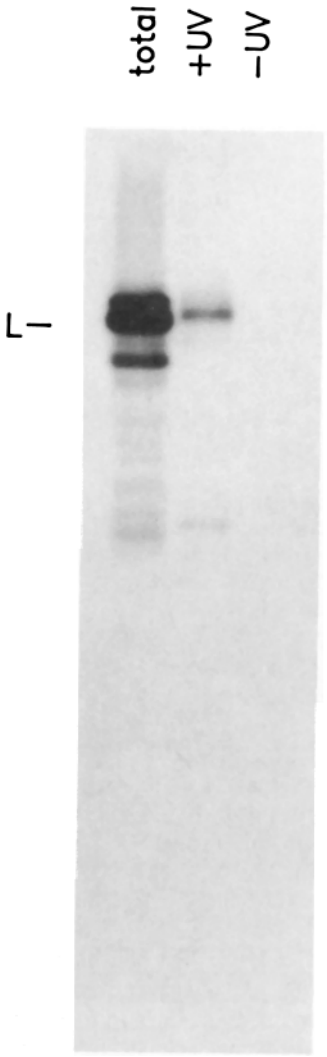

Figure 3. Immunoblot analysis of L protein crosslinked in vivo to poly(A)$^+$ RNA by UV irradiation of intact cells. HeLa cells grown in monolayer were irradiated with UV light and the poly(A)$^+$ RNA was isolated as described in the text and digested with RNAses. The released proteins were resolved by SDS-PAGE, blotted onto nitrocellulose paper, and probed with monoclonal antibody 4D11. Lane *total*, total HeLa proteins; lane *+UV*, proteins cross-linked to poly(A)$^+$ RNA by UV light; lane *−UV*, proteins from samples treated as in lane *+UV*, except that UV irradiation of the cells was omitted.

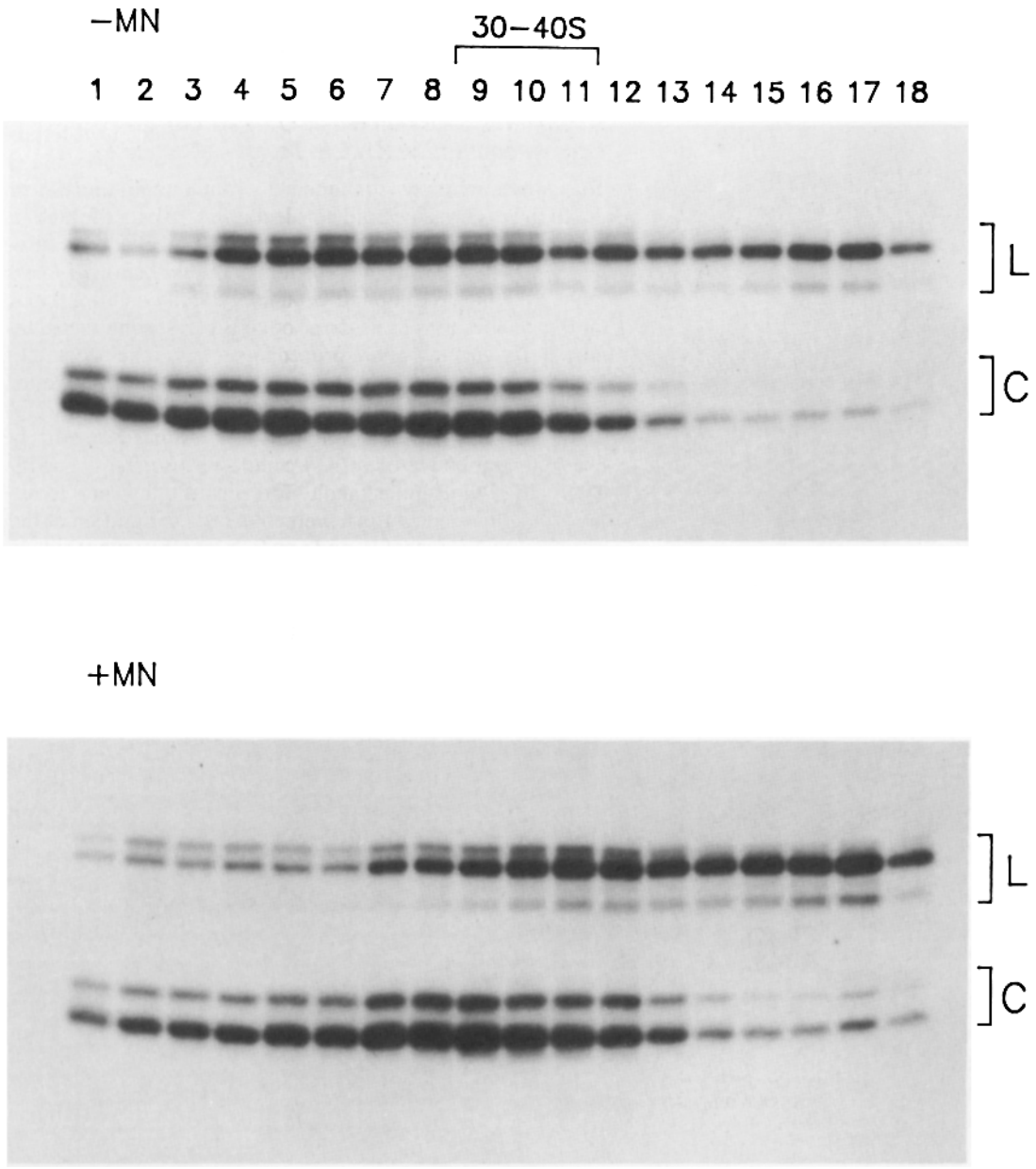

Figure 4. Sucrose gradient sedimentation behavior of the L protein and of the C proteins. Nucleoplasm from HeLa cells was sedimented on 10–30% sucrose gradients as described in the text, after digestion with micrococcal nuclease at 0 (*−MN*) or 2 (*+MN*) U/ml. 600-μl fractions were collected from the bottom. The proteins in each fraction were precipitated with trichloroacetic acid, and the distribution of the L and C proteins was analyzed by SDS-PAGE and immunoblotted with the monoclonal antibodies 4D11 and 4F4 on the same gels. Lane *1* corresponds to the bottom fraction.

poly(A)-containing RNA in the living cell, and thus further supports the conclusion that it is an authentic hnRNP protein. The mobility shift of L in the UV(+) lane (Fig. 3), as well as the diffuse nature of the signal obtained, are characteristic of proteins that have become cross-linked to RNA, and are the result of residual nucleotides that remain covalently bound to the protein even after exhaustive nuclease digestion. The specificity of the cross-linking is further substantiated by the fact that no 4D11-reactive material is detected if the cells are not exposed to UV light before the RNA isolation procedure (Fig. 3, *UV−*). The lower molecular mass bands seen in the total lane in Fig. 3 are most likely proteolysis products of L since their presence is variable between preparations.

L Protein Is Associated with RNAse-sensitive Structures

Given the indications that considerable amounts of L may be found outside of hnRNP complexes that can be immunopurified with antibodies to the C proteins (Fig. 2), the association of the L protein with RNA-containing structures and its relationship to other hnRNP proteins was examined by sucrose gradient sedimentation. The association of hnRNA with hnRNP proteins results in the sedimentation of hnRNP complexes in a heterodisperse manner on sucrose gradients (Samarina et al., 1968). Mild digestion of the RNA by endogenous nucleases or by added RNases results in a shift of the labeled RNA and associated proteins entering the gradient

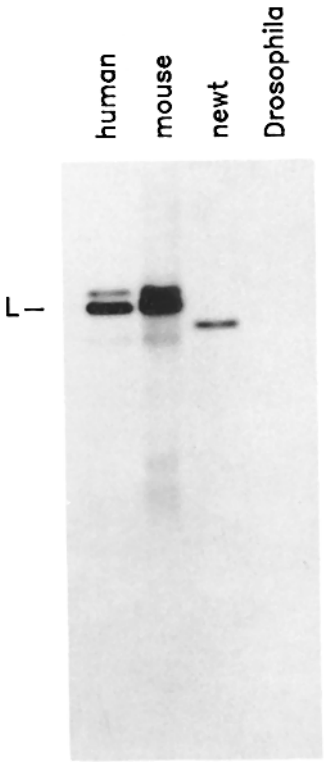

Figure 5. Conservation of the L protein among vertebrate cells. Total cellular proteins from the indicated species were resolved by SDS-PAGE, transferred to nitrocellulose, and probed with the anti-L protein monoclonal antibody 4D11. The mouse proteins were obtained from cultured MEL cells, and the newt proteins were obtained from *N. viridescens* GV.

to a peak at ~30 S, presumably reflecting the generation of hnRNP monoparticles, due to digestion of RNAse-sensitive regions of the RNA-connecting monoparticles in large hnRNP complexes. The 30 S peak contains hnRNA fragments and the bulk of the proteins in these complexes. We analyzed the presence of the L proteins in sucrose gradient fractions after various digestion conditions. This was done by SDS-PAGE of the fractions from the gradients, followed by immunoblotting with the monoclonal antibodies 4D11 and 4F4. As shown in the corresponding immunoblots in Fig. 4 (*−MN*), the L protein sediments throughout the gradient if nuclease treatment of the nucleoplasm is omitted before sedimentation. There appears to be a considerable amount of the protein sedimenting near the top of the gradient, in addition to L protein in hnRNP complexes, that is either free or is part of other structures. The association of the L protein with RNA-containing structures is indicated by the shift in their sedimentation towards the top of the gradient after mild RNAse digestion of the nucleoplasm (Fig. 4, *+MN*). There also appears to be a strong bias towards the preferential appearance of a minor band of L, with a higher apparent molecular mass, towards the bottom of the gradient. This higher molecular mass form of L is also apparent by immunopurification with 4D11 (see Fig. 1 *right* and Fig. 2). The relationship between these two forms of the protein is at present unclear, but it appears that they may associate differentially with RNA-containing complexes, with the higher molecular mass band exhibiting a preferential association with faster sedimenting structures. As explained in the previous section, the lower molecular mass bands that react with mAb 4D11, and which are especially prominent in the +MN panel, are most likely proteolytic fragments of L. As a reference, the sedimentation pattern of the well-characterized hnRNP C proteins is shown by immunodetection with the mAb 4F4 on the same blots.

L Protein Is Conserved in Vertebrate Cells

Because L is a novel hnRNP protein, we wanted to ascertain that it is a general component of hnRNP complexes rather than a protein unique to HeLa cells. Immunoblotting of material from a variety of vertebrate cells using the anti-L monoclonal antibody 4D11 (Fig. 5) indicates that the L protein is found in various vertebrates including *Xenopus laevis* (data not shown) and the newt *Notophthalmus viridescens*. No detectable signal was found in cells of *Drosophila melanogaster* or in the yeast *Saccharomyces cerevisiae* (data not shown). The apparent molecular mass of the signal obtained with 4D11 across such a wide range of organisms is also remarkably conserved. The high degree of cross-reactivity of the 4D11 antibody across vertebrate species makes it possible to carry out studies in other organisms that would otherwise be difficult to perform in human cells, such as those described below using the lampbrush chromosomes of *N. viridescens*.

L Protein Is Found in Nucleoplasm and in Unidentified Nuclear Structures

Immunofluorescence microscopy on human JW36 cells with the mAb 4D11 demonstrates the nuclear localization of these proteins and their absence from the cytoplasm (Fig. 6 *B*). The pattern of staining is similar to that obtained with monoclonal antibodies against other hnRNP proteins, such as with the anti-A1 protein mAb 4B10 (see Fig. 6 *A*), in that the overall staining is nucleoplasmic with the exclusion of nucleoli. However, 4D11 also strongly stains one to three discrete (usually two) nonnucleolar structures that are apparent in all cells. The precise identity of these structures is at present unknown, but they are not observed with anti-C, anti-A1, and anti-U antibodies which show only nucleoplasmic staining. Identical patterns of staining have also been obtained with additional monoclonal antibodies against L, as well as with polyclonal antisera raised against an L-β galactosidase fusion protein (data not shown). We have also observed similar staining in mouse tissue culture cells and tissue sections.

Distribution of the L Protein on Nascent Transcripts

To examine the intranuclear distribution of L in greater detail, we stained lampbrush chromosome preparations of the newt *N. viridescens* with mAb 4D11. The great majority of the lateral loops bound the antibody, suggesting that L is associated with most nascent transcripts on the chromosomes. Binding of 4D11 to the cluster of giant loops near the centromere of chromosome 2 was especially striking (Fig. 7, *A* and *B*). Even when allowance is made for the greater thickness of the loop matrix on the giant loops, it seems probable that the concentration of L is higher in them than in typical loops. The staining of the giant loops by 4D11 is unusual in another respect. Other antibodies that stain typical lampbrush chromosomes stain the giant loops only faintly or not at all (Roth and Gall, 1987). For example, mAb Y12, which is directed against the Sm epitope of snRNPs (Lerner et al., 1981), stains most lampbrush chromosome loops, but leaves the giant loops only slightly above background level (Fig. 7, *C* and *D*).

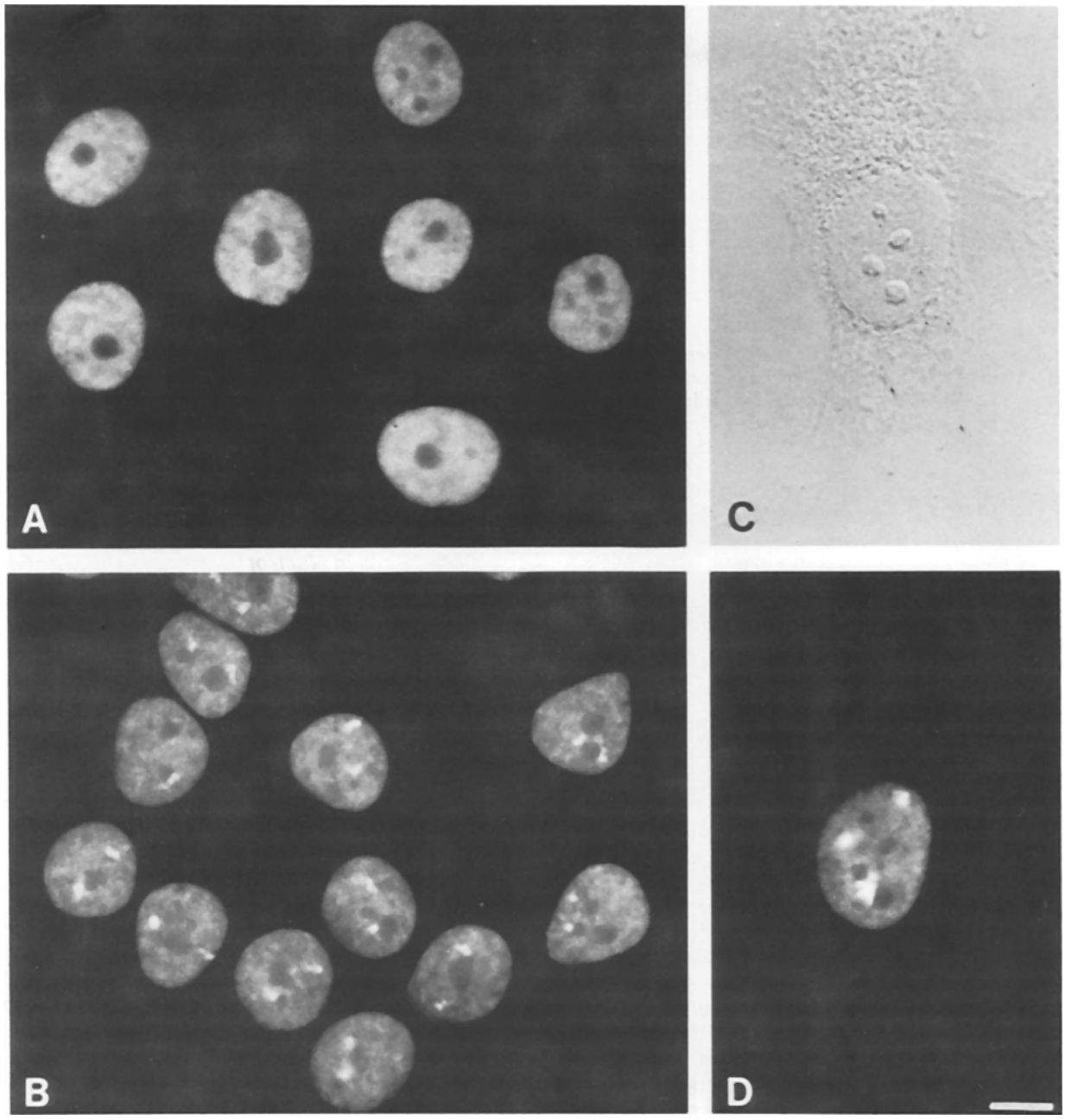

Figure 6. Immunofluorescence microscopy with monoclonal antibodies 4D11 and 4B10 on human cells. Immunofluorescence microscopy of human JW36 cells stained with the anti-A1 monoclonal antibody 4B10 (*A*) and the anti-L monoclonal antibody 4D11 (*B*). *D* shows an individual cell stained with 4D11, with the corresponding Nomarski optics image shown in *C*. Bar, 10 μm.

In the intact germinal vesicle (GV) the giant loops on chromosome 2 are surrounded by a cloud of small, irregularly shaped granules. During centrifugation of the lampbrush chromosomes for cytological analysis, these granules come to lie in the general vicinity of the giant loops. They also stain intensely with mAb 4D11 (Fig. 7, *A* and *B*), suggesting that products from the giant loops are being shed regularly into the nucleoplasm. The hundreds of extrachromosomal nucleoli and a variety of other particulates in the nucleoplasm fail to stain with 4D11. Sections of immature newt ovary were fixed by freeze-substitution, embedded in paraffin, and sectioned at 4 μm. After staining with mAb 4D11, the germinal vesicle contents were more or less uniformly stained except for the nucleoli. In some oocytes an intensely fluorescent mass within the GV was easily recognizable above the general level of staining (not shown). We presume that this mass contains the giant loops.

Isolation of cDNA Clones for the L Protein

Coincident with the isolation of the mAb 4D11 monoclonal antibody, the mouse antiserum was used for isolation of cDNA clones. Fig. 8 (lane *total*) shows that the serum of one immunized mouse reacted with several hnRNP proteins in-

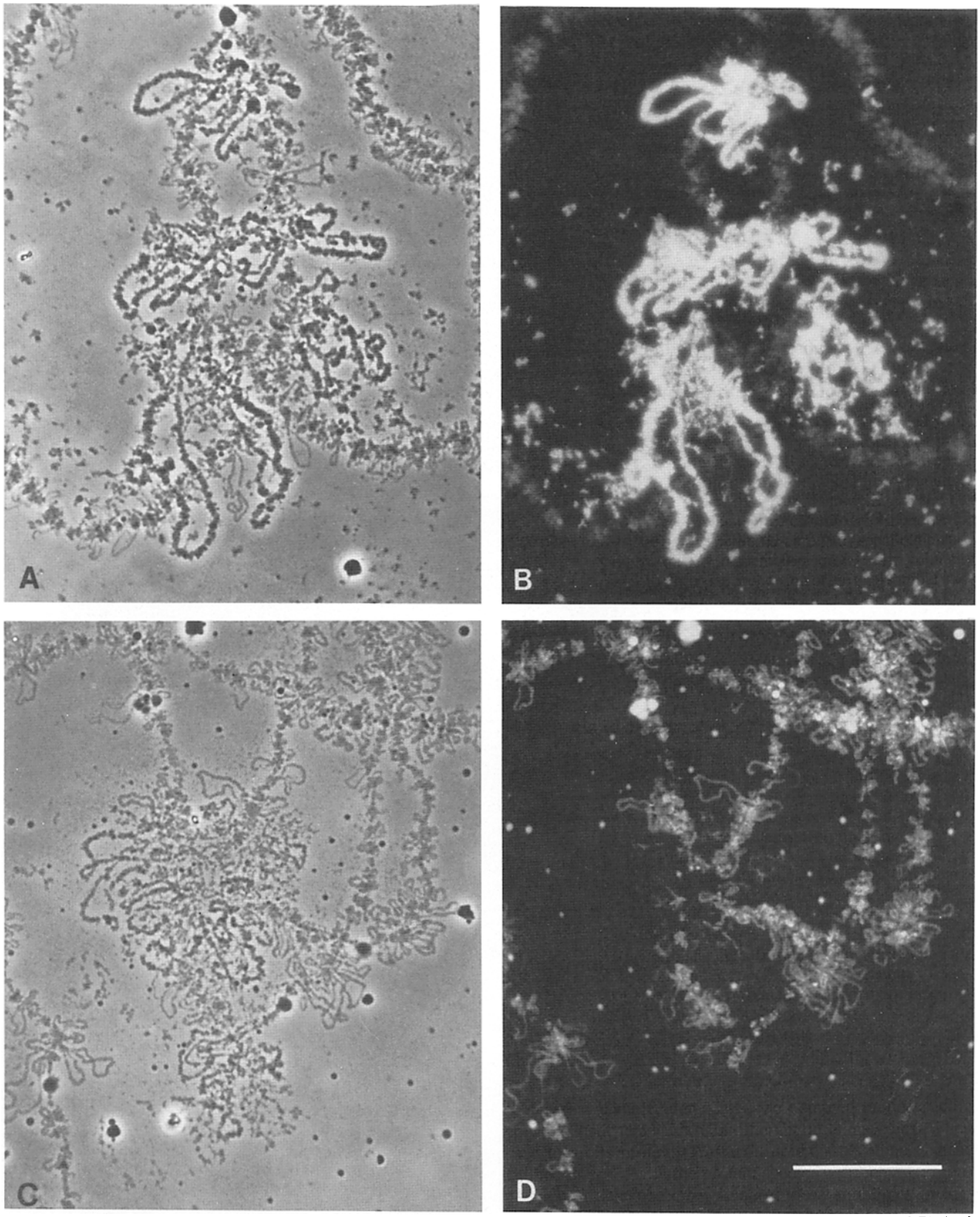

Figure 7. Immunofluorescence microscopy with monoclonal antibodies 4D11 and Y12 on newt lampbrush chromosomes. (*A* and *B*) Antibody 4D11. Portion of lampbrush chromosome 2 from the newt *Notophthalmus viridescens* showing the giant loops near the centromere. (*A*) Phase-contrast and (*B*) fluorescence images of the same region after staining with 4D11 and rhodamine-labeled second antibody. The giant loops are intensely stained, as are numerous extrachromosomal granules that regularly accompany these loops. The majority of typical loops are also stained by the antibody, but less intensely. (*C* and *D*) Antibody Y12. (*C*) Phase-contrast and (*D*) fluorescence images of the same region of chromosome 2 after staining with mAb Y12, which is directed against the Sm epitope of snRNP proteins. Most typical loops are well stained, but the giant loops are barely detectable. Y12 also stains numerous small, spherical granules in the nucleoplasm, but not the irregular granules that accompany the giant loops (note the unstained patches to the left of the giant loops). Bar, 50 μm.

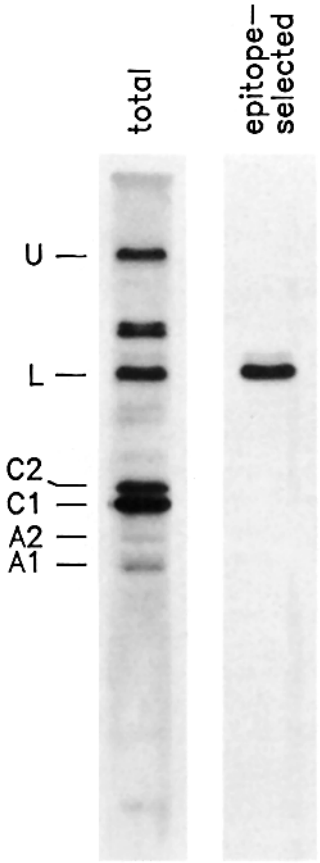

Figure 8. Immunoblots using total and affinity-selected antisera. Total HeLa cell proteins were fractionated by SDS-PAGE, transferred to nitrocellulose, and probed with either total mouse antisera raised against a 2 M NaCl fraction from ssDNA agarose (lane *total*) or antibodies selected by immobilized fusion proteins (lane *epitope-selected*) expressed by the purified λgt11 clone, phcL1.

cluding A1, A2, C1, C2, L, and U and an unidentified protein of ~90 kD. This serum was used to directly screen a HeLa cell λgt11 cDNA library, and positive plaques were selected and purified. As previously described for the isolation of the yeast mRNA poly(A)-binding protein (Adam et al., 1986), these purified phage were used to epitope-select and purify antibodies directed against specific proteins. As also illustrated in Fig. 8 (lane *epitope-selected*) one of these clones, hcL1, selected antibodies that specifically recognized the hnRNP L protein. Subsequently, the mAb 4D11 was shown to also recognize the fusion protein produced by this phage (data not shown). To further demonstrate that the hcL1 clone encoded the L protein, this clone was used to hybrid-select mRNA from total poly(A)$^+$ RNA, and the specifically selected RNA was translated in vitro. As Fig. 9 shows, the hcL1 clone hybrid-selected mRNA which translated into an ~68-kD protein (panel *total*, lane *hcLl*). This protein was specifically immunopurified with 4D11 (Fig. 9, panel *immunopurified*, lane *hcLl*) and it was not found in the nitrocellulose control without DNA (lanes *C* in both panels). As a positive control, and for comparison purposes, a clone for the hnRNP C proteins (Swanson et al., 1987) was used to hybrid-select mRNA for the C1 and C2 proteins (Fig. 9, lanes *hcl2*).

Nucleotide Sequence of the cDNA for the L Protein and the Deduced Amino Acid Sequence

Once identification of the hcL1 clone had been confirmed, this cDNA was used to screen an RNA blot to determine the size and complexity of the mRNA. Fig. 10 shows that the L protein is encoded by a single-sized poly(A)-containing RNA of ~2.3 kb. Using the hcL1 clone as a hybridization probe, a cDNA clone containing the entire protein coding region was isolated, phcL3 (Fig. 11 *A*). This clone codes for the entire L protein by two different criteria. (*a*) In vitro transcription and translation of pHCL3 yields a protein that comigrates with the hybrid selection/translation product shown in Fig. 9 and with authentic L protein isolated from HeLa cells by immunopurification with the mAb 4D11. (*b*) Other L protein cDNA clones that contain sequence information upstream of the 5′ end of the cDNA reported in Fig. 11 *B* contain stop codons in all three reading frames up-

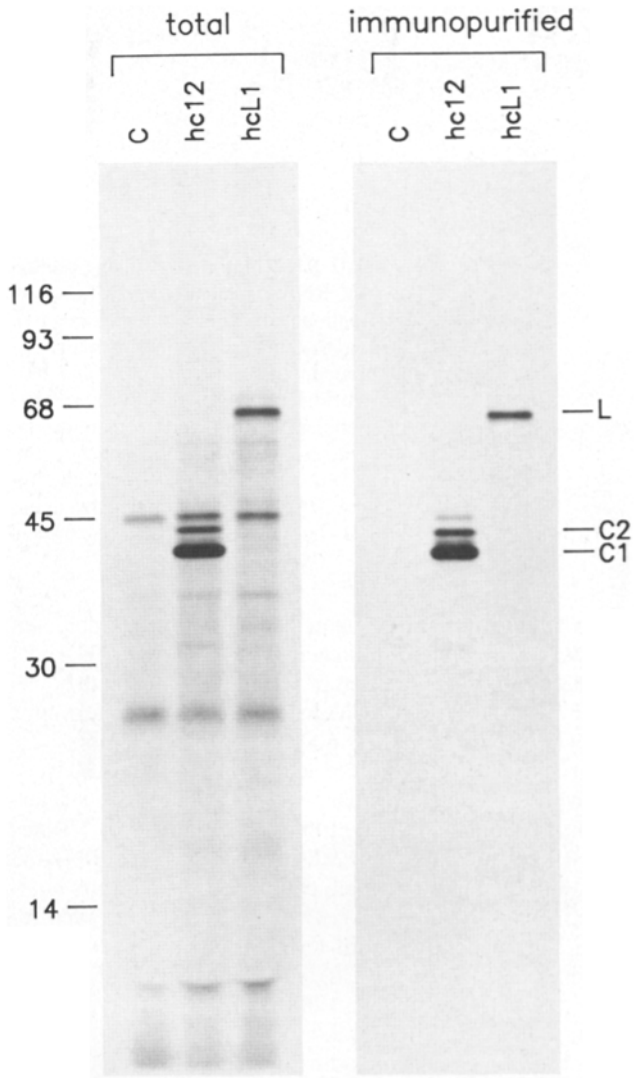

Figure 9. Hybrid selection and in vitro translation using phcL1. HeLa poly(A)$^+$ RNA was hybridized to filter-bound DNAs, eluted, and translated in vitro in a rabbit reticulocyte lysate in the presence of [^{35}S]methionine. The filter-bound DNAs were either a cDNA clone for the hnRNP C proteins (lanes *hcl2*), or the expression clone for the L proteins (lanes *phcLl*), or a control (lanes *C*) in which no DNA was bound to the filter. The proteins produced from these poly(A)$^+$ RNAs by in vitro translation were either directly fractionated by SDS-PAGE (lanes *total*) or first immunopurified (lanes *immunopurified*) with the anti-C protein monoclonal antibody 4F4 (lane *immunopurified*, *hcl2*) or the anti-L protein monoclonal 4D11 (lane *immunopurified*, *hcLl*).

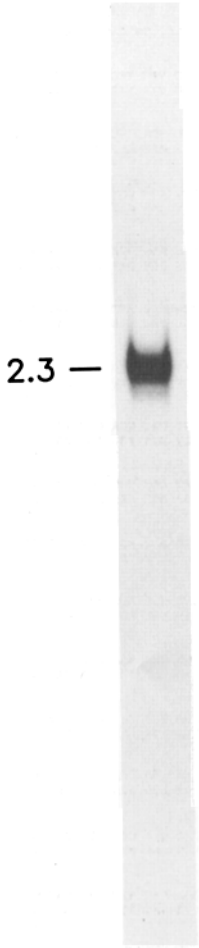

Figure 10. RNA blot analysis using phcL1. Poly(A)$^+$ RNA, isolated from HeLa cells, was fractionated by electrophoresis on a 1.4% agarose gel containing formaldehyde and blotted onto nitrocellulose. The blot was hybridized with both the expression clone, phcL1, and the clone containing the entire protein coding region, phcL3; the results were identical. The sizes of the hybridizing RNAs were estimated by using Hind III–digested λDNA as size marker.

stream of the initiation codon used in pHCL3 (data not shown). The predicted amino acid sequence of L (Fig. 11 *B*) indicated a protein composed of 558 amino acids with a molecular mass of 60,187 D, which contains an amino terminus in which 29 out of the first 60 residues are glycine.

A structural feature common to the only other hnRNP proteins whose complete primary structures have been determined, A1 (Cobianchi et al., 1986; Riva et al., 1986; Haynes et al., 1987; Buvoli et al., 1988; Biamonti et al., 1989) and C1 (Swanson et al., 1987), are the highly related RNA binding domains of ~90 amino acids whose most conserved region is an octapeptide motif termed the ribonucleoprotein consensus sequence (RNP-CS) (Adam et al., 1986; Swanson et al., 1987; Dreyfuss et al., 1988). The RNP-CS type RNA binding domain (CS-RBD) is a true RNA binding domain since protein segments including primarily this region, produced either by proteolysis (Herrick and Alberts, 1976; Bugler et al., 1987) or by in vitro transcription/translation (Bandziulis et al., 1989; Query et al., 1989), are able to bind single-stranded polynucleotides and, in some cases, discriminate between specific types of RNAs. The L protein contains two segments of ~80 amino acids each (amino acids 63–143 and 155–237 in Fig. 11, *B* and *C*) which are weakly related to each other and to the CS-RBDs of the other hnRNP and snRNP proteins (20% amino acid identity when optimally aligned in Fig. 11 *C*). As shown in Fig. 11 *C*, the most highly conserved segment of this weakly repeating region is an octapeptide, which is intriguingly similar to the RNP-CS octapeptides found in the snRNP U1A and U2B″ proteins (RGQAFVIF in domain 1 and HDIAFVEF in domain 2 of both proteins) (Sillekens et al., 1987). One of the most highly conserved positions in the entire 90 amino acid CS-RBD is the fifth position of the RNP-CS octapeptide (underlined above in the U1A and U2B″ sequences) that is usually a phenylalanine which in the case of A1 has been shown to readily cross-link to DNA oligonucleotides (Merrill et al., 1988). In L, this position is a leucine or methionine which clearly distinguishes this protein from the normal RNP-CS although other amino acids have been found in this position, as is the case for the first domain of the *Drosophila* sex-lethal protein that contains a serine residue at this position (Bandziulis et al., 1989).

Computer searches of the protein data banks did not reveal significant similarities with any known proteins although some sequence similarity was detectable between the amino terminus of the L protein and the carboxy domain of the hnRNP A1 protein (22.5% identity in an 89 amino acid overlap), and a variety of proline-rich and other glycine-rich proteins also shared a limited degree of sequence similarity. The predicted secondary structure of the L protein, obtained using the ChouFas and PlotChou of the UWGCG programs, suggests a protein that contains a small number of short alpha-helical domains and several long stretches of predicted large hydrophobic moment.

Discussion

We describe here a novel constituent of hnRNP complexes, the L protein. L is the first non-A, -B, -C type hnRNP protein that has been extensively characterized. The L protein is a bona fide hnRNP protein, and it is a constituent of the same hnRNP complexes that can be immunopurified with antibodies to the A1 and C proteins. Its abundance in immunopurified hnRNP complexes is comparable to that of the B1, B2, and C2 proteins (Piñol-Roma et al., 1988). Along with other hnRNP and snRNP proteins, it is localized on the majority of lampbrush chromosome loops of the newt *Notophthalmus viridescens* and is, therefore, probably associated with most nascent transcripts. However, L exhibits several properties that set it apart from other hnRNP proteins for which immunological probes and sequence data are available, indicating that it represents a new type of hnRNP protein. Among the unique characteristics of L is its occurrence also outside of the previously defined hnRNP complex. This is evident from the immunopurification experiments and from the sedimentation profiles of L in sucrose gradients, but it is most vividly apparent from immunofluorescent microscopy on somatic nuclei and spread amphibian lampbrush chromosomes. The distinct distribution of L by all these criteria contrasts with that observed for other hnRNP proteins such as A1, C, and U.

The analysis of the distribution of the L protein on lampbrush chromosomes is particularly instructive. Among the antibodies that stain lampbrush chromosomes, 4D11 is the only one that stains typical loops and the giant loops on chromosome 2. In Fig. 7 *B* the typical loops appear to be poorly stained, but this figure was purposely underexposed in order not to wash out detail in the brilliantly fluorescent giant loops. In fact, 4D11 stains typical loops about as brightly as

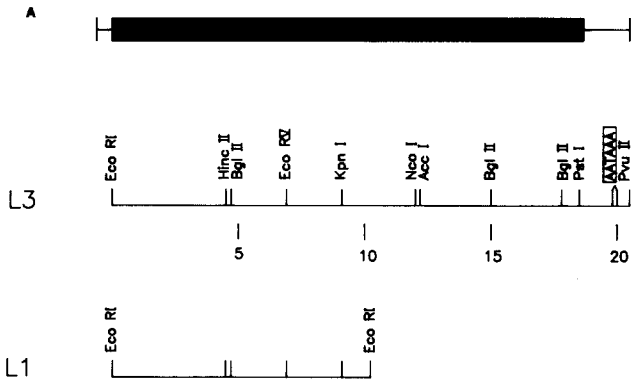

Figure 11. Structure of the L protein. (*A*) Restriction map of phcL1 and phcL3. The protein coding region of the cDNAs is indicated above the restriction maps of the expression clone, phcL1, and the full-length clone, phcL3, as a black box. The polyadenylation signals at the 3′ end of the phcL3 cDNA are also indicated (*AATAAA*). The Pvu II site at the 3′ end of L3 is not part of the cDNA but was used for directional cloning. (*B*) The nucleotide and deduced amino acid sequence of the L protein. Both the phcL1 and phcL3 clones were completely sequenced on both strands using overlapping restriction fragments subcloned into M13 vectors. The only open reading frame sufficiently large to encode the L protein is shown below the nucleotide sequence. (*C*) Sequence alignment of the two ~80 amino acid regions within the L protein which share a limited degree of sequence similarity to each other and to the RNP-CS RNA-binding domain (~20%). Identical amino acids are indicated with an asterisk, and the more highly conserved octapeptide is marked by a stippled box.

B

```
                                                                GGACGAGCAGCGGAGGCCGGTCGGGAGCCG

ATG GTG AAG ATG GCG GCG GCG GGC GGC GGA GGC GGC GGT GGC CGC TAC TAC GGC GGC GGC
MET VAL LYS MET ALA ALA ALA GLY GLY GLY GLY GLY GLY GLY ARG TYR TYR GLY GLY GLY
                                        10                                  20

AGT GAG GGC GGC CGG GCC CCT AAG CGG CTC AAG ACT GAC AAC GCC GGC GAC CAG CAC GGA
SER GLU GLY GLY ARG ALA PRO LYS ARG LEU LYS THR ASP ASN ALA GLY ASP GLN HIS GLY
                                        30                                  40

GGC GGC GGC GGT GGC GGT GGA GGA GCC GGG GCG GCG GGC GGC GGC GGC GGT GGG GAG AAC
GLY GLY GLY GLY GLY GLY GLY GLY ALA GLY ALA ALA GLY GLY GLY GLY GLY GLU ASN
                                        50                                  60

TAC GAT GAC CCG CAC AAA ACC CCT GCC TCC CCA GTT GTC CAC ATC AGG GGC CTG ATT GAC
TYR ASP ASP PRO HIS LYS THR PRO ALA SER PRO VAL VAL HIS ILE ARG GLY LEU ILE ASP
                                        70                                  80

GGT GTG GTG GAA GCA GAC CTT GTG GAG GCC TTG CAG GAG TTT GGA CCC ATC AGC TAT GTG
GLY VAL VAL GLU ALA ASP LEU VAL GLU ALA LEU GLN GLU PHE GLY PRO ILE SER TYR VAL
                                        90                                 100

GTG GTA ATG CCT AAA AAG AGA CAA GCA CTG GTG GAG TTT GAA GAT GTG TTG GGG GCT TGC
VAL VAL MET PRO LYS LYS ARG GLN ALA LEU VAL GLU PHE GLU ASP VAL LEU GLY ALA CYS
                                       110                                 120

AAC GCA GTG AAC TAC GCA GCC GAC AAC CAA ATA TAC ATT GCT GGT CAC CCA GCT TTT GTC
ASN ALA VAL ASN TYR ALA ALA ASP ASN GLN ILE TYR ILE ALA GLY HIS PRO ALA PHE VAL
                                       130                                 140

AAC TAC TCT ACC AGC CAG AAG ATC TCC CGC CCT GGG GAC TCG GAT GAC TCC CGG AGC GTG
ASN TYR SER THR SER GLN LYS ILE SER ARG PRO GLY ASP SER ASP ASP SER ARG SER VAL
                                       150                                 160

AAC AGT GTG CTT CTC TTT ACC ATC CTG AAC CCC ATT TAT TCG ATC ACC ACG GAT GTT CTT
ASN SER VAL LEU LEU PHE THR ILE LEU ASN PRO ILE TYR SER ILE THR THR ASP VAL LEU
                                       170                                 180

TAC ACT ATC TGT AAT CCT TGT GGC CCT GTC CAG AGA ATT GTC ATT TTC AGG AAG AAT GGA
TYR THR ILE CYS ASN PRO CYS GLY PRO VAL GLN ARG ILE VAL ILE PHE ARG LYS ASN GLY
                                       190                                 200

GTT CAG GCG ATG GTG GAA TTT GAC TCA GTT CAA AGT GCC CAG CGG GCC AAG GCC TCT CTC
VAL GLN ALA MET VAL GLU PHE ASP SER VAL GLN SER ALA GLN ARG ALA LYS ALA SER LEU
                                       210                                 220

AAT GGG GCT GAT ATC TAT TCT GGC TGT TGC ACT CTG AAG ATC GAA TAC GCA AAG CCT ACA
ASN GLY ALA ASP ILE TYR SER GLY CYS CYS THR LEU LYS ILE GLU TYR ALA LYS PRO THR
                                       230                                 240

CGC TTG AAT GTG TTC AAG AAT GAT CAG GAT ACT TGG GAC TAC ACA AAC CCC AAT CTC AGT
ARG LEU ASN VAL PHE LYS ASN ASP GLN ASP THR TRP ASP TYR THR ASN PRO ASN LEU SER
                                       250                                 260

GGA CAA GGT GAC CCT GGC AGC AAC CCC AAC AAA CGC CAG AGG CAG CCC CCT CTC CTG GGA
GLY GLN GLY ASP PRO GLY SER ASN PRO ASN LYS ARG GLN ARG GLN PRO PRO LEU LEU GLY
                                       270                                 280

GAT CAC CCC GCA GAA TAT GGA GGG CCC CAC GGT GGG TAC CAC AGC CAT TAC CAT GAT GAG
ASP HIS PRO ALA GLU TYR GLY GLY PRO HIS GLY GLY TYR HIS SER HIS TYR HIS ASP GLU
                                       290                                 300

GGC TAC GGG CCC CCC CCA CCT CAC TAC GAA GGG AGA AGG ATG GGT CCA CCA GTG GGG GGT
GLY TYR GLY PRO PRO PRO PRO HIS TYR GLU GLY ARG ARG MET GLY PRO PRO VAL GLY GLY
                                       310                                 320

CAC CGT CGG GGC CCA AGT CGC TAC GGC CCC CAG TAT GGG CAC CCC CCA CCC CCT CCC CCA
HIS ARG ARG GLY PRO SER ARG TYR GLY PRO GLN TYR GLY HIS PRO PRO PRO PRO PRO PRO
                                       330                                 340

CCA CCC GAG TAT GGC CCT CAC GCC GAC AGC CCT GTG CTC ATG GTC TAT GGC TTG GAT CAA
PRO PRO GLU TYR GLY PRO HIS ALA ASP SER PRO VAL LEU MET VAL TYR GLY LEU ASP GLN
                                       350                                 360

TCT AAG ATG AAC GGT GAC CGA GTC TTC AAT GTC TTC TGC TTA TAT GGC AAT GTG GAG AAG
SER LYS MET ASN GLY ASP ARG VAL PHE ASN VAL PHE CYS LEU TYR GLY ASN VAL GLU LYS
                                       370                                 380

GTG AAA TTC ATG AAA AGC AAG CCG GGG GCC GCC ATG GTG GAG ATG GCT GAT GGC TAC GCT
VAL LYS PHE MET LYS SER LYS PRO GLY ALA ALA MET VAL GLU MET ALA ASP GLY TYR ALA
                                       390                                 400

GTA GAC CGG GCC ATT ACC CAC CTC AAC AAC AAC TTC ATG TTT GGG CAG AAG CTG AAT GTC
VAL ASP ARG ALA ILE THR HIS LEU ASN ASN ASN PHE MET PHE GLY GLN LYS LEU ASN VAL
                                       410                                 420

TGT GTC TCC AAG CAG CCA GCC ATC ATG CCT GGT CAG TCA TAC GGG TTG GAA GAC GGG TCT
CYS VAL SER LYS GLN PRO ALA ILE MET PRO GLY GLN SER TYR GLY LEU GLU ASP GLY SER
                                       430                                 440

TGC AGT TAC AAA GAC TTC AGT GAA TCC CGG AAC AAT CGG TTC TCC ACC CCA GAG CAG GCA
CYS SER TYR LYS ASP PHE SER GLU SER ARG ASN ASN ARG PHE SER THR PRO GLU GLN ALA
                                       450                                 460

GCC AAG AAC CGC ATC CAG CAC CCC AGC AAC GTG CTG CAC TTC TTC AAC GCC CCG CTG GAG
ALA LYS ASN ARG ILE GLN HIS PRO SER ASN VAL LEU HIS PHE PHE ASN ALA PRO LEU GLU
                                       470                                 480

GTG ACC GAG GAG AAC TTC TTT GAG ATC TGC GAT GAG CTG GGA GTG AAG CGG CCA TCT TCT
VAL THR GLU GLU ASN PHE PHE GLU ILE CYS ASP GLU LEU GLY VAL LYS ARG PRO SER SER
                                       490                                 500

GTG AAA GTA TTC TCA GGC AAA AGT GAG CGC AGC TCC TCT GGA CTG CTG GAG TGG GAA TCC
VAL LYS VAL PHE SER GLY LYS SER GLU ARG SER SER SER GLY LEU LEU GLU TRP GLU SER
                                       510                                 520

AAG AGC GAT GCC CTG GAG ACT CTG GGC TTC CTG AAC CAT TAC CAG ATG AAA AAC CCA AAT
LYS SER ASP ALA LEU GLU THR LEU GLY PHE LEU ASN HIS TYR GLN MET LYS ASN PRO ASN
                                       530                                 540

GGT CCA TAC CCT TAC ACT CTG AAG TTG TGT TTC TCC ACT GCT CAG CAC GCC TCC TAA
GLY PRO TYR PRO TYR THR LEU LYS LEU CYS PHE SER THR ALA GLN HIS ALA SER  *
                                       550                             558

TTAGGTGCCTAGGAAGAGTCCCATCTGAGCAGGAAGACATTTCTCTTTCCTTTATGCCATTTTTTGTTTTTGTTATTTG
CAAAAGATCTTGTATTCCTTTTTTTTTTTTTTTTTTTTTTAAATGCTAGGTTTGTAGAGGCTTACTTAACCTTAATGGAA
ACGCTGGAAATCTGCAGGGGGAGGGAGAGGGGAACTGTTATCTCCCAAGATTAACCTTCACTTTTAAAAAATTATTGTA
CATGTGATTTTTTTTTTTTCCTGTTCATACATTTGTGCTGCCCATGTACTCTTGGCACATTTCAATAAAATTGTTTGGA
AAATAAACACAGCAAAAAAAAAAAAAAAA
```

C

AMINO ACID

```
63-143    DPHKTPASPVVHIRGLIDGVVEADLVEALQEFGPISYVVVMPK. KRQALVEF EDVLGACNAVNYAADNQIYIAGHPAFVNYS
          *              *   *        *        **     *   *     ** ***   *   *   *          **          *
156-237   DSRSVNSVLLFTILNPIYSITTDVLYTICNPCGPVQRIVIFRKN GVQAMVEF DSVQSAQRAKASLNGADIYSGCCTLKIEYA
```

Y12, shown in Fig. 7 *D*. All other antibodies that stain typical loops, including some against known hnRNP proteins, stain the giant loops very weakly or not at all (Roth and Gall, 1987; unpublished observations). Antibodies that show this pattern include mAb iD2 against A and B proteins (Leser et al., 1984), mAb 3G6 against the U protein (Dreyfuss et al., 1984*b*), and mAbs SE5 and UA5 against newt GV proteins (Roth and Gall, 1987). The anti-snRNP antibody Y12 (Lerner et al., 1981) shows the same pattern of loop staining, but also stains the structures known as "spheres" and numerous smaller nucleoplasmic granules (Fig. 7, *C* and *D*). It is abundantly clear, therefore, that the giant loops are deficient in a set of common hnRNP and snRNP proteins found on typical loops. The giant loops also contain unique associated antigens not present at detectable levels in the typical loops. This is shown by their staining with two mAbs that do not stain typical loops, mAb A1 (Lacroix et al., 1985) and mAb TH2 (Roth and Gall, 1987). Thus, the transcripts from the giant loops are associated with the L protein and at least two other antigens, perhaps in the form of specialized hnRNP complexes.

Only limited information is available concerning transcription on the giant loops. Transcription on them, like that on typical loops, is inhibited by α-amanitin at 0.5 μg/ml, and is presumably carried out by RNA polymerase II (Schultz et al., 1981). The efficiency of incorporation of the four ribonucleotides is rather different from that seen in typical loops, suggesting that the giant loop RNA has an unusual nucleotide composition, high in cytidine and low in guanine (25% A, 27% U, 39% C, and 9% G) (Hartley and Callan, 1978). An unusual nucleotide composition is also indicated by the fact that the DNA axis of the giant loops is not cut by the restriction endonuclease Hae III (Gould et al., 1976), whereas most loops are readily digested by this enzyme (a "four-cutter" that recognizes the sequence GGCC). The giant loops are cut by other restriction enzymes and by DNase I, indicating that their DNA axis is generally accessible to enzymes. One interpretation of the incorporation and restriction enzyme data is that the giant loops contain a simple, repeated sequence that happens to lack GGCC. The transcription of repeated sequences ("satellite DNA") on lampbrush chromosome loops is well documented by in situ hybridization (Varley et al., 1980; Diaz et al., 1981). It will be of great interest to identify the RNA sequences with which the L protein is associated in the giant loops.

Immunofluorescence microscopy on fixed somatic cells (Fig. 6) shows one to three (generally two) loci of high concentration of L in the nucleus in addition to a general nucleoplasmic localization. What are these intensely staining structures? It is possible that they represent simply a pool of free L protein in the nucleoplasm, in excess of what is bound to hnRNP complexes. A more interesting alternative, suggested by the immunofluorescence observations on newt lampbrush chromosomes, is that L is a component of specialized hnRNP complexes, located at discrete chromosomal loci. According to this interpretation, the bright regions seen in somatic nuclei correspond to the quite similar bright regions in sections of newt GVs, and these in turn correspond to the giant loops and their associated granules in lampbrush chromosome spreads. In other words, the bright granules in the somatic nuclei may represent concentrations of specific hnRNP complexes still associated with the chromosome regions from which they arose. Evidence for this hypothesis could be obtained by examining lampbrush chromosomes and somatic nuclei from a variety of amphibians (the high degree of cross reaction already seen with mAb 4D11 makes it probable that such observations will be possible). A correlation between the (maximum) number of granules in somatic nuclei and the number of brightly staining lampbrush loci would suggest that the same loops are active in somatic and germinal nuclei.

L is a new type of hnRNP protein also in its amino acid sequence. One of the remarkable features of the sequence of L is that it contains two 80 amino acid domains only distantly related to the consensus sequence RNA binding domain (Dreyfuss et al., 1988; Bandziulis et al., 1989). We have been unable to identify any additional significant homologies between L and other RNA-binding proteins, including hnRNP proteins, except for some limited sequence similarity between the glycine-rich amino terminus of the L protein and the carboxy domain of the hnRNP A1. Further analysis of the relationship between protein sequence characteristics and the distribution of specific hnRNP proteins on hnRNA will probably advance from further studies on many other of the >20 proteins in immunopurified hnRNP complexes (Piñol-Roma et al., 1988) that must still be analyzed.

In summary, the distribution of the L protein in the cell, its association with the bulk of hnRNP complexes as well as with unique transcripts, and the primary structure of the L protein deduced from cDNA cloning, all indicate that L represents a new and unique type of hnRNP protein. The findings reported here and the availability of the antibodies and cDNA clones for L open the way for a number of exciting investigations. These include isolation and characterization of the transcripts of giant loops, and the general question of what are the signals that direct specific proteins, such as L, to specific loci on chromosomes.

We thank Michael Matunis, Dawn Sailer, and Janet Settle for assistance with the isolation of the L protein and the cDNA clones, David Miller for help with sequence database searches, and members of the laboratory for helpful discussions and comments on the manuscript. We are also grateful to Dr. Joan Steitz for a gift of the monoclonal antibody Y12.

This work was supported by grants from the National Institutes of Health to J. Gall and G. Dreyfuss. G. Dreyfuss is an Established Investigator of the American Heart Association.

Received for publication 26 May 1989 and in revised form 9 August 1989.

References

Adam, S. A., T. Y. Nakagawa, M. S. Swanson, T. Woodruff, and G. Dreyfuss. 1986. mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell. Biol.* 6:2932–2943.

Bandziulis, R. J., M. S. Swanson, and G. Dreyfuss. 1989. RNA-binding proteins as developmental regulators. *Genes Dev.* 3:431–437.

Beyer, A. L., M. E. Christensen, B. W. Walker, and W. M. LeStourgeon. 1977. Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell.* 11:127–138.

Biamonti, G., M. Buvoli, M. T. Bassi, C. Morandi, F. Cobianchi, and S. Riva. 1989. Isolation of an active gene encoding human human hnRNP A1: evidence for alternative splicing. *J. Mol. Biol.* 207:491–503.

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

Buvoli, M., G. Biamonti, P. Tsoulfas, M. T. Bassi, A. Ghetti, S. Riva, and C. Morandi. 1988. cDNA cloning of human hnRNP protein A1 reveals the existence of multiple mRNA isoforms. *Nucleic Acids Res.* 16:3751–3770.

Choi, Y. D., and G. Dreyfuss. 1984*a*. Isolation of the heterogeneous nuclear ribonucleoprotein complex (hnRNP): a unique supramolecular assembly.

Proc. Natl. Acad. Sci. USA. 81:7471–7475.

Choi, Y. D., and G. Dreyfuss. 1984*b*. Monoclonal antibody characterization of the C proteins of heterogeneous nuclear ribonucleoprotein complexes in vertebrate cells. *J. Cell Biol.* 99:1997–2004.

Choi, Y. D., P. J. Grabowski, P. A. Sharp, and G. Dreyfuss. 1986. Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science (Wash. DC).* 231:1534–1539.

Chung, S. Y., and J. Wooley. 1986. Set of novel, conserved proteins fold premessenger RNA into ribonucleosomes. *Proteins.* 1:195–210.

Cobianchi, F., D. N. SenGupta, B. Z. Zmudzka, and S. H. Wilson. 1986. Structure of rodent helix-destabilizing protein revealed by cDNA cloning. *J. Biol. Chem.* 261:3536–3543.

Diaz, M. O., G. Barsacchi-Pilone, K. A. Mahon, and J. G. Gall. 1981. Transcripts from both strands of a satellite DNA occur on lampbrush chromosome loops of the newt *Notophthalmus. Cell.* 24:649–659.

Dreyfuss, G. 1986. Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. *Annu. Rev. Cell Biol.* 2:459–498.

Dreyfuss, G., S. A. Adam, and Y. D. Choi. 1984*a*. Physical change in cytoplasmic messenger ribonucleoproteins in cells treated with inhibitors of mRNA transcription. *Mol. Cell. Biol.* 4:415–423.

Dreyfuss, G., Y. D. Choi, and S. A. Adam. 1984*b*. Characterization of hnRNA-protein complexes *in vivo* with monoclonal antibodies. *Mol. Cell. Biol.* 4:1104–1114.

Dreyfuss, G., M. S. Swanson, and S. Piñol-Roma. 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.* 13:86–91.

Gall, J. G., E. C. Stephenson, H. P. Erba, M. O. Diaz, and G. Barsacchi-Pilone. 1981. Histone genes are located at the sphere loci of newt lampbrush chromosomes. *Chromosoma (Berl.).* 84:159–171.

Gould, D. C., H. G. Callan, and C. A. Thomas. 1976. The actions of restriction endonucleases on lampbrush chromosomes. *J. Cell Sci.* 21:303–313.

Hartley, S. E., and H. G. Callan. 1978. RNA transcription on the giant lateral loops of the lampbrush chromosomes of the american newt *Notophthalmus viridescens. J. Cell Sci.* 34:279–288.

Haynes, S. R., M. L. Rebbert, B. A. Mozer, F. Forquignon, and I. G. Dawid. 1987. *pen* repeat sequences are GGN clusters and encode a glycine-rich domain in *Drosophila* homologous to the rat helix destabilizing protein. *Proc. Natl. Acad. Sci. USA.* 84:1819–1823.

Herrick, G., and B. Alberts. 1976. Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. *J. Biol. Chem.* 251:2124–2132.

Jacob, M., G. Devilliers, J. P. Fuchs, H. Gallinaro, R. Gattoni, C. Judes, and J. Stevenin. 1981. Isolation and structure of the ribonucleoprotein fibrils containing heterogeneous nuclear RNA. *In* The Cell Nucleus. H. Busch, editor. Vol. 8. Academic Press, Inc., New York. 194–246.

Karn, J., G. Vidali, L. C. Boffa, and V. G. Allfrey. 1977. Characterization of the non-histone nuclear proteins associated with rapidly labeled heterogeneous nuclear RNA. *J. Biol. Chem.* 252:7307–7322.

Lacroix, J. C., R. Azzouz, D. Boucher, C. Abbadie, C. K. Pyne, and J. Charlemagne. 1985. Monoclonal antibodies to lampbrush chromosome antigens of *Pleurodeles waltlii. Chromosoma (Berl.).* 92:69–80.

Laskey, R. A., and A. D. Mills. 1975. Quantitative film detection of ^{3}H and ^{14}C in polyacrylamide gels by fluorography. *Eur. J. Biochem.* 56:335–341.

Lehrach, H., D. Diamond, J. M. Wozney, and H. Boedtker. 1977. RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry.* 16:4743–4751.

Lerner, E. A., M. R. Lerner, C. A. Janeway, and J. A. Steitz. 1981. Monoclonal antibodies to nucleic acid-containing cellular constituents: probes for molecular biology and autoimmune disease. *Proc. Natl. Acad. Sci. USA.* 78:2737–2741.

Leser, G. P., J. Escara-Wilke, and T. E. Martin. 1984. Monoclonal antibodies to heterogeneous nuclear RNA-protein complexes. The core proteins comprise a conserved group of related polypeptides. *J. Biol. Chem.* 259:1827–1833.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratories, Cold Spring Harbor, New York. 545 pp.

Martin, T. E., P. Billings, A. Levey, S. Ozarsian, I. Quinlan, H. Swift, and L. Urbas. 1974. Some properties of RNA-protein complexes from the nucleus of eukaryotic cells. *Cold Spring Harbor Symp. Quant. Biol.* 42:899–909.

Mayrand, S., B. Setyono, J. R. Greenberg, and T. Pederson. 1981. Structure of nuclear ribonucleoprotein: identification of proteins in contact with poly(A)+ heterogeneous nuclear RNA in living HeLa cells. *J. Cell Biol.* 90:380–384.

Merrill, B. M., K. L. Stone, F. Cobianchi, S. H. Wilson, and K. R. Williams. 1988. Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 hnRNP protein. *J. Biol. Chem.* 263:3307–3313.

Nakagawa, T. Y., M. S. Swanson, B. J. Wold, and G. Dreyfuss. 1986. Molecular cloning of cDNA for the nuclear ribonucleoprotein particle C proteins: a conserved gene family. *Proc. Natl. Acad. Sci. USA.* 83:2007–2011.

O'Farrell, P. Z., H. M. Goodman, and P. H. O'Farrell. 1977. High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell.* 12:1133–1142.

Pandolfo, M., O. Valentini, G. Biamonti, P. Rossi, and S. Riva. 1987. Large-scale purification of hnRNP proteins from HeLa cells by affinity chromatography on ssDNA-cellulose. *Eur. J. Biochem.* 162:213–220.

Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA.* 85:2444–2448.

Pederson, T. 1974. Proteins associated with heterogeneous nuclear RNA in eukaryotic cells. *J. Mol. Biol.* 83:163–183.

Piñol-Roma, S., Y. D. Choi, M. J. Matunis, and G. Dreyfuss. 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev.* 2:215–227.

Query, C. C., R. C. Bentley, and J. D. Keene. 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell.* 57:89–101.

Rigby, P. W. J., M. Diekmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237–251.

Riva, S., C. Morandi, P. Tsoulfas, M. Pandolfo, G. Biamonti, B. Merrill, K. R. Williams, G. Multhaup, K. Bayreuther, H. Werr, B. Henrich, and K. P. Schafer. 1986. Mammalian single-stranded DNA binding protein UP1 is derived from the hnRNP core protein A1. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:2267–2273.

Roth, M. B., and J. G. Gall. 1987. Monoclonal antibodies that recognize transcription unit proteins on newt lampbrush chromosomes. *J. Cell Biol.* 105:1047–1054.

Samarina, O. P., E. M. Lukanidin, J. Molman, and G. P. Georgiev. 1968. Structural organization of nuclear complexes containing DNA-like RNA. *J. Mol. Biol.* 33:251–263.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.

Schultz, L. D., B. K. Kay, and J. G. Gall. 1981. In vitro RNA synthesis in oocyte nuclei of the newt *Notophthalmus. Chromosoma (Berl.).* 82:171–187.

Sierakowska, H., W. Szer, P. J. Furdon, and R. Kole. 1986. Antibodies to hnRNP core proteins inhibit *in vitro* splicing of human β globin pre-mRNA. *Nucleic Acids Res.* 14:5241–5254.

Sillekens, P. T. G., W. J. Habets, R. P. Beijer, and W. J. van Venrooij. 1987. cDNA cloning of the human U1 snRNA-associated A protein: extensive homology between U1 and U2 snRNP-specific proteins. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:3841–3848.

Snyder, M., S. Elledge, D. Sweeter, R. A. Young, and R. W. Davis. 1987. λgt11: gene isolation with antibody probes and applications. *Methods Enzymol.* 154:107–128.

Swanson, M. S., and G. Dreyfuss. 1988*a*. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell. Biol.* 8:2237–2241.

Swanson, M. S., and G. Dreyfuss. 1988*b*. RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO (Eur. Mol. Biol. Organ.) J.* 11:3519–3529.

Swanson, M. S., T. Y. Nakagawa, K. LeVan, and G. Dreyfuss. 1987. Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA and pre-rRNA-binding proteins. *Mol. Cell. Biol.* 7:1731–1739.

van Eekelen, C. A., T. Riemen, and W. van Venrooij. 1981. Specificity in the interaction of hnRNA and mRNA with proteins as revealed by *in vivo* cross linking. *FEBS (Fed. Eur. Biochem. Organ.) Lett.* 130:223–226.

Varley, J. M., H. C. Macgregor, and H. P. Erba. 1980. Satellite DNA is transcribed on lampbrush chromosomes. *Nature (Lond.).* 283:686–688.

Wilk, H. E., H. Werr, D. Friedrich, H. H. Kiltz, and K. P. Schaefer. 1985. The core proteins of 35S hnRNP complexes: Characterization of nine different species. *Eur. J. Biochem.* 146:71–81.