4210–4217 *Nucleic Acids Research, 1993, Vol. 21, No. 18* © *1993 Oxford University Press*

hnRNP G: sequence and characterization of a glycosylated RNA-binding protein

Michel Soulard, Véronique Della Valle, Mikiko C.Siomi[1], Serafín Piñol-Roma[1], Patrice Codogno[2], Chantal Bauvy[2], Michel Bellini[3], Jean-Claude Lacroix[3], Guillaume Monod, Gideon Dreyfuss[1] and Christian-Jacques Larsen*

INSERM U-301, Institut de Génétique Moléculaire, 27 rue J.Dodu, 75010 Paris, France, [1]Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, PA 19104-6148, USA, [2]INSERM U-239, Faculté de Médecine Xavier Bichat, 16 rue H.Huchard, 75018 Paris and [3]Laboratoire de Génétique du Développement, Université P. et M.Curie, Pl. Jussieu, 75005 Paris, France

Received June 14, 1993; Revised and Accepted September 28, 1993 EMBL accession no. Z23064

ABSTRACT

The autoantigen p43 is a nuclear protein initially identified with autoantibodies from dogs with a lupus-like syndrome. Here we show that p43 is an RNA-binding protein, and identify it as hnRNP G, a previously described component of heterogeneous nuclear ribonucleoprotein complexes. We demonstrate that p43/hnRNP G is glycosylated, and identify the modification as O-linked N-acetylglucosamine. A full-length cDNA clone for hnRNP G has been isolated and sequenced, and the predicted amino acid sequence for hnRNP G shows that it contains one RNP-consensus RNA binding domain (RBD) at the amino terminus and a carboxyl domain rich in serines, arginines and glycines. The RBD of human hnRNP G shows striking similarities with the RBDs of several plant RNA-binding proteins.

INTRODUCTION

In eukaryotic cells, the processing of heterogeneous nuclear RNAs (hnRNAs), the precursors of messenger RNAs, takes place as they are in heterogeneous nuclear ribonucleoprotein complexes or hnRNPs (for reviews, see 11, 12, 13, 32). Immunopurification of hnRNPs with monoclonal antibodies and two-dimensional (2-D) gel electrophoresis revealed at least 20 different abundant hnRNP polypeptides between 34kDa and 120kDa molecular weights and designated A1 through U (32). Many of these proteins have been characterized including A1 (3, 5, 7, 36), A2/B1 (4), and C1/C2, (4, 32), L (33), K (29), I (16), M (9) and U (23).

Most, if not all, of the hnRNP proteins are RNA-binding proteins (32, see also 13). Many of the hnRNP proteins for which amino acid sequence is available share a highly conserved RNA-binding domain or RBD, which consists of 80–100 amino acids with two hallmark consensus sequences: RNP-1 (8 aa) and RNP-2 (6 aa) (2, 12, 22), and which has RNA-binding activity (for review, see 13 and 22). In addition, another motif, the RGG box, has been found in the hnRNP U proteins and has been shown to have RNA-binding activity (23). We have recently reported the isolation of a novel glycoprotein designated p43 in the nucleus of a variety of mammalian cells, and initially detected by circulating autoantibodies from dogs with lupus-like syndrome (38). Previous data had suggested that p43 antigen associates with hnRNP complexes in HeLa cells.

In this report, we show that p43 is the hnRNP G protein, a basic polypeptide that is a component of hnRNP complexes (32). We demonstrate that hnRNP G contains O-linked N-acetylglucosamine (GlcNac) which explains its reactivity to wheat germ agglutinin (WGA, 38). The sequence of a cDNA clone encoding the entire amino acid sequence of the human protein is presented. Finally, the presence of antigenically similar proteins reacting with canine antibodies to hnRNP G in divergent animal species suggests that it is conserved through evolution.

MATERIALS AND METHODS

Cell cultures, labeling and cell fractionation

HeLa S3 cells were grown in agitated suspension as previously described (38). They were labeled for 48 hrs with 11μc/ml of 6-[^{3}H]glucosamine (specific activity: 0.37–1.11TBl/mM; Centre Energie Atomique, Saclay, France). HeLa cells were labeled with 20μCi of [^{35}S]methionine per ml for 20hrs in Dulbecco's modified Eagle's medium (DMEM) containing $^1/_{10}$ the normal methionine levels and 5% calf thymus.

Preparation of cell extracts and immunoprecipitation procedure

Cell fractionation, immunoprecipitation and Western blotting of nuclear extracts were as previously reported (38). Immunopurification of hnRNP complexes with the anti-C proteins monoclonal

* To whom correspondence should be addressed

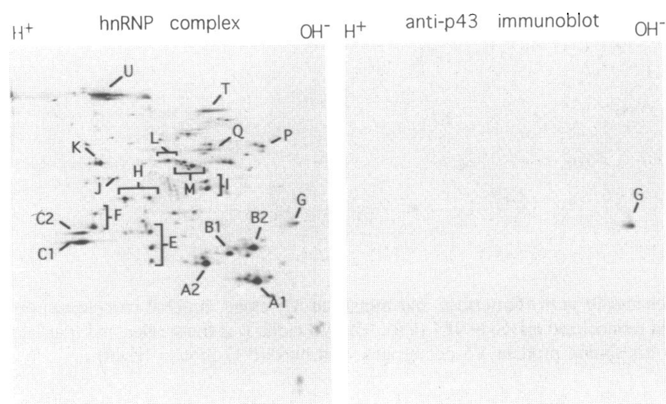


Figure 1. Two-dimensional gel electrophoresis of immunopurified hnRNP complexes and immunoblot with anti-p43. hnRNP complexes were immunopurified from the nucleoplasm of [^{35}S]methionine-labeled HeLa cells, and the proteins were resolved by two-dimensional gel electrophoresis (NEPHGE in the first dimension and SDS-PAGE in the second dimension), transferred to nitrocellulose, and detected by autoradiography (left panel). The same blot was then probed with anti-p43 antiserum and the antigen detected by chemiluminescence (right panel). Direct superimposition of the two autoradiographs shows that p43 comigrates precisely with the previously identified hnRNP G (33).

antibody 4F4 and analysis by two-dimensional gel electrophoresis were as previously described (38). Where indicated, the proteins were transferred to nitrocellulose following two-dimensional gel electrophoresis. p43 was detected by immunoblotting with anti-p43 antisera (38) as previously described (10) and developed using the ECL chemiluminescence kit (Amersham, UK) following the manufacturer's instructions.

Isolation of cDNA clones

A high titer canine antiserum monospecific for p43 was used to screen (1:200 dilution) a λgt11 ZS-75 cell cDNA library (Clontech Lab. Inc., Palo Alto, California). The presence of p43 in these cells had been verified by immunoblot prior to the screening. One positive plaque was isolated, whose clone was subcloned into the Bluescript plasmid (Stratagene Cloning System, La Jolla, California). Since the insert did not encode a full length cDNA, the partial fragment it contained was used as a hybridization probe to rescreen the library. One positive clone, V5, was isolated whose insert was amplified by Polymerase Chain Reaction using two amplifiers complementary to each phage arm. The amplified product was then subcloned in a KS$^+$/EcoRV bluescript vector (Stratagene) modified by T nucleotide addition at the end of the linearized vector.

In vitro transcription, *in vitro* translation

The plasmid, pV5, was linealized with Hind III to provide template for in vitro transcription using T7 RNA polymerase (Promega Corporation, Madison, Wisconsin). The RNA was translated in nuclease treated rabbit reticulocyte lysate in the presence of [^{35}S]methionine according to the manufacturer's instructions.

DNA sequence determination and analysis

Sequencing of the immunological clone was carried out on single stranded DNA subclones prepared by exonuclease III treatment For sequencing the V5 clone. The insert was subcloned in pGEM-blue (Promega) and the double stranded DNA was used. All sequencing reaction was performed by the dideoxy chain termination method (37). Sequencing analysis and database searches for sequencing similarities were performed by using NBRF and SwissProt. Sequencing analysis and database searches for sequencing similarities were performed by using the BLAST network service at the National Center for Biotechnology Information (1).

Protein-RNA cross-linking studies and ssDNA affinity chromatography

A procedure adapted from Dreyfuss et al. (10) was employed. Twenty millions HeLa cells suspended in a volume of 10ml were irradiated under UV light. Nucleoplasmic extracts were boiled and loaded to oligo(dT) cellulose columns at 0.5M NaCl. After washes with 0.5M NaCl, fractions bound to the columns were eluted with buffer devoid of NaCl. The eluted cross-linked complexes were incubated with RNAse A (200U/100μl, at 37°C for 30 min.) to digest unprotected RNA. Samples were submitted to Western blot analysis and revealed by anti-p43 serum or the 4F4 monoclonal antibody against hnRNP C1/C2 protein.

Affinity chromatography on ssDNA columns was performed as described elsewhere (32). ssDNA sepharose was replaced by ssDNA ultrogel (IBF, Paris).

Release of glycans from [^{3}H]-labeled p43 antigen by alkaline borohydride treatment

Mild alkaline-borohydride treatment of [^{3}H]glucosamine-labeled p43 antigen was performed with 0.05M NaOH/1.0M NaBH4 for 16 hours at 45°C (6). The reaction mixture was cooled to 4°C and adjusted to pH 5.0 by addition of 2.0M acetic acid. The sample was desalted through a column of AG-50-X8 (H^+ form, 200–400 mesh, Biorad, Richmond, USA). The borate excess was removed by addition of methanol followed by rotary evaporation of methylborate. The resulting material was loaded onto a column (1.4cm×80cm) of Bio-gel P2 [200–400 mesh, BioRad), and eluted with water at a constant 8ml/hr flow rate. Bovine serum albumin and mannose were used to determine the void volume and the included volume respectively. Radioactivity of the collected fractions (1ml) was measured in a β-scintillation counter. Bio-gel P2 chromatography was also performed before mild alkaline borohydride treatment in order to estimate the efficiency of [^{3}H]-labeled glycans bound to p43. Upon concentration, ^{3}H-labeled material released by mild alkaline borohydride treatment was separated by chromatography for 40 hours on Schleicher and Schuell 2043b paper in the butan-1-ol/pyridine/water (6:4:3, by vol.) solvent (42). Prior to application of the sample, the chromatography paper was immersed in a solution of 0.57% $Na_2B_4O_7$ (10 H_2O)+10mM NaCl, and dried. Standards put in guide strips consisted of N-acetyl-galactosamine, N-acetylglucosamine, N-acetyl-galactosaminitol, and N-acetylglucosaminitol. Both N-acetyl-hexosaminitols were obtained by NaBH4 reduction of acetyl-hexosamines. The standards were stained with $AgNO_3$ reagents (40). Radioactivity was determined on 1.0cm strips cut out of the chromatogram.

Immunofluorescence staining of lampbrush chromosomes

The procedure for preparation of lampbrush chromosomes has been described elsewhere (27). A dog serum selected for its ability to react specifically with HeLa cell p43 was used at a 1:250 dilution. The antigen-antibodies complexes were revealed by a

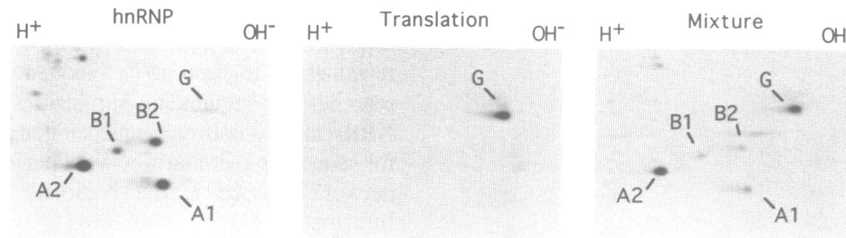


Figure 2. Two-dimensional gel electrophoresis of immunopurified hnRNP complexes and *in vitro* transcribed and translated V5 clone. hnRNP complexes were immunopurified from [^{35}S]methionine labeled HeLa nucleoplasm with the anti-C protein monoclonal antibody 4F4 (left). The V5 clone was transcribed and translated in the presence of [^{35}S] methionine in vitro (middle). The in vitro transcription and translation product V5 comigrates with hnRNP G protein (right).

biotinylated rabbit anti dog IgG (Jackson Immuno Research laboratories, West Grove, Pennsylvania, dilution 1:50), and Streptavidin Texas Red (Amersham U.K., dilution 1:50).

RESULTS

p43 is hnRNP G

Previous results had indicated an association of p43 with hnRNP complexes isolated from HeLa cells (38). In order to confirm this association and to unambiguously identify the hnRNP protein recognized by anti-p43 antisera, we isolated hnRNP complexes from [^{35}S]methionine-labeled HeLa cell nucleoplasm and resolved the hnRNP polypeptides by two-dimensional gel electrophoresis. The proteins were then blotted onto nitrocellulose and detected by autoradiography. The same blot was subsequently probed with anti-p43 antiserum and p43 was detected by chemiluminescence as described in Materials and Methods. This allowed the direct superimposition of the autoradiographs to identify the hnRNP polypeptide corresponding to p43. As Figure 1 shows, only one of the polypeptides of the hnRNP complex (left panel) is recognized by the antisera, and this polypeptide comigrates precisely with hnRNP G (right panel). Several of the hnRNP proteins in the left panel are labeled following the nomenclature of Piñol-Roma et al. (32) for clarity. The spot observed at the acidic end of the gel is probably an artifact of the detection procedure, as it can also be observed when similar experiments are carried out with a variety of antibodies to other hnRNP proteins. Longer exposures also revealed very weak reactivity of the anti-p43 antiserum with the hnRNP A and B proteins (data not shown).

Cloning and sequencing of hnRNP G cDNA

The dog antiserum p43 was used to screen a λgt11 library from the human breast carcinoma cell line ZS-75. A clone was isolated as described in Materials and Methods and its nucleotide sequence was determined in both orientations. As it did not encode a full length hnRNP G, hybridization screening with the clone was carried out, and a cDNA of 1894 nucleotides, V5, was obtained which encodes the partial insert hnRNP G as judged by in vitro transcription/translation on 2D gels (Fig. 2). V5 contains an open reading frame potentially encoding 437 amino acids. The sequence around the first ATG codon at position 12 (5′-AAAAATGGTT-3) provides a favorable context for translation initiation with a G at +4 and an A at −3 (24). The predicted amino acid sequence of hnRNP G is shown in Figure 3. The deduced amino acid sequence consists of 437 residues with a predicted molecular mass of 47,753 daltons and a calculated

```
                RNP2                                                  RNP1
  1 MVEADRPGK[illegible]NTETN EKALEAVFGK YGRIVEVLLM KDRETNKS[illegible]ESPA
 61 DAKDAARDMN GKSLDGKAIK VEQATKPSFE SGRRGPPPPP RSRGPPRGL[illegible]SGGTR
121 GPPS[illegible]HMD DGGYSMNFNM SSSRGPLPVK RGPPPRSGGP PPKRSAPSGP VRSSSGMGGR
181 APVSRGRDSY GGPPRREPLP SRRDVYLSPR DDGYSTKDSY SSRDYPSSRD TRDYAPPPRD
241 YTYRDYGHSS SRDDYPSRGY SDRDGYGRDR DYSDHPSGGS YRDSYESYGN SRSAPPTRGP
301 PPSYGGSSRY DDYSSSRDGY GGSRDSYSSS RSDLYSSGRD RVGRQERGLP PSMERGYLLH
361 VIPTAVQAAD SQEVVAVEEA DLIEGEAEAD TRNKQNFGPK SQFKETKSGN YSIITTQGLL
421 KGKIVLLFLN SLLSSPP 437
```

Figure 3. Predicted amino acid sequence of the hnRNP G/p43 protein encoded by clone V5. The open reading frame is depicted with the one letter code. The RNA-binding domain is boxed, the RNP1 and RNP2 motifs of the RBD are black-boxed, and the three RGG motifs are stipple-boxed. Two putative O-glycosylation sites are underlined. A GYGGS sequence (see text) is double underlined.

pI of 9.56. These values are in close agreement with the migration of hnRNP G on 2D gels.

Primary structure features of hnRNP G protein

The primary structure of the protein displays two main features: presence of one amino-terminal 80-amino acid consensus sequence RNA-binding domain (RBD) and a carboxy-terminal basic auxiliary domain. The RBD is a typical one as it contains perfect matches to RNP1 and RNP2 motifs (2, 12, 22) and a series of highly conserved segments which can be perfectly aligned with those of other RBDs from proteins of divergent organisms. Interestingly, the hnRNP G has only one RBD. Among other known hnRNP proteins, hnRNP C1 is the only other protein with only one RBD. Several other RNA-binding proteins that contain only one RBD have been described. Of those 1×RBD proteins, the hnRNP G protein RBD shares additional amino acid sequence motifs with several glycine-rich RNA-binding proteins from Maize, Sorghum and Arabidopsis (8, 17, 28, 41). These RBDs can be aligned starting from the initial methionine and the additional strongly conserved sequences, DRET and MNGKXLDG, are boxed in Figure 4.

The carboxy-terminal part of hnRNP G can be divided into two different segments: the first one extends approximately from amino acids 87 through 356. This domain, which is basic and rich in serine (52/270 aa, 19.3%), arginine (44/270 aa, 16.3%), and glycine (44/270, 16.3%) contents has an overall amino acid content typical of the auxiliary domain of a number of RNA-binding proteins. Beside high numbers of SR or RS repeats, we also note the presence of SSRD or SSSRD repeats which are

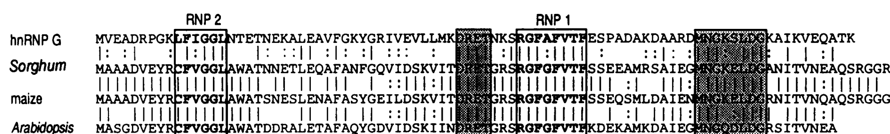


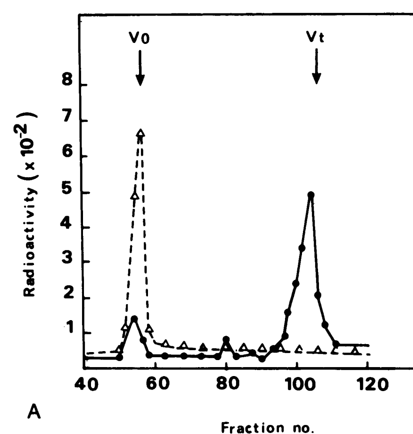
Figure 4. Alignment of RBDs of RNA-binding proteins. p43 is at the top line. Other sequences: SORVU, glycine-rich protein of Sorghum vulgare (8); Abscisic acid-inducible glycine-rich protein of Maize (17); glycin-rich protein of Arabidopsis thaliana (41). The RNP1 and RNP2 motifs are boxed. Additional sequences with perfect homology are black-boxed. The sequences have been recovered from the NBRF and SwissProt data Banks.

not frequently found in arginine-serine-rich domains (43). This domain also contains three RGG repeats surrounded by other RG/GR repeats at amino acids 94 through 127. This regions is similar to the RGG box, a novel RNA-binding motif first identified in the hnRNP U protein (23). RG(P)nRS (n = 3−5) repeats can also be found all throughout this domain. The sequence GYGGS at amino acid 320 appears to be conserved in a variety of other hnRNP proteins including human proteins A1, A2/B1 (GYGGG, GYGSG), Drosophila proteins hjrp40.1, and 36.1 (30, 31). Especially, in the auxiliary domains of the glycine-rich proteins from plants mentioned above, the sequences are frequently repeated. The second part of the carboxy-terminal domain of hnRNP G corresponds to amino acids 357 through 437, and is rich in leucine (10/81 aa, 12.4%).

hnRNP G contains O-linked N-acetylglucosamine

Previous studies have shown that hnRNP G binds Wheat Germ Agglutinin (WGA) and incorporates radioactive [^{14}C]glucosamine upon metabolic labeling of cells (38). These data raised the possibility that hnRNP G was post-translationally modified by O-linked N-acetylglucosamine, as reported for other nuclear glycoproteins (20). To examine this further, [^{3}H]glucosamine-labeled hnRNP G was immunoprecipitated with anti p43, the recovered material was submitted to mild alkaline borohydride treatment in order to release O-linked carbohydrate moieties from polypeptide chains (6), and chromatographed on a Bio-gel-P2 column (Fig. 5A). Upon this treatment, 80% of the loaded radioactivity was eluted in a peak with a chromatographic behavior similar to that of monosaccharides (mannose). The remaining radioactive material (20%) eluting at the position of the untreated p43 sample was shown to result from incomplete release of O-linked sugars (data not shown). In a parallel control experiment, we verified that the radioactivity associated with p43 was not released by a mild acetic acid treatment (data not shown). Our data strongly suggested that the radioactivity born by p43 was associated with a N-acetylhexosamine and not with sialic acid which is sensitive to mild acid treatment.

To analyze the nature of the carbohydrate molecules, the radioactive material released by mild alkaline borohydride treatment was submitted to paper chromatography (19). Under the conditions used, only 10% of the spotted material did not migrate (fig. 5B). Most of the radioactivity had the mobility of authentic acetylglucosaminitol. A minor peak accounting for 13% of the migrating radioactivity was not identified according to the standards used. This material was not further characterized. These results show: 1) the efficiency of the alkaline conditions since the major part of the radioactivity was identified as an N-acetylhexose; 2) the exclusive presence of the O-linked-N-acetylglucosamine as no N-acetylgalactosaminitol was detected. Moreover, these results are consistent with experiments showing



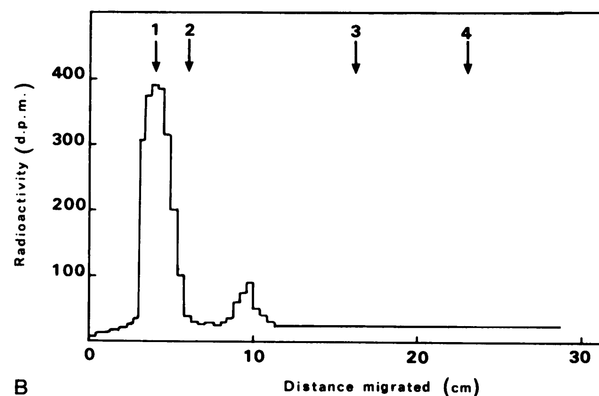


Figure 5. O-glycosylated status of p43/hnRNP G protein. **A:** Chromatographic behavior of [^{3}H]glucosamine-labeled p43 before and after mild alkaline borohydride treatment. p43 labeled with tritiated glucosamine was applied to a column of Bio-gel P2 (80×1.4cm) equilibrated with water, before (△—△) and after (■—■) alkaline borohydride treatment. One ml fractions were collected. V_o and V_t represent the excluded volume and the total volume respectively. **B:** Identification of [^{3}H]glucosamine-labeled material released from p43 by mild alkaline borohydride treatment. The radioactive material was subjected to paper chromatography (see materials and methods). Arrows indicate the migration of standards: 1: N-acetylgalactosaminitol; 2: N-acetylglucosaminitol; 3: N-acetylgalactosamine; 4: N-acetylglucosamine.

that treatment of HeLa cells with 2mM tunicamycin, an inhibitor of N-glycosylation, did not appear to modify the electrophoretic mobility of radioactivity bound to hnRNP (data not shown). The hydroxyamino acids (serine, threonine) that are O-glycosylated are usually adjacent to an acidic amino acid and within at most two residues of a proline (20). The sequence ESP is located close to the RNP1 motif at position 180. Also a PSRDT sequence is found starting at nucleotide 687. However, other O-glycosylated

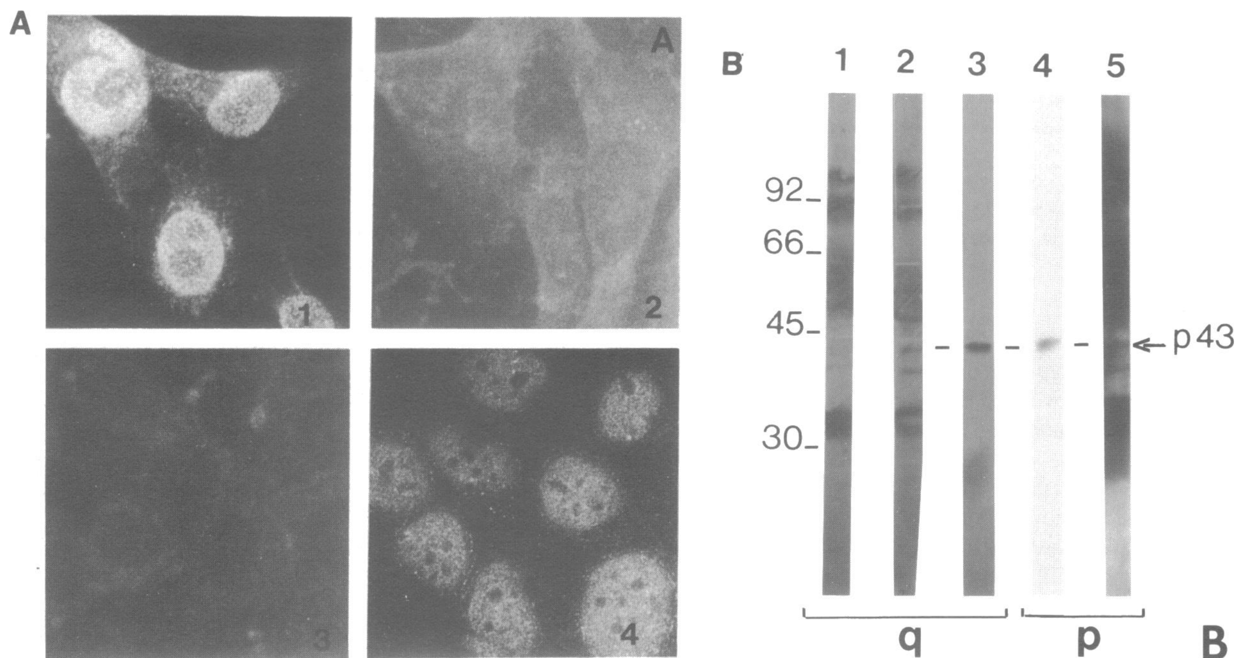


Figure 6. Glycosylation status of p43/hnRNP G protein in resting and proliferating HOS cells. A: Indirect immunofluorescence microscopy. Confluent HOS cells were deprived by incubation for three days in medium supplemented with 0.5% of fetal calf serum. Cell growth arrest was ascertained by flow cytometry analysis showing that up to 90% of cells were in G0/G1 on day 2 of serum deprivation Cells were stained with canine and anti-p43 antibodies (dilution 1:400, A and D) or a pool of normal dog serum (B and C) at the same dilution detection was with FITC-conjugated goat anti dog immunoglobulin. A, B: proliferating cells; C, D: resting cells. Cytoplasmic staining with nonimmune serum is completely non-specific and was not reproducible from one experiment to another one. **B:** p43 is glycosylated in serum deprived cells. Anti-p43 immune complexes were recovered from protein extracts of resting (q) and proliferating (p) cells. They were resolved by SDS-PAGE, transferred to nitrocellulose membrane which were assayed with anti-p43 antibodies or WGA. Detection of p43 in immune and nonimmune complexes was performed with protein A coupled to peroxidase, using the ECL procedure (Amersham). Glycosylated bands were revealed by WGA coupled to peroxidase as previously reported (38). Lanes 1, 2 and 5: WGA coupled to peroxidase. Lane 1: nonimmune complex with normal dog serum. Lanes 2 and 5: immune complex with anti-p43 antibodies. Lane 3 and 4: anti-p43 immune complex with anti-p43 antibodies. The background variations reflect the different times of exposures for revealing the bands. The presence of multiple signals in nonimmune and immune complexes results from the enrichment for Ig proteins by protein A.

sites that differ from this arrangement have also been described for cytoplasmic glycoproteins, (18).

hnRNP G is present as a glycoprotein in serum-deprived cells

We have previously shown that p43/hnRNP G is located in the nucleus of actively proliferating interphase cells and in the extrachromosomal area in mitotic cells (38). We investigated the fate of the protein in non-proliferating cells. Immunofluorescence analysis was carried out on HOS cells (a human carcinoma cell line) arrested by serum deprivation (Fig. 6A). We noticed that the fluorescence staining appeared to be significantly diminished in resting cells, compared to that of growing cells (compare patterns 1 and 4 in Fig. 6), although its nucleoplasmic localization excluding nucleoli persisted.

In order to examine whether the glycosylated status of hnRNP G varied between resting and proliferating cells, hnRNP G was immunoprecipitated from extracts of actively proliferating or serum-deprived HOS cells with anti-p43 antibodies. The immune complexes were electrophoresed, transferred on nitrocellulose and revealed either by anti-p43 antibodies or by WGA coupled to peroxidase (Fig. 6B). As expected, only one band was revealed by anti-p43 antibodies in those lanes corresponding to resting cell extracts (lane 3) and proliferating cell extracts (lane 4). Peroxidase-WGA revealed a number of non specific bands in the lane corresponding to anti-p43 immune complex as well as to the normal serum immune complex. Most of these correspond to glycosylated heavy (50–60kD range) and light (25–35kD range) immunoglobulin chains which are recovered by protein A. One band comigrating with hnRNP G/p43 was repeatedly observed in the anti-p43 antibody complex (lanes 2 and 5) but not in the non-immune complex (lane 1, the significance of the lower band that is not detected by antibody.is unknown) These data indicate that hnRNP G is glycosylated both in actively proliferating and resting cells. Although the signal with WGA is stronger in growing cell, this may merely be due to the greater amount of hnRNP G in these cells.

hnRNP G is bound to RNA *in vivo*

In order to confirm the association of hnRNP G with RNA *in vivo*, living HeLa cells were irradiated with UV light for different times and complexes of poly(A)-containing RNA-proteins were boiled with SDS and β-mercaptoethanol in order to eliminate non-covalent associations, and chromatographed onto oligo(dT) cellulose columns. The eluted fractions were digested with RNase A and submitted to gel electrophoresis (SDS-PAGE). Figure 7A shows that the anti-C1/C2 4F4 monoclonal antibody revealed the presence of a band at the expected position for these proteins. Similarly, anti-p43 serum reacted with a band at the position of hnRNP G. These bands were not observed in control samples which had not been submitted to UV irradiation.

Consistent with this result, the hnRNP G protein was retained on a column of single-stranded DNA and eluted with 0.4M NaCl (Fig. 7B). This binding was sensitive to heparin that completely released the protein in the flow-through. Parallel experiments with

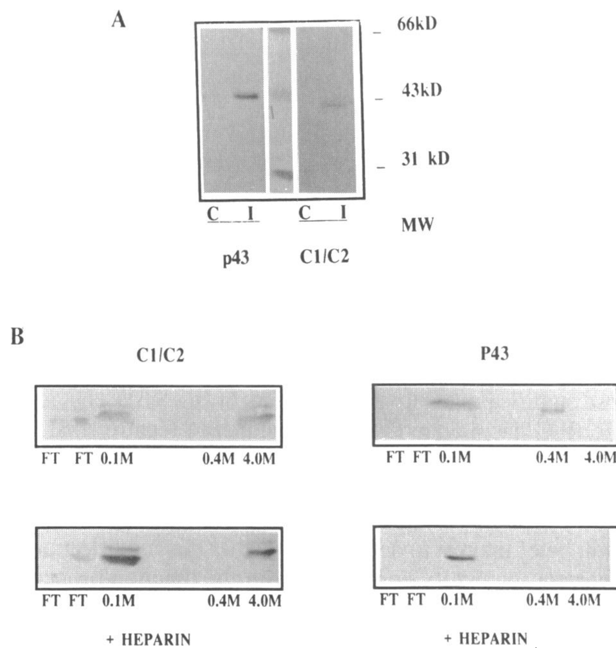


Figure 7. Immunoblot analysis of HeLa cell p43/hnRNP G protein cross-linked to RNA and single-stranded DNA ultrogel chromatography. **A**: Protein RNA complexes eluted from oligo(dT) cellulose were treated with RNAse A and submitted to SDS-PAGE. The blots were probed with anti-p43 serum or 4F4 monoclonal antibody to C1/C2 proteins. C: non-irradiated control sample. I: irradiated sample. Note that a longer time of UV exposure was required to stabilize the p43-RNA complex (3 min.) than the C1/C2 complex (1.5 min.). **B**: HeLa cell nuclear fractions suspended in buffer containing 0.1M NaCl were loaded onto a column of ssDNA-ultrogel column. Following several washes with the same salt molarity, elution was carried out by increasing concentrations of NaCl. Heparin (1mg/ml) was added in the loading solutions.

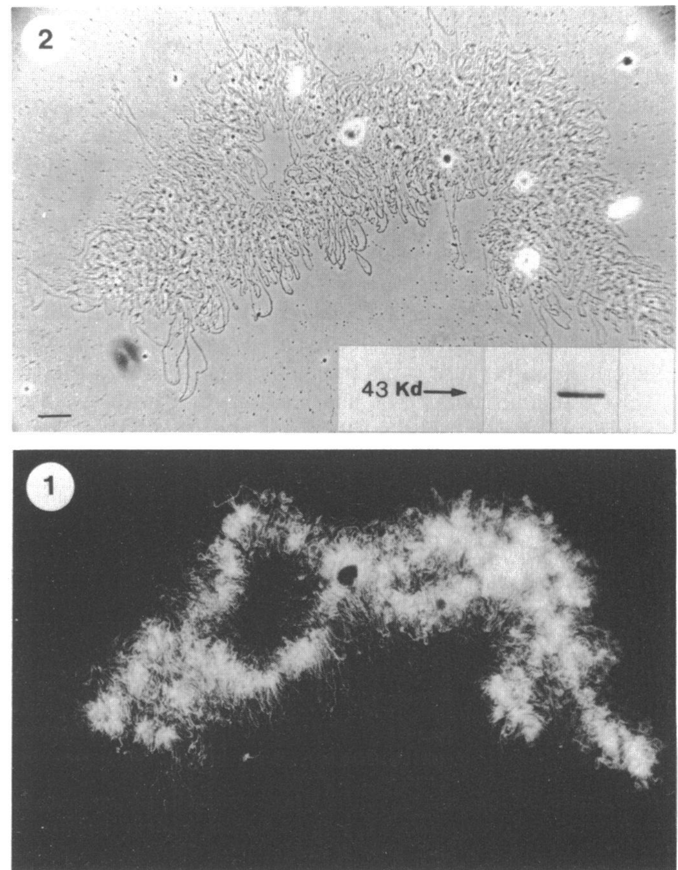


Figure 8. Immunofluorescence microscopy of Pleurodeles waltl lampbrush chromosomes stained with anti-p43 antibodies. Phase contrast (**A**) and fluorescence (**B**) images of the same chromosome area after treatment with anti-p43 serum and staining with Texas Red. Note that the majority of lateral loops are stained. No staining was observed with non-immune antibodies. Bar:10μm. Inset: immunoblot of an oocyte germinal vesicle extract showing the presence of a band running at 43kD (lane 2). A non-immune serum (lane 1) or another dog serum with another specificity (lane 3) gave a negative result.

the C1/C2 protein confirmed, as previously noted (33), that the DNA-protein interaction of this protein was more stable, as 4.0 M NaCl was needed to elute it. Taken together, these data confirm that hnRNP G is a single-stranded nucleic acid binding protein and that it is bound to poly(A)$^+$ RNA in living cells.

Distribution of hnRNP G in amphibian oocyte lampbrush chromosomes

We have previously shown that the anti-p43 canine autoantibodies reacted positively with different mammalian cell species by indirect immunofluorescence assay. The distribution of hnRNP G was also examined in oocytes of the newt, *Pleurodeles waltl.* The protein was abundant in the germinal vesicle (nucleus, inset Fig. 8) but was not detected in the cytoplasm. Similarly, staining of *X.laevis* embryos showed that hnRNP G was present in the nuclei of all the cells at all stages examined (data not shown). As antibodies to other hnRNP proteins have previously shown differential association of hnRNP proteins with nascent transcripts (30, 33), we used anti-p43 antibodies to examine the distribution of hnRNP G on lampbrush chromosomes. Anti-p43 antibodies also stained very strongly the lampbrush chromosomes (Fig. 8). All loops exhibited a homogeneous pattern of labeling consistent with staining of the transcriptional complexes. The nucleoplasm also contained small labeled particles (not visible on this figure) that are likely to represent RNP particles. Lampbrush chromosome axis where DNA condenses in the chromomeres as well as nucleoli did not appear to be stained.

DISCUSSION

We have characterized and determined the amino acid sequence of a novel autoantigen, p43, and have found that it corresponds to a component of hnRNP complexes, the G protein. The identification of p43 as hnRNP G was ascertained by several lines of evidence: first, the anti-p43 antibodies recognized a protein in immunopurified hnRNP complexes with the same molecular mass and pI as hnRNP G. Second, the in vitro transcription and translation product of a full length clone comigrated with the authentic hnRNP G protein on two-dimensional gels. In addition, the primary structure of the p43/hnRNP G protein contains one RBD (RNA-binding domain) which is a characteristic of many RNA-binding proteins. Interestingly, the RBD of hnRNP G showed amino acid sequence similarities with some glycine-rich RNA-binding proteins from plants. Though the significance of the homologous regions outside of RNP1 and RNP2 is still unknown, these regions of similarity may confer similar RNA-binding specificity. In addition to having only one RBD, the

auxiliary domains of these plant proteins are rich in glycine and have GYGGX motifs like hnRNP G. The presence of similar structures in these proteins suggests that they may have similar functions in RNA biogenesis. The cluster of three RGG repeats downstream of the RBD suggests that this part of the hnRNP G protein exhibits RNA-binding activity, like other single-stranded DNA-binding proteins containing RGG boxes (23). Alignment of the p43 RGG cluster with the consensus sequence of the RGG box motif shows no evidence of conservation. In addition, while this region is glycine-rich (41%), it contains no asparagine nor phenylalanine residues at proximity of the arginine residues, two significant features of the RGG box (23). Nevertheless, because of its basic character, this p43 region is likely to interact with RNA and justifies experiments to prove formally its binding capacity. Beside glycine, the auxiliary domain of hnRNP G exhibits high arginine and serine contents similar to other hnRNP proteins and contains unique motifs as yet not described in other RNA-binding proteins.

Regions containing interspersed aromatic residues in glycine-rich domains have been shown to contribute to cooperative binding to single-stranded nucleic acids (25). While it is possible that the auxiliary domain of hnRNP G is functionally analogous to that of other RNA-binding proteins because of its overall basic and hydrophobic character, the presence of repeated motifs suggests a particular function, for instance an association with specific RNA or with other proteins. For this reason, it will be interesting to isolate other RNA-binding proteins containing the same or closely related motifs in their C-terminal domains.

Perhaps the most notable feature of hnRNP G is that it is glycosylated. Evidence for the existence of a variety of nuclear glycoproteins is extensively documented (for a review see 20). Strikingly, a number of these glycoconjugates bear GluNAc residues through an O-linkage to serine or threonine. However, unequivocal characterization has been achieved only for a few individual molecular species. Beside the glycoproteins of the nuclear pore complex (NPC), several human RNA-polymerase II transcription factors including Sp1, AP1/jun and c-fos have also been shown to bear multiple monosaccharide O-GluNAc residues (21). For instance, transcriptional stimulation by Sp1, a zinc finger transcriptional factor, has been reported to be significantly reduced when the factor was preincubated with WG-A (21), although a specific function for the GluNAc residues could not be deduced from these experiments. A role for the O-linked proteins of the nuclear pore complex in mediated nuclear protein import is supported by findings that nuclear import is inhibited by WGA or monoclonal antibodies recognizing these glycoproteins (14, 15). Recent studies have suggested the association of a lectin, CBP35, in hnRNPs (26). It will be of interest to investigate the possibility that hnRNP proteins other than the hnRNP G species are glycosylated.

It has been recently shown that the hnRNP A1 protein shuttles between the nucleus and the cytoplasm in a transcription-dependent fashion, while others (C1/C2) do not appear to be dependent on RNA Pol II-mediated transcription for transport (34, 35). It will be of interest to determine to which class of these hnRNP G belongs. If a transcription-dependent shuttle of hnRNP G would exist, one should expect a direct association with cytoplasmic poly(A)-RNA. In addition, it will be important to examine whether the glycosylation status of hnRNP G bears any relation to its intracellular transport and/or location.

The staining pattern of lampbrush chromosomes of Pleurodeles indicates that hnRNP G/p43 is associated with nascent transcripts. Furthermore, the homogeneous pattern of staining involving all loops is clearly different of that of hnRNP L protein which also concerns all loops, with a more intense staining of giant loops (33). These data suggest that the p43 protein is likely to be systematically associated with all hnRNP transcription complexes, while the L protein might be, at least in part, a component of specialized hnRNP complexes. Contrasting with this putative role in the general process of transcription, p43 is a low abundance hnRNP protein. (32, 38). In fact, we have previously shown that the solubility of the protein is strongly dependent on the isolation conditions from nuclear extracts (38).

To summarize, our results show that the p43/hnRNP G protein shares some common properties with other hnRNP proteins, the most significant being the presence of an RBD. In addition, hnRNP G has unique C-terminal sequence features that clearly distinguish it from other hnRNP proteins sequences. Perhaps more importantly, to our knowledge hnRNP G is the first glycosylated hnRNP (or RNA-binding protein) described to date. It will be of interest to determine whether this modification bears important regulatory consequence on the function/properties of hnRNP G. The availability of the antisera and of cDNA clones opens the possibility of extensive studies on the significance of properties of hnRNP G described here.

ACKNOWLEDGEMENTS

This work was supported by a Grant from the 'association pour la Recherche contre le Cancer de Villejuif' (ARC) to CJL and by the Howard Hughes Medical Institute and grants from the NIH to GD.

REFERENCES

1. Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J. (1992) *J. Mol. Biol.*, 215, 403–410.
2. Bandziulis,R.J., Swanson,M.S., and Dreyfuss, G. (1989) *Genes Dev.*, 3, 431–437.
3. Biamonti,G., Buvoli,M., Bassi,M.T., Morandi,C., Cobianchi,R., and Riva,S. (1989). *J. Mol. Biol.*, 207, 491–503.
4. Burd,C.G., Swanson,M.S., GÖrlach,M., and Dreyfuss,G. (1989) *Proc. Natl. Acad. Sci. USA*, 86, 9788–9792
5. Buvoli,M., Biamonti,G., Tsoulfas,P., Bassi,M.T., Ghetti,A., Riva,S., and Morandi,C. (1988) *Nucleic Acids Res.*, 16, 3751–3770.
6. Carlson,D.M. (1968) *J. Biol. Chem.*, 243, 616–626.
7. Cobianchi,F., SenGupta,D.N., Zmudzka,B.Z., and Wilson,S.H. (1986 *J. Biol. Chem.*, 261, 3536–3543.
8. Cretin,C., and Puigdomènech,P. (1990) *Plant Mol. Biol.*, 15, 783–785.
9. Datar,K.V., Dreyfuss,G., and Swanson,M.S. (1993) *Nucleic Acids Res.*, 21, 439–446.
10. Dreyfuss,G., Choi,Y.D., and Adam,S.A. (1984) *Mol. Cell. Biol.*, 4,1104–1114.
11. Dreyfuss,G. (1986) *Annu. Rev. Cell Biol.*, 2, 459–498.
12. Dreyfuss,G., Swanson,M.S., and Piñol-Roma,S. (1988) *Trends in Biochem. Sci.*, 13, 86–91.
13. Dreyfuss,G., Matunis,M.J., Piñol-Roma,S., and Burd,C. (1993) Annu. Review Biochem., *in press*.
14. Featherstone,C., Darby,M.K., and Gerace,L. (1988) *J. Cell Biol.*, 107, 1289–1297.
15. Finlay,D.R., Newmeyer,D.D., Price,T.M., and Forbes,D.J. (1987) *J. Cell Biol.*, 104, 189–200.
16. Ghetti,A., Piñol-Roma,S., Michael,W.J., and Dreyfuss,G. (1992) *Nucleic Acid Res.*, 20, 3671–3678.
17. Gòmez,J., Sanchez-Martínez,D., Stiefel,V., Rigau,J., Puigdomènech,P., and Pagès,M. (1988) *Nature*, 334, 262–264.
18. Hagmann,J., Grob,M., and Burger,M.M. (1992) *J. Biol. Chem.*, 267, 14424–14428.
19. Hanover,J.A., Cohen,C.K., Willingham,M.C., and Park,M.K. (1987) *J. Biol. Chem.*, 262, 9987–9994.

20. Hart,G., Matliwanger, R., Holt,G. and Kelly,W. (1989) *Annu. Rev. Biochem.*, 58, 841–874.
21. Jackson,S., and Tjian,R. (1988) *Cell*, 55, 125–133.
22. Kenan,D.J., Query,C.C. and Keene,J.D. (1991) *Trends in Biochem. Sci.*, 16, 214–220.
23. Kiledjian,M. and Dreyfuss,G. (1992) *EMBO J.*, 11, 2655–2664.
24. Kozak,M., (1986) Cell, 44, 283–292.
25. Kumar,A., Casas-Finet,A.J., Luneau,C.J., Karpel,R.L., Merrill,B.M., Williams, K.R., and Wilson,S.H. (1990) *J. Biol. Chem.*, 265, 17094–17100.
26. Laing,J.G. and Wang,J.L. (1988) *Biochemistry*, 27, 5329–5334.
27. Lacroix,J-C., Azzouz,A., Boucher,D., Abbadie,C., Pyne,K. and Charlemagne,J. (1985) *Chromosoma (Berlin)*, 92, 69–80.
28. Ludevid,M.D., Freire,M.A., Gùmez,J., Burd,C.G., Giralt,A.E., Dreyfuss,G., and Pagès,M. (1992) *Plant J.*, 2, 999–1003.
29. Matunis,M.J., Michael,W.M., and Dreyfuss,G. (1992) *Mol. Cell. Biol.*, 12, 164–171.
30. Matunis,M.J., Matunis,E.L., and Dreyfuss,G. (1992) *J. Cell Biol.*, 116, 245–255.
31. Matunis,E.L., Matunis,M.J. and Dreyfuss,G. (1992) *J. Cell Biol.*, 116, 257–269.
32. Piñol-Roma,S., Choi,D.Y., Matunis,M.J., and Dreyfuss,G. (1988) *Genes Dev.*, 2, 215–227.
33. Piñol-Roma,S., Swanson,M.S., Gall,J.G., and Dreyfuss,G. (1989) *J. Cell Biol.*, 109, 2575–2587.
34. Piñol-Roma,S., and Dreyfuss,G. (1991) *Science*, 253, 312–314.
35. Piñol-Roma,S., and Dreyfuss G. (1992) *Nature*, 355, 730–732.
36. Riva,S., Morandi,P., Tsoulfas,M., Pandolfo,G., Biamonti,B., Merril,K.R., Williams,G., Multhaup,K., Bayreuther,H., Werr,B.H., and Schafer,K.P. (1986) *EMBO J.*, 5, 2267–2273.
37. Sanger,F., Nicklen,S., and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, 74, 5463–5467.
38. Soulard,M., Barque,J-P., Della Valle,V., Hernandez-Verdun,D., Masson,C., Danon,F., and Larsen,C-J. (1991) *Exp. Cell Res.*, 193, 59–71.
39. Swanson,M.S., Nakagawa,T.Y., LeVan,K., and Dreyfuss,G. (1987) Mol. Cell. Biol., 7, 1731–1739.
40. Trevelyan,W.E., Procter,D.P., and Harrison,J.S. (1950) *Nature*, 166, 444–445.
41. Van Nocker,S., and Vierstra,R.D. (1993) *Plant Mol. Biol.*, 21, 695–699.
42. Youakim,A., and Herscovics,A. (1985) *Cancer Res.*, 45, 5505–5511.
43. Zahler,A.M. Stolk,J.A., Lane,W.S., and Roth,M.B. (1992) *Genes Dev.*, 6, 837–847.