

# cDNA cloning and sequencing of human fibrillarin, a conserved nucleolar protein recognized by autoimmune antisera

(evolutionary conservation/scleroderma/RNA-binding motif)

JOHN P. ARIS AND GÜNTER BLOBEL

Laboratory of Cell Biology, Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10021

Contributed by Günter Blobel, October 12, 1990

**ABSTRACT** We have isolated a 1.1-kilobase cDNA clone that encodes human fibrillarin by screening a hepatoma library in parallel with DNA probes derived from the fibrillarin genes of *Saccharomyces cerevisiae* (*NOPI*) and *Xenopus laevis*. RNA blot analysis indicates that the corresponding mRNA is  $\approx 1300$  nucleotides in length. Human fibrillarin expressed *in vitro* migrates on SDS gels as a 36-kDa protein that is specifically immunoprecipitated by antisera from humans with scleroderma autoimmune disease. Human fibrillarin contains an amino-terminal repetitive domain  $\approx 75$ –80 amino acids in length that is rich in glycine and arginine residues and is similar to amino-terminal domains in the yeast and *Xenopus* fibrillarins. The occurrence of a putative RNA-binding domain and an RNP consensus sequence within the protein is consistent with the association of fibrillarin with small nucleolar RNAs. Protein sequence alignments show that 67% of amino acids from human fibrillarin are identical to those in yeast fibrillarin and that 81% are identical to those in *Xenopus* fibrillarin. This identity suggests the evolutionary conservation of an important function early in the pathway for ribosome biosynthesis.

Fibrillarin is a component of a nucleolar small nuclear ribonucleoprotein (snRNP) particle thought to participate in the first step in processing preribosomal RNA. In humans, fibrillarin is associated with the U3, U8, and U13 small nuclear RNAs (1). Several lines of evidence indicate a role for fibrillarin in this initial processing step, which in human is the endonucleolytic removal of the  $\approx 414$  nucleotide (nt) long 5' externally transcribed spacer (2). A role for fibrillarin has been demonstrated *in vitro* (3). The U3 RNA has been shown to be involved by psoralen cross-linking (4, 5). Immunolocalization suggests that fibrillarin occurs at the 5' end of pre-rRNA transcripts in the familiar "Christmas tree" pattern (6). In eukaryotic cell types examined fibrillarin occurs in the interphase nucleolus (7–9), where ribosome synthesis takes place. However, composition of the fibrillarin-containing snRNP may differ between species. In the yeast *Saccharomyces cerevisiae* up to 9 small nuclear RNAs may associate with fibrillarin (10).

Fibrillarin was originally identified in *Physarum polycephalum* by Christensen *et al.*, and was termed fibrillarin by Ochs *et al.* because of its ultrastructural localization to the fibrillar region of the mammalian nucleolus (11, 12). Fibrillarins from different species share several biochemical characteristics: a molecular mass from 33 to 38 kDa, basic isoelectric point (pI 8.5–10), and abundance of glycine ( $\approx 20$  mol%) and  $N^G, N^G$ -dimethylarginine ( $\approx 4$  mol%). Antisera from  $\approx 8\%$  of humans with the autoimmune disease scleroderma recognize fibrillarin and may detect this protein across the fungi and animal kingdoms (7–9, 13–15). The genes for fibrillarin from *S. cerevisiae* and *Xenopus laevis* have been isolated and se-

quenced (14–16). The yeast gene, which we termed *NOPI* (nucleolar protein 1), is essential for viability (ref. 16, R. Henriquez, G.B., and J.P.A., unpublished data). The primary structure of human fibrillarin reported herein shows a high degree of homology with fibrillarins from yeast and *Xenopus*, which agrees with the detection of fibrillarin in all eukaryotes examined from yeast to human.

## MATERIALS AND METHODS

**Screening a Human cDNA Library.** Two probes were used in parallel to screen  $\approx 1 \times 10^6$  plaques of a HepG2 library in bacteriophage  $\lambda$ ZAP (Stratagene). The 443-base-pair (bp) *Bst*XI-*Nde* I restriction fragment from the yeast *NOPI* gene (14) and the 1.19-kilobase (kb) *Ava* I fragment from the *Xenopus* Xomfib cDNA clone (15) were purified and labeled as described (14, 17). Hybridizations and washes were done according to standard methods under conditions of low stringency to achieve a melting temperature ( $t_m$ ) =  $-35^\circ\text{C}$  for the *NOPI* probe and a  $t_m$  =  $-30^\circ\text{C}$  for the Xomfib probe, assuming an average labeled probe length of 100 bp (18, 19). Although the probes hybridized to a large number of plaques, only a small number hybridized to both yeast and *Xenopus* fibrillarin probes. Seventeen isolates were purified, and their inserts were retrieved by *in vivo* excision (Stratagene) to yield plasmids containing inserts. The plasmid retrieved from clone  $\lambda 20$  hybridized under high stringency in the presence of tetramethylammonium chloride (17) to a degenerate mixture of complementary oligonucleotides based on the amino-terminal sequence MKPGFS (single-letter code) of rat fibrillarin (8). Phage DNA was prepared from  $\lambda 20$ , and the insert was ligated into the *Eco*RI site of pBluescript SK+ to generate plasmid p20J. Forty-eight base pairs of untranslated sequence at the 5' end of the cDNA insert in p20J, including an ATG codon out of frame with the initiator ATG, were removed to generate plasmid p20J2.

**DNA Sequence Analysis.** Both DNA strands of the insert in p20J2 were sequenced by using double-stranded DNA templates and synthetic primers with the dideoxynucleotide chain-termination method and modified T7 polymerase (Sequenase) with dGTP and, for ambiguous sequences, with 2'-deoxyinosine-5'-triphosphate (United States Biochemical).

**RNA Blot Analysis.** Total and poly(A)<sup>+</sup> RNA were prepared from *S. cerevisiae* cultures and HeLa cells as described (14, 20) and electrophoresed in a 1% agarose/formaldehyde gel, transferred to nitrocellulose membrane, and processed according to standard methods (17). The genomic *Eco*RI fragment containing *NOPI* or the *Eco*RI restriction fragment from p20J2 was isolated and labeled as described (14). The *NOPI* probe was incubated with the blot under standard

conditions and washed twice with  $\times 0.125$  standard sodium citrate (SSC) buffer ( $1\times$  SSC is 0.15 mM sodium chloride/0.015 mM sodium citrate, pH 7) at 55°C (17). The blot was placed into  $\approx 100^\circ\text{C}$  water, cooled for 5 min, and exposed to film to verify complete removal of the *NOP1* probe. The blot was then hybridized to the human fibrillarin probe from p20J2 under the same conditions as above but washed with  $1\times$  SSC buffer at 25°C to achieve conditions of lower stringency.

**In Vitro Protein Synthesis.** Cell-free translations were conducted for 1 hr at 25°C in a 50- $\mu\text{l}$  reaction volume containing 20  $\mu\text{l}$  of rabbit reticulocyte lysate (Promega Biotec), 100  $\mu\text{Ci}$  (1 Ci = 37 GBq) of [ $^{35}\text{S}$ ]methionine, 5 units of RNasin, and  $\approx 500$  ng of mRNA, as described (21). Synthetically capped human or yeast fibrillarin mRNA was prepared by transcription of plasmid p20J2 linearized with *Bam*HI or plasmid pNOP1-RV linearized with *Eco*RI, respectively. Plasmid pNOP1-RV was constructed by digesting yeast fibrillarin clone 32 (14) with *Eco*RV restriction enzyme followed by blunt-end ligation using standard methods (17). Transcription reactions (25  $\mu\text{l}$ ) were done with T7 RNA polymerase (10 units) with an RNA transcription kit (Stratagene).

**Immunoprecipitation.** Translations were diluted with 950  $\mu\text{l}$  of 1% Triton X-100 and 1 mM iodoacetamide in IP buffer (50 mM Tris, pH 7.5/150 mM NaCl/1 mM EDTA/2 mM  $\text{NaN}_3$ ). The diluted translation (200  $\mu\text{l}$ ) was combined with 20  $\mu\text{l}$  of reconstituted antinucleolar antiserum (ANA) (equiva-

lent to 1.0  $\mu\text{l}$  of antiserum), or 0.5  $\mu\text{l}$  of antiserum 538, or no antiserum. After incubation on ice for 2 hr and centrifugation for 5 min at  $12,000\times g$  at 4°C, the supernatant was transferred to a microcentrifuge tube containing 15  $\mu\text{l}$  of protein G-Sepharose (Pharmacia LKB Biotechnology) previously washed with IP buffer/1% Triton X-100. The Sepharose beads were briefly centrifuged and washed sequentially at  $\approx 25^\circ\text{C}$  with IP buffer/1% Triton X-100/2 M urea, IP buffer/1% Triton X-100, IP buffer, and 10 vol of water. Samples were boiled for 3 min in SDS/PAGE sample buffer containing 50 mM dithiothreitol, followed by the addition of 1/10th vol of freshly made 1 M iodoacetamide and analyzed by SDS/PAGE using 12% gels as described (14).

## RESULTS

**cDNA Cloning and Sequencing.** A human hepatoma library in bacteriophage  $\lambda$  was screened in parallel with two DNA probes derived from the yeast fibrillarin gene *NOP1* (14) and the *Xenopus* fibrillarin clone *Xomfib* (15). The *NOP1* probe did not include that part of the 5' end of the gene consisting of a long G+C-rich repetitive sequence, and the *Xomfib* probe included only  $\approx 50$  bp of the equivalent region. Probes were prepared in this way because a previous screening of the hepatoma library with full-length probes generated a number of false-positive clones, apparently due to the G+C-rich

-28	CCC	GAG	CCG	CAC	AA	ACC	AGG	CGC	TG	CC	ATG	AAG	CCA	GGA	TTC	AGT	CCC	CGT	GGG	GGT	GGC	TTT	GGC	GGC	CGA	GGG	GGC	TTT	GGT	GAC		
1													MET	LYS	PRO	GLY	PHE	SER	PRO	ARG	GLY	GLY	GLY	PHE	GLY	GLY	ARG	GLY	GLY	PHE	GLY	ASP
61	CGT	GGT	GGT	CGT	GGA	GGC	CGA	GGG	GGC	TTT	GGC	GGG	GGC	CGA	GGT	CGA	GGC	GGA	GGC	TTT	AGA	GGT	CGT	GGA	CGA	GGA	GGA	GGA	GGA	GGA		
21	ARG	GLY	GLY	ARG	GLY	GLY	ARG	GLY	GLY	PHE	GLY	GLY	GLY	ARG	GLY	ARG	GLY	GLY	GLY	PHE	ARG	GLY	ARG	GLY	ARG	GLY	GLY	GLY	GLY	GLY		
142	GGT	GGA	GGC	GGC	GGC	GGC	GGT	GGA	GGA	GGA	GGA	AGA	GGT	GGT	GGA	GGC	TTC	CAT	TCT	GGT	GGC	AAC	CGG	GGT	CGT	GGT	CGG	CGG	CGG			
48	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	GLY	ARG	GLY	GLY	GLY	GLY	PHE	HIS	SER	GLY	GLY	ASN	ARG	GLY	ARG	GLY	ARG	GLY	ARG			
223	GGA	GGA	AAA	AGA	GGA	AAC	CAG	TCG	GGG	AAG	AAT	GTG	ATG	GTG	GAG	CCG	CAT	CGG	CAT	GAG	GGT	GTC	TTC	ATT	TGT	CGA	GGA	GGA	GGA			
75	GLY	GLY	LYS	ARG	GLY	ASN	GLN	SER	GLY	LYS	ASN	VAL	MET	VAL	GLU	PRO	HIS	ARG	HIS	GLU	GLY	VAL	PHE	ILE	CYS	ARG	GLY	GLY	GLY			
304	AAG	GAA	GAT	GCA	CTG	GTC	ACC	AAG	AAC	CTG	GTC	CCT	GGG	GAA	TCA	GTT	TAT	GGA	GAG	AAG	AGA	GTC	TCG	ATT	TCG	GAA	GGA	GGA	GGA			
102	LYS	GLU	ASP	ALA	LEU	VAL	THR	LYS	ASN	LEU	VAL	PRO	GLY	GLU	SER	VAL	TYR	GLY	GLU	LYS	ARG	VAL	SER	ILE	SER	GLU	GLY	GLY	GLY			
385	GAT	GAC	AAA	TTT	GAG	TAC	CGA	GCC	TGG	AAC	CCC	TTC	CGC	TCC	AAG	CTA	GCA	GCA	GCA	ATC	CTG	GGT	GGT	GTG	GAC	CAG	ATC	ATC	ATC			
129	ASP	ASP	LYS	PHE	GLU	TYR	ARG	ALA	TRP	ASN	PRO	PHE	ARG	SER	LYS	LEU	ALA	ALA	ALA	ILE	LEU	GLY	GLY	VAL	ASP	GLN	ILE	ILE	ILE			
466	CAC	ATC	AAA	CCG	GGG	GCT	AAG	GTT	CTC	TAC	CTC	GGG	GCT	GCC	TCG	GGC	ACC	ACG	GTC	TCC	CAT	GTC	TCT	GAC	ATC	GTT	GGT	GGT	GGT			
156	HIS	ILE	LYS	PRO	GLY	ALA	LYS	VAL	LEU	TYR	LEU	GLY	ALA	ALA	SER	GLY	THR	THR	VAL	SER	HIS	VAL	SER	ASP	ILE	VAL	GLY	GLY	GLY			
547	CCG	GAT	GGT	CTA	GTC	TAT	GCA	GTC	GAG	TTC	TCC	CAC	CGC	TCT	GGC	CGT	GAC	CTC	ATT	AAC	TTG	GCC	AAG	AAG	AGG	ACC	AAC	AAC	AAC			
183	PRO	ASP	GLY	LEU	VAL	TYR	ALA	VAL	GLU	PHE	SER	HIS	ARG	SER	GLY	ARG	ASP	LEU	ILE	ASN	LEU	ALA	LYS	LYS	ARG	THR	ASN	ASN	ASN			
628	ATC	ATT	CCT	GTG	ATC	GAG	GAT	GCT	CGA	CAC	CCA	CAC	AAA	TAC	CGC	ATG	CTC	ATC	GCA	ATG	GTG	GAT	GTG	ATC	TTT	GCT	GAT	GAT	GAT			
210	ILE	ILE	PRO	VAL	ILE	GLU	ASP	ALA	ARG	HIS	PRO	HIS	LYS	TYR	ARG	MET	LEU	ILE	ALA	MET	VAL	ASP	VAL	ILE	PHE	ALA	ASP	ASP	ASP			
709	GTG	GCC	CAG	CCA	GAC	CAG	ACC	CGG	ATT	GTG	GCC	CTG	AAT	GCC	CAC	ACC	TTC	CTG	CGT	AAT	GGA	GGA	CAC	TTT	GTG	ATT	TCC	TCC	TCC			
237	VAL	ALA	GLN	PRO	ASP	GLN	THR	ARG	ILE	VAL	ALA	LEU	ASN	ALA	HIS	THR	PHE	LEU	ARG	ASN	GLY	GLY	HIS	PHE	VAL	ILE	SER	SER	SER			
790	ATT	AAG	GCC	AAC	TGC	ATT	GAC	TCC	ACA	GCC	TCA	GCC	GAG	GCC	GTG	TTT	GCC	TCC	GAA	GTG	AAA	AAG	ATG	CAA	CAG	GAG	AAC	AAC	AAC			
264	ILE	LYS	ALA	ASN	CYS	ILE	ASP	SER	THR	ALA	SER	ALA	GLU	ALA	VAL	PHE	ALA	SER	GLU	VAL	LYS	LYS	MET	GLN	GLN	GLU	ASN	ASN	ASN			
871	ATG	AAG	CCG	CAG	GAG	CAG	TTG	ACC	CTT	GAG	CCA	TAT	GAA	AGA	GAC	CAT	GCC	GTG	GTC	GTG	GGA	GTG	TAC	AGG	CCA	CCC	CCC	CCC	CCC			
291	MET	LYS	PRO	GLN	GLU	GLN	LEU	THR	LEU	GLU	PRO	TYR	GLU	ARG	ASP	HIS	ALA	VAL	VAL	VAL	VAL	GLY	VAL	TYR	ARG	PRO	PRO	PRO	PRO			
952	AAG	GTG	AAG	AAC	TGA	AGTTC	ACG	CTG	TCA	GAG	ATT	GCG	AG	AG	AT	G	T	G	T	G	T	A	T	A	A	A	A	A	A	A		
318	LYS	VAL	LYS	ASN	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		

FIG. 1. cDNA and protein sequences of human fibrillarin. Numbers indicate the positions of nucleotides (above) and amino acids (below) in the cDNA and protein sequences. The poly(A) addition signal is underlined. Features discussed in the text are as follows: the GDRGG motif is shaded; underlined arginines may be  $\text{N}^G, \text{N}^G$ -dimethylated; underlined lysines are near the end of the amino-terminal domain; the putative RNP consensus sequence is boxed; and a potential  $\alpha$ -helix is delimited by arrows.

repetitive sequences (data not shown). Seventeen clones were obtained that hybridized to both probes under low-stringency hybridization conditions. One positive clone hybridized under conditions of high stringency to a degenerate mixture of oligonucleotides based on the six amino-terminal residues of rat fibrillar, MKPGFS (single-letter code) (data not shown; ref. 8). The cDNA insert was subcloned to yield plasmid p20J2 and sequenced in both directions (see *Materials and Methods*).

The human fibrillar cDNA insert in p20J2 is 1104 bp in length and contains a single long open reading frame that begins with an ATG near one end (Fig. 1). The nucleotide sequence TCGCC upstream of this ATG agrees with the consensus sequence (CCRCC) found at most translation initiation sites (22). The amino-terminal sequence beginning with this ATG is identical to 30 of the 31 chemically determined amino-terminal residues from rat fibrillar (ref. 8 and see below). Thus, we infer that the ATG starting at +1 in Fig. 1 is the initiation codon. The open reading frame extends for 963 bp and encodes a 321-amino acid protein of 33,823 Da with a predicted isoelectric point of  $\approx 10$  (Fig. 1). The 5' end of the open reading frame is rich in purine nucleosides, with the 210 bp from codon 6 (Ser) to codon 75 (Gly) having a G+C content of 70.5%. The entire open reading frame contains 57.6% G+C. The open reading frame ends at a TGA stop codon, followed by an 80-bp 3' untranslated region and a poly(A) stretch. A polyadenylation signal ATTAAA in the 3' flanking region is located 13 bp away from the poly(A) tract.

**RNA Blot Analysis.** A Northern (RNA) blot of human and yeast RNAs was analyzed to identify the human fibrillar mRNA and reveal cross-hybridization with other mRNAs (Fig. 2). When analyzed with the human fibrillar probe, an RNA  $\approx 1300$  nt is detected in the human poly(A)<sup>+</sup> enriched sample but not in total human RNA (Fig. 2, lanes 2 and 3). The 1300-nt size of this mRNA is consistent with the size of the human fibrillar cDNA, considering the length of the poly(A) tail. The human fibrillar probe reveals cross-hybridizing mRNAs under the conditions of stringency used to analyze the blot (see *Materials and Methods*). In yeast RNA samples, the human probe detects a 1300-nt band found in the poly(A)<sup>+</sup>-selected RNA but not in total RNA (Fig. 2, lanes 4 and 5). Hybridization of the same blot under high-stringency conditions with a yeast fibrillar (*NOP1*) probe demonstrates that the 1300-nt band is the yeast fibrillar mRNA (Fig. 2, lanes 6 and 7). This result agrees with our previous estimate of the size of the yeast fibrillar mRNA (14). In the human RNA samples, the human probe weakly detects a human RNA of  $\approx 2700$  nt in the poly(A)<sup>+</sup> lane (Fig. 2, lane 3). This mRNA species may correspond to the human

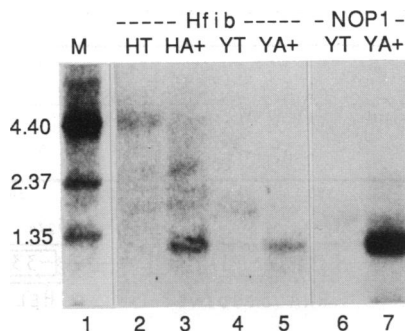


FIG. 2. Northern blot analysis. Total (T, 5  $\mu$ g) or poly(A)-selected (A<sup>+</sup>, 2  $\mu$ g) RNAs from human (H) or yeast (Y) were analyzed according to standard procedures. Hybridization with human fibrillar (Hfib) or yeast fibrillar probe (NOP1) detects the  $\approx 1300$ -nt fibrillar mRNAs. RNA standards are shown with sizes in kb (M).

nucleolin mRNA. Nucleolin is a nucleolar protein of 100 kDa that contains a repeated amino acid sequence rich in glycine and arginine near the carboxyl terminus (23). A similar cross-hybridization pattern was reported in a Northern blot analysis of *Xenopus* RNAs (15).

**Immunoprecipitation of Fibrillar with Autoimmune Antisera.** Human fibrillar was expressed *in vitro* to test for the capacity to be recognized by autoimmune antisera directed against fibrillar (see *Materials and Methods*). Synthesis of human fibrillar in a cell-free system results in the production of a protein that typically migrates as a doublet of 37 kDa and 36 kDa on SDS/PAGE gels. Treatment of electrophoresis samples with dithiothreitol followed by iodoacetamide results in the appearance of a major band at 36 kDa and a minor band at 37 kDa (Fig. 3, lane 2). Without reduction and alkylation treatment, the 37-kDa and 36-kDa bands are present in apparently equal amounts (data not shown); this suggests that an intramolecular disulfide bond causes fibrillar synthesized *in vitro* to migrate in SDS gels as a 37-kDa band. The observed molecular mass of fibrillar synthesized *in vitro* agrees well with the calculated value of 33,823 Da.

Two sera from patients with scleroderma were incubated with fibrillar in the presence of the nonionic detergent Triton X-100. We tested a recent lot of ANA because fibrillar was originally identified by Ochs *et al.* by using an ANA antiserum from Sigma (12). The antiserum 538 is virtually monospecific for yeast fibrillar as assayed by immunoblotting and indirect immunofluorescence (14, 24). Both ANA and 538 antisera recognize the *in vitro*-synthesized human fibrillar by immunoprecipitation (Fig. 3, lanes 3 and 4). The differential immunoprecipitation of the  $\approx 36$ -kDa band may reflect a preference of the antisera for a conformation present in this cell-free synthesis product. In

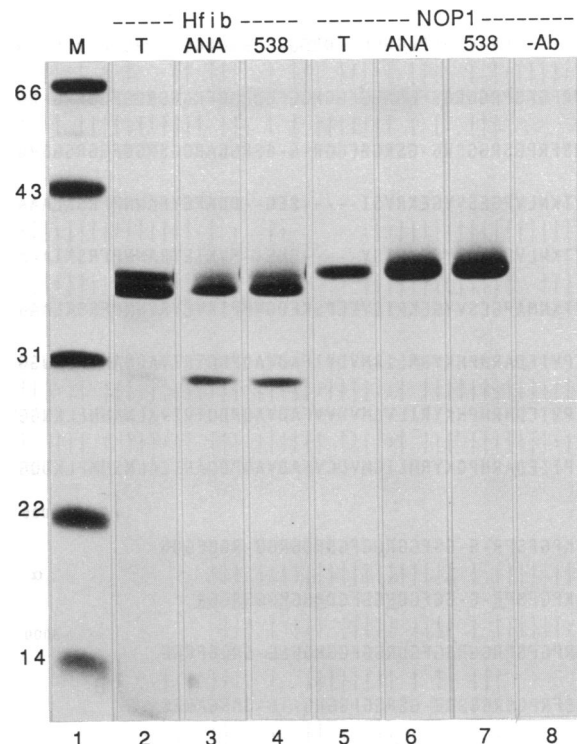


FIG. 3. Immunoprecipitation of human and yeast fibrillar with autoimmune antisera. Human fibrillar (Hfib) or yeast fibrillar (NOP1) was translated in a rabbit reticulocyte lysate system, and 1  $\mu$ l of the total translation (T) was subjected to SDS/PAGE on a 12% gel. Immunoprecipitates obtained with antiserum ANA or 538 or no antiserum (-Ab) from aliquots (10  $\mu$ l) of the translation were analyzed by SDS/PAGE. Molecular mass markers (<sup>14</sup>C-labeled, Bio-Rad) are shown in kDa (M).

addition, a protein of  $\approx 30$  kDa is immunoprecipitated from the translation mixture, in which it is synthesized in low amount. This protein may be a translation product for which initiation occurs at Met-87.

Yeast fibrillarlin synthesized *in vitro* migrates as a single band of  $\approx 38$  kDa (Fig. 3, lane 5). Both the ANA and 538 antisera immunoprecipitate yeast fibrillarlin (Fig. 3, lanes 6 and 7). In control experiments in which no antiserum was incubated with either human or yeast fibrillarlin synthesized *in vitro*, nonspecific precipitation of fibrillarlin was not seen (Fig. 3, lane 8). Immunoprecipitation of the human or yeast fibrillarlin with the two autoimmune antisera was abolished by boiling the translation mixture in SDS before dilution in Triton X-100-containing buffer, presumably due to irreversible denaturation of fibrillarlin (data not shown).

**Fibrillarlin Is a Highly Conserved Protein.** Fibrillarins from human, *X. laevis*, and *S. cerevisiae* show a striking degree of evolutionary conservation of primary structure (Fig. 4A). The human and *Xenopus* fibrillarlin sequences are 81% identical and 90% homologous, when conservative amino acid replacements are taken into account. The human and yeast sequences are 67% identical and 82% conserved.

Human fibrillarlin begins with an amino-terminal domain rich in glycine and arginine residues, which accounts for the abundance of glycine (20.2 mol%) and arginine (4.2 mol%) in the protein. The sequence of this amino-terminal domain is repetitive in nature and similar to that of *Xenopus* and to that of yeast (Fig. 4A). However, the human sequence contains a stretch of 17 glycines interrupted by one arginine. The amino-terminal domain ends near Lys-77 or Lys-84, whereas in both *Xenopus* and yeast it ends near Lys-85 (Fig. 4A). The amino-terminal domain in human fibrillarlin contains one acidic residue, Asp-20, which occurs in the sequence

GDRGG. Three aspartate residues are present in the *Xenopus* fibrillarlin amino-terminal domain, and each occurs in the sequence GDRGG (Fig. 4A). The arginine residues in this amino-terminal domain are in similar relative positions as arginines known to be modified by methylation in fibrillarins from rat and *Physarum* (8, 26). This suggests that these arginines in human fibrillarlin may occur as  $N^G, N^G$ -dimethylarginine residues (Fig. 4B).

Centrally located in human fibrillarlin is an  $\approx 90$ -amino acid stretch extending from Glu-133 to Lys-222, which bears significant resemblance to domains that participate in RNA binding (30, 31). This putative RNA-binding domain may be aligned with domains from various snRNP proteins (data not shown, see ref. 15). A sequence of eight amino acids beginning with Gly-185, GLVYAVEF, resembles the octameric RNP consensus sequence (Fig. 4A). With one exception this sequence is conserved between human, *Xenopus*, and yeast. However, the fibrillarlin RNA-binding domain does not conform to the consensus sequence as well as do other RNA-binding proteins (see refs. 30 and 31 for compilations).

At the carboxyl-terminal end of human fibrillarlin is a stretch extending between Ser-274 and His-306 that may adopt an  $\alpha$ -helical conformation (Fig. 4C). Predictions of secondary structure using the methods of Robson and co-workers and Chou and Fasman (28, 29) indicate that this 33-amino acid span displays the highest potential to form an  $\alpha$ -helix and the lowest potential to assume an extended  $\beta$  conformation. Helical wheel projections suggest that an  $\alpha$ -helix located from Ser-274 to His-306 would not be amphipathic in nature (data not shown). The locations of the amino-terminal domain, the potential RNA-binding domain, and the putative  $\alpha$ -helical region are shown diagrammatically in Fig. 4D. Although the diagram represents a map of possible

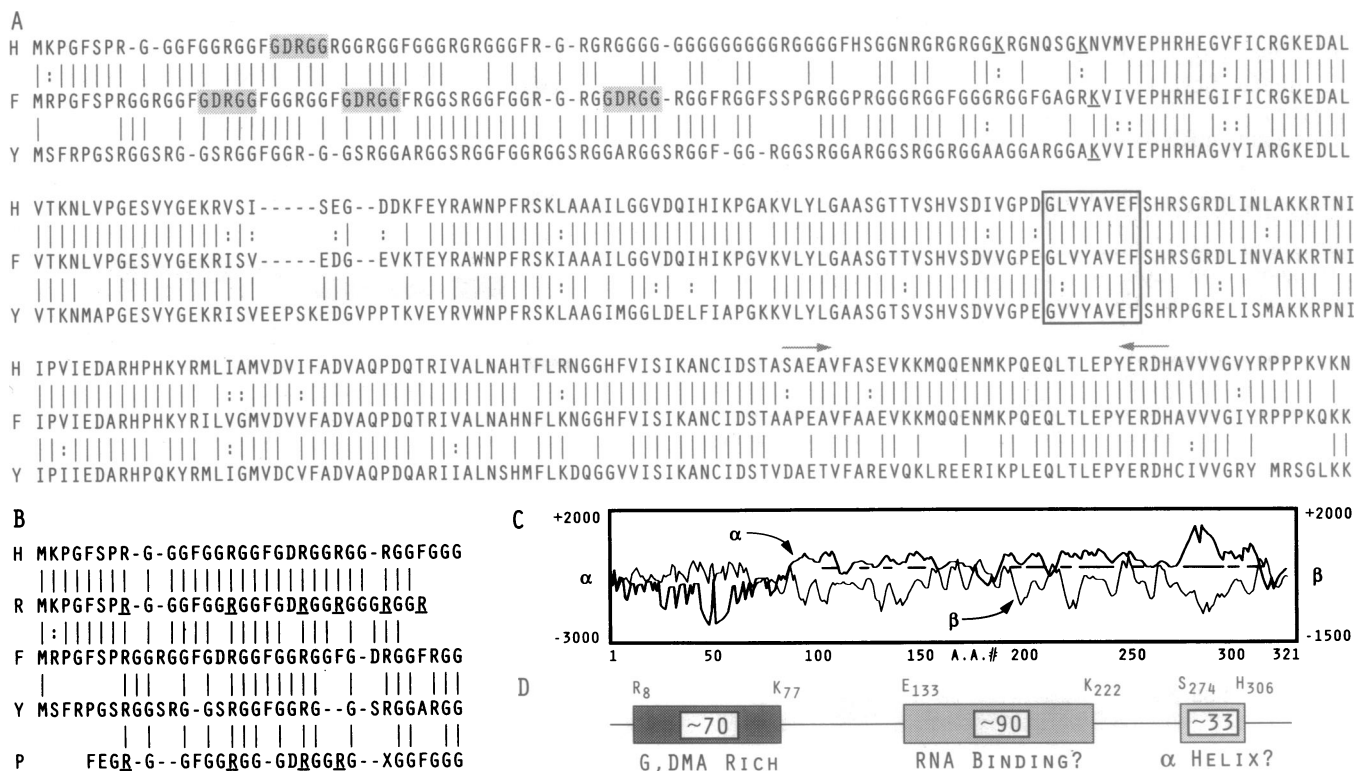


FIG. 4. (A) Three-way alignment (25) between the sequences of fibrillarins from human (H), *X. laevis* (F, from ref. 15), and *S. cerevisiae* (Y, from ref. 14). Identities (!) and conservative replacements (:) are shown. See Fig. 1 for additional explanations. (B) Alignments between the amino-terminal residues of fibrillarins from human, rat (R, from ref. 8), frog, yeast, and the slime mold *P. polycephalum* (P, from ref. 26) generated as described (27). Underlined arginines are known to be  $N^G, N^G$ -dimethylated. (C) Robson structural prediction values (28) for helix ( $\alpha$ ) and extended conformations ( $\beta$ ) plotted as a function of amino acid number (A.A.#). Stretches of amino acids predicted to form an  $\alpha$ -helix by two prediction methods (28, 29) are designated by the horizontal bars. (D) Putative domain structure of human fibrillarlin, showing length and boundaries of domains, in register with C. DMA,  $N^G, N^G$ -dimethylarginine.

domain structure in human fibrillarin, the occurrence of similar domains in *Xenopus* and yeast suggests that they may be general features of fibrillarins from other species as well.

### DISCUSSION

With the molecular cloning of fibrillarin genes from *S. cerevisiae* and *X. laevis* it has become clear that fibrillarins from different species comprise a highly conserved family of nucleolar proteins (14–16). We isolated a human fibrillarin cDNA clone by virtue of its hybridization to sequences from the yeast and frog clones to determine the primary structure of a mammalian fibrillarin and to open the door to studies on the autoimmune response to this protein in humans with scleroderma. The two autoimmune antisera that we have tested immunoprecipitate both human and yeast fibrillarins, suggesting the presence of a common epitope in these proteins. The striking conservation between human and yeast fibrillarins, 67% identity of amino acids in optimized alignments, is similar to the degree of homology in proteins, such as actin and tubulin. Recent experiments establish a role for fibrillarin in the initial cleavage of preribosomal RNA transcripts, a processing event likely to occur in all eukaryotes. The evolutionary conservation of fibrillarin structure is consistent with a role in such an evolutionarily conserved function.

The portion of fibrillarin that accounts for most variation in primary structure between human, amphibian, and yeast is the 70–80 residues at the amino-terminal end, in which glycines and arginines are prevalent. The sequence divergence at the amino terminus may be explained if this amino-terminal domain functions to nonspecifically enhance binding of fibrillarin to RNA(s). A similar glycine and arginine-rich domain at the carboxyl-terminal end of Chinese hamster nucleolin increases the affinity of nucleolin for RNA by 10-fold but without any sequence specificity (32). The center portion of human fibrillarin contains a sequence of  $\approx 90$  amino acids that bears significant resemblance to the RNA-binding domain involved in, or essential for, interactions with RNA and/or single-stranded DNA in many, but not all, RNA-binding proteins (30, 31). At the predicted location within this 90-residue stretch are eight amino acids that may be the equivalent of the octameric motif known as the RNP consensus sequence. Interaction(s) between human fibrillarin and small nuclear RNAs, or pre-rRNA, may be mediated by these sequences.

The nucleolar snRNP particle containing human fibrillarin is thought to carry out a processing step that is the first to shorten the preribosomal RNA transcript at the 5' end. The elucidation of the primary structure of human fibrillarin offers the potential to address this processing step and RNA-binding functions, among others, at the molecular level. The dissection of these functions may also shed some light on the molecular basis for the autoimmune response raised against fibrillarin in humans with scleroderma.

The human antiserum 538 was a gift of Drs. Michael Snyder and Charles Yang. We thank Dr. Eckart Bartnik for providing HeLa cell total and poly(A)<sup>+</sup> selected RNAs. Dr. Eduard Hurt kindly exchanged with us DNA sequence data before publication. Drs. Nilabh Chaudhary, Jean-Claude Courvalin, Gary Greenburg, Ruben Hen-

riquez, Giovanni Migliaccio, Christopher Nicchitta, Westley Reeves, Danny Schnell, Gregory Shelness, Norbert Schuelke, Jun Sukegawa, and Richard Wozniak gave us invaluable advice and criticism at many stages of this project.

1. Tyc, K. & Steitz, J. A. (1989) *EMBO J.* **8**, 3113–3119.
2. Kass, S., Craig, N. & Sollner-Webb, B. (1987) *Mol. Cell. Biol.* **7**, 2891–2898.
3. Kass, S., Tyc, K., Steitz, J. A. & Sollner-Webb, B. (1990) *Cell* **60**, 897–908.
4. Maser, R. L. & Calvet, J. P. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6523–6527.
5. Stroke, I. L. & Weiner, A. M. (1989) *J. Mol. Biol.* **210**, 497–512.
6. Scheer, U. & Benavente, R. (1990) *BioEssays* **12**, 14–21.
7. Aris, J. P. & Blobel, G. (1988) *J. Cell Biol.* **107**, 17–31.
8. Lischwe, M. A., Ochs, R. L., Reddy, R., Cook, R. G., Yeoman, L. C., Tan, E. M., Reichlin, M. & Busch, H. (1985) *J. Biol. Chem.* **260**, 14304–14310.
9. Guiltinan, M. J., Schelling, M. E., Ehtesham, N. Z., Thomas, J. C. & Christensen, M. E. (1988) *Eur. J. Cell Biol.* **46**, 547–553.
10. Tollervey, D. & Hurt, E. C. (1990) *Mol. Biol. Rep.* **14**, 103–106.
11. Christensen, M. E., Beyer, A. L., Walker, B. & LeStourgeon, W. M. (1977) *Biochem. Biophys. Res. Commun.* **74**, 621–629.
12. Ochs, R. L., Lischwe, M. A., Spohn, W. H. & Busch, H. (1985) *Biol. Cell* **54**, 123–134.
13. Tan, E. M. (1989) *J. Clin. Invest.* **84**, 1–6.
14. Henriquez, R., Blobel, G. & Aris, J. P. (1990) *J. Biol. Chem.* **265**, 2209–2215.
15. Lapeyre, B., Mariottini, P., Mathieu, C., Ferrer, P., Amadi, F., Amalric, F. & Caizergues-Ferrer, M. (1990) *Mol. Cell. Biol.* **10**, 430–434.
16. Schimmang, T., Tollervey, D., Kern, H., Frank, R. & Hurt, E. C. (1989) *EMBO J.* **8**, 4015–4024.
17. Ausubel, F. A., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1990) *Current Protocols in Molecular Biology* (Greene Publishing and Wiley-Interscience, New York).
18. Beltz, G. A., Jacobs, K. A., Eickbush, T. H., Cherbas, P. T. & Kafatos, F. C. (1983) *Methods Enzymol.* **100**, 266–285.
19. Wahl, G. M., Berger, S. L. & Kimmel, A. R. (1987) *Methods Enzymol.* **152**, 399–407.
20. Wozniak, R. W., Bartnik, E. & Blobel, G. (1989) *J. Cell Biol.* **108**, 2083–2092.
21. Nicchitta, C. V. & Blobel, G. (1989) *J. Cell Biol.* **108**, 789–795.
22. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
23. Lapeyre, B., Amalric, F., Ghaffari, S. H., Venkatarama Rao, S. V., Dumber, T. S. & Olson, M. O. J. (1986) *J. Biol. Chem.* **261**, 9167–9173.
24. Yang, C. H., Lambie, E. J., Hardin, J., Craft, J. & Snyder, M. (1989) *Chromosoma* **98**, 123–128.
25. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
26. Christensen, M. E. & Fuxa, K. P. (1988) *Biochem. Biophys. Res. Commun.* **155**, 1278–1283.
27. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
28. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
29. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45–148.
30. Query, C. C., Bentley, R. C. & Keene, J. D. (1989) *Cell* **57**, 89–101.
31. Bandziulis, R. J., Swanson, M. S. & Dreyfuss, G. (1989) *Genes Dev.* **3**, 431–437.
32. Ghisolfi, L., Joseph, G., Erard, M., Escoubas, J.-M., Mathieu, C. & Amalric, F. (1990) *Mol. Biol. Rep.* **14**, 113–114.