Primary Structure of a Human Small Nuclear Ribonucleoprotein Polypeptide as Deduced by cDNA Analysis*

Luis A. Rokeach, Mehrdad Jannatipour, Jeanne A. Haselby, and Sallie O. Hoch‡
From The Agouron Institute, La Jolla, California 92037

Anti-Sm is an antibody specificity present in patients affected with the autoimmune disease systemic lupus erythematosus. The polypeptides Sm-B'/B (estimated molecular mass 27 and 26 kDa, respectively) are primary targets of Sm antibodies. Sm-B'/B are part of the core polypeptides of small ribonucleoprotein particles (snRNPs) involved in pre-mRNA splicing. Sm-B'/B share the same amino-terminal sequence as we determined by microsequence analyses of the purified polypeptides. Oligonucleotide probes based on that sequence were used to isolate seven clones from a human lymphoblastoid cDNA library in λgt10. The clones contained a single coding region for a protein of ~25 kDa. The predicted amino-terminal sequence was identical to that of the isolated Sm-B'/B polypeptides. In vitro translation experiments produced a protein immunoreactive with human polyclonal anti-Sm antibodies. The isolation of only one unique cDNA sequence suggests that Sm-B'/B may be post-translational variants encoded by a single message. The specific structural features which distinguish Sm-B' from Sm-B have yet to be determined. Northern blot analysis confirmed the diverse tissue and species distribution expected for these immunologically conserved polypeptides. The Sm-B'/B primary sequence is rich in proline (20%) and glycine (15%) residues. The polypeptides are concentrated in the carboxyl-terminal half of the protein and display a repetitive unit that is shared with other snRNPs and nucleic acid binding proteins. Analysis of these arrays suggests an eight residue proline-rich consensus sequence with potential as either an RNA binding domain, or as a site of protein/protein interaction.

Anti-Sm is an antibody specificity present in patients affected with the autoimmune disease systemic lupus erythematosus (reviewed in Sharp and Alspaugh (1)). Sm antibodies precipitate the small ribonucleoprotein complexes (snRNPs)\(^1\) U1, U2, U4, U5, and U6, which are involved in pre-mRNA splicing. Determination of the primary sequence obtained from these snRNPs has hence a 2-fold relevance, as antigens in autoimmune disease and as functional components of the intricate splicing machinery. At least six polypeptides, designated as the core proteins, are shared by all Sm snRNPs: B' (27 kDa), B (26 kDa), D (13 kDa), E/F (11-kDa doublet), and G (<10 kDa). Moreover, there are additional polypeptides unique to each snRNP. For instance, 70 kDa, A (32 kDa), and C (18.5 kDa) are associated with the U1 snRNP; A' (31 kDa) and B'' (26.5 kDa) belong to the U2 snRNP (4, 5). Antibodies of Sm specificity have been found to react principally with the B', B, and D proteins. Sm-B'/B and Sm-D share at least one common epitope as the Y12 anti-Sm monoclonal antibody reacts with all three proteins (6). On the other hand, Sm-B' and Sm-B are closely related because chymotrypsin mapping of these polypeptides produced very similar peptide patterns (7).

Thus far, no enzymatic activity has been assigned to the snRNP proteins. Some models (2, 3, 8) favor the RNA component as the catalytical element of snRNPs and ascribe to the protein moieties the role of scaffolding units within each snRNP and in the splicing complex, mediating interactions between snRNP/snRNP and snRNP/pre-mRNA. Either enzymatic or scaffolding, the function of the snRNP proteins is nonetheless essential since anti-Sm and anti-RNP antibodies (which recognize the U1 snRNP) inhibit the splicing reaction (9, 10).

In the past 3 years a considerable number of cDNAs encoding snRNP proteins have been isolated and sequenced: 70K (11, 12), A (13), A', B' (14), C (15, 53), D (16), and E (17). Determination of the primary sequence obtained from these cDNAs is an important step in the elucidation of how these proteins are involved in splicing and in autoimmune disease. Furthermore, these cDNAs will allow for the production of sufficient amounts of wild type and mutant proteins for epitope mapping, studies of the three-dimensional structures, and the reconstitution of snRNPs, which studies ultimately are necessary for a detailed understanding of pre-mRNA splicing.

In this paper we present the amino acid sequence of at least one of the Sm-B'/B polypeptides as deduced from seven independent cDNA isolates. The predicted primary structure was analyzed and compared to other snRNP and RNA binding proteins.

MATERIALS AND METHODS

Isolation of Sm-B' and Sm-B Polypeptides—Sm snRNP were isolated from HeLa extracts by immunofinity chromatography as described (18, 19). The individual snRNP polypeptides were fractionated by SDS-polyacrylamide preparative gel electrophoresis. The polypeptides were isolated in microgram quantities by electroelution according to the protocol of Hunkapiller et al. (20), with modification of the staining procedure. Since a number of polypeptides were closely spaced, the Coomassie Blue concentration was decreased to 0.05% solution.

\(\text{REFERENCES}

1. This work was supported by Research Grants AI 21083 and AR 36387 from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

2. The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J00169.

3. To whom correspondence should be addressed: The Agouron Institute, 505 Coast Blvd. S., La Jolla, CA 92037.

* The abbreviations used are: snRNP, small nuclear ribonucleoprotein; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; PIPES, 1,4-piperazinediethanesulfonic acid.

\(\text{*}

‡ L. Fresco and J. Keene, personal communication.
from the original protocol to allow for clear visualization of bands before excision from the gel. After electroelution of individual bands, each was checked for homogeneity by silver stain and for immunoreactivity by protein blot (21, 22). The amino-terminal sequence analyses of electrophoresed Sm-B' and Sm-B were carried out at the Protein Chemistry Laboratory, Washington University, St. Louis, MO, employing a gas-phase peptide microsequencer (Applied Biosystems).

Oligonucleotide Probes—Since the Sm-B' and Sm-B polypeptides share identical amino-terminal sequences (see "Results") we utilized the same probes for the screening of cDNAs encoding Sm-B' and Sm-B. Based on the amino-terminal sequences of Sm-B' and Sm-B we designed three oligonucleotide probes depicted in Fig. 1. Probes B1 and B3 were 23 nucleotides long. The B1 probe was designed using deoxynucleosine at the third position of the most ambiguous codons (23). Accordingly, the B1 probe was synthesized as a mixture of 64 different oligonucleotides. For the B3 probe, all four nucleotides were included at the third position of redundant codons. Based upon human codon usage (24), we considered only cytosine at the first position and deoxynosine at the third position of the most ambiguous codons (23). As a result, the total number of molecular species for the B3 probe was reduced to 192. In the case of the B4 probe, a single 54-mer species was designed based upon human codon usage (24). The probes were labeled to a specific activity of \(1.2 \times 10^8\) cpdm/\(\mu g\), employing \([\gamma^3P]ATP\) (6000 Ci/mmol; Du Pont-New England Nuclear) and T4 polynucleotide kinase (United States Biochemical Corp.).

Screening—A 7gt10 cDNA library made from human B-lymphocyte poly(A)+ RNA was purchased from Clontech Laboratories (Palo Alto, CA). The 7gt10 library was plated (approximately 5000 plaques/85-mm plate) on Escherichia coli Y1090 (25); plaques were transferred in duplicate to nitrocellulose filters as described (26). Hybridization to the probes was carried out for 16 h in 6 \(\times\) SSC/5 \(\times\) Denhardt's solution (1 \(\times\) Denhardt's = 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin, 0.02% Ficoll; 1 \(\times\) SSC = 150 mM NaCl, 15 mM sodium citrate). probes were added to \(10^6\) cpdm/ml. The filters were washed in 6 \(\times\) SSC for 2 h with three changes of buffer at the hybridization temperature. Hybridizations with probe B1 were carried out at 46 °C, with probes B2, B3, and B4 at 40 °C and with probe B4 at 37 °C, 50, and 65 °C. The filters were exposed to autoradiography. Plaques identified as potential positives were purified by two successive rounds of hybridization.

DNA Sequencing—The EcoRI cDNA inserts from positive \(\lambda\) recombinants were cloned in the M13mp18 or pUC18 vectors (27). The nucleotide sequence was determined by the dideoxy chain termination method (28) employing \([\alpha^32P]dATP\) (Du Pont-New England Nuclear) and Klenow enzyme (Bethesda Research Laboratories) or Sequenase (United States Biochemical Corp.). To determine the complete nucleotide sequence in the coding strand of all seven Sm-B'/B cDNA clones, the universal primer and three synthetic oligonucleotide primers were used in a "walk along" fashion. This procedure allowed us to perform the parallel analysis of all seven cDNAs in short time and determine that the seven clones shared identical coding sequences and differed only in the length of their non-coding sequences. Four of the seven EcoRI clones were further analyzed as follows. The nucleotide sequence of the non-coding strand of clone B3-1 was determined after a series of overlapping deletions was made by the method of Dale et al. (29), using the Cyclone System (IBI, New Haven, CT). Overlapping restriction fragments from clones B409-2, B445-2, and B445-2 were inserted into the polynucleotide of pUC18 and/or M13mp18 and M13mp19. Coding and non-coding strands of these subclones were subsequently sequenced.

Expression of the Cloned B443-2 cDNA in Vitro—The 988-base pair HindIII-EcoRI B443-2 cDNA fragment spanning the Sm-B'/B coding sequence was inserted into pTZ19R (30). The resulting plasmid was linearized with EcoRI and transcribed using \(E. coli\) RNA polymerase. Four \(\mu g\) of mRNA synthesized in vitro was translated using a rabbit reticulocyte system (Promega Biotec) and \([35S]\)methionine (Amerham Corp.). Immunoprecipitation was carried out essentially as described (26) with modifications in the buffers. The immunoprecipitated proteins were analyzed by electrophoresis on 10% SDS-PAGE gels, in 6 \(\times\) SSC, pH 6.8, 50 mM sodium phosphate, pH 7.5, 100 mM NaCl, 1 mM EDTA, at 65 °C. The filters were washed twice in 3\% SDS, 1 \(\times\) SSC, and twice in 0.1\% SDS, 0.2 \(\times\) SSC at 65 °C. The probe spanned the whole EcoRI fragment of the B443-2 cDNA.

RESULTS

Amino-terminal Sequence of the Sm-B'/B Polypeptides—The individual polypeptides of the Sm snRNP were successfully fractionated by SDS-PAGE and electroeluted following the protocol (and apparatus) of Hunkapiller et al. (20). After elution each polypeptide was verified for homogeneity by silver stain and for immunoreactivity by protein blot; the latter probe to be the more sensitive detecting see bands not seen by silver stain. Electroeluted Sm-B' and Sm-B were subjected to gas-phase amino acid sequencing, which yielded information on 19 residues for Sm-B and 20 for Sm-B'. Both polypeptides share an identical amino-terminal sequence which is presented in Fig. 1. These results are in agreement with the suggestion that the Sm-B'/B polypeptides are closely related (7).

Isolation of cDNAs Encoding the Sm-B'/B Proteins—In order to isolate cDNAs encoding the Sm-B'/B antigens we screened a human B-lymphocyte (Raji) \(\lambda\)gt10 library. Raji cells contain approximately equivalent amounts of Sm-B' and Sm-B (not shown). Approximately 250,000 independent \(\lambda\)gt10 recombinants were plated and transferred to nitrocellulose filters. The filters were screened by hybridization with the \(32P\)-radiolabeled B1 probe (see Fig. 1). From this first round of screening, 19 potentially positive plaques were identified. An additional screening of another 250,000 independent recombinant plaques was performed with probe B3. In this latter screening, 16 plaques were identified as hybridizing to the \(32P\)-labeled to be probe. The nine strongest positive clones from the screening with the B1 probe and all 16 positive clones from the screening with the B3 probe were subjected to two more cycles of plaque purification, screening each cycle with the B3 probe. After two such additional cycles of hybridization, a total of 24 positive \(\lambda\)clones were obtained. DNA from these clones was prepared and digested with EcoRI. The inserts were classified into six groups which ranged in lengths from 0.80 to 2.0 kilobase pairs. In order to limit the number of clones to be sequenced, the 24 clones were screened further with a longer consensus probe, the B4 probe designed on the basis of most of the known amino acid sequence. Hybridizations and washings were carried out at three different temperatures. From this last screening, seven recombinant clones were identified as hybridizing to the \(32P\)-labeled B4 probe at all three temperatures over a clear background. The seven clones were designated: AB3-1, AB22-2, AB407-1, AB419-2, AB409-2, AB43-2, and AB445-2 (see Fig. 2).

From the seven positive clones identified with the B4 probe, six belonged to the 1.3-1.4 kilobase pairs size group. The positive clone AB419-2 had an insert size of 1.1 kilobase pairs (cf. Fig. 2). The cDNA EcoRI inserts from recombinants AB3-1, AB22-2, AB407-1, AB419-2, and the non-coding strand of \(\lambda\)B443-2 were subcloned into M13mp18. The EcoRI fragments from clones AB409-2, AB445-2, and \(\lambda\)B443-2 could not be cloned into M13 vectors, probably due to their long poly(A)+ tails (cf. Fig. 2) and were therefore inserted into pUC18. The nucleotide sequence of the cDNA inserts was determined by employing the dioxyde method as described under "Materials and Methods." All seven clones were found to contain a DNA
sequence encoding the known sequence of 20 amino acids in the Sm-B'/B polypeptides.

Nucleotide and Predicted Amino Acid Sequences of cDNAs Encoding Sm-B'/B—The nucleotide sequence of all seven clones was revealed to be the same throughout the coding region. In addition, the DNA sequences at the 5' and 3'-non-coding regions were identical except that the length of such regions differed from clone to clone (see Figs. 2 and 3). This indicates that each of the clones arose from an independent cDNA synthesis event. The identity of coding sequences and non-coding sequences in the seven clones isolated from a large cDNA synthesis event. The identity of coding sequences and their respective lengths. Unique restriction sites are shown; each cDNA fragment is bracketed by EcoRI sites (not shown). kb, kilobase pair(s).

The composite DNA sequence of all seven clones as well as the predicted amino acid sequence is presented in Fig. 3. Starting after a 481-nucleotide, 5'-untranslated leader region, the Sm-B'/B open reading frame is capable of encoding a 240-residue protein with a deduced molecular mass of 24,614 daltons. The first methionine codon of this open reading frame (nucleotides +1 to +3), is embedded within the sequence CGCAATCATGA resembling the consensus CGCCACCATGG for methionine initiator codons (32). Therefore, translation is likely to start at this ATG codon, and the methionine encoded by this triplet is subsequently removed since it is absent in the mature protein. Following the TAG termination triplet at positions 721–723, is a 138-nucleotide, 3'-untranslated region spanning five putative polyadenylation signals (33) at nucleotides 818–824 (ATTATA), 821–826 (AATAAT), 831–836 (AATAAT), 834–839 (AA-TAAA), and 850–856 (AATTAAA).

According to the predicted amino acid sequence, the polypeptide encoded by these seven cDNAs is remarkably rich in proline (48 residues), glycine (36 residues), alanine (24 residues), arginine (20 residues), methionine (12 residues), and lysine (9 residues). The majority of proline residues are localized in the carboxyl-terminal half of the protein and are organized in tandemly repeated arrays.

Predicted Secondary Structure of Sm-B'/B—Fig. 4 shows the predicted secondary structure of Sm-B'/B using the Chou and Fasman algorithm (34). The structure of Sm-B'/B would be rather elongated, and there is a potential to form two short helical regions: one around residues 49-65 which are positively charged, and the other situated between residues 150 and 160 which is hydrophobic and especially rich in alamines. In general, hydrophobic and hydrophilic regions alternate throughout the molecule. Sm-B'/B should have a marked propensity to fold in ß-turns, especially in the carboxyl-terminal half of the protein; such turns would be induced by a series of (Pro-Pro-Gly) repeats which are punctuated by Met, Ile, or Arg residues. This positively charged coiled-coil structure consisting of repetitive units is likely to be exposed to the surface of the molecule and be immunogenic.

Immunoreactivity of Protein Synthesized in Vitro—To confirm the identity of the cDNA clones, B443-2 mRNA synthesized in vitro was used in a rabbit reticulocyte in vitro translation system employing [35S]methionine. Direct analysis of the translation products by fluorography after SDS-PAGE (Fig. 5, panel A) showed a major band comigrating with unlabeled Sm-B'/B from HeLa snRNP. This major translation product was immunoprecipitable with different human anti-Sm sera (Fig. 5, panel B). It should be noted that we consistently observed a limited amount of heterogeneity around the primary translation product.

Northern Blot Analysis—In order to ascertain the tissue distribution of Sm-B'/B, the entire B443-2 cDNA was used to hybridize to poly(A)⁺ RNA from different species (Fig. 6). In Raji lymphoblastoid cells, the source of RNA from which the cDNA library was made, there was a major hybridization band of approximately 1300–1400 nucleotides, consistent with the length of our cDNA clones. In HeLa cells (Fig. 6, lane 2), the B'/B mRNA is slightly shorter, 1100–1200 bases (with its abundance varying with the preparation, not shown). The B443-2 probe also reacted with RNA from an African Green monkey kidney cell line, rabbit liver, and mouse pancreas, indicating the strong conservation of Sm-B'/B as is expected for polypeptides involved in a fundamental biological process such as splicing. Cross-species immunoprecipitation of
Primary Structure of Human Sm-B'/B

Fig. 3. Composite nucleotide sequence of human Sm-B'/B cDNA clones, and deduced amino acid sequence. Extraneous sequences added during cDNA synthesis and cloning processes are shown in lower case characters. Clone B3-1 extended from nucleotide -465 to +841 and had 15 C residues at the 5' untranslated end and no poly(A) tail. Clone B407-2 extended from nucleotide -447 to +840 and had 17 C residues at the 5' end and no poly(A) tail. Clone B409-2 extended from nucleotide -464 to +847 and had 17 C residues at the 5' end and no poly(A) tail. Clone B443-2 extended from nucleotide -447 to +861 and had 2 C residues at the 5' end and no poly(A) tail. Clone B445-2 extended from nucleotide -447 to +861 and had 14 C residues and a 32-nucleotide poly(A) tail. Clone B445-2 extended from nucleotide -348 to +861 and had 17 C residues at the 5' end and no poly(A) tail. Clone B445-2 extended from nucleotide -464 to +847 and had 17 C residues at the 5' end and no poly(A) tail. The length of C stretches at the 5' end included the C of the EcoRI linker. The termination codon TAG is indicated by an asterisk. Underlined nucleotides represent potential poly(A) signals (33).

snRNP particles with anti-Sm antibodies have been observed previously (35, 36).

**DISCUSSION**

In this report we have described the isolation of seven cDNA clones capable of encoding at least one of the human Sm-B'/B snRNP core proteins. The identity of the cDNAs was established on the basis of the following criteria: 1) the deduced amino-terminal amino acid sequences are identical to those obtained by protein sequencing of Sm-B' and Sm-B; 2) mRNA synthesized in vitro used in a rabbit reticulocyte system directed the synthesis of a major 32S-labeled polypeptide band that comigrated with Sm-B' isolated from human snRNPs; 3) the polypeptide synthesized in vitro was immunoprecipitable with various Sm antisera.

We favor the idea that Sm-B' and Sm-B are the translational products of the same mRNA and a post-translational event(s), such as a covalent modification or proteolytic cleavage, produces one species from the other. Several lines of

Fig. 4. Chou-Fasman plot (34) of the predicted secondary structure of Sm-B'/B. NH2 symbolizes the amino terminus and COOH the carboxyl terminus of Sm-B'/B. Amino acid positions are numbered every 50 residues. The hydrophobic (diamonds) and charged residues (octagons) are indicated.

Fig. 5. In vitro translation and immunoprecipitation of Sm-B'/B encoded by B443-2 cDNA. The B443-2 cDNA was subcloned into an expression vector and transcribed in vitro with T7 RNA polymerase (see "Materials and Methods"). A, lane 1, translation with no exogenous RNA. Lane 2, the synthetic B443-2 mRNA was translated in vitro employing a rabbit reticulocyte lysate system and [35S]methionine. B, 32P-Radiolabeled proteins in the in vitro translation reaction were immunoprecipitated with antibodies from different sources. Lanes 1, 3, and 4, anti-Sm human sera; lane 2, normal human serum. Radiolabeled polypeptides were fractionated by SDS-PAGE and visualized by fluorography. Positions of Coomassie Blue-stained size markers (in kilodaltons) are shown on the left of each panel, and positions of stained HeLa Sm snRNP proteins are given on the right of each panel.
were exposed for different periods of time: 

by hybridization with 32P-labeled B443-2 cDNA. Hybridized filters very similar (7); 3) B' and B share the same amino acid evidence support our proposal: 1) B and B' are immunologically closely related (6); 2) the chymotrypsin maps of both are 

cation on an agarose/formaldehyde gel and transferred onto a 

markers in kilobases are indicated on the 

pendent clones we have isolated; 

the cDNA library to obtain the single cDNA sequence de-

sequence over a stretch of 19 residues at the amino terminus; 

determine, at the molecular level, what the differences 

differences may have in their function. 

evidence at this time would indicate whether Sm-B' is modi-

Sm snRNP from HeLa or Raji, both of which contain Sm-

lated peptide/calcitonin alternative splicing pathways. We 

smaller than the apparent molecular weight of Sm-B' and 

the Sm-B'/B cDNAs is 24,614 daltons, which is slightly 

Human Sm-B'/B protein (42). Identities are indicated 

that at least two alignments; 

lower case letters 

structurally significant number of plaques (500,000) were screened from 

primary RNA transcript. 

consisting of @-strands and @turns. High frequency of pro-

residues conserved in at least 5 out of 7 alignments; 

represent residues present in at least two alignments. 

McAllister et al. (38) have reported an enriched polypeptide in rat brain which is recognized by anti-Sm sera. This polypeptide, designated N, has an apparent molecular mass slightly higher than Sm-B'. The expression of N appears to be tissue-restricted. N was proposed as a splicing factor since its abundance correlated with a cell line's ability to express 

as a probe, demonstrated that the Sm-B'/B mRNA was found in every tissue and cell line examined (see Fig. 6). Furthermore, Southern hybridization (not shown) of human genomic DNA, using the same probe, suggested that the Sm-B'/B coding gene is present at a limited number of copies (probably one) per genome. Therefore it seems likely that all the reported variants of Sm-B'/B are post-translational modifications or possibly alternative splicing products of the same primary RNA transcript. 

The predicted molecular mass of the protein encoded by the Sm-B'/B cDNAs is 24,614 daltons, which is slightly smaller than the apparent molecular weight of Sm-B' and Sm-B on SDS-PAGE. The deduced amino acid sequence showed a bias for Pro, Gly, Arg, Met, and Lys residues. The carboxyl-terminal half of the molecule is particularly rich in prolines, arranged in repeated units (e.g. Pro-Pro-Pro-Gly) interspersed with hydrophobic and basic residues. The abundance of Pro and Gly residues is reminiscent of the primary structure of collagen- and proline-rich proteins (39, 40). Perhaps of more significance are the similarities we found between Sm-B'/B and the (U1)snRNP-specific A and C proteins (13, 15, 53) and the hnRNP C protein (41). A common motif, PP/aPGMR/IP, as seen in Fig. 7A, is present in these four proteins. This conserved sequence may have a role in protein-protein interactions in RNP complexes. Alternatively, it may be suggestive of a possible RNA binding domain distinct from the so-called RNP consensus (41, 42) which has been demonstrated in a number of hnRNP and poly(A)+ binding proteins, as well as the snRNP proteins, A (13), B' (14), and 70K (11, 12). The RNP consensus is however absent in the core proteins Sm-D (16), Sm-E (17), and Sm-B'/B, as well as the unique U1-associated protein C (15, 53). Thus, for the prolinc consensus, the series of prolines interspersed with hydrophobic and basic residues could dictate the folding of the carboxyl-terminal part of Sm-B'/B into a &-type structure consisting of &-strands and &-turns. High frequency of prolines occupying second position on type I and II &-turns was discussed by Richardson et al. (43). Such &-turns can stack to form a pseudo-30-helix, which could fit into and interact with the deep groove of double stranded RNA. Conversely, Sm-B'/B may form, in solution, an oligomeric protein which can fold into a collagen-like helix, around which RNA could wrap. 

A) 

mRNA; in parentheses is the range 

poly(A) BP (422-439) ATAAAMAA--AGMPGQFMP 

Comparison of the predicted amino acid sequence 

of Sm-B'/B with other nucleic acid binding proteins; in parentheses is the range of nucleotides compared. A, comparison with the A and C human snRNP proteins (13, 15); and human hnRNP C protein (41). Capital letters represent residues conserved in at least 5 out of 7 alignments; lower case letters represent residues present in at least two alignments. B, comparison with a yeast poly(A)+ binding protein (42). Identities are indicated by bars; conservative changes are denoted by dots. Comparisons were carried out using programs in the UWGGC (University of Wisconsin Genetics Computer Group) and NBRF packages (51, 52). The NBRF database release 15.0 was used in these similarity searches. 

Fig. 7. Comparison of the predicted amino acid sequence (in the single-letter code) of Sm-B'/B with other nucleic acid binding proteins; in parentheses is the range of nucleotides compared. A, comparison with the A and C human snRNP proteins (13, 15); and human hnRNP C protein (41). Capital letters represent residues conserved in at least 5 out of 7 alignments; lower case letters represent residues present in at least two alignments. B, comparison with a yeast poly(A)+ binding protein (42). Identities are indicated by bars; conservative changes are denoted by dots. Comparisons were

Fig. 6. Interspecies Northern analysis of Sm-B'/B transcripts. Poly(A)+ selected RNAs from different sources were fractionated on an agarose/formaldehyde gel and transferred onto a Nytran membrane by capillarity. Sm-B'/B transcripts were detected by hybridization with 32P-labeled B443-2 cDNA. Hybirdized filters were exposed for different periods of time: lane 1, Raji; lane 2, HeLa; lane 3, African Green monkey cells (COS-7); lane 4, rabbit liver; lane 5, mouse pancreas; (see "Materials and Methods"). Molecular size markers in kilobases are indicated on the left. Lanes 1-5 were exposed for 16 h at room temperature; lanes 4-5, for 70 h at -58 °C. 

Fig. 5.1 

B'/B (189-201) MAPPPGMRPPMG 

B'/B (214-226) GMPPPPGMRPMP 

B'/B (229-240) GPAPPMPGMRPM 

A (163-175) YMPPMPGMRPPPL 

C (46-58) PPPAPAMIPPPPS 

C (91-103) VGAPMPGMRPPG 

hnRNP C (131-143) PPRPPPIARAVPS 

Consensus pPpPpMPMPppq g a i n g 

B'/B (150-169) AAAAVATAAGTAGPQYYPP 

poly(A) BP (422-439) ATAAAAGAA---AGPGQBP
It can be noted that two of the proteins demonstrating this proline-rich sequence, hnRNP C and snRNP C have recently been shown to directly bind RNA (44, 45). Sm-B'/B also shows similarities with a yeast poly(A)" binding protein (42; see Fig. 7B). Similar alanine-rich stretches have been found in other nucleic acid binding proteins; for example, several eukaryotic and prokaryotic "A" ribosomal proteins (46) and tobacco rattle virus coat protein (47). It is premature to assess whether such a poly-alanine sequence has any specific function. Nevertheless, one could speculate that if this common sequence resided on the surface of the molecule, due to its amphiphilic nature, it could mediate protein-protein contacts in a multi-component system.

The common amino acid sequences found in Sm-B'/B and in the U1 snRNP-specific A and C proteins due to the probe content have a high propensity to be on the surface of the molecules. This is of special interest from the immunological point of view in the study of auto-epitopes. A logical prediction from the alignments suggested in Fig. 7A would be that subsets of Sm antibodies should cross-react with the A and C proteins; and, conversely, (U1)RNP antibodies should recognize Sm-B'/B. This has been observed by different investigators; e.g. relatively monospecific anti-(U1)RNP antibodies reacted weakly with B' and B (6), and affinity-purified anti-C antibodies reacted with the B'/B polypeptides (15).

Although Sm-B'/B and Sm-D are recognized by several anti-Sm monoclonal antibodies (6, 48), we did not find any striking similarity between these proteins other than certain short alignments, one of which overlaps with the Sm-D (Gly-Arg) repeated motif (16). In common with Sm-D (but at a different site), Sm-B'/B showed similarities with the Epstein-Barr nuclear antigen, EBNA-1 (49); Sm-B'/B residues 98-119 can be aligned with the A-like repetitive sequence of EBNA-1. This region is one of the major epitopes of the Epstein-Barr nuclear antigen (50), and the similarity may have relevance in terms of the etiology of lupus as we have previously suggested (16). The use of recombinant DNA technology will facilitate the determination of the common epitope(s) between Sm-B'/B and Sm-D. With this purpose, heterologous expression of immunoreactive Sm-B'/B has recently been achieved in our laboratory. The comparative study of snRNP proteins also bears significance in terms of evolution. The similarities between the different snRNP polypeptides and other RNA binding proteins involved in mRNA metabolism point to common ancestors and/or domain shuffling. As these proteins are involved in an indispensable biological process, a cross-species examination of snRNP proteins would be a most useful tool to further our understanding of molecular evolution.

Acknowledgments—We would like to express our gratitude to Drs. Krysztof Appelt, James Hoch, and Richard Ogden for helpful discussions and manuscript. We are grateful to Dr. Richard Spritz for providing HeLa and COS-7 poly(A)" RNA. We also thank Leticia Wolpert for preparation of the manuscript.

REFERENCES
Primary Structure of Human Sm-B'/B