cDNA cloning of a human autoimmune nuclear ribonucleoprotein antigen

(small nuclear RNAs/systemic lupus erythematosus/autoantibodies/Sm antigen/autoimmunity)

ERIC D. WIEBEN*, ANNE M. ROHLDER*, JOAN M. NENNINGER†, AND THORU PEDERSON‡

*Department of Cell Biology, Mayo Foundation, Rochester, MN 55902; and †Cell Biology Group, Worcester Foundation for Experimental Biology, Shrewsbury, MA 01545

Communicated by Paul Zamecnik, August 2, 1985

ABSTRACT  Sera from patients with systemic lupus erythematosus and other autoimmune disorders contain autoantibodies against nuclear proteins. One such autoantibody system, known as Sm, reacts with antigens associated with small nuclear RNA molecules. In this paper we report the use of Sm autoantibodies to isolate a cDNA clone for the mRNA of one of these nuclear antigens. A HeLa cell cDNA library was screened by message selection followed by autoantibody reaction of cell-free translation products. This led to the identification of a cDNA clone, p281, containing sequences complementary to mRNA for an Sm autoantibody-reactive, 11,000 M, protein. This cloned Sm antigen comigrated with the small nuclear RNA-associated protein known as “E” and reacted with four out of four Sm autoantibodies that precipitate E protein from total mRNA translation products. RNA gel blot hybridization with clone p281 DNA revealed a poly(A)+ mRNA of ~600 nucleotides in human and marmoset (New World primate) cells. Southern blot hybridization of HeLa cell and human lymphocyte DNA indicated the presence of ~6–10 copies of p281-homologous sequences. Similar copy numbers were observed with genomic DNA from baboon, cat, and mouse, indicating that the Sm antigen mRNA sequence represented in p281 is conserved across three classes of the Mammalia (primates, carnivores, and rodents). However, no cross-hybridization of p281 was observed with frog or Drosophila DNA. In light of existing evidence that the mammalian Sm antigen E is a weaker autoantigen than other small nuclear RNA-associated proteins, these results suggest a possible correlation between a protein’s capacity to serve as an autoantigen during breakdown of the host’s immunological tolerance and its extent of evolutionary conservation, whereas the inverse relationship applies to conventional immunity. We suspect, as have others, that this is a clue to the mechanism of autoimmunity.

Patients with systemic lupus erythematosus and a number of other autoimmune disorders have circulating antibodies against nuclear antigens (1). Among these are proteins complexed with the small nuclear RNAs U1, U2, U4, U5, and U6 (2). There is mounting evidence that these small nuclear ribonucleoprotein (snRNP) complexes are cofactors for mRNA processing reactions (2). We and others have studied the extent to which the RNA and protein components of the snRNPs have been conserved in evolution (2–9). It emerges that each of the small nuclear RNAs contains remarkably conserved elements, embedded in more highly divergent sequences (e.g., refs. 4, 5, 9, 10). These conserved RNA domains include sequences implicated in small nuclear RNA functions (5, 9, 11) as well as snRNP protein binding sites (8, 10, 12, 13). As expected from this, the snRNP proteins themselves also show strong evolutionary conservation. In contrast to the RNAs, however, for which direct sequence information is available, the evidence for evolutionary conservation of snRNP proteins is based entirely on immunochromatrical criteria. The human Sm autoantibody system (1) reacts with snRNPs in diverse organisms, such as dinoflagellate (14), insect (3, 5, 7, 15), sea urchin (16), and frog (6). In immunoblot analysis of human snRNP proteins, Sm autoantibodies and a monoclonal Sm antibody react with three and sometimes four proteins, termed B, B', D, and E (17), and it is presumably the presence of one or more of these proteins (or antigenic sites) that accounts for the remarkable cross-reaction of Sm antibody with nonmammalian snRNPs. Indeed, proteins with electrophoretic mobilities similar to the B, D, and E proteins have been observed in Drosophila (7) and Xenopus snRNPs (18), and some of these react with Sm antibodies in immunoblots (19, 20).

Resolution of the immunological homology between the human Sm antigens and their counterparts in other species will ultimately require protein sequence information. As a first step, we here describe the molecular cloning of cDNA for one of the human Sm antigens and report data on its mRNA expression and the extent of DNA sequence conservation.

MATERIALS AND METHODS

Poly(A)+ cytoplasmic RNA from HeLa cells was isolated as described (8). A cDNA library to mRNA from HeLa cells was constructed by the method of Wood and Lee (22) as modified by Zain et al. (22). Briefly, mRNA-cDNA hybrids were synthesized under standard first-strand cDNA conditions by using oligo(dT)12-18 primer (23). Following chromatography on Sephadex G-100, the hybrids were A-tailed by using terminal transferase and dATP. They were then hybridized with EcoRI- and BamHI-digested pBR322, which had been tailed with TTP and terminal transferase, and transformed into Escherichia coli RR1 cells.

Clones were screened by hybridization selection of HeLa poly(A)+ cytoplasmic RNA by using DNAs derived from pools of individual clones. Initially, six 500-ml bacterial cultures were inoculated with 50 different recombinant clones each. Fifty micrograms of plasmid DNA prepared from each pool was immobilized on nitrocellulose and then used for mRNA hybridization (13, 24). Selected mRNAs were translated in exogenous mRNA-dependent rabbit reticulocyte lysate (25) in the absence of exogenous small nuclear RNAs, and translation products were analyzed by Sm autoantibody selection and electrophoresis (26). Pools of plasmid DNA that selected mRNAs for snRNP proteins were further subdivided, and the analysis was repeated until an individual clone was identified.

Abbreviations: snRNP, small nuclear ribonucleoprotein; bp, base pair(s); kb, kilobase(s).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
RESULTS

Isolation of a cDNA Clone for a snRNP Protein. Approximately 0.2–0.4% of the [35S]methionine-labeled translation products synthesized in an exogenous message-dependent rabbit reticulocyte lysate programmed with HeLa cytoplasmic poly(A)+ mRNA are selected by Sm autoantibody (8). To further enrich for snRNP protein mRNAs, polyadenylated HeLa cytoplasmic RNA was focused on sucrose gradients (Fig. 1). The positions of the mRNAs for snRNP proteins throughout the gradient were determined by cell-free translation of RNA from selected fractions, followed by reaction with Sm autoantibody. This indicated that 2–4% of the translatable mRNAs in the 9–11S region of the gradient encoded proteins that react with Sm autoantibody. Analysis of the autoantibody-selected translation products revealed that these fractions were enriched in mRNAs coding for the four smallest snRNP proteins, D–G (Fig. 1). As expected, mRNAs for the larger snRNP proteins sedimented faster, near the peak of total mRNA (see also ref. 27).

A cDNA library was constructed in pBR322 from gradient fractions enriched in the mRNAs for the smallest snRNP protein mRNAs (“cDNA,” Fig. 1) by a mRNA–cDNA hybrid cloning method (21, 22). Three hundred clones were selected for the presence of sequences complementary to snRNP protein mRNAs by sequential rounds of hybridization selection, using pools of plasmid DNAs (see Materials and Methods). RNAs selected by the pooled cDNAs were translated in a reticulocyte lysate, and the products were selected with an Sm autoantibody determined previously to react with the snRNP proteins B, B', D, and E in total mRNA translation products. One of the cloned DNAs (p281) contained a 350-base-pair (bp) insert that selected a mRNA coding for a single 11,000–12,000 M₅ protein (see arrow, lane 1, Fig. 2). (The other proteins in Fig. 2, lanes 1–3, are background reticulocyte mRNA products.) The translation product of the p281-selected mRNA reacted with Sm autoantibody but not with nonimmune human IgG control serum (compare lanes 6 and 7 of Fig. 2). No comparable translation product was detected when another cDNA clone (p280) was used for the hybridization-selection (lanes 2, 4, and 5, Fig. 2). These data show that p281 contains sequences that are complementary to the mRNA for one of the Sm antigens.

Electrophoresis of the p281-specific translation product in parallel with authentic HeLa cell snRNP proteins demonstrated that the selected protein comigrated with the 11,000 M₅ E protein (lanes 3 and 4, Fig. 3). We observed this with Sm autoantibodies from four different patients. We wish to emphasize, however, that these results do not unequivocally establish the translation product as the snRNP E protein, as opposed to some other 11,000 M₅ Sm antigen against which there are also antibodies in all four anti-Sm sera (see also Discussion).

Fig. 1. Sucrose gradient fractionation of HeLa snRNP mRNAs. Fifty micrograms of HeLa cell poly(A)+ cytoplasmic RNA was centrifuged on 5–20% sucrose gradients containing 70% (vol/vol) formamide, 3 mM EDTA, and 10 mM Tris-HCl (pH 7.5) (Beckman SW41 rotor, 44 hr, 25°C). The sedimentation values indicated were derived from A₂₆₀ profiles of total cytoplasmic RNA run in parallel. mRNA was recovered from gradient fractions by ethanol precipitation and was translated for 1 hr in an exogenous message-dependent reticulocyte lysate containing [35S]methionine. Translation products were allowed to react with Sm autoantibody, and antigenic material was recovered on protein A-Sepharose and analyzed by electrophoresis and fluorography (7). Shown are the fluorographic results with the gel lanes loaded with translation products corresponding to the mRNA gradient size classes indicated. cDNA indicates the region of the gradient from which mRNA was taken for cDNA synthesis.

Fig. 2. Hybrid-selection of mRNA by p281. Twenty micrograms of HeLa poly(A)+ cytoplasmic RNA was hybridized with 4 µg of Pst I-digested, nitrocellulose-immobilized DNA from clones p280 and p281. The hybrids were washed and eluted as described (24) and the hybridized RNA was recovered by ethanol precipitation with 2 µg of tRNA carrier prior to translation in reticulocyte lysate. Antibody reactions, electrophoresis, and fluorography were as in Fig. 1. Lanes 1–3, total translation products (no antibody selection). Lane 1, p281-selected mRNA; arrow indicates major product of exogenous message. Lane 2, p280-selected mRNA. Lane 3, no added mRNA. Lanes 4–7, antibody (Ab)-selected translation products. Lane 4, p280-selected mRNA, nonimmune human IgG. Lane 5, p280-selected mRNA, Sm antibody. Lane 6, p281-selected mRNA, Sm antibody. Lane 7, p281-selected mRNA, nonimmune human IgG.


Biochemistry: Wieben et al.
Confirmation of the mRNA hybridization specificity of p281 DNA comes from analysis of the Sm autoantibody-reactive proteins synthesized in vitro from the mRNAs not selected by p281. Both of the two other major antigens recognized by Sm autoantibody ("B" and "D"; ref. 17) were present in this fraction (lane 2, Fig. 3). Therefore, these other mRNAs remained intact throughout the hybridization-selection procedure but were not selected by p281 under the conditions used.

**RNA Blot Hybridization of Sm Antigen mRNA.** Plasmid p281 DNA hybridized to a single band of HeLa cell cytoplasmic polyadenylated mRNA (lane 2, Fig. 4 Left, and lane 3, Fig. 4 Right). Relative to 18S and 28S rRNA markers, this mRNA migrated with an apparent Mr of ≈200,000 on 1.5% agarose gels after denaturation with glyoxal. This corresponds to a RNA molecule of ≈600 nucleotides, which is about twice the minimum size required to code for the observed 11,000 Mr protein. This difference presumably reflects the presence of 5' and 3' untranslated regions and the poly(A) tail (see Discussion).

RNA gel blot analysis of RNAs from other organisms revealed that RNAs similar in size to the human mRNA are detectable in a marmoset (a New World primate) cell line (lane 6, Fig. 4 Left) and mouse erythroleukemia cells (not shown). No higher molecular weight hybridizing RNAs were detected reproducibly in the poly(A)+ or total nuclear RNA of any species under the conditions used (for example, see lane 1, Fig. 4 Right).

To our surprise, cross-hybridizing RNAs were not detected in Drosophila Kc cells, macaque (primate) liver, sea urchin eggs (lanes 3–5, Fig. 4 Left), or Xenopus oocytes (lane 4, Fig. 4 Right). These results could reflect either a lack of conservation of the Sm antigen's mRNA sequence in p281 or, alternatively, low expression of this message in these cell types. To obtain further information concerning evolutionary conservation, these sequences were probed in genomic DNA by Southern blot analysis.

**Analysis of Sm Antigen DNA Sequences.** Hybridization of p281 to restriction endonuclease digests of HeLa cell DNA revealed a small set of bands. With EcoRI digestion nine bands were seen (lane 2, Fig. 5 Left), ranging from ≈1.8 to ≈15 kilobases (kb) in size (the uppermost band was resolved as a doublet on longer gels). In longer exposures, two additional bands of <1.5 kb were observed (e.g., lane 1, Fig. 5 Right). BamHI digestion generated four or five bands (lane 3, Fig. 5 Left), with the one at ≈7.5 kb probably including more than one fragment. Each of the nine major EcoRI bands corresponded in intensity to p281 DNA standards that were loaded and run in parallel calibrated to represent one copy per haploid human genome equivalent. Similar analysis of EcoRI-digested human lymphocyte DNA (not shown) revealed six p281-hybridizing bands. Though not precisely defining their multiplicity or arrangement, these data indicate that DNA sequences represented in clone p281 are present at only a relatively small number of copies in the human genome.

Low copy-number DNA sequences hybridizing with p281 were observed in baboon, cat, and mouse (lanes 2–4, Fig. 5 Right) and guinea pig (lane 1, Fig. 5 Left). Thus, nucleotide sequences in the human form of this Sm antigen's mRNA are conserved in mammals. However, no significant hybridization of p281 was detected with Xenopus or Drosophila DNA (not shown), providing evidence for evolutionary divergence of this Sm antigen.

**DISCUSSION**

**Characteristics of the snRNP Protein cDNA.** Clone p281 carries a HeLa cDNA insert of ≈350 nucleotides, of which 60 nucleotides are poly(dA)(dT) tails used in cloning (see Materials and Methods). Therefore, p281 contains about 290 nucleotides of the non-poly(A) portion of this mRNA. Since the protein encoded by our cDNA clone has a Mr of 11,000, the mRNA's coding region is ≈330 nucleotides (11,000 daltons + 100 daltons/average amino acid × 3 nucleotides/ amino acid = 330 nucleotides). The mRNA's total length estimated from RNA gel blot hybridization is 600

**Fig. 4.** RNA gel blot hybridization. RNA prepared as described (8) was glyoxalated, separated by electrophoresis on 1.5% agarose gels, transferred to nitrocellulose, and hybridized to 32P-labeled p281 DNA. The hybridizations were carried out at 42°C in 0.75 M NaCl/75 mM sodium citrate, 50 mM sodium phosphate buffer (pH 6.5), 0.5% formamide, 0.5 mg of salmon testes DNA per ml, Denhardt's solution (0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin), and 10 µg of poly(A) per ml for 16–20 hr. The blots were washed at 42°C in 30 mM NaCl/3 mM sodium citrate (Left) or 15 mM NaCl/1.5 mM sodium citrate (Right). (Left) Lane 1, 15 µg of HeLa poly(A)+ cytoplasmic RNA. Lane 2, 3 µg of HeLa poly(A)+ cytoplasmic RNA. Lane 3, 15 µg of Drosophila Kc cell RNA. Lane 4, 15 µg of Macaca fascicularis liver RNA. Lane 5, 15 µg of Strongylocentrotus purpuratus egg cytoplasmic RNA. Lane 6, 15 µg of marmoset cell RNA. (Right) Lane 1, 18 µg of HeLa nuclear RNA. Lane 2, 15 µg of HeLa cytoplasmic RNA. Lane 3, 3 µg of HeLa poly(A)+ cytoplasmic RNA. Lane 4, 3 µg of Xenopus laevis oocyte poly(A)+ RNA.

**Fig. 5.** Hybridization of genomic DNAs with p281. Approximately 10 µg of high molecular weight DNA was digested as indicated, separated on 1% agarose gels, transferred to nitrocellulose, and hybridized with 32P-labeled p281 DNA. Hybridization was at 65°C in 0.45 M NaCl/45 mM sodium citrate, 0.1% NaDodSO4, Denhardt's solution, 10 µg of poly(A) per ml, and 50 µg of salmon testes DNA per ml for 18 hr. The blots were washed in 0.45 M NaCl/45 mM sodium citrate at 65°C (Left), 0.1% EcoRI-digested guinea pig DNA. Lane 2, EcoRI-digested HeLa DNA. Lane 3, BamHI-digested HeLa DNA. (Right) All EcoRI-digested DNAs. Lane 1, HeLa. Lane 2, baboon. Lane 3, cat. Lane 4, mouse.
nucleotides, of which 470 is assumed to be non-poly(A) [the average length of HeLa mRNA poly(A) = 130 nucleotides; refs. 28 and 29]. From this it follows that there is a total of 140 nucleotides (470 - 330) of untranslated sequence in this mRNA. Even if this were all 3' of the termination codon (a most unlikely possibility), the cDNA clone p281 would contain at least 120 nucleotides of coding sequence (290 - 140). Sequence data (not shown) suggest that p281 contains only one open reading frame consistent with coding for an 11,000-Mr protein. This reading frame is 244 nucleotides long. It is clear from these considerations that p281 contains sufficient protein-coding sequence to constitute an appropriate hybridization probe for addressing the evolutionary conservation of a substantial portion of this nuclear autoantigen's mRNA.

The criterion by which we have identified the mRNA sequence that is hybrid-selected by clone p281 is the specific reaction of its translation product with human Sm autoantibodies. This is not as definitive as a direct biochemical characterization of the protein, but given the present complete absence of such information on the Sm antigens, the immunochemical criterion remains the best one available. Two lines of evidence indicate that the autoantigen we have cloned corresponds to the small nuclear RNA-associated E protein. The first is its electrophoretic comigration with authentic E protein (Fig. 3). The second follows from consideration of the natural variation of the human Sm autoantibody system. In immunoblots, Sm sera from most patients react only with the B, B', and D small nuclear RNA-associated proteins, but some react well with the E protein (17). In the initial cDNA clone screening, we used a particular patient serum with Sm antibody reactivity of the latter type. After the cDNA clone was isolated, we found that the hybrid-selected mRNA's translation product reacted with three additional Sm patient sera that also were known to react with E protein in total mRNA cell-free translation products. Thus, these observations strongly support the designation of the autoantigen we have cloned as the small nuclear RNA-associated E protein. This conclusion must be regarded as tentative, however, until more biochemical information becomes available for the Sm autoantibody-reactive proteins.

mRNA Expression. The lack of p281 DNA hybridization to HeLa cell nuclear RNA sequences larger than the identified cytoplasmic mRNA (~600 nucleotides) raises the possibility that the expressed snRNP E protein gene, or genes, for this Sm antigen may lack introns. However, the mRNA species that we have cloned is only of moderate cytoplasmic prevalence (see Results) and intron-containing nuclear precursors of this abundance class of cytoplasmic mRNA (30) are usually below the level of detection of RNA transfer blot hybridization. The cDNA we have obtained has facilitated isolation of genomic clones (unpublished data), which will answer this and other important questions about the organization and regulation of these genes.

The Southern blot results with human DNA indicate that there are probably not more than 10 copies of p281-homologous sequences in the genome. On the one hand, this eliminates the possibility that we have inadvertently cloned a moderately or highly repeated sequence, and on the other it strongly suggests that the copy number is more than one per haploid genome. In keeping with this, six distinct genomic clones have now been isolated by using p281 DNA as a probe (unpublished data). Determination of the existence of multiple gene copies for this Sm antigen and the further possibility of developmental regulation await further investigation; preliminary data indicating tissue-specific differences in Xenopus snRNP proteins have been reported recently (20).

Another clue to Sm antigen mRNA regulation comes from the RNA transfer blot results (Fig. 4). Given the strong cross-hybridization of the human cloned cDNA with marmoset mRNA, the complete absence of a signal with macaque, an Old World monkey, is of interest. Since p281 strongly hybridizes with genomic DNA from baboon, a member of the same primate family as macaque (31), the lack of mRNA hybridization in the latter undoubtedly reflects a low level of expression. Since the macaque RNA was from liver, whereas the human and marmoset RNAs were from exponentially growing cultured cells, this suggests that mRNA for this Sm antigen is more prevalent in dividing cells. This is but one example of the kinds of new questions that can now be directly approached with this cDNA clone.

Extent of Sm Antigen Evolutionary Conservation. Our results demonstrate extensive nucleotide sequence homology in the cloned Sm antigen's mRNA between human and marmoset, a New World primate, and in the genomic sequences for this mRNA among human, baboon, cat, guinea pig, and mouse (encompassing three orders of the Mammalia: Primates, Carnivora, and Rodentia). The absence of significant p281 hybridization with RNA in amphibian oocytes, echinoderm eggs, and Drosophila cultured cells is understandable, given the Sn antigens' demonstrated lack of sequence homology with Xenopus and Drosophila genomic DNA sequences. This indicates that if these latter organisms indeed contain functional homologues of the cloned Sm antigen, then the portion of the mammalian mRNA sequence contained in our cDNA has diverged significantly during vertebrate evolution. This is somewhat surprising, viewed against the remarkable cross-reaction of Sm antibody with snRNP antigens in phylogenetically diverse organisms. Since the B and D snRNP proteins of Xenopus and Drosophila react with Sm antibodies in immunoblots (19, 20), conserved antigenicity of the E protein is obviously not required to explain the observed antibody precipitation of snRNPs in these organisms. It is also of interest that although Xenopus snRNP contains a polypeptide whose molecular weight corresponds to the mammalian E protein (18), this component did not react with human Sm antibody in the one study reported (20).

There is substantial evidence that nonimmunogenic regions of proteins correspond to highly conserved amino acid sequences (32-34), and a possible relationship of this phenomenon to rates of protein evolution has been suggested (35). A remarkable aspect of autoimmunity is that antibodies are often elicited against macromolecular complexes (e.g., nucleoproteins) whose components are notoriously poor conventional immunogens. Among such molecules are nucleic acids and extremely conserved proteins such as histones. Our data suggest that the snRNP E protein is less conserved than the B and D proteins. It also elicits a weaker autoimmune response (17). This is consistent with the possibility that there is a threshold of evolutionary conservation below which certain proteins or nucleoproteins never become autoantigens during breakdown of the host's immunological tolerance or do so only weakly. Continued molecular analysis of strong vs. weak autoantigens and their immunological and evolutionary homologies may provide further clues to the important, related problems of immunological surveillance and autoimmunity.

Note. A brief account of this work was presented at the American Society for Cell Biology meeting, Nov. 12-16, 1984. This is paper no. 34 in a series from the laboratory of T.P. entitled "Ribonucleoprotein Organization of Eukaryotic RNA."

We thank Naomi Rothfield (University of Connecticut Health Sciences Center) for patient serum and Michael Getz (Mayo Foundation), Ronald Desrosiers (New England Regional Primate Research Center, Harvard Medical School), and Joel Richter and Samuel Wadsworth (Worcester Foundation) for DNA and RNA sam-
ples. This research was supported by American Cancer Society Grant CD-126 to T.P., by National Institute of Health Postdoctoral Fellowship CA-06751 to E.D.W., and in part by the Mayo Foundation.