Molecular Cloning, Expression, and Chromosome 19 Localization of a Human Ro/SS-A Autoantigen

Daniel P. McCauliffe, F. Anthony Lux, Tau-San Lieu, Ignacio Sanz, Jeffrey Hanke, Marianna M. Newkirk, Linda L. Bachinski, a Yasuhiko Itoh, b Michael J. Siciliano, a Morris Reichlin, c Richard D. Sontheimer, and J. Donald Capra

Departments of Dermatology, Internal Medicine, and Microbiology, University of Texas, Southwestern Medical Center, Dallas, Texas 75235; *M. D. Anderson Hospital and Tumor Institute, Houston, Texas 77030; and Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104

Abstract

Ro/SS-A antibodies are found in a number of human autoimmune disorders including Sjögren's syndrome and several systemic lupus erythematosus–related disorders. These heterogeneous autoantibodies are known to recognize several distinct cellular antigens. With synthetic oligonucleotides corresponding to amino acid sequence information we have isolated a full-length cDNA clone which encodes a human Ro ribonucleoprotein autoantigen. The 1,890-base pair clone contains an open reading frame that encodes a 417-amino-acid, 48-kD polypeptide that migrates abnormally at 60 kD by SDS-PAGE. Rabbit antibodies raised against this protein's recently described amino-terminal epitope react with a previously identified 52-kD human Ro protein and immunoprecipitate the human cytoplasmic RNAs. Ultraviolet light cross-linking studies suggest that this Ro protein binds each of the four major human cytoplasmic RNAs. The deduced amino acid sequence is 63% homologous to an Onchocerca volvulus antigen. Southern filter hybridization analysis indicates that this gene is not highly polymorphic and exists as a single copy in the human genome. Chromosomal localization studies place this gene on the short arm of chromosome 19 near the gene encoding the low density lipoprotein receptor. (J. Clin. Invest. 1990: 85:1379–1391) autoantigen • human cytoplasmic RNA • onchocerciasis • ribonucleoprotein • Sjögren's syndrome • systemic lupus erythematosus

Introduction

The Ro autoantigens are of clinical interest in that antibodies directed against them are found in the majority of patients with primary Sjögren's syndrome, subacute cutaneous lupus erythematosus, neonatal lupus erythematosus, anti–nuclear antibody-negative lupus erythematosus and systemic lupus erythematosus–like disease secondary to homozygous C2 or C4 complement deficiency (1–6). The significance of these antibodies is uncertain, but there is substantial evidence that they have a major role in the pathogenesis of disease (7).

In 1969 Clark et al. (8) first demonstrated the presence of these antibodies in sera from patients with systemic lupus erythematosus and Sjögren's syndrome. These sera precipitated antigens from cell and tissue extracts by immunodiffusion. They designated these the Ro antibody and Ro antigen, respectively. In 1975 Alspaugh and Tan (9) similarly described the presence of three types of precipitating antibodies in Sjögren's syndrome patient sera; these authors referred to the three types as SS-A, SS-B, and SS-C. SS-A antibodies were later shown to be immunologically equivalent to the Ro antibodies (10).

In 1981 Lerner et al. (11) demonstrated that human Ro antisera precipitated a novel class of small RNAs designated the human (h)1 cytoplasmic (Y) RNAs, although there is now evidence that the hY RNAs may also be intranuclear (12). The immunoprecipitation of the hY RNAs required the presence of protein; thus it was deduced that the hY RNAs are bound to a protein antigen which in 1984 was shown to be a 60-kD protein (13). We subsequently isolated a 60-kD protein from a human B-cell line which reacts with human Ro antisera. The amino terminus of this protein was sequenced and synthetic peptides corresponding to this sequence are reactive to Ro antisera (14). These data as well as a significant body of other experimental data at that time suggested that Ro antibodies were directed at a single 60-kD ribonucleoprotein (RNP) that was designated the Ro protein, Ro RNP, or Ro (auto)antigen. Recently, however, it has been shown that Ro antisera react with at least four immunologically distinct “Ro” proteins of three different molecular masses: a 52- and 54-kD protein and two 60-kD proteins (15). It is not known whether all of these proteins bind the hY RNAs or if they are structurally or functionally related. Further molecular characterization of these proteins will help address these questions and should further clarify the cellular function(s) and the pathogenic role(s) of these protein autoantigens. Here we report the molecular characterization of a human Ro RNP through the cloning and analysis of its cDNA.

Methods

Protein purification and sequence analysis

The Ro protein was purified from the human Wil-2 cell line (an Epstein-Barr virus–transformed lymphoblastoid B-cell line) as previously described (14). Staphylococcus aureus V8 (Boehringer Mannheim Biochemicals, Indianapolis, IN) and cyanogen bromide (Sigma Chemical

1. Abbreviations used in this paper: CHO, Chinese hamster ovary; h, human; LDLR, LDL receptor; poly-A, polyadenylated; RNP, ribonucleoprotein; UVC, short wavelength ultraviolet light; Y, cytoplasmic.
Co., St. Louis, MO) cleavage fragments were generated according to established protocols (16) and sequenced using a model 470A protein sequence/model 120A PTH analyzer (Applied Biosystems, Inc., Foster City, CA), as previously described (14).

**Synthetic oligonucleotide construction**

A codon utilization table was employed to convert the amino acid sequence into its most probable nucleic acid sequence (17). The oligonucleotides were synthesized using a model 380B DNA synthesizer (Applied Biosystems).

**cDNA library construction**

Total RNA was isolated from the Wil-2 cell line by the guanidinium method and enriched for the polyadenylated (poly-A) fraction with an oligo(dT)-cellulose column (18). cDNA was made from the poly-A-enriched fraction with the cDNA synthesis system (Bethesda Research Laboratories, Gaithersburg, MD). The cDNA was dG-tailed with dGTP and terminal transferase and ligated into similarly dC-tailed pGEM plasmid DNA with T4 DNA ligase (19). DH5 Escherichia coli competent cells (Bethesda Research Laboratories) were transfected with the cDNA-pGEM ligation mixture and a cDNA library was constructed (18). A human hybridoma cDNA library was similarly constructed.

The enzymes used in the various recombinant nucleic acid techniques were obtained from Promega Biotec, Madison, WI, or Pharmacia, Inc., Piscataway, NJ, unless stated otherwise.

**cDNA isolation**

The synthetic oligonucleotides were radiolabeled and hybridized with nitrocellulose filters to which the cDNA-containing bacterial colonies had been fixed (18). A single colony containing a 1.2-kb cDNA insert was isolated. Later this 1.2-kb cDNA was radiolabeled and used to screen a human hybridoma cell cDNA library, and a single 1.9-kb cDNA insert was isolated.

**cDNA characterization**

**Restriction enzyme analysis.** The 1.9-kb cDNA was digested with various restriction enzymes and the restriction fragments were analyzed by Southern filter hybridization with radiolabeled synthetic oligonucleotides (19).

**Sequence**. Several of the cDNA restriction fragments were electroeluted from a 1% agarose gel and subcloned into M13mp18 and M13mp19 plasmid vectors (Boehringer Mannheim Biochemicals) and single-stranded DNA complementary to both strands of cDNA was generated (20). This DNA was sequenced by the Sanger dideoxy method with [α-35S]dATP and modified T7 DNA polymerase (Sequenease) according to the manufacturer's recommendations (United States Biochemical Corp., Cleveland, OH).

**Northern filter hybridization.** Total RNA and poly-A-enriched RNA from several human white blood cell lines (obtained as outlined above) were electrophoresed in a 1% agarose-formaldehyde gel, electrophoretically transferred to Zeta-Probe nylon-reinforced support membrane according to the manufacturer's guidelines (Bio-Rad Laboratories, Richmond, CA), hybridized with radiolabeled cDNA, and then washed at 65°C in 0.25× SSC (1× SSC is 0.15 M NaCl and 0.015 M sodium citrate, pH 7.0) and 0.1% sodium dodecyl sulfate (SDS).

**Southern filter hybridization.** 15 μg of human genomic DNA was digested with various restriction enzymes, separated by 0.6% agarose gel electrophoresis, and transferred to nitrocellulose support membrane where it was hybridized with radiolabeled full- or partial-length cDNA. Membranes were washed in a 0.5× SSC, 0.1% SDS solution at 65°C (19).

**cDNA expression analysis**

The bacteriophage T7 RNA polymerase/promoter system developed by Tabor and Richardson (21) was utilized for cDNA expression. The 1.9-kb cDNA was amplified with the polymerase chain reaction with two degenerate 24-base oligonucleotides in order to incorporate an Nde I restriction enzyme site at the ATG translation start site and 64 base pairs beyond the stop codon (22). The resultant polymerase chain reaction-derived fragment was then ligated into the Nde I site of the pT7-7 expression vector which was obtained through Dr. S. Tabor, Department of Biological Chemistry, Harvard Medical School, Boston, MA. TG1 E. coli cells were transformed with the resultant ligation mixture and a colony which had the cDNA insert in the correct orientation with preservation of the Nde I sites, as determined by restriction enzyme analysis, was chosen for further study. Cells from this colony were subsequently transformed with the pGPI-2 vector. The cDNA was translated in the presence of rifampicin and [35S]methionine, the cells were harvested and samples of bacterial lysate were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (21).

**Radiolabeling and autoradiography**

Synthetic oligonucleotides were end labeled with [γ-32P]ATP using T4 polynucleotide kinase (19). cDNA was radiolabeled using the heximer extension method with heximer primers (Pharmacia, Inc.). [α-32P]-dCTP and E. coli DNA polymerase I (Klenow fragment) (19). Radiolabeled cDNA were obtained from New England Nuclear Corp., Boston, MA.

Filters and dried gels were exposed to X-OMAT-AR film (Eastman Kodak Co., Rochester, NY) between intensifying screens for an optimal period of time, and the film was then developed on a QX-60A film processor (Konica Medical Corp., Wayne, NJ).

**Chromosomal localization**

Somatic cell hybrid clone panels were formed by polyethylene glycol-mediated fusion of human lymphocytes to Chinese hamster ovary (CHO) cell lines defective for various DNA repair capabilities. Cytogenetic analysis was used to determine the presence or absence of human chromosomes in each of the hybrid clones. Due to frequent human chromosomal alterations in these clones, the human chromosomes were more definitively detected by analysis of isoenzyme and DNA markers (23, 24). Probes for complement component 3 (C3) and low density lipoprotein receptor (LDLR) were used to identify the short arm of chromosome 19.

**Computer-based sequence analysis**

The 1.9-kb cDNA nucleic acid sequence and its deduced amino acid sequence were analyzed for homologies to other published sequences. This was done with the University of Wisconsin Computer Genetics Group’s Genetics Analysis software and the FASTA/FASTP programs. The nucleic acid sequence was compared to the European Molecular Biology Lab database-Version 13 (April 1988) and the Genbank database-Version 56 (July 88). The protein sequence was compared to the National Biomedical Research Foundation database-Version 13 (March 1988) (25).

**Deglycosylation analysis**

The purified Ro protein was digested with neuraminidase, endo-α-N-acetylglactosaminidase, and glycopeptidase F according to the manufacturer’s recommendations (Boehringer Mannheim Biochemicals), and then subjected to SDS-PAGE.

**Rabbit antiserum immunopurification and immunoblotting**

A female New Zealand White rabbit was immunized with a synthetic peptide corresponding to amino acids 6–19 of the mature Ro protein as previously described (14). 10% polyacrylamide gels with 4.5% polyacrylamide stacking gels were run using the discontinuous buffer method previously described (26). Samples were denatured by boiling for 10 min in the presence of 5% 2-mercaptoethanol (Sigma Chemical Co.). The 1.5-mm gels were run at 20 mA constant current on a mini-gel apparatus model SE-200 (Hoefer Scientific Instruments, San Francisco, CA). Protein that had been electrophoresed was blotted onto nitrocellulose membranes and incubated with antibodies in the Western blot procedure as previously described (27). For elution of
affinity-purified antibodies, a vertical electrophoresis apparatus (14 \times 16 \text{ cm}) (model 2001, Pharmacia LKB, Piscataway, NJ) was used. The method for the elution of antibodies has been described (27). Briefly, the strips were washed thoroughly and the antibody eluted with 3 M NaSCN. The eluate was dialyzed to PBS (Tris-buffered saline with Tween 20 for Western immunoblotting or IPP buffer [10 mM Tris-HCl, 500 mM NaCl, 0.1% NP-40, pH 8.0] for RNA immunoprecipitation) by repeated dilution and concentration using a Centriprep concentration device as recommended by the manufacturer (Amicon Corp., Lexington, MA).

RNA binding studies
Immunoprecipitation. Immunoprecipitation, electrophoresis, and silver staining of the hY RNAs was performed according to a well-established protocol (28).

Short wavelength ultraviolet light (UVC) cross-linking. The purified Ro protein was irradiated with 25–125 ml/cm² of UVC using four Phillips TUV-15W germicidal lamps (Gulf Coast Electric, Houston, TX). The Ro protein was analyzed by Western immunoblotting as previously described (14). A portion of the cross-linked purified extract was also incubated with ribonuclease A (1 mg/ml overnight at room temperature) before immunoblot analysis.

Results
Amino acid sequencing and synthetic oligonucleotide construction. We isolated a 60-kD protein with Ro antigenic activity from the Epstein-Barr virus–transformed human Wil-2 B-cell line and subjected it to a limited S. aureus V8 protease digestion. This produced 23- and 37-kD fragments which were identified by SDS-PAGE. The amino-terminal end of the 60-kD protein and its 23- and 37-kD fragments were sequenced and this information was used to construct two nondegenerate synthetic oligonucleotides (Fig. 1). The amino terminus of the 23-kD fragment was identical to that of the 60-kD protein.

cDNA isolation, bacterial expression, and sequence analysis. A single 1.2-kb cDNA clone was isolated from the Wil-2 cell cDNA library with the two synthetic oligonucleotides. This clone was characterized by restriction enzyme analysis and sequenced. The cDNA encoded the previously determined amino acid sequences, however, this reading frame contained no termination codon, indicating that the cDNA was truncated at the 3′ end. Northern filter hybridization with this cDNA (Fig. 2) identified a 2-kb RNA species but no 1.2-kb species, confirming that our cDNA was abbreviated. A human B-cell hybridoma cDNA library was subsequently screened with the 1.2-kb cDNA, and a single 1.9-kb cDNA clone was isolated and sequenced. The first 1,238 basepairs of this clone contains 1,890 basepairs which include a single 1,251-base open reading frame beginning with an AUG start site at position 67 as part of a putative Kozak ribosomal translation initiation site and ending with the termination codon UAG (Fig. 3) (29). The sequence AUUAAA (Fig. 3) is a putative polyadenylation signal (30), but there is not a typical poly-A sequence between this signal and the end of the cDNA sequence, suggesting that this 1.9-kb cDNA may be minimally truncated.

The deduced polypeptide has a molecular mass of 48 kD which includes a 17-amino acid hydrophobic leader segment that is not present in the purified mature protein. The M₁ of the deduced polypeptide without the leader segment is ~ 14 kD less than that of the “60”-kD protein as measured by SDS-PAGE. The amino acid sequence contains one potential site for NH₂-linked glycosylation (amino acid position 344 in Fig.

![Figure 2](image_url)
cctactgcagccgctgccggtgcttttaaagggcccgctttgcctttggctttctggtgctggcttccgctttcctcgtttttctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
3) but deglycosylation analysis of the purified 60-kD protein shows no evidence of NH2- or COOH-linked glycosylation (data not shown). Bacterial expression of the cDNA produced a protein which migrates at ~60 kD by SDS-PAGE, which is inclusive of the ~2-kD hydrophobic leader segment (data not shown). A similar M, discrepancy between that encoded by the cDNA and that measured by SDS-PAGE, has been reported in the cloning of several other proteins as shown in Table 1. These proteins have a highly charged region in common which may cause retarded gel migration and thus an overestimation of the M,‘s by SDS-PAGE. Our protein has a highly charged region between residues 358 and 408 where 47 of 51 residues are strongly charged (36 are negatively charged and 11 are positively charged). The calculated isoelectric point of this polypeptide is 4.14 which closely approximates our value of 4.67 measured from the native purified protein (36).

This protein contains three different sets of repeating sequences (Fig. 4) which may have arisen from internal replications and may be of functional importance. The first set has 82% of its nucleic acid sequence and 82% of its amino acid sequence conserved. The second set has 78% nucleic acid and 71% amino acid sequence conservation, and the third set has 83% nucleic acid and 80% amino acid sequence conservation. This protein also contains several PEST regions (Fig. 4, upper panel) as proposed by Rogers et al. (37): these so-called PEST regions are rich in the amino acids proline (P), glutamic acid (E), serine (S), and/or threonine (T), and to a lesser extent aspartic acid (D). These regions are thought to make a protein susceptible to rapid intracellular degradation.

The primary structure of this protein is unique but computer-based analysis of the nucleic and amino acid sequence shows striking 63% amino acid sequence homology to an antigen recently partially cloned from Onchocerca volvulus (Fig. 5) (38). 14 of the 15 amino acids at the amino terminus of the mature protein are identical with the 15-amino terminal residues recently deduced from the purified rabbit calregulin protein (39). The negatively charged carboxy-terminal region has some minor amino acid sequence homology with a number of other proteins of diverse origin and function. The most striking of these homologies is with residues 44-80 of the 17-kD subunit of yeast ubiquinone cytochrome c reductase (40), where 18 of 36 residues are identical and 9 of the remaining 18 residues are a Asp for Glu or a Gln for Asp substitution with Ro residues 379-415. Within this negatively charged region is a sequence (residues 384-395) which has some homology to previously described calcium binding domains (41), although this sequence is not flanked by hydrophobic residues as in the classic “EF hand” calcium-binding domain (42). The carboxy-terminal sequence Lys-Asp-Glu-Leu (KDEL) follows the negatively charged region (Fig. 4, upper panel) and is identical to the carboxy signal sequence which has been shown to be crucial for the retention of several proteins in the endoplasmic reticulum (43). These other proteins likewise have a highly negatively charged region just proximal to the KDEL sequence. The 17-amino acid hydrophobic leader sequence (Fig. 4, upper panel) is similar to that of a number of other precursor proteins and indicates that this protein is transported into the endoplasmic reticulum (44).

There is no striking sequence similarity to other RNA-binding proteins, including another recently sequenced Ro cDNA (45). There is no major homology to the RNP consensus sequence (46) and there are no zinc finger (47) or leucine zipper (48) nucleic acid-binding motifs.

Chou-Fasman computer-based secondary structure analysis (49) predicts a complex secondary structure (Fig. 6), which includes several helix-turn-helix units centered around residues 57, 70, 210, 233, and 246. Three of these units are found within the sequence triplications between residues 207 and 300 (Fig. 4, upper panel), one unit per each sequence repeat. There are also several β-sheet-rich areas between residues 1 and 17, 144 and 186, and 285 and 333. The carboxy-terminal residues 349-417 are predicted to have an α-helical array.

Kyte-Doolittle hydrophatic analysis (50) predicts a strongly hydrophobic leader segment and several smaller regions of hydrophobicity, including an area just proximal to the negatively charged carboxy-terminal residues which could be a membrane-spanning region. This analysis also predicts several strongly hydrophilic domains particularly between amino acids 210-300 and 350-417. Residues 210-300 include the first two sets of sequence triplications and residues 350-417 span the negatively charged carboxy end which includes the set of sequence duplications.

Jameson-Wolf antigenicity analysis (51) (Fig. 6) predicts the location of several potential epitopes including the previously characterized epitope at the amino terminus of this polypeptide, (synthetic peptide 6-19) (14), and a recently characterized epitope (synthetic peptide 171-194) corresponding to residues 171-194 (52).

Southern filter hybridization analysis and chromosomal localization. Southern filter hybridization of Eco R1 digested genomic DNA from 10 normal individuals shows a single

---

**Table 1. Proteins with Highly Charged Regions and Retarded SDS-PAGE Gel Migration**

<table>
<thead>
<tr>
<th>Protein</th>
<th>cDNA encoded</th>
<th>SDS-PAGE determined*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amphibian nucleoplasmin</td>
<td>22</td>
<td>33</td>
<td>32</td>
</tr>
<tr>
<td>Amphibian N1/N2 histone-binding protein</td>
<td>65</td>
<td>110</td>
<td>34</td>
</tr>
<tr>
<td>Bovine chromogranin A</td>
<td>53</td>
<td>75</td>
<td>33</td>
</tr>
<tr>
<td>Yeast GCN4 transcription activator</td>
<td>31</td>
<td>45</td>
<td>31</td>
</tr>
<tr>
<td>Human nuclear RNP particle C2</td>
<td>32</td>
<td>40-44</td>
<td>57</td>
</tr>
<tr>
<td>Human U1-70-kD small nuclear RNP</td>
<td>52</td>
<td>70</td>
<td>35</td>
</tr>
</tbody>
</table>

* Without any carbohydrate moiety.

---

**Figure 3.** The 1.9-kb Ro cDNA nucleic acid and encoded amino acid sequence. The 1,890-base coding strand encodes a 417-amino acid polypeptide which includes four previously determined amino acid sequences (underlined) from sequencing the native protein and cyanogen bromide and S. aureus V8 cleavage products. The eukaryotic ribosomal consensus sequence for the initiation of translation is boxed and the putative polyadenylation signal is overlined. These sequence data are available from EMBL/GenBank/DDBJ under accession number M32294.

**Cloning and Chromosome 19 Localization of a Human Ro/SS-A Autoantigen**
were 224 KI...HLI...1384 McCauliffe 19 (8%) chromosome...between...human
dancy with human chromosomes. Similarly...human chromosome 19...is...human chromosome...long arm of this small, slightly submetacentric chromosome. Therefore, the three clones (1HL14, 9HL9, and 24HL8) discordant for the chromosome 19 markers and Ro suggest that Ro might be on the short arm of the chromosome. This hypothesis was tested by examining...three hybrids for the presence of known chromosome 19 short-arm markers, C3 and LDLR. The results are summarized in Table III. Ro is perfectly...LDLR in this set, clearly placing the gene on the short arm of chromosome 19. The discordancy of LDLR with C3-PEPD-GPI in clone 24HL8 is consistent with the linkage data placing LDLR distal to C3 (53).

**Immunoblotting.** Rabbit anti-synthetic peptide 6–19 (SP 6–19) antiserum bound to the 60-kD protein and a previously identified 52-kD Ro protein on the same immunoblot (data not shown). When the rabbit antibodies are eluted from the 52-kD protein they react with the 60-kD protein.

**RNA binding studies.** Previous studies have demonstrated that human Ro antisera immunoprecipitate the hY RNAs which are noncovalently bound to a 60-kD protein to which the antisera reacts (13). Recently the hY RNAs were similarly immunoprecipitated with immunoaffinity-purified Ro antisera directed against either a 52- or a 60-kD protein (54). With immunopurified rabbit anti–SP 6–19 antibodies we were able to immunoprecipitate the hY RNAs from HeLa cells (Fig. 9,
left panel). We have previously demonstrated the presence of RNA in our purified Ro protein product by ultraviolet absorbance analysis (14). To determine whether this was hY RNA, the purified protein was irradiated with UVC to cross-link any associated RNA to the Ro protein so that a shift in molecular mass could be detected by SDS-PAGE analysis. This method of cross-linking protein to intimately associated RNA molecules has been well established (55). In this manner we demonstrated that the purified Ro protein which normally associates at 60 kD by SDS-PAGE, migrates at four different higher molecular masses between ~86 and 96 kD after UVC cross-linking (Fig. 9, right panel). The UVC-induced molecular mass could be attenuated by digesting the cross-linked sample with ribonuclease before SDS-PAGE (data not shown). Rabbit anti-SP 6–19 serum and a human Ro antiseraum both react to the 60-kD protein and each of the four higher molecular mass cross-linked species by immunoblot analysis (data not shown).

Discussion

We have isolated, characterized, and expressed a cDNA clone which encodes a 46-kD Ro RNP autoantigen which migrates at 60 kD by SDS-PAGE. The deduced amino acid sequence although novel is 63% homologous with an *Onchocerca volvulus* antigen (On). The cDNA sequence was truncated at the 5' end and thus the encoded amino terminus of this antigen is incomplete (incomplete NH2 end) for full homology comparison. The numbers correspond to the Hu amino acid sequence position. Symbols: (·) represents a conservative amino acid substitution; (-) represents a semiconservative substitution; (--) indicates a gap between amino acids; (+) denotes the end of the On protein. These sequence data are available from EMBL/GenBank/DDBJ under accession number M32294.

Figure 5. Amino acid sequence homology between the human Ro protein (Hu) and an *Onchocerca volvulus* antigen (On). The On cDNA sequence was truncated at the 5' end and thus the encoded amino terminus of this antigen is incomplete (incomplete NH2 end) for full homology comparison. The numbers correspond to the Hu amino acid sequence position. Symbols: (·) represents a conservative amino acid substitution; (-) represents a semiconservative substitution; (--) indicates a gap between amino acids; (+) denotes the end of the On protein. These sequence data are available from EMBL/GenBank/DDBJ under accession number M32294.
The fact that Ro antibodies specific for a sequence encoded by our cDNA recognize our 60-kD protein from Wil-2 cells and a 52-kD protein from chronic lymphocytic leukemia cells, on the same immunoblot, suggests that both proteins have one or more shared epitopes. The two proteins might be separate gene products, or the 52-kD protein might be a degradation product of the 60-kD protein. Similarly the 60-kD protein might result from a posttranslational modification of the 52-kD protein. It is unlikely that these two proteins result from different transcripts from the same gene as only one molecular weight species of RNA was detected on Northern filter hybridization. We found no evidence that the 60-kD protein is glycosylated or ubiquitinated (personal observation). Differences in purification techniques or cell/tissue source might explain why two different Mr proteins were isolated.

The absence of a typical poly-A tail at the 3' end of our 1.9-kb cDNA suggests that it is truncated as was the 1.2-kb clone. This may have arisen from aberrant cDNA synthesis or from subsequent deletion of the poly-A tail after cDNA synthesis. Another explanation would be that the Ro mRNA is not poly-A tailed, like histone mRNA. However, there is no comparable 3' end processing signal sequence as found in the histones (56) and the 1.9-kb clone does have a poly-A signal sequence.

No major similarities were found between the RNP consensus sequence and the deduced amino acid sequence of our 1.9-kb cDNA. We have, however, detected a 20-29-kb hybridizing Hind III fragment in the human (HeLa) cell line in comparison to a 5.7 hybridizing fragment in the hamster (CHO) cell line. Human × hamster hybrid clones 24HL10 and 1H15 contain that portion of human chromosomal DNA where this Ro gene resides whereas clone 1H14 does not.
Table II. Human Chromosomes and DNA Sequences Present or Absent in Hybrids

<table>
<thead>
<tr>
<th>Hybrid clone</th>
<th>Chromosomes</th>
<th>Ro probe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1HL14</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>1HL15</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>3HL2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3HL3</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3HL4</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>3HL10</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3HL11</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>5HL14</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>5HL15</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>5HL16</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5HL17</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>5HL20</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>9HL3</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>9HL7</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>9HL9</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>9HL10</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9HL11</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>9HL12</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>20HL1</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>20HL4</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>20HL6</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>20HL7</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>20HL21-4</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>24HL5</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>24HL7-5</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>24HL8</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>24HL10</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>41HL7</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>41HL10</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>135HL3</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>135HL6</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>135HL7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>135HL9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>135HL11</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>135HL12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>135HL15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3B2</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>5A3</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Percentage discordant with Ro (%) | 42 | 61 | 49 | 53 | 45 | 53 | 57 | 34 | 47 | 54 | 50 | 42 | 57 | 37 | 42 | 65 | 61 | 49 | 8 | 50 | 50 | 47 | 45

Symbols: +, present; -, absent. Blank spaces were not read.
Table III. Chromosome 19 Markers in Hybrid Clones with Discordancies

<table>
<thead>
<tr>
<th>Hybrid clone</th>
<th>Ro</th>
<th>LDLR</th>
<th>C3</th>
<th>PEPD</th>
<th>GPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>24HL8</td>
<td>+</td>
<td>+</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9HL9</td>
<td></td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>1HL14</td>
<td></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Symbols: +/−, presence/absence of marker hybridization.

cDNA (46). However, the RNP consensus sequence is not necessarily a requirement nor a universal property of RNA binding proteins for it is absent in ribosomal proteins, in many viral RNA-binding nucleocapsid proteins and in the Sm-D RNP autoantigen (57, 58). The three duplications between residues 207 and 255 each have a helix-turn-helix configuration and may be a site of RNA binding. However, these helix-turn-helix units have a larger turn component than those described in other nucleic acid binding proteins (59). It is of interest to note that another RNA-binding protein, the human 70-kD small nuclear RNP, also resides on chromosome 19 and has a similar M, disparity, though it has no significant sequence homology with Ro protein (35).

The role of this protein in cellular function and its precise cellular location are unknown. Analysis of the amino acid sequence gives some insight into these issues. The hydrophobic leader segment of the polypeptide suggests that this protein undergoes transmembrane transport. This sequence may serve to transport this protein across the endoplasmic reticulum for modification, however there is no evidence of glycosylation. The KDEL carboxy-signal sequence suggests that this protein may reside in the endoplasmic reticulum.

The amino terminal amino acid sequence similarity with rabbit calregulin, a calcium-binding protein with an M, and pI similar to this 60-kD Ro protein (39), and a region of putative calcium binding in our protein suggest that this protein may have a similar function or may even be the human equivalent to rabbit calregulin. This is even more interesting in light of studies which have shown that calregulin is sequestered in the endoplasmic reticulum or another membrane-bound cytoplasmic organelle (60). Thus the 60-kD Ro protein might have a hydrophobic leader to allow transport into a membrane bound organelle where it might bind calcium, and the KDEL sequence would insure retention of the protein within that organelle.

Indirect immunofluorescence microscopy on cultured mouse L cells (fibroblasts) and human Hep-2 cells (epitheloid cells) with rabbit antiserum raised against SP 6–19, reveals predominantly perinuclear and cytoplasmatic staining (data not shown), similar to the pattern reported by Hendrick and colleagues with Ro antiserum (61), and also similar to the pattern seen with antibodies directed against calregulin (60). Other investigators have reported predominantly intranuclear localization of the Ro antigens by indirect immunofluorescence (12). Whether this discrepancy in subcellular localization is related to the method of cell fixation, cell substrate, or the type of Ro antiserum used, needs to be further investigated. Isolation of antibodies specific for each Ro antigen should allow a more precise subcellular localization of each antigen and these results may help explain the discrepancies encountered thus far with immunofluorescence staining.

The Ro polypeptide does contain the sequence PPKIKDPD (residues 203–212 in Fig. 4, upper panel) which is very similar to nuclear targeting signals of other nuclear proteins (62). This sequence might facilitate transport of this protein into the nucleus. Histone-binding proteins similarly have nuclear targeting signals and highly negatively charged regions (32, 34).

There has been mounting evidence that foreign microbial antigens may trigger an inappropriate immune response against self-antigens through molecular mimicry (63). Initial computer search for sequence homology to microbial agents has not been fruitful. As the Ro epitopes become better defined it may become more apparent whether microbial agents play a role in the pathogenesis of this autoimmune response. The sequence homology with the *Onchocerca volvulus* antigen does suggest the possibility that a foreign protein homologous to a self-protein might trigger an immune response which reacts with the self protein. *Onchocerca volvulus* is a filarial nematode which causes river blindness, sclerosing lymphadenitis, and dermatologic disease in humans residing in parts of Africa and Central America (64). Studies are underway to determine whether sera from patients with this disease contain Ro antibodies.

The relationship between this Ro protein and the others is unknown. They may be structurally and (or) functionally related as are the antigenic U series of RNP (65). Taken to-
gether, our data and the hY RNA-binding data from other investigators suggest that a 52-kD and two 60-kD Ro proteins bind hY RNA. Ro antisera specific for a 52- or a 60-kD protein have been shown to immunoprecipitate the hY RNAs from cellular extracts. However, whenever the 52-kD specific antibodies were used in this study, the hY RNAs and 52-kD protein were precipitated along with a small amount of 60-kD protein (13, 54). Thus it is not certain whether the 52-kD species binds the hY RNA directly or indirectly through its association with a 60-kD hY RNA-binding protein. Binding of a hY RNA has also been demonstrated in reconstitution studies with another recently characterized 60-kD Ro RNP, though the efficiency of binding was reportedly quite low (45). Immunopurified rabbit antibodies directed at our 60-kD protein's amino-terminal amino acid sequence immunoprecipitate the hY RNAs and also recognize a recently characterized 52-kD protein, making it uncertain whether the immunoprecipitated hY RNA was bound to the 52-kD protein, the 60-kD protein, or both. The UVC cross-linking studies give more direct evidence that the "60"kD protein binds each of the four major hY RNAs. Each of the four higher molecular mass species is consistent with the addition of one of the four major hY RNAs (13). These findings support the concept of only one hY RNA molecule bound per Ro molecule as suggested by Wolin and Steitz (13), who demonstrated that the hY RNA–protein complexes sediment at ~ 7 S (equivalent to 93 kD) in sucrose gradients (13). This is consistent with the Mo of one 60-kD protein molecule complexed with one of the hY RNA molecules which range from 28 to 38 kD in size (an average of 33 kD).

Ro was a term first used to define a soluble cytoplasmic antigen which formed a unique line of precipitation in double-immunodiffusion studies with sera from patients with systemic lupus erythematosus and Sjogren's syndrome (8). From the work of Steitz and co-workers, Ro has been further defined as a cytoplasmic hY RNA binding protein which migrates at 60 kD by SDS-PAGE (13). Like the originally described Ro protein, the protein we describe measures 60 kD by SDS-PAGE and is reactive by Western immunoblotting to prototypal monospecific human Ro antisera from several different laboratories including the laboratory of the Center for Disease Control (AF/CDC7). Rabbit antiserum raised against the amino-terminal portion of this protein (anti–SP 6-19) demonstrates a cytoplasmic pattern with indirect immunofluorescence staining, and contains antibodies that immunoprecipitate the hY RNAs. These data suggest that this Ro autoantigen may be the originally described 60-kD, cytoplasmic, hY RNA-binding protein.

Now that several proteins with Ro antigenicity have been identified, including two different 60-kD proteins, a system of classification needs to be developed so that each Ro antigen gets a more unique designation. This should be accomplished as the antigens become better characterized.

Whether or not patients with Ro antibodies can be clinically categorized by which of the Ro proteins or epitopes their sera recognize and whether or not this is related to their HLA type has not yet been determined. The ability to categorize patients based on Ro epitope recognition could have great clinical utility if the patient's clinical course and/or response to therapy could be predicted by these results. The characterization of the various Ro cDNAs and their encoded epitopes should be helpful in this regard and should also provide a means to further clarify the functional and pathologic roles of these protein autoantigens.

Acknowledgments

We are grateful to the assistance from Peter Roome and Dr. Robert Fisher in the computer-based sequence analysis and to Jeff Wilson and Carol Williams for their excellent technical assistance.

These studies were in part supported by grants AI-12127, GM-31689, AR-19101, and AR-07341 from the National Institutes of Health and by support from the Dallas Biomedical Corporation. Dr. Sontheimer is the recipient of National Institutes of Health Research Career Development Award AR01784.

References

57. Swanson, M. S., T. Y. Nakagawa, K. LeVan, and G. Dreyfuss. 1987. Primary structure of human nuclear ribonucleoprotein C pro-


