

## Cloning of a complementary DNA coding for the 100-kD antigenic protein of the PM-Scl autoantigen.

Q Ge, ... , C O'Brien, I N Targoff

*J Clin Invest.* 1992;**90**(2):559-570. <https://doi.org/10.1172/JCI115895>.

### Research Article

Anti-PM-Scl antibodies are associated with polymyositis-scleroderma overlap or either disease alone. Among sera from 39 patients with anti-PM-Scl, 23 recognized the 100-kD band in immunoblot against HeLa cell extract, 16 of which also stained the 70-kD band. A human thymocyte lambda gt11 cDNA expression library was screened with anti-PM-Scl serum, and two clones were identified whose products reacted with 33 and 37 of 39 anti-PM-Scl sera, respectively, but none of 26 negative control sera. Affinity-purified antibody reacting specifically with plaques of the clone stained the 100-kD band on immunoblot, reacted with nucleoli of HEp-2 cells, and immunoprecipitated the PM-Scl protein complex. Partial sequences of both inserts were identical. One insert was fully sequenced, and additional 5' and 3' sequence was obtained using a gene-specific primer to form a cDNA with HeLa cell RNA as template followed by PCR. The complete nucleotide sequence included 2,739-bp coding for a predicted full-length protein of 98,088 D. There was no homology with the PM-Scl 75-kD protein and no significant homology with other proteins. A mixed-charge cluster was identified, with 22 charged amino acids of 37. In conclusion, the full-length cDNA sequence was determined coding for the PM-Scl 100-kD protein, the most commonly antigenic protein of the PM-Scl complex.

Find the latest version:

<https://jci.me/115895/pdf>



# Cloning of a Complementary DNA Coding for the 100-kD Antigenic Protein of the PM-Scl Autoantigen

Qun Ge,\* Mark Barton Frank,\* Charles O'Brien,\* and Ira N. Targoff,\*\*

\*Arthritis and Immunology Program, Oklahoma Medical Research Foundation 73104; and †Department of Medicine, University of Oklahoma Health Sciences Center 73190; and Veterans Affairs Medical Center, Oklahoma City, Oklahoma 73104

## Abstract

Anti-PM-Scl antibodies are associated with polymyositis-scleroderma overlap or either disease alone. Among sera from 39 patients with anti-PM-Scl, 23 recognized the 100-kD band in immunoblot against HeLa cell extract, 16 of which also stained the 70-kD band. A human thymocyte  $\lambda$ gt11 cDNA expression library was screened with anti-PM-Scl serum, and two clones were identified whose products reacted with 33 and 37 of 39 anti-PM-Scl sera, respectively, but none of 26 negative control sera. Affinity-purified antibody reacting specifically with plaques of the clone stained the 100-kD band on immunoblot, reacted with nucleoli of HEP-2 cells, and immunoprecipitated the PM-Scl protein complex. Partial sequences of both inserts were identical. One insert was fully sequenced, and additional 5' and 3' sequence was obtained using a gene-specific primer to form a cDNA with HeLa cell RNA as template followed by PCR. The complete nucleotide sequence included 2,739-bp coding for a predicted full-length protein of 98,088 D. There was no homology with the PM-Scl 75-kD protein and no significant homology with other proteins. A mixed-charge cluster was identified, with 22 charged amino acids of 37. In conclusion, the full-length cDNA sequence was determined coding for the PM-Scl 100-kD protein, the most commonly antigenic protein of the PM-Scl complex. (*J. Clin. Invest.* 1992. 90:559-570.) Key words: polymyositis • scleroderma • autoantibodies • anti-PM-Scl • nucleolar proteins

## Introduction

Autoantibodies reacting with the PM-Scl antigen have been found in sera from patients with polymyositis or scleroderma or from patients with overlap of these two conditions (1-5). They were identified in ~ 8% of a large group of myositis patients and ~ 3% of a group of scleroderma patients (1, 4). However, it is very rare for this antibody to be found in patients with other conditions, making it a valuable disease marker, and suggesting that it may have links to fundamental disease mechanisms in those patients in whom it is found. Anti-PM-Scl is the unique specificity derived from anti-PM-1 (1, 6).

The PM-Scl antigen is found in the nucleolus (7, 8), and all

Portions of this work were presented at the 54th and 55th Annual Meetings of the American College of Rheumatology, 27 October-1 November 1990 and 17-21 November 1991.

Address correspondence to Ira N. Targoff, M. D., Arthritis and Immunology Program, Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, OK 73104.

Received for publication 8 August 1991 and in revised form 2 March 1992.

The Journal of Clinical Investigation, Inc.  
Volume 90, August 1992, 559-570

sera with anti-PM-Scl stain the nucleolus in indirect immunofluorescence. Most sera with the antibody also stain the nucleus, but antibody could be affinity purified that reacted only with the nucleolus (7). The role of the antigen in cell physiology is not known, but it may be involved in preribosomal formation, a function of the nucleolus, since compounds that interfere with ribosomal formation also interfere with its production (8).

Immunoprecipitation of <sup>35</sup>S-labeled HeLa cell extracts with anti-PM-Scl shows a group of 11-16 proteins that appear to be a complex (8, 9). A prominent band is seen at 100 kD, but bands are present from 20 to 110 kD. No nucleic acids have been associated with the antigen by immunoprecipitation. Most sera with the antibody react with a 100-110-kD protein in immunoblot studies (9), and some also react with a second protein migrating in the range of 70-75 kD (10). Occasional sera react primarily with this second protein. The gene encoding the second protein has recently been cloned and sequenced and has been found to produce a protein of 39.2 kD that migrated aberrantly on PAGE, apparently because of a region rich in acidic residues in the carboxyl half of the molecule (10). The 100-110-kD protein that is commonly antigenic has not been analyzed in detail. To study further the larger antigenic protein of the PM-Scl antigen on a molecular level, we have isolated a complementary DNA (cDNA)<sup>1</sup> that codes for this protein and have determined its sequence.

## Methods

**Sera.** Sera were drawn from clinical screening at the Oklahoma Medical Research Foundation or from sera referred for testing for myositis-associated autoantibodies. Sera from 39 patients were found to contain anti-PM-Scl by immunodiffusion and confirmed by immunoprecipitation. 20 myositis sera from this collection found to be negative for anti-PM-Scl were tested as controls, some of which had other autoantibodies. Six sera from healthy laboratory workers were also tested.

**Immunoblotting.** Each serum was tested by immunoblotting against whole HeLa cell lysate as antigen source to determine its component reactivity. Antibodies that were affinity purified from *Escherichia coli*-expressed PM-Scl fusion protein as below were tested against whole HeLa cell lysates or against immunoprecipitates prepared with anti-PM-Scl serum from HeLa cell lysates. The size of the PM-Scl molecules expressed in *E. coli* were also determined with immunoblot assays against *E. coli* lysate containing the recombinant protein.

The HeLa lysate was prepared by sonication in 0.15 M NaCl in 0.05 M Tris-HCl buffer at pH 7.4 whereas *E. coli* lysates were prepared by lysis with  $\lambda$ gt11 bacteriophage with or without the cloned insert or with the sample buffer containing SDS for use in electrophoresis (11). The lysates and immunoprecipitates were electrophoresed in 10% SDS-PAGE and blotting was carried out as previously described (11), ex-

1. Abbreviations used in this paper: cDNA, complementary DNA; IPTG, isopropylthio- $\beta$ -galactoside; PCR, polymerase chain reaction; TBST, 0.05 M Tris buffer pH 8.0 with 0.15 M NaCl, 0.05% Tween-20.

cept that 8% nonfat milk was used for blocking. Sera were diluted 1:100 for routine testing, and affinity purified antibody was used undiluted.

**Isolation of PM-Scl cDNA.** It was previously known that the PM-Scl antigen was present in thymus tissue by immunodiffusion against calf thymus extract (1) and in HeLa cells by immunoprecipitation (8). Therefore, human thymocyte cDNA and HeLa cell cDNA  $\lambda$ gt11 expression libraries from Clontech Laboratories (Palo Alto, CA) were screened. Serum from patient J.H. with diffuse scleroderma, shown to contain anti-PM-Scl by immunodiffusion and to react with both the 100- and 70-kD proteins by immunoblotting, was used for screening, diluted 1/500 in 0.05 M Tris buffer at pH 8.0 with 0.15 M NaCl and 0.05% Tween-20 (TBST) with 2% BSA and 10 mM sodium azide. Recombinant phage were plated with *E. coli* Y1090 on LB agar, incubated at 42°C for 3–4 h, overlaid with nitrocellulose membranes soaked in isopropylthio- $\beta$ -galactoside (IPTG), and incubated for another 3 h at 37°C. The membranes were then washed, blocked, incubated with diluted patient serum, and developed as for immunoblotting, using alkaline-phosphatase-coupled anti-human IgG conjugate and 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium substrate. A total of  $2 \times 10^6$  plaques from each of the libraries were screened. Positive clones were plaque purified.

**Affinity purification of antibodies.** Specific antibodies were affinity purified by elution from proteins derived from plaques of the recombinant PM-Scl bacteriophage. To reduce the background attributable to anti-*E. coli* antibodies, sera were first preabsorbed with nitrocellulose membranes coated with proteins from *E. coli* that had been lysed with wild-type  $\lambda$ gt11 bacteriophage as described above. LB plates were prepared with a high concentration of either plaque-purified recombinant PM-Scl bacteriophage or wild-type bacteriophage. The plates were overlaid with IPTG-soaked nitrocellulose membranes, which were then incubated with absorbed anti-PM-Scl or other serum overnight at 4°C, as described above for screening. After five TBST washes, membrane-bound antibodies were eluted with 0.2 M glycine-HCl at pH 2.5 for 1.5 min. Collected eluates were neutralized immediately by adding 1 M Tris base. The activity of the affinity-purified antibody was confirmed by testing against recombinant and wild-type bacteriophage proteins on nitrocellulose as above.

**Testing of patient sera.** To determine whether patient sera with anti-PM-Scl or controls reacted with the protein product of the clone, plaque-purified recombinant bacteriophage were mixed 5:1 with wild-type  $\lambda$ gt11 and plated with *E. coli* Y1090, as for screening. Pieces of nitrocellulose carrying plaque proteins were incubated with 5 ml of an individual serum diluted 1:600 and developed. Reactions were considered positive if significant staining of ~20% of the plaques was seen. Intensity of staining was estimated on a 0 (negative) to 4+ (strongest) scale. Three anti-PM-Scl and two normal sera were also tested against IPTG-induced *E. coli* lysed with recombinant bacteriophage in immunoblot as described above.

**Immunoprecipitation.** Immunoprecipitation from  $^{35}$ S-methionine-labeled HeLa cell lysate was performed as previously described (12). 1 mg protein A-Sepharose (Pharmacia LKB, Uppsala, Sweden)

(10  $\mu$ l of swelled bead suspension) was coated with 20  $\mu$ l of serum or 500  $\mu$ l of affinity-purified antibody. Immunoprecipitates were analyzed on 10% SDS-PAGE. Preparation of unlabeled immunoprecipitates for immunoblot analysis of affinity-purified antibodies was similar but the immunoprecipitates were treated as for protein analysis. Immunoprecipitation from unlabeled HeLa cell lysate for nucleic acid analysis was also performed as previously described (12). One 150-cm<sup>2</sup> flask of HeLa cells was used for three samples. 4 mg of protein-A-Sepharose were coated with 20  $\mu$ l of serum. Phenol-extracted immunoprecipitates were analyzed on 7 M urea, 10% PAGE, and developed with silver stain (Bio-Rad Laboratories, Richmond, CA).

**Northern blot analysis.** Total RNA was isolated from HeLa cells by the procedure of Chomczynski and Sacchi (13), and poly-(A)<sup>+</sup> RNA was isolated from HeLa cells using a commercial kit (MicroFastTrack; Invitrogen, San Diego, CA). Both were electrophoresed in a denaturing 1% agarose gel containing 2 M formaldehyde and were transferred onto Nytran membrane (Schleicher & Schuell, Inc., Keene, NH) (14). Prehybridization with salmon sperm DNA and hybridization with radiolabeled cDNA probe (15, 16) were performed in 2 $\times$  prehybridization and hybridization buffers from 5 Prime-3 Prime Inc. (Boulder, CO) according to their instructions, with formamide at 42°C. Membranes were washed under high stringency conditions with 0.1 $\times$  standard saline citrate and 0.1% SDS at 65°C for 30 min and were developed with autoradiography.

**DNA sequencing.** DNA from the plaque-purified recombinant  $\lambda$ gt11 bacteriophage was prepared from liquid lysates and digested with EcoRI to excise the cDNA insert. The insert was isolated by fractionation in an agarose gel and subcloned into the plasmid pUC18 (17). After restriction enzyme mapping of the cDNA, individual restriction fragments were isolated from agarose gels. These were ligated into M13mp18 and mp19 (17) and were used to transform *E. coli* strain XL1 Blue. The complete nucleotide sequences of both strands of the cDNA were determined by using the dideoxynucleotide chain-termination method (18), using  $^{35}$ S-labeled dATP and T7 DNA polymerase [United States Biochemical, Cleveland, OH (Sequenase Version 2.0) or Pharmacia LKB Biotechnology, Piscataway, NJ] (19). Compression regions were resolved by using deoxyinosine triphosphate.

**Determination of the 5' and 3' ends of the cDNA sequence.** Additional cDNA sequence for the gene for the PM-Scl 100-kD protein not included in the original bacteriophage insert was obtained by using the methods of Loh et al. (20) and Frohman et al. (21).

For amplification of the 5' end of the cDNA, 15  $\mu$ g of total HeLa cell RNA and 2  $\mu$ g of HeLa cell poly-(A)<sup>+</sup> RNA were separately reverse transcribed using the cDNA Cycle kit (Invitrogen) with gene-specific primer 1 (Table I). After cDNA synthesis, 20  $\mu$ g of ribonuclease A were added to the reverse transcribed cDNA pool and incubated at 37°C for 30 min. Excess primer and deoxynucleotide triphosphates were removed by three rounds of spin filtration in a Centricon-30 (Amicon Corp., Scientific Sys. Div., Danvers, MA), in 5 mM Tris-HCl buffer at pH 8.0 with 0.5 mM EDTA. The sample was concentrated in a vacuum centrifuge and adjusted to 23  $\mu$ l with H<sub>2</sub>O. A terminal deoxynucleoti-

Table I. Gene Specific Primers Used for cDNA Synthesis and Polymerase Chain Reaction

Primer	Sequence	Length	Region	Use
		nt		
1	5'-GGCTTCATCCAGTAAAATACC-3'	21	431–411*	cDNA synthesis for 5' extension
2	5'-CCAGATCTGGTACTGCATTACTCTGC-3'	26	310–292**	5' End amplification
3	5'-CATCTAGAAGCTTCAGGTACAACCTGGC-3'	27	2593–2610 <sup>§</sup>	3' End amplification
4	5'-TAGGATCCGTCCTGTCCGGCACCAGC-3'	26	69–86 <sup>  </sup>	5' Primer for PCR for expression
5	5'-CAAGATCTGCCAGTTGTACCTGAAG-3'	25	2610–2594**	3' Primer for PCR for expression

Bold sequences were added for restriction enzyme sites. Numbered regions correspond to sequence shown in Fig. 9. nt, = nucleotides; \* Complementary to coding sequence. <sup>†</sup> Underlined sequence is BglIII restriction site. <sup>§</sup> Underlined sequence is XbaI restriction site. <sup>||</sup> Underlined sequence is BamHI restriction site.

dyl-transferase tailing reaction was then performed with dATP according to instructions from Promega Corp. (Madison, WI).

The resulting extended, tailed 5' end was amplified using the polymerase chain reaction (PCR). The reaction was performed in 100  $\mu$ l, containing the following in final concentrations: 1 $\times$  Taq DNA polymerase buffer (Promega Corp.); 200  $\mu$ M deoxynucleotide triphosphates; 0.25  $\mu$ M gene-specific primer 2; and 0.25  $\mu$ M oligo-d(T)<sub>17</sub> with restriction-site adapter. The primer used was different from that used for first-strand cDNA synthesis in order to increase the specificity for this gene and avoid amplification of tailed first-strand primer. The cDNA template was denatured at 95°C for 5 min followed by chilling on ice. It was centrifuged briefly, and then 2.5 U of Taq DNA polymerase (Promega Corp.) and two drops of mineral oil were added. A total of 35 cycles were performed, consisting of 94°C for 50 s, 54°C for 1 min, and 72°C for 1 min, with a final elongation step at 72°C for 5 min.

The 3' end was amplified by a similar procedure. Total RNA and poly-(A)<sup>+</sup> RNA from HeLa cells were reverse transcribed using an oligo-d(T) primer. After RNase digestion, the deoxynucleotide triphosphates were removed by using a G-50 spin column (5 Prime-3 Prime Inc.). The 3' end was amplified from the cDNA pool in a PCR using gene-specific primer 3 (Table I) and the oligo-d(T) primer.

The 5' and 3' extended cDNA products were analyzed by electrophoresis in 8% PAGE. To confirm that the products were specific, Southern blotting was performed (22), in which the gel products were electroblotted onto Hybond-N membrane (Amersham Corp., Arlington Heights, IL), and hybridization was performed with a <sup>32</sup>P-labeled cDNA insert as probe. Specific products were isolated from the gel by diffusion (14). A portion of the products were subcloned into the plasmid vector pUC18 after digestion of the 5' products with PstI and BglII, or digestion of the 3' products with XbaI and PstI. Other PCR products were subcloned into the plasmid vector PCR1000 (Invitrogen) according to manufacturer's instructions without previous treatment. All extended products were sequenced from a double-stranded DNA template with T7 DNA polymerase (Pharmacia LKB) as above.

**Expression of the recombinant protein.** Expression of the protein coded for by the sequence including the extended 5' end was achieved from a new cDNA prepared by reverse transcription from total HeLa

Table II. Comparison of Reaction in Immunoblot with Reaction with JH4 Anti-PM-Scl Clones for Anti-PM-Scl Sera

Serum no.	100 kD	70 kD	Other	JH4B <sub>1</sub>	JH4C <sub>1</sub>
1	+	+	+	4+	4+
2	+	+	+	3+	3+
3	+	+	-	3+	4+
4	+	+	-	3+	3+
5	+	+	-	4+	3+
6	+	+	+	2+	3+
7	+	+	+	3+	3+
8	+	+	-	4+	3+
9	+	+	+	4+	4+
10	+	+	+	4+	4+
11	+	+	+	3+	2+
12	+	+	+	3+	2+
13	+	+	+	2+	3+
14	+	+	+	2+	3+
15	+	+	+	2+	2+
16	+	+	-	4+	3+
17	+	-	-	1+	3+
18	+	-	-	4+	4+
19	+	-	-	4+	4+
20	+	-	-	-	1+
21	+	-	+	4+	4+
22	+	-	+	4+	3+
23	+	-	+	4+	4+
24	-	+	+	2+	1+
25	-	-	+	3+	2+
26	-	-	+	1+	1+
27	-	-	+	4+	3+
28	-	-	+	-	-
29	-	-	+	-	-
30	-	-	+	3+	4+
31	-	-	-	4+	3+
32	-	-	-	1+	2+
33	-	-	-	1+	2+
34	-	-	-	3+	4+
35	-	-	-	-	1+
36	-	-	-	-	1+
37	-	-	-	3+	3+
38	-	-	-	1+	1+
39	-	-	-	-	2+

All 39 sera were positive for anti-PM-Scl by immunodiffusion and were tested for reaction with the 100- and 70-kD bands on immunoblot against whole HeLa cell extract. Reaction in immunoblot was assessed as positive (+) or negative (-), whereas positive reaction against the two JH4 clones was estimated as 1+ through 4+. Other refers to reaction in immunoblot with bands other than the 100- and 70-kD bands.

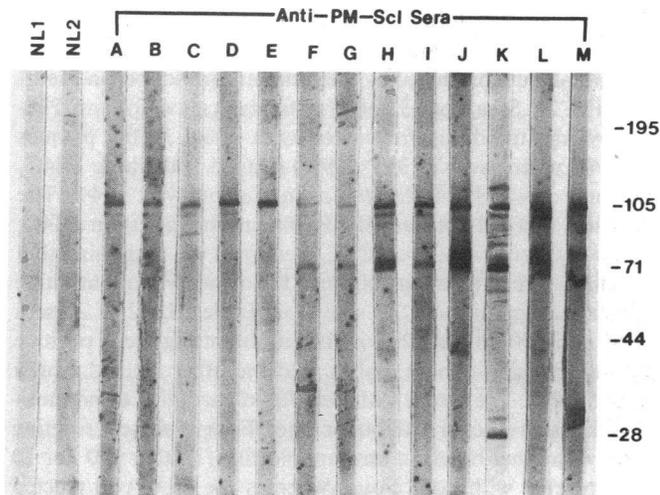


Figure 1. Immunoblot with anti-PM-Scl patient sera. Whole HeLa cell extract as antigen was electrophoresed in a 10% SDS-PAGE, transferred to nitrocellulose, and developed with anti-PM-Scl sera (A-M) or normal (NL) sera. The position of molecular mass markers in kD is shown at the right. The spots are artifacts appearing after original gel development. Lanes developed with selected sera are shown, illustrating the patterns observed: staining of the 100-kD band (A-E), staining of both the 100- and 70-kD bands (F-M), and staining of other bands (A, C, F, H-M).

cell RNA using an oligo-d(T) primer, followed by PCR amplification using gene-specific primer 4 (Table I) beginning 31 bp after the starting ATG with a BamHI site attached and gene-specific primer 5 beginning 8 bp upstream of the stop codon with a BglII site attached. Initial attempts using a primer beginning 5' to the starting site were unsuccessful.

Constructs of the pQE expression plasmids prepared using the QIAexpress kit from Qiagen, Inc. (Chatsworth, CA) were used as vector for protein expression. Purified PCR product was digested with

BamHI and BglII to generate compatible ends and subcloned into the type III pQE vector constructs. The vector was transformed into *E. coli* strain M15(pREP4). Colonies were selected by BamHI-BglII digestion of plasmid DNA and were analyzed for optimal protein expression.

**Other methods.** Ouchterlony immunodiffusion was performed by standard methods using concentrated calf thymus extract as antigen compared with a standard anti-PM-Scl serum (1) or a serum verified against the standard. Indirect immunofluorescence against HEP-2 cell substrate was performed using a commercial kit (INOVA Diagnostics, Inc., San Diego, CA), following the instructions of the manufacturer. Computer analysis of nucleotide and amino acid sequences was performed with the Sequence Analysis Software Package of the Genetics Computer Group of the University of Wisconsin (23), using a VAX 8250 computer, and also using a supercomputer through the Pittsburgh Supercomputing Facility as previously described (24). Tests for charged runs and clusters in protein sequences were performed using the PROPAT program (25), kindly provided by Dr. V. Brendel, Department of Mathematics, Stanford University, Stanford, CA. Protein motifs were examined using a database obtained from Dr. A. Bairoch, Centre Medical Universitaire, Geneva, Switzerland and adapted for use on a VAX computer by M. Barton Frank.

## Results

**Immunoblotting.** The 39 sera identified as positive for anti-PM-Scl were tested by immunoblotting against whole HeLa cell extract to determine the antigenic proteins for each serum (shown for 13 patient sera in Fig. 1). 23 sera (59%) stained a 100-kD band, and 16 of these 23 sera (41% of total) also stained a 70-kD band (Tables II and III). Only 1 serum stained the 70-kD band and not the 100-kD band, although 1 of the 23 that stained the 100-kD band stained the 70-kD band much more strongly. The 70- and 100-kD bands appeared to correspond to the 75- and 110-kD bands reported by others (9, 10). The 100-kD band appeared as a doublet in some runs, apparently due to degradation of the 100-kD band since the second band was not consistently present. 21 sera recognized bands other than the 100- and 70-kD bands, 6 of which were negative against the 100- and 70-kD bands. Among the other bands, bands of 30- and 32-kD were seen with 2 sera (K and M in Fig. 1); a 50-kD band with 3; and a weak, inconsistent 85-kD band with 10, which was similar to a band reported by others (10). Other bands were unique to individual sera. Nine sera had negative immunoblots. Staining of the 100- and 70-kD bands by all positive sera was weak in each of the several runs performed, even by sera with high anti-PM-Scl titers, but some of the other bands were stained strongly. The relationship of the other bands to the PM-Scl antigen complex remains unclear.

Table III. Summary of Anti-PM-Scl Sera Reactions

WB-100 kD	WB-70 kD	CLONE JH4C <sub>1</sub>	
		+	-
		%	%
+	+	16 (41)	0 (0)
+	-	7 (18)	0 (0)
-	+	1 (2.6)	0 (0)
-	-	13 (32.5)	2 (5.1)

Summary of reactions of all 39 sera found to be positive for anti-PM-Scl in immunodiffusion as described in the table above.

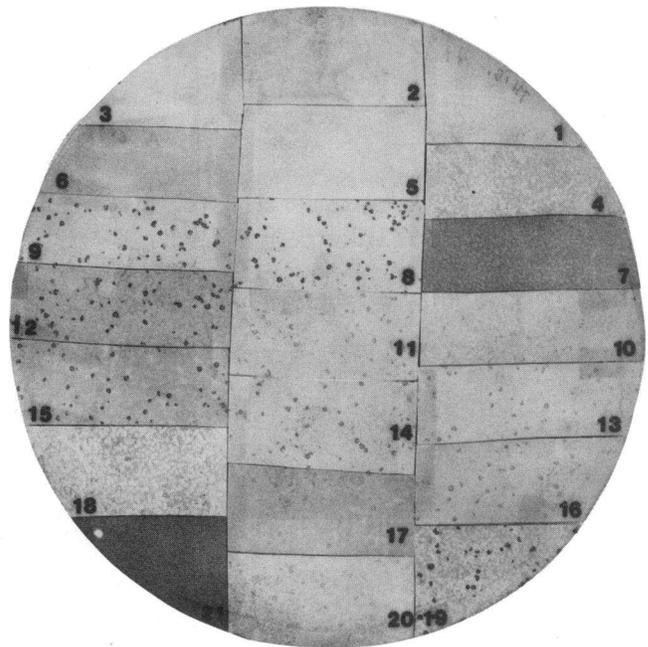


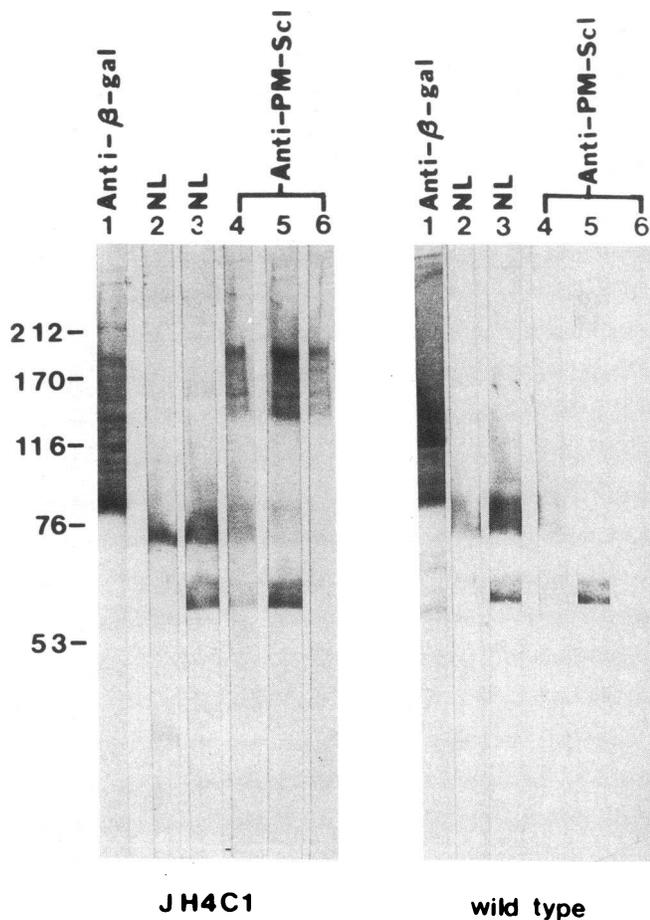
Figure 2. Testing of patient sera for immunoreactivity with expressed protein. *E. coli* was infected with a 5:1 mixture of wild-type and JH4B<sub>1</sub>  $\lambda$ gt11 and induced with IPTG. Plaque proteins were transferred to nitrocellulose, which was divided into segments, and developed with different sera. Sera used for development: 1 and 2: normal sera; 3: no serum; 4-7: myositis sera with autoantibodies other than anti-PM-Scl; 8-21: anti-PM-Scl sera (identified by immunodiffusion). In segments developed with anti-PM-Scl, approximately one fifth of the plaques present were positively stained whereas segments developed with other sera showed no positive plaques.

**Isolation of PM-Scl cDNA clone.** Two  $\lambda$ gt11 clones that persistently produced immunoreactive plaques, designated JH4B<sub>1</sub> and JH4C<sub>1</sub>, were identified from the thymocyte library, and none were identified from the HeLa cell library. Multiple anti-PM-Scl positive and negative sera were tested for reactivity with plaques produced by these clones as described in Methods (Fig. 2). Sera from 33 of 39 patients positive for anti-PM-Scl by immunodiffusion reacted with clone JH4B<sub>1</sub> plaques (84.6%) whereas 37 of 39 (94.9%) reacted with clone JH4C<sub>1</sub> plaques, of which 26 (66.7%) stained strongly at 3-4+. The four sera that did not stain JH4B<sub>1</sub> plaques but did stain JH4C<sub>1</sub> plaques generally stained JH4C<sub>1</sub> weakly (1+), suggesting that the difference in reactivity between the clones was quantitative rather than qualitative. As shown in Table III, 22 of 23 sera with anti-100-kD activity in immunoblotting were positive against plaques of both clones, and the other reacted faintly with JH4C<sub>1</sub>. The serum with anti-70-kD activity but not anti-100-kD also reacted with both clones. Four of six sera reacting only with other bands in immunoblotting (neither 100 nor 70 kD) reacted with both clones whereas the other two reacted with neither clone, the only 2 of the 39 tested that were negative against JH4C<sub>1</sub>. Six of nine sera that were negative in immunoblot reacted with both clones, and the other three reacted weakly with JH4C<sub>1</sub>. None of the sera from 20 patients with other autoantibodies, nor sera from 6 normal controls reacted with plaques of either clone (not shown).

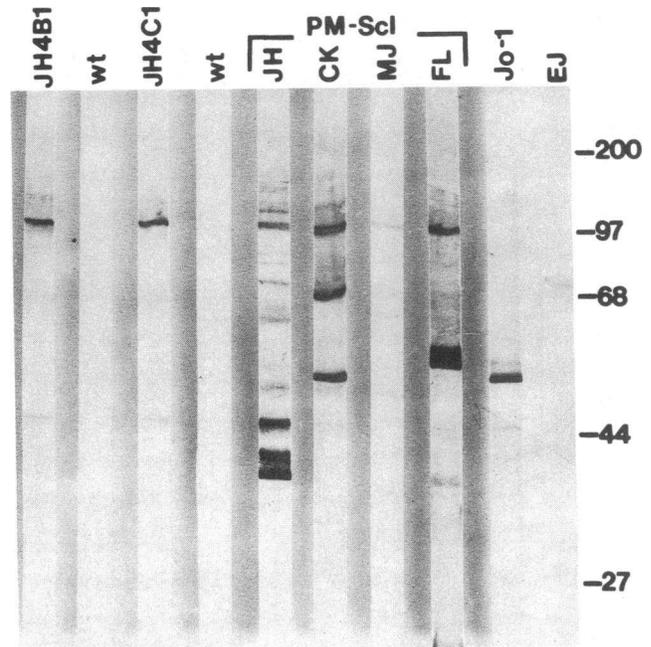
To confirm that this reactivity was with fusion protein, further testing in immunoblotting was performed. Reaction against a lysate of *E. coli* produced by the JH4C<sub>1</sub>  $\lambda$ gt11 bacterio-

phage clone was compared with reaction with a lysate produced by wild-type  $\lambda$ gt11 (Fig. 3). A protein of  $\sim 200$  kD was recognized in JH4C<sub>1</sub> lysate by all three anti-PM-Scl tested but was not found in wild-type lysate. The two normal sera did not recognize any proteins in this region from either lysate. Thus, proteins carrying antigenic determinants recognized specifically by most anti-PM-Scl sera were expressed by these clones, suggesting that the inserts coded for an antigenic protein of PM-Scl.

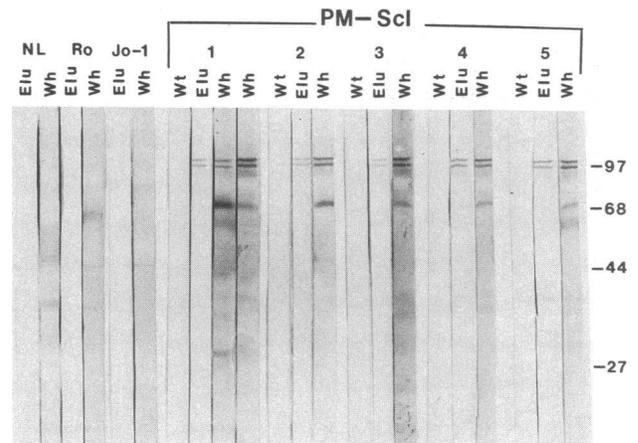
**Confirmation of the identification of the clones.** The immunological specificity of the expressed protein of the two clones was further confirmed by testing affinity-purified antibody prepared, as described in Methods, by elution from plaque proteins of the JH4 clone or wild-type  $\lambda$ gt11 after they had been incubated with anti-PM-Scl serum. Affinity-purified antibody from anti-PM-Scl positive JH serum using either of the JH4 clones showed strong activity against plaques of both clones, demonstrating that the recombinant proteins of the two clones shared antigenic determinants with each other, and also confirming that the affinity-purified antibody remained immunologically active. Eluates from wild-type plaques after incuba-



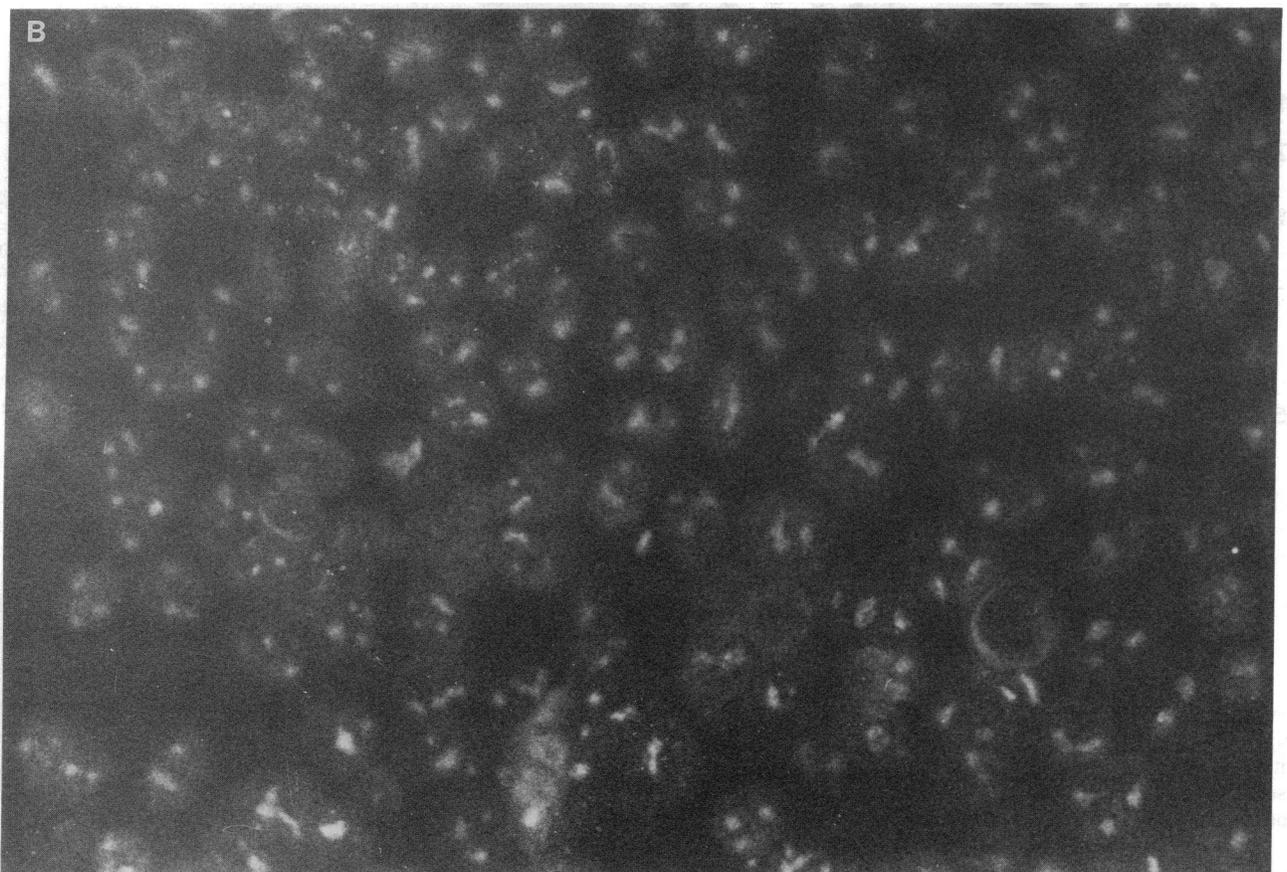
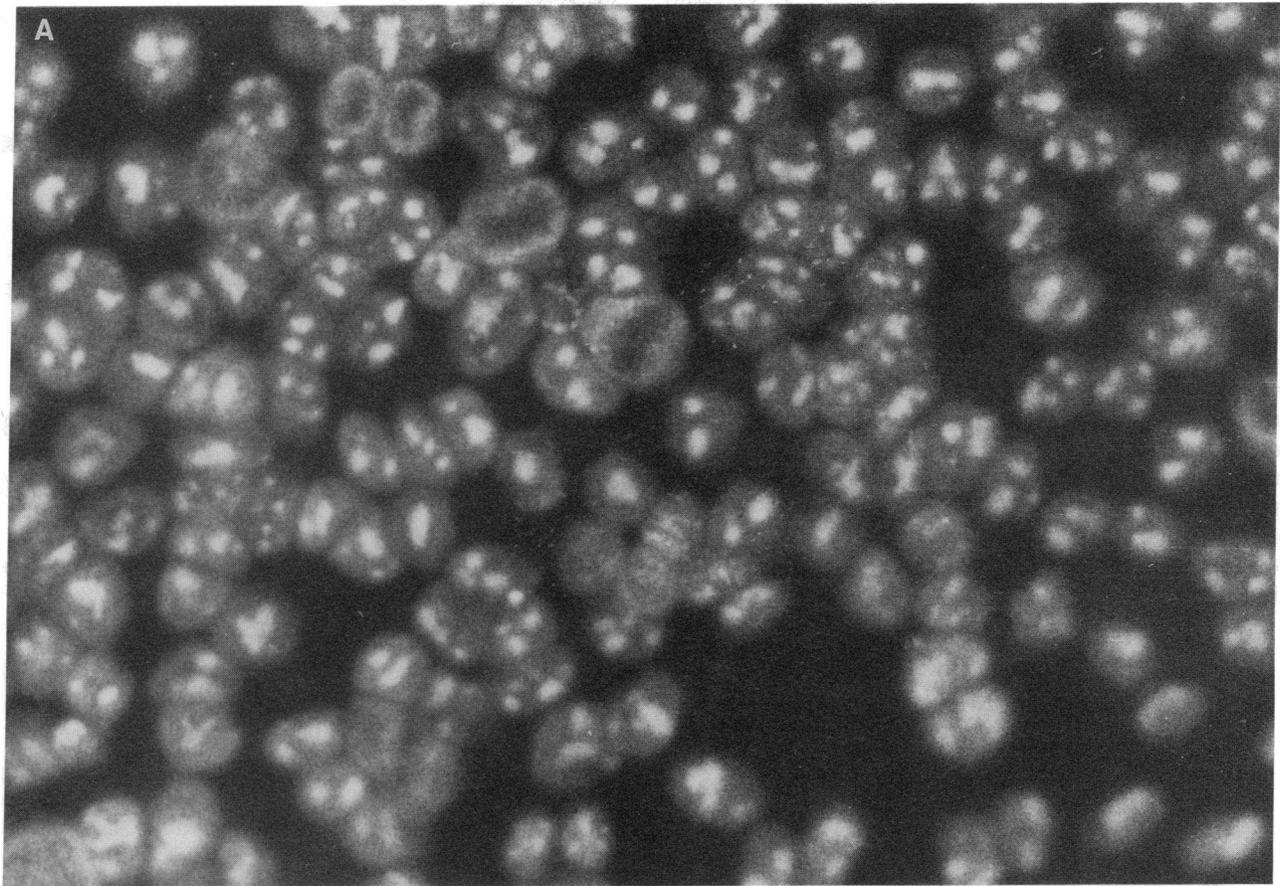
**Figure 3.** Immunoblot of *E. coli* lysates. Lysates of *E. coli* prepared either with JH4C<sub>1</sub>  $\lambda$ gt11 (left) or wild-type (right) were electrophoresed in a 10% SDS-PAGE and developed with the sera indicated: 1, mouse anti- $\beta$ -galactosidase (developed with anti-mouse conjugate); 2 and 3, normal human sera; 4-6, anti-PM-Scl sera. The position of molecular mass markers in kilodaltons is shown at the left.



**Figure 4.** Immunoblot with affinity-purified antibodies. Whole HeLa cell extract as antigen was electrophoresed in a 10% SDS-PAGE in all lanes, transferred to nitrocellulose, and developed with affinity-purified antibodies from plaques of JH4B<sub>1</sub> clone, JH4C<sub>1</sub> clone, or wild-type (*wt*)  $\lambda$ -gt11, anti-PM-Scl sera (JH, CK, MJ, FL), or control sera with anti-Jo-1 (*Jo-1*) or anti-EJ (*EJ*). Although whole JH serum stains multiple bands of HeLa cell extract, affinity-purified antibodies from either clone stain only the 100-kD protein. The position of molecular mass markers in kilodaltons is shown at the right.



**Figure 5.** Immunoblot with affinity-purified antibodies against anti-PM-Scl immunoprecipitates. Immunoprecipitates prepared from HeLa cell extract using JH serum were electrophoresed in all lanes of a 10% SDS-PAGE and transferred to nitrocellulose. The positions of molecular mass markers in kilodaltons are shown to the right. Each strip was developed with either whole serum (*Wh*) at 1:00 dilution or affinity-purified antibodies prepared from that serum by elution from plaques of JH4B<sub>1</sub> clone (*Elu*) or wild-type  $\lambda$ gt11 (*Wt*). The serum in each case is indicated in the upper line as: NL, normal serum; Ro, anti-Ro/SSA; Jo-1, anti-Jo-1; and anti-PM-Scl sera, labeled 1-5. For serum 1 (JH), lane Wh\* was developed with serum that had been incubated with plaques of JH4B<sub>1</sub>, showing decreased staining of the 100-kD doublet, without decrease in the staining of the 70-kD band.



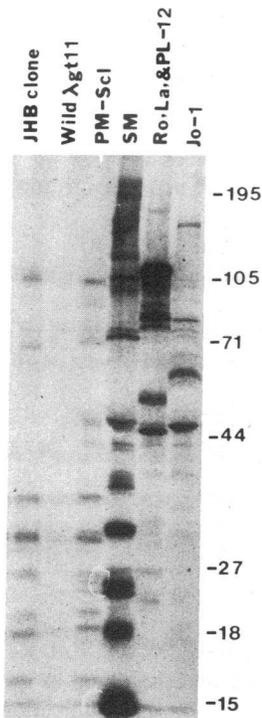
*Figure 6.* Indirect immunofluorescence on HEp-2 cells. (A) Whole JH serum used at 1:00 dilution. (B) Affinity-purified antibody (antibody from JH4B<sub>1</sub> plaques) used undiluted. Antibody from wild-type plaques was completely negative (not shown).

tion with JH serum were negative against the plaques of both clones.

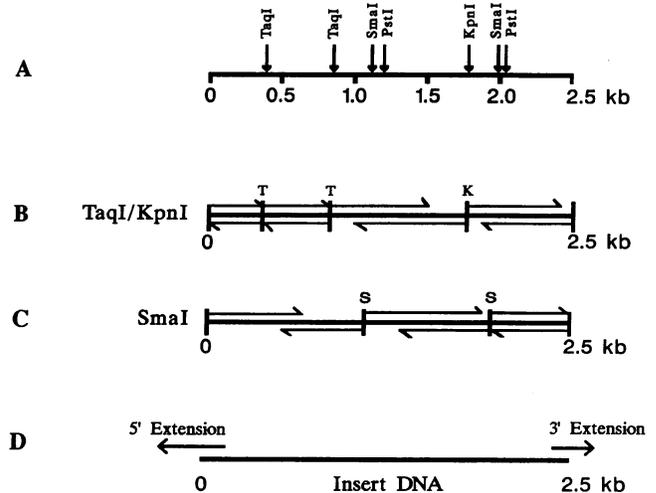
Affinity-purified antibody from JH serum using clone JH4B<sub>1</sub> stained a 100-kD protein in immunoblotting against whole HeLa cell extract, similar in size to the 100-kD protein stained by whole JH serum (Fig. 4). However, other proteins stained in immunoblotting by whole JH serum, including the 70-kD protein, were not seen with affinity-purified antibody. Wild-type eluates were negative in immunoblotting.

Further testing was performed to determine whether this result was typical of other anti-PM-Scl sera. Anti-PM-Scl immunoprecipitates from HeLa cell extract prepared using JH serum were used as antigen rather than whole HeLa cell extract. Whole JH serum recognized a 100-kD doublet and the 70-kD band whereas affinity-purified antibody from JH serum recognized only the 100-kD doublet (Fig. 5). After incubation with JH4B<sub>1</sub> plaque proteins, staining of the 100-kD doublet by JH serum was substantially decreased but staining of the 70-kD band was not. Antibody was similarly affinity purified from four of the other anti-PM-Scl sera that had reacted with both the 100- and 70-kD bands, using both JH4B<sub>1</sub> and wild-type  $\lambda$ gt11. In each case, the affinity-purified antibody from the JH4 clone reacted with the 100-kD doublet of the immunoprecipitate but not the 70-kD band. The wild-type eluates were all negative (Fig. 5).

JH serum antibodies affinity purified from the JH4B<sub>1</sub> clone were also found to stain the nucleoli of HEp-2 cells, in a pattern similar to that seen with whole JH serum (Fig. 6) whereas wild-type eluates were negative. Affinity-purified antibody was also active in <sup>35</sup>S-immunoprecipitation. In addition to the 100-kD band, affinity-purified antibody immunoprecipitated the smaller bands (under 40 kD) typical of the immunoprecipitation pattern of anti-PM-Scl sera, apparently due to their affinity for the 100-kD band as part of the PM-Scl complex (Fig. 7).



**Figure 7.** Immunoprecipitation with affinity-purified antibody. Immunoprecipitates from <sup>35</sup>S-methionine-labeled HeLa cell extracts using anti-PM-Scl serum (*PM-Scl*), affinity-purified antibody eluted from JH4B<sub>1</sub> (*JHB clone*), wild-type eluate (*Wild λgt11*), or control sera with anti-Sm, anti-PL-12 with anti-Ro/SSA and anti-La/SSB, and anti-Jo-1, were electrophoresed in a 10% SDS-PAGE and autoradiographed. The position of molecular mass markers in kD is shown at the right. The entire PM-Scl complex seen with whole anti-PM-Scl serum is immunoprecipitated by the JH4B<sub>1</sub> eluates, except for the artifactual bands surrounding the 44-kD marker. The PM-Scl pattern is very faintly seen with control eluates but was much stronger with JH4B<sub>1</sub>.



**Figure 8.** Sequencing strategy for cDNA encoding the PM-Scl 100-kD protein. Bar A shows a restriction enzyme map of the cleavage sites on the JH4B<sub>1</sub> cloned cDNA. In bars B and C, the arrows show the maximum sequenced length and direction for each fragment. Bar D shows the extended ends.

Antibody eluted from wild-type plaques also showed this pattern, apparently because of background binding of patient antibody to the plaques, but it was much weaker than that seen with antibody eluted from the clone. These studies indicated that anti-100-kD anti-PM-Scl antibodies reacted with the protein expressed by the JH4 clones.

**Sequencing of PM-Scl cDNA.** The DNAs of clones JH4B<sub>1</sub> and JH4C<sub>1</sub> were digested with EcoRI. Agarose gel electrophoresis of the products showed a similar sized single insert of ~ 2.5 kb for each clone. Inserts from both clones were subcloned into the EcoRI site of pUC18. Partial sequencing from the double-stranded plasmid DNA showed that both clones had identical terminal sequences.

A restriction map of the JH4B<sub>1</sub> insert was obtained by single and double digestions with selected restriction enzymes (Fig. 8). Four fragments were obtained by digestion with KpnI and TaqI, and three fragments were obtained following SmaI digestion. All seven fragments of JH4B<sub>1</sub> were then isolated from agarose gels, subcloned into M13mp18, and sequenced. Subclones from each of these strategies were used to verify the positions of the five restriction sites shown in Fig. 8.

The sequence of both strands of the JH4B<sub>1</sub> insert was thus determined. Including EcoRI linkers (not shown), it contained 2,479 nucleotides, with a single long open-reading frame of 2,331 nucleotides from the first ATG codon at position 288 in Fig. 9 to a stop codon at position 2,619. The predicted protein from the insert cDNA included 777 amino acids, with a predicted molecular mass of 89 kD, smaller than the observed mass of the antigenic protein in this study (100 kD) or that reported in other studies (110 kD). No poly-(A) tail was detected in the cDNA insert.

**Northern blot analysis.** Northern blot against both total RNA and poly-(A)<sup>+</sup> RNA from HeLa cells showed hybridization of <sup>32</sup>P-labeled JH4B<sub>1</sub> cDNA insert with a single band of 2.7 kb (Fig. 10). Similar results showing hybridization of a single 2.7-kb band were obtained using total RNA from an SKHep1 cell line from human liver adenocarcinoma (kindly provided by Dr. P. Kincade, Oklahoma Medical Research Foundation)

1 **GACAAGCTCTCGCGAGACGAGCCGTCAGGCTGAAAAAATGGCGCCACCCAGTACCCGGGAGCCCAGGGTCTCTGTCGGG**  
MetAlaProProSerThrArgGluProArgValLeuSerAla 14

81 **ACCAGCGCAACCAATCCGACGGAGAGATGGTGCTGCCAGGCTTCCCGGACGCCGACAGCTTTGTGAAGTTTGTCTTGGG**  
ThrSerAlaThrLysSerAspGlyGluMetValLeuProGlyPheProAspAlaAspSerPheValLysPheAlaLeuGly 41

162 **TCCGTGGTGGCAGTCACCAAGGCATCTGGGGCCCTACCACAGTTTGGCGATGAGTATGATTTTTACCGAAGTTTTCTGGC**  
SerValValAlaValThrLysAlaSerGlyGlyLeuProGlnPheGlyAspGluTyrAspPheTyrArgSerPheProGly 68

243 **TTCCAAGCATTTTGCAGAACACAGGGAGACAGGTTGCTTTCAGTGCATGAGCAGAGTAATGCAGTACCATGGGTGTGCGAGC**  
PheGlnAlaPheCysGluThrGlnGlyAspArgLeuLeuGlnCysMetSerArgValMetGlnTyrHisGlyCysArgSer 95

324 **AACATTAGGATCGAAGTAAAGTGACTGAGCTGGAAGACAAGTTTGATTTACTAGTTGATGCCAATGATGTAATTCTGGAG**  
AsnIleLysAspArgSerLysValThrGluLeuGluAspLysPheAspLeuLeuValAspAlaAsnAspValIleLeuGlu 122

405 **AGAGTGGGTATTTTACTGGATGAAGCCTCAGGTGTAAACAAGAATCAACAGCCTGTCCTCCCTGCCGGCTTGCAGTCCCC**  
ArgValGlyIleLeuLeuAspGluAlaSerGlyValAsnLysAsnGlnGlnProValLeuProAlaGlyLeuGlnValPro 149

486 **AAAACGGTAGTGTCCAGCTGGAACCGTAAGGCAGCAGAATATGGCAAAAAGCAAATCTGAACTTTCGGCTGCTTCAT**  
LysThrValValSerSerTrpAsnArgLysAlaAlaGluTyrGlyLysLysAlaLysSerGluThrPheArgLeuLeuHis 176

567 **GCAAAAAATATCATCCGACCTCAGCTCAAGTTTCGAGAGAAGATTGACAATCCAACACACCATTCTTCTTAAATCTTC**  
AlaLysAsnIleIleArgProGlnLeuLysPheArgGluLysIleAspAsnSerAsnThrProPheLeuProLysIlePhe 203

648 **ATCAAACCCAAATGCTCAGAAACCTCTCCCTCAAGCTCTCTTAAGGAAAGCGGGAACGCCACAGGATCGTCTGAGGAC**  
IleLysProAsnAlaGlnLysProLeuProGlnAlaLeuSerLysGluArgArgGluArgProGlnAspArgProGluAsp 230

729 **TTGGACGTCCCCCTGCACTGGCTGATTTTCATCCATCAGCAGAGAACCAGCAGGTTGAGCAAGACATGTTTGCACATCCT**  
LeuAspValProProAlaLeuAlaAspPheIleHisGlnGlnArgThrGlnGlnValGluGlnAspMetPheAlaHisPro 257

810 **TATCAATATGAACTAAATCACTTACCCAGCAGATGCAGTGCTTCAAAGCCACAACCCAGTTATACAGACCTATAGAA**  
TyrGlnTyrGluLeuAsnHisPheThrProAlaAspAlaValLeuGlnLysProGlnProGlnLeuTyrArgProIleGlu 284

891 **GAGACACCATGCCATTTTCATATCCTCCCTGGATGAACTCGTGGAATCAACGAAAAGCTCTTGAATTGTCAGGAATTTGCA**  
GluThrProCysHisPheIleSerSerLeuAspGluLeuValGluLeuAsnGluLysLeuLeuAsnCysGlnGluPheAla 311

972 **GTTGACTTGGAGCACCCTCTTACAGGAGCTTCTGGGACTGACCTGCCTGATGCAAATTTCTACTCGGACGGAAGACTTC**  
ValAspLeuGluHisHisSerTyrArgSerPheLeuGlyLeuThrCysLeuMetGlnIleSerThrArgThrGluAspPhe 338

1053 **ATCATTGACACCCCTCGAGCTTCGAAGTGACATGTACATTCTCAATGAGAGCCTCACAGACCCAGCCATCGTTAAGGTCTTT**  
IleIleAspThrLeuGluLeuArgSerAspMetTyrIleLeuAsnGluSerLeuThrAspProAlaIleValLysValPhe 365

1134 **CATGGTGTGATTGAGACATAGAATGGCTACAGAAAGACTTTGGGTGTATGTAGTAAACATGTTTGATACTCATCAGGCA**  
HisGlyAlaAspSerAspIleGluTrpLeuGlnLysAspPheGlyLeuTyrValValAsnMetPheAspThrHisGlnAla 392

1215 **GCACGCCTTCTTAACCTGGGCAGGCACTCACTCGATCATCTCTGAACTCTACTGCAACGTGGACTCAAACAAGCAATAT**  
AlaArgLeuLeuAsnLeuGlyArgHisSerLeuAspHisLeuLeuLysLeuTyrCysAsnValAspSerAsnLysGlnTyr 419

1296 **CAGCTGGCTGATTGGAGAATACGCCCTCTGCCCGAGGAGATGCTCAGCTACGCCCGGATGACACCCATTACCTGCTATAT**  
GlnLeuAlaAspTrpArgIleArgProLeuProGluGluMetLeuSerTyrAlaArgAspAspThrHisTyrLeuLeuTyr 446

1377 **ATCTATGACAAAATGAGGCTGGAGATGTGGGAGCGCGCAACGGGCAGCCCGTGCAGCTGCAGGTGGTGTGGCAACGGAGC**  
IleTyrAspLysMetArgLeuGluMetTrpGluArgGlyAsnGlyGlnProValGlnLeuGlnValValTrpGlnArgSer 473

Figure 9. The complete nucleotide and predicted amino acid sequence of the PM-Scl 100-kD protein. The sequence in regular type was derived from the JH4B<sub>1</sub> cDNA whereas the sequences in bold were derived from the extended ends. The asterisk indicates the stop codon. The superscripted dots mark every 10th nucleotide. The underline indicates the polyadenylation signal. Nucleotide position is shown at the left and amino acid position is shown on the right.

(not shown). This indicated that the JH4 cDNA inserts, at < 2.5 kb, included most but not all of the full length of the corresponding mRNA.

*Determination of the 5' and 3' ends of the full-length cDNA.* The ends of the cDNA were extended and amplified as described in Methods. The products of 5' end amplification

1458 AGGGACATCTGCCTCAAGAAATTCATCAAACCTATCTTCACGGATGAGTCCTACCTTGAACCTCTATAGGAAGCAGAAGAAG  
 ArgAspIleCysLeuLysLysPheIleLysProIlePheThrAspGluSerTyrLeuGluLeuTyrArgLysGlnLysLys 500

1539 CACCTTAACACACAGCAGTTGACAGCCTTTCAGCTGCTGTTGCTGGAGGGATAAAACAGCTCGCAGGGAAGATGAAAGT  
 HisLeuAsnThrGlnGlnLeuThrAlaPheGlnLeuLeuPheAlaTrpArgAspLysThrAlaArgArgGluAspGluSer 527

1620 TACGGATATGTACTGCCAAACCACATGATGCTGAAAATAGCTGAAGAACTGCCTAAGGAACCTCAGGGCATCATAGCTTGC  
 TyrGlyTyrValLeuProAsnHisMetMetLeuLysIleAlaGluGluLeuProLysGluProGlnGlyIleIleAlaCys 554

1701 TGCAACCCAGTACCGCCCCTTGTGCGGCAGCAGATCAACGAAATGCACCTTTTAAATCCAGCAGGCCGAGAGATGCCCTG  
 CysAsnProValProProLeuValArgGlnGlnIleAsnGluMetHisLeuLeuIleGlnGlnAlaArgGluMetProLeu 581

1782 CTCAAGTCTGAAGTTGACGCCGAGTGAAGAAGAGCGGACCGCTGCCAGTGCTGAGAGATTGGAGAATGTTCTCTTTGGA  
 LeuLysSerGluValAlaAlaGlyValLysLysSerGlyProLeuProSerAlaGluArgLeuGluAsnValLeuPheGly 608

1863 CCTCAGACTGCTCCCATGCCCTCCGGATGGCTATCCAATCATCCAACAGTGGATCTGTGCCAGTTCAGAAGCAGGCG  
 ProHisAspCysSerHisAlaProProAspGlyTyrProIleIleProThrSerGlySerValProValGlnLysGlnAla 635

1944 AGCCTCTCCCTGATGAAAAAGAAGATAACTTGCTGGGTACCACATGCCTGATTGCCACAGCTGTCATCACGTTATTTAAT  
 SerLeuPheProAspGluLysGluAspAsnLeuLeuGlyThrThrCysLeuIleAlaThrAlaValIleThrLeuPheAsn 662

2025 GAACCTAGTGTGAAGACAGTAAAAAGGGTCCATTGACAGTTGCACAGAAAAAGCCAGAACATCATGGAGTCTTTGAA  
 GluProSerAlaGluAspSerLysLysGlyProLeuThrValAlaGlnLysLysAlaGlnAsnIleMetGluSerPheGlu 689

2106 AATCCATTTAGGATGATCAGCAACCGTTGGAAGCTGGCCAGGTACAAGTACAAAAAGACTCTAAAGAAGCTGTCAAGAAG  
 AsnProPheArgMetIleSerAsnArgTrpLysLeuAlaGlnValGlnValGlnLysAspSerLysGluAlaValLysLys 716

2187 AAGGCAGCTGAGCAAACAGCTGCCCGGAACAGGCAAAGGAGGCGTGCAAAGCTGCAGCAGAACAGGCCATCTCCGTCCGA  
 LysAlaAlaGluGlnThrAlaAlaArgGluGlnAlaLysGluAlaCysLysAlaAlaAlaGluGlnAlaIleSerValArg 743

2268 CAGCAGTCTGCTAGAAAATGCTGCAAAGAAGAGAGCGAGCAACAAGCGACCCAAGGACCACAGAACAGAAACAAGAG  
 GlnGlnValValLeuGluAsnAlaAlaLysLysArgGluArgAlaThrSerAspProArgThrThrGluGlnLysGlnGlu 770

2349 AAGAAACGACTCAAATTTCCAAGAAGCCAAAGGCCAGAGCCACCAGAAAAAGAGTTTACGCCTTACGACTACAGCCAG  
 LysLysArgLeuLysIleSerLysLysProLysAspProGluProProGluLysGluPheThrProTyrAspTyrSerGln 797

2430 TCAGACTCAAGGCTTTTGTGGAACAGCAAATCCAAAGTTTCTTCTCAGTTTGATCCAAATAAACAGACCCCGTCTGGC  
 SerAspPheLysAlaPheAlaGlyAsnSerLysSerLysValSerSerGlnPheAspProAsnLysGlnThrProSerGly 824

2511 AAGAAATGCATTGCAGCCAAAAAATTAACAGTCGGTGGAAACAAAAGCATGTCCCTTCCAAGTGGAAAGTCAGACAGA  
 LysLysCysIleAlaAlaLysLysIleLysGlnSerValGlyAsnLysSerMetSerPheProThrGlyLysSerAspArg 851

2592 GGCTTCAGGTACAACCTGGCCACAGAGATAGTCTCGAAGACACGTGGCGCTGTGGACCGGAAGCACCAATGCTGGTGCT  
 GlyPheArgTyrAsnTrpProGlnArg \* 860

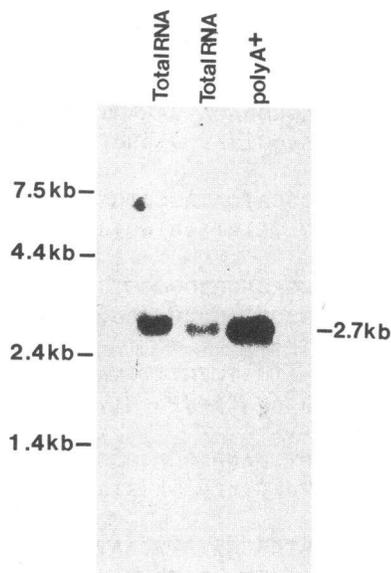
2673 GCTTTGTACATACATATTTTTAAACCATTAAAATCTTCCTGAAGAAAAAAAAAAAAAAAAAAAAA

Figure 9 (Continued)

showed a predominant band ranging from 320 to 340 bp, whether total or poly-(A)<sup>+</sup> RNA was used for the template for cDNA synthesis. In some PCR amplifications using the same cDNA as template, analysis of the products showed multiple bands in addition to the predominant 340-bp band. However, after electroblotting of electrophoresed PCR products onto a nylon membrane, only the 340-bp band hybridized with <sup>32</sup>P-labeled JH4B<sub>1</sub> insert.

To identify the longest extensions present, a range of products were isolated from the polyacrylamide gel by diffusion. Aliquots of PCR products were subcloned into pUC18 or PCR1000 (see Methods). 19 subclones were obtained and 11

were selected and sequenced from plasmid mini-preps. These 11 were derived from 5 different PCRs that were performed using cDNA pools from three reverse transcription reactions. The sequences showed that all of the clones started from the gene-specific primer, contained the 88 bp of known sequence from the JH4B<sub>1</sub> cDNA (discounting the EcoRI linker), extended 180–210 nucleotides beyond the 5' end of the cDNA, and then showed the poly-(T) expected from the tailing reaction. The first initiation codon of the maximum-length-extended 5' portion was at position 39 (Fig. 9), outside the original cDNA. All of the 5' extension subclones contained this codon. Therefore, all PCR-produced 5' ends gave the same



**Figure 10.** Northern blot of HeLa cell RNA. 40  $\mu$ g (left) or 20  $\mu$ g (middle) of total HeLa RNA, or 8  $\mu$ g of isolated poly-(A)<sup>+</sup> HeLa RNA, were electrophoresed in a 1% agarose gel, transferred to Nytran membrane, and hybridized with <sup>32</sup>P-labeled PM-Scl 100-kD cDNA insert isolated from clone JH4B<sub>1</sub>. A single band was identified in all three preparations. The position of size markers is shown at the left.

putative translational starting site, which differed from that predicted from the initial cDNA insert by 249 nucleotides (83 amino acids). None of these subclones contained an in-frame stop codon upstream of the first ATG.

Two slight discrepancies were noted among the 11 sequenced 5' extensions: 1 of 11 showed a change from a T to a C at position 96 of the 5' extension whereas another showed a change from a G to an A at position 191. Although it is possible that one of these represents a different allele, their low frequency suggests that they more likely represent mutant Taq polymerase products.

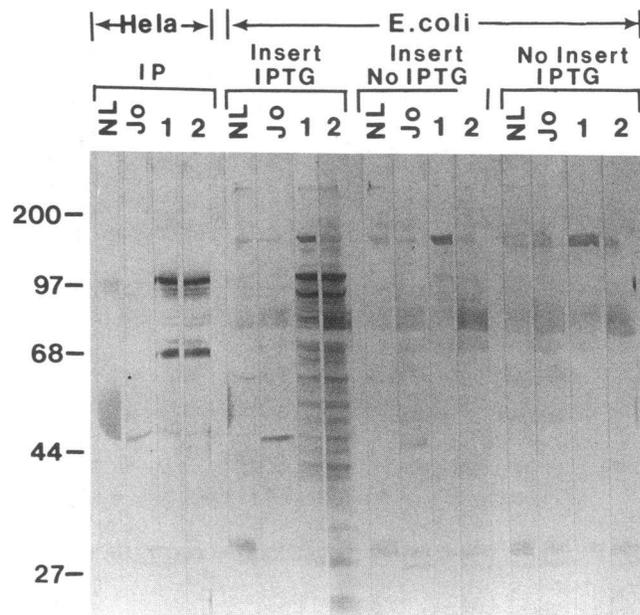
Although a stop codon was present in the open-reading frame of the cDNA insert (Fig. 9, position 2,619), the poly-(A) tail was not present, and therefore the 3' end was amplified as well. PAGE of PCR amplification products of the 3' extensions showed a predominant band at 180 bp, but multiple other bands were identified. Southern blotting and hybridization with a <sup>32</sup>P-labeled cDNA insert, however, showed hybridization only with the 180-bp band. An aliquot of PCR products was digested with XbaI and PstI and subcloned into pUC18 and PCR1000. Eight subclones were selected for sequencing. The sequences of all of the clones started from the gene-specific primer 99 bp upstream from the 3' end of the original cDNA insert, included the 90 bp of known sequence up to the EcoRI linker, extended 32 nucleotides further, and ended with a poly-(A) tail that usually contained ~ 20 bp. Two of the eight sequenced 3' extensions showed an absence of the three nucleotides 2,716–2,718 just upstream of the poly-(A) tail. A putative polyadenylation signal, ATTTAAA, was located 19 nucleotides from the poly-(A) tail within the extended region.

**Expression of the recombinant protein.** *E. coli* M15-(pREP4) was transformed with type III constructs of pQE plasmid containing PCR-amplified insert coding for the PM-Scl 100-kD protein as described in Methods. The expressed protein lacked the amino-terminal 10 amino acids and the carboxyl-terminal 3 amino acids of the predicted protein, but the primers and the expression system added 15 unrelated amino acids, so that the expressed protein was predicted to be 2 amino acids longer and 381 D greater in molecular mass than the native protein. Anti-PM-Scl sera recognized a protein in lysate

of induced *E. coli* transformed with vector containing this insert, but not in controls without insert or with recombinant plasmids without IPTG induction (Fig. 11). The mobility of the *E. coli*-expressed protein in PAGE was slightly larger, by ~ 2–4 kD, than the native HeLa cell immunoprecipitated 100-kD protein.

**Analysis of the full-length sequence.** The full-length nucleotide sequence of this PM-Scl gene including the extended 5' and 3' ends was 2,739 bp long. This sequence consists of 38 nucleotides in the 5' untranslated region, 2,580 bp in the predicted coding region from the first methionine codon, and 100 nucleotides from the stop codon up to the 21 nucleotide poly-(A) tail (Fig. 9). The deduced amino acid sequence was composed of 860 residues with a predicted molecular mass of 98,088 D, very close in size to the observed relative molecular weight in PAGE of the PM-Scl 100-kD antigen.

Several possible sites of posttranslational modification were observed. Two potential *N*-glycosylation sites were identified at residues 353 and 839. A number of potential phosphorylation sites were identified, including those for protein kinase C (at residues 849, 823, and 846), casein kinase II (at residues 18, 104, 292, 293, 333, 370, and 665), and tyrosine kinase (at 523). There were also amidation sites identified at residues 163 and 823. However, the presence of these sites does not necessar-



**Figure 11.** Immunoblot of expressed recombinant protein. Immunoprecipitate from HeLa lysate using serum JH (*HeLa*) or lysates of recombinant *E. coli* were electrophoresed in a 10% SDS-PAGE, transferred to nitrocellulose, and developed with either normal serum (*NL*), anti-Jo-1 serum (*Jo*), or one of two anti-PM-Scl sera (lanes 1 and 2). Insert/IPTG: *E. coli* transformed with plasmid vector containing the PCR-amplified PM-Scl 100-kD insert, induced with IPTG; Insert/no IPTG: *E. coli* transformed with plasmid containing insert but not induced; No Insert/IPTG: *E. coli* transformed with plasmid without insert and induced. The 100-kD band recognized in immunoprecipitates by anti-PM-Scl sera (near the 97-kD marker) was similar in size to the full-length expressed protein. Multiple weaker additional bands of expressed protein reacting with anti-PM-Scl appear to represent incomplete, partial-length expressed protein or degradation products.

ily indicate that posttranslational modification occurs at these sites, and they are often observed in proteins that are not modified. There has been no evidence of posttranslational processing of the PM-Scl 100-kD protein reported.

One report indicated that a number of lupus and scleroderma autoantigens showed a higher frequency than other proteins of charge runs (a long series of consecutive charged amino acids) and charge clusters (a high frequency of charged amino acids, not necessarily consecutive, within a given stretch of amino acids) (25). Although no charge runs were observed in this PM-Scl protein, a "mixed charge cluster" (consisting of both positive and negative residues) was found between residues 753 and 789, with 14 positive and 8 negatively charged residues among 37 amino acids, for a total of 59% charged amino acids within this region.

There was no significant homology with the sequence of the 75-kD PM-Scl protein (10), and there were no sequences of more than four consecutive amino acids shared between the two proteins. Independence of the two major PM-Scl antigenic proteins on the amino acid level is consistent with the observation noted above that they were antigenically independent, since affinity-purified JH4B<sub>1</sub> eluates stained the 100-kD protein but not the 70-kD protein. Alderuccio et al. (10) had similarly noted specific reactivity with the expressed 75-kD protein without cross-reaction with the 110-kD protein. No significant homology with other eukaryotic proteins in the data banks was identified.

## Discussion

In this study, the full-length nucleotide and predicted amino acid sequence for the 100-kD antigenic protein of the PM-Scl antigen was determined. The 100-kD protein was found to be antigenic for the great majority of patients with anti-PM-Scl antibodies, even if no activity was seen on routine immunoblotting. This protein appears to be independent of the 70-kD protein immunologically, consistent with the lack of primary amino acid sequence similarity between their predicted polypeptides.

Strong support was obtained for the conclusion that the cloned cDNA encodes a protein that reacts with anti-PM-Scl antibodies. First, extensive serum testing demonstrated that reaction with recombinant protein was specific for sera with anti-PM-Scl. No reaction was seen with sera containing other autoantibodies or normal human sera. Second, the affinity-purified antibodies that reacted with the fusion protein demonstrated similar specificity to that of antibodies to the 100-kD PM-Scl protein. The elution from the recombinant plaques resulted in affinity purification of antibodies reactive with the 100-kD protein, but not the antibodies to the 70-kD protein from the same sera, confirming the specific reactivity of the fusion protein. Affinity-purified antibodies also reacted in immunofluorescence and in immunoprecipitation as expected for anti-PM-Scl. Much higher activity was seen with JH4B<sub>1</sub> eluates than wild-type eluates in immunoprecipitation, and the background reactivity of wild-type eluates was probably due to nonspecific binding. Because the patient serum contained a high titer of antibody, specific anti-PM-Scl would be a significant proportion of any nonspecific antibody binding that occurred.

Evidence was also provided to support the conclusion that

the cDNA codes for the same 100-kD PM-Scl protein that is part of the immunoprecipitated PM-Scl antigen complex. The size of the expressed recombinant protein and the immunoprecipitated antigenic protein were shown to be similar. Affinity-purified antibodies recognized the 100-kD protein immunoprecipitated from HeLa cell extract by anti-PM-Scl sera. However, the possibility that the cDNA codes for a separate PM-Scl 100-kD protein that cross-reacts with the 100-kD antigenic protein of the immunoprecipitated PM-Scl complex has not been unequivocally excluded, since direct amino acid sequence of the protein was not available. Of interest was that in some cases, anti-PM-Scl sera stained a doublet at 100 kD in immunoblotting. The two forms of this protein shared epitopes since both reacted equally with affinity-purified antibodies. This may represent a degradation product as suggested by its intermittent appearance but other possibilities, such as a difference in posttranslational modifications, or, less likely, two different but closely related proteins, cannot be excluded.

There is also strong support for the conclusion that this is the full length of the cDNA sequence. The mRNA detected by Northern blot was the same size as the original cDNA plus the extended ends. The predicted protein from the sequence is within 2% of the estimated molecular weight (100 kD) of the native antigenic PM-Scl protein observed by SDS-PAGE. The recombinant protein appeared even larger than the immunoprecipitated protein probably because of differences in processing or posttranslational modifications. Others have reported the protein to be 110 kD. If 110 kD is the true physical molecular mass, the difference in size from the predicted protein may result from posttranslational modifications at one of the potential sites identified above. The sequence preceding the first ATG is consistent with commonly observed starting sites in that an A is present in the third position before the ATG whereas it is usually not found in the unusual event of an ATG codon upstream of the starting ATG (26). There is no A three nucleotides upstream to the second ATG, which is found within the original insert, suggesting it was not the starting site. Supporting the identification of the 3' end, a clear stop codon is present and the extended end contains the expected poly-(A) tail. Although the most common polyadenylation signal is AATAAA, ATTAAA is a recognized naturally occurring variant found in 12% of mRNAs and provides almost as much polyadenylation activity as the AATAAA signal (27).

Although the 100-kD protein was the protein most commonly recognized by anti-PM-Scl positive sera in immunoblotting against whole HeLa cell extract, reaction was not seen with sera from 41% of patients with anti-PM-Scl. Although Gelpi et al. (9) found that all of 10 anti-PM-Scl sera reacted with the 110-kD protein, Alderuccio et al. (10) found results similar to ours, with only half of anti-PM-Scl sera reacting in immunoblot against the 100–110-kD protein. However, we have found that sera from as many as 94.9% of patients with anti-PM-Scl reacted with the protein expressed in IPTG-induced plaques of the clone. The most likely explanation is that some sera recognized exclusively conformational epitopes that were lost by denaturation during SDS-PAGE but were present on the fusion protein. Although the recombinant portion of fusion proteins is often denatured or otherwise presented in a nonnative state, this does not exclude the possibility of conformation. Alternatively, low abundance of the 100-kD protein in the whole HeLa extract compared with the plaques may also have accounted for the difference.

Sera of two patients did not react with plaques of either clone or with the 100-kD protein in immunoblot, suggesting that those two sera did not recognize the 100-kD protein in any form. Since they showed lines of identity with standard anti-PM-Scl sera in immunodiffusion and showed the typical PM-Scl complex in immunoprecipitation, they clearly react with PM-Scl. They may react with a conformational epitope not present on the fusion protein; they may react with the native 75-kD protein only; or they may react exclusively with other components of the complex. Both of the two sera did react with other bands in immunoblot, but it is not yet known if these bands are components of the complex.

Anti-PM-Scl is an example in which a protein complex is targeted by autoantibodies. As demonstrated in this and other studies, antibody to the 100-kD protein is the most common specificity found. This suggests that the 100-kD band may be the main target of the anti-PM-Scl response. However, reaction with at least one other complex component is commonly seen, that of 70–75 kD. In this study, specific antibodies reacting with the 100-kD protein in immunoblot did not cross-react with the 70-kD protein. Others have also found that antibodies to the two proteins are independent (10). Independent reaction with multiple components suggests that the complex as a whole has become antigenic, which supports the antigen-driven nature of the autoantibody response. It is possible that some of the other protein bands recognized by anti-PM-Scl sera in immunoblotting are also complex components.

Preliminary results of Bluthner et al. (28), presented in abstract form, described cDNA cloning of a small portion of the PM-Scl 100-kD protein and indicated that all anti-PM-Scl sera reacted with that small portion, suggesting a predominant epitope. The sequence of that fragment has not yet been published, and we are therefore not able to compare it to our sequence. Identification of the location of the epitopes of the PM-Scl 100-kD protein will be facilitated by the availability of the full-length sequence determined in this study. Analysis of the sequence did not reveal motifs that might suggest a specific function, but the availability of this recombinant protein will assist in further studies of the cellular function and cell biology of PM-Scl.

## Acknowledgments

The authors thank Dr. Russell Rother, Dr. Kazuko Itoh, and Sally Bleichinger for helpful discussions; Drs. Morris Reichlin, Frank C. Arnett, Paul H. Plotz, and Frederick W. Miller for referral of sera; and Edward P. Trieu for expert technical assistance. The oligonucleotide primers were synthesized at the Molecular Biology Resource Facility of the Saint Francis Hospital of Tulsa Research Institute at the University of Oklahoma Health Sciences Center.

This work was supported by NIH grants AR-32214, AI-27181, and AI-21568; a grant from the Pittsburgh Supercomputer Center through cooperative agreement U41-RR04154 from the National Institutes of Health Division of Research Resources; and by Department of Veterans Affairs medical research funds.

## References

1. Reichlin, M., P. J. Maddison, I. N. Targoff, T. Bunch, F. C. Arnett, G. Sharp, E. Treadwell, and E. M. Tan. 1984. Antibodies to a nuclear/nucleolar antigen in patients with polymyositis—overlap syndrome. *J. Clin. Immunol.* 4:40–44.
2. Bernstein, R. M., C. C. Bunn, G. R. V. Hughes, A. M. Francoeur, and M. B.

- Mathews. 1984. Cellular protein and RNA antigens in autoimmune disease. *Mol. Biol. Med.* 2:105–120.
3. Treadwell, E. L., M. A. Alspaugh, J. F. Wolfe, and G. C. Sharp. 1984. Clinical relevance of PM-1 antibody and physicochemical characterization of PM-1 antigen. *J. Rheumatol.* 11:658–662.
4. Reimer, G., V. D. Steen, C. A. Penning, T. A. Medsger, Jr., and E. M. Tan. 1988. Correlates between autoantibodies to nucleolar antigens and clinical features in patients with systemic sclerosis (scleroderma). *Arthritis Rheum.* 31:525–532.
5. Genth, E., R. Mierau, P. Genetzky, C. A. Von Muhlen, S. Kaufmann, H. von Wilmowsky, M. Meurer, T. Krieg, H.-J. Pollmann, and P. W. Hartl. 1990. Immunogenetic associations of scleroderma-related antinuclear antibodies. *Arthritis Rheum.* 33:657–665.
6. Wolfe, J. F., E. Adelstein, and G. C. Sharp. 1977. Antinuclear antibody with distinct specificity for polymyositis. *J. Clin. Invest.* 59:176–178.
7. Targoff, I. N., and M. Reichlin. 1985. Nucleolar localization of the PM-Scl antigen. *Arthritis Rheum.* 28:226–230.
8. Reimer, G., U. Scheer, J.-M. Peters, and E. M. Tan. 1986. Immunolocalization and partial characterization of a nucleolar autoantigen (PM-Scl) associated with polymyositis/scleroderma overlap syndromes. *J. Immunol.* 137:3802–3808.
9. Gelpi, C., A. Alguero, M. Angeles Martinez, S. Vidal, C. Juarez, and J. L. Rodriguez-Sanchez. 1990. Identification of protein components reactive with anti-PM-Scl autoantibodies. *Clin. Exp. Immunol.* 81:59–64.
10. Alderuccio, F., E. K. L. Chan, and E. M. Tan. 1991. Molecular characterization of an autoantigen of PM-Scl in the polymyositis/scleroderma overlap syndrome. A unique and complete human cDNA encoding an apparent 75-kD acidic protein of the nucleolar complex. *J. Exp. Med.* 173:941–952.
11. Targoff, I. N., F. C. Arnett, L. Berman, C. O'Brien, and M. Reichlin. 1989. Anti-KJ: a new antibody associated with the myositis/lung syndrome that reacts with a translation-related protein. *J. Clin. Invest.* 84:162–172.
12. Targoff, I. N. 1990. Autoantibodies to aminoacyl-transfer RNA synthetases for isoleucine and glycine: two additional synthetases are antigenic in myositis. *J. Immunol.* 144:1737–1743.
13. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid-guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162:156–159.
14. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 6.46–6.48.
15. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6–13.
16. Feinberg, A. P., and B. Vogelstein. 1984. Addendum: a technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 137:266–267.
17. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene (Amst.)* 33:103–119.
18. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.
19. Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA.* 84:4767–4771.
20. Loh, E. Y., J. F. Elliott, S. Cwirla, L. L. Lanier, and M. M. Davis. 1989. Polymerase chain reaction with single-sided specificity: analysis of T cell receptor delta chain. *Science (Wash. DC).* 243:217–220.
21. Frohman, M. A., M. K. Dush, and G. R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA.* 85:8998–9002.
22. Southern, E. M. 1978. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.
23. Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387–395.
24. Itoh, K., Y. Itoh, and M. B. Frank. 1991. Protein heterogeneity in the human Ro/SSA ribonucleoproteins. The 52- and 60-kD Ro/SSA autoantigens are encoded by separate genes. *J. Clin. Invest.* 87:177–186.
25. Brendel, V., J. Dohlman, B. E. Blaisdell, and S. Karlin. 1991. Very long charged runs in systemic lupus erythematosus-associated antigens. *Proc. Natl. Acad. Sci. USA.* 88:1536–1540.
26. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12:857–872.
27. Wickens, M. 1990. How the messenger got its tail: addition of poly-(A) in the nucleus. *Trends Biochem. Sci.* 15:277–281.
28. Bluthner, M., E. Genth, and F. A. Bautz. 1989. Cloning of a cDNA-fragment coding for an epitope recognized by anti-PM-Scl-autoantibodies. *In Molecular and Cell Biology of Autoantibodies and Autoimmunity: Abstracts of the First International Workshop*. E. K. F. Bautz, J. R. Kalden, M. Homma, and E. M. Tan, editors. Springer-Verlag, Heidelberg, FRG. p. 2.