Molecular Definition and Sequence Motifs of the 52-kD Component of Human SS-A/Ro Autoantigen

Edward K. L. Chan, John C. Hamel, Jill P. Buyon,* and Eng M. Tan

The W. M. Keck Autoimmune Disease Center, Department of Molecular and Experimental Medicine, Scripps Clinic and Research Foundation, La Jolla, California 92037; and *Hospital for Joint Diseases Orthopedic Institute, New York 10003

Abstract

Serum SS-A/Ro autoantibodies are commonly found in patients with Sjogren's syndrome, systemic lupus erythematosus, neonatal lupus, and subacute cutaneous lupus. Two proteins of 60 and 52 kD have been described as targets for these autoantibodies. To define the 52-kD component unambiguously, cDNA clones were isolated from human HepG2 and MOLT-4 cell cDNA libraries. The identity of cDNA was established by (a) the specificity of the antibody affinity purified from the recombinant protein, (b) the reactivity of the purified recombinant protein with prototype SS-A/Ro sera in immunoblot and ELISA, and (c) two-dimensional gel comigration of MOLT-4 cell 52-kD protein and the recombinant protein. A 1.9-kb cDNA encoded the complete 52-kD protein containing 475 amino acids (M, 54,082). Putative zinc-finger domains and a leucine zipper motif were identified in the amino-terminal half of the 52-kD protein, implicating its possible association with DNA/RNA. Sequence homology detected between the 52-kD protein and human ret transforming protein, and mouse T cell gene expression down-regulatory protein rpt-1, may provide leads to the functional role of the 52-kD protein in addition to the possibility that these proteins might constitute members of a subfamily of finger proteins. (J. Clin. Invest. 1991. 87:68-76.) Key words: autoantibody • Sjogren's syndrome • zinc finger protein • systemic lupus erythematosus • autoimmunity

Introduction

The molecular characteristics of SS-A/Ro¹ antigens are of special clinical interest because of the relatively high frequency of

Address reprint requests to Edward K. L. Chan, Ph.D., W. M. Keck Autoimmune Disease Center, Scripps Clinic and Research Foundation, 10666 N. Torrey Pines Road, La Jolla, CA 92037.

The original nucleotide and amino acid sequence data have been submitted to EMBL/Genbank under accession No. M35041.

Received for publication 13 June 1990 and in revised form 1 August 1990.

J. Clin. Invest.

the SS-A/Ro autoantibody in patients with several rheumatic diseases including systemic lupus erythematosus, Sjogren's syndrome, neonatal lupus syndrome with congenital heart block, and subacute cutaneous lupus (for current reviews, see references 1-3). Various molecular species ranging from 50 to 150 kD have been described for the antigenic component(s) of SS-A/Ro (summarized in reference 4). In 1984 Wolin and Steitz (5) showed that autoantibodies to SS-A/Ro immunoprecipitated small ribonucleoprotein particles composed of hY-RNAs and a 60-kD protein component. The 60-kD protein was defined as the major autoimmune target because SS-A/Ro autoimmune sera did not recognize hY-RNAs alone. Recent work of Boire and Craft (6) reported a subpopulation of autoantibodies recognizing an antigenic epitope restricted to the intact ribonucleoprotein particle composed of hY5 RNA and the 60kD SS-A/Ro protein. Besides the 60-kD protein, Ben-Chetrit et al. (4) found that most SS-A/Ro-positive sera as defined by immunoprecipitation in Ouchterlony double diffusion assay also recognized another protein of 52 kD. It was not clear whether the 52-kD protein interacted with hY-RNAs directly or via the association with the 60-kD protein that was known to have in vitro RNA-binding activity (5, 7). Indirect immunofluorescence using these specific anti-52-kD reagents gave nuclear punctate staining similar to those observed for antibodies specific to the 60-kD SS-A/Ro component (4). It was also shown by antibody affinity purification that patient sera contained distinct, noncross-reacting antibodies directed to the 60- and 52-kD components (4). Autoantibodies to the 52-kD component were in fact very common in SS-A/Ro autoimmune sera analyzed (> 80%) and the reason the 52-kD component had not been detected earlier was probably a result of coexisting autoantibodies to SS-B/La and to the incomplete separation of the 47-kD SS-B/La antigen from the 52-kD component in the usual gel separation system (8). These two SS-A/Ro components were found in all human cell lines tested including HeLa, MOLT-4, Raji, and Wil-2 (4). Rader et al. (9) also described two human SS-A/Ro species of 60 and 52 kD from lymphocytes and, in addition, two red cell-specific SS-A/Ro components of 60 and 54 kD that were immunologically related to the lymphocyte 60- and 52-kD species, respectively. To define the 52-kD component of SS-A/Ro protein unambiguously, we have cloned the full-length cDNA encoding the 52-kD protein. This report describes the isolation and characterization of cDNA clones using a human autoimmune serum with specific autoantibody to the 52-kD SS-A/Ro protein and the special features of the deduced polypeptide that might provide insights

Methods

Cell extracts. MOLT-4 cells (human T cell lymphoblastic leukemia, ATCC CRL 1582) were cultured in DMEM containing 10% calf serum

into the functional role of the protein.

Preliminary reports in abstract form were presented at the First International Workshop on the Molecular and Cell Biology of Autoantibodies and Autoimmunity, Heidelberg, Germany, 27–29 July 1989, and at the Workshop on Structure and Function of Eukaryotic RNP, Delphi, Greece, 1–5 April 1990 (1990. *Mol. Biol. Rep.* 14:53 [Abstr.]).

^{1.} Abbreviations used in this paper: CD^4 , cluster designation 4; IPTG, isopropyl- β -thiogalactopyranoside; PBS-T, phosphate buffered saline with 0.05% Tween-20; SS-A/Ro, nuclear antigen A of Sjogren's Syndrome; SS-B/La, nuclear antigen B of Sjogren's syndrome.

[©] The American Society for Clinical Investigation, Inc. 0021-9738/91/01/0068/09 \$2.00 Volume 87, January 1991, 68-76

at 37°C in an 8% CO₂ incubator. Cultures were supplemented with 2.5 μ g/ml gentamicin sulfate and maintained at 10⁶ cells per ml. Cells were harvested and extracted in buffer A (150 mM NaCl, 10 mM Tris-HCl, pH 7.2, 0.5% Nonidet P-40) and freed of cell nuclei by centrifugation at 12,000 g for 15 min.

Immunoblotting. For high-resolution separation of the 60-kD, 52kD, and 47-kD SS-A/Ro-SS-B/La proteins, cell extracts were separated by PAGE as described by Laemmli (10) using 15% gel slabs (20×13 \times 0.1 cm) and the ratio of acrylamide:bis-acrylamide was changed to 172.4:1 in the separating gel (8). It was shown recently that the standard gel composition used by many laboratories was not optimal for the separation of the 52-kD SS-A/Ro and 47-kD SS-B/La proteins (8). Proteins were transferred from gel onto nitrocellulose using a constant voltage of 60 V for 2 h (11). The nitrocellulose sheet was air-dried, cut into strips, and incubated for 30 min in 3% nonfat milk diluted in PBS to block nonspecific binding sites. The nitrocellulose strips were incubated with 1 ml of a 1:100 dilution of sera for 1 h and then washed in PBS with 0.05% Tween 20 (PBS-T). ¹²⁵I-protein A (ICN Biochemicals, Irvine, CA) was used to detect bound human antibodies. The following protein standards (Bio-Rad Laboratories, Richmond, CA) were used to determine molecular weights: phosphorylase B, 92,500; BSA, 66,200; ovalbumin, 45,000; carbonic anhydrase, 31,000; soybean trypsin inhibitor, 21,500; and lysozyme, 14,400.

Screening of phage libraries for SS-A/Ro cDNA. A HepG2 cell λ Zap cDNA library was a kind gift from Dr. Frank R. Jirik, University of British Columbia, Vancouver, BC, Canada. A high titer SS-A/Ro human serum Bo was used for immunoscreening of 10⁶ recombinants. ¹²⁵I-protein A was used to detect bound human antibodies. One positive clone Cl was identified after multiple screenings and was subcloned in vivo into pBluescript plasmid pCl using R408 helper phage (Stratagene, La Jolla, CA) as recommended in the manufacturers' instructions.

Human T cell lymphoblastoma (MOLT-4) λ gt-ll cDNA library was constructed by Dr. K. Ogata and Dr. D. J. Noonan, Scripps Clinic and Research Foundation. The MOLT-4 library was screened for fulllength SS-A/Ro cDNA clones by DNA hybridization. Two partially complementary synthetic oligonucleotides (5'-TGTGCAGTGCAT-GGAGAGAGACTTCACCTG-3' and 5'-TCTTTCTCACAGAAC-AGGTGAAGTCTCTCT-3') were designed based on the 5' sequence of pCl cDNA insert. They were mixed and labeled with [α^{32} P]-ATP using the standard fill-in reaction of Klenow polymerase (12). All screenings were carried out with duplicate filters and positive phages were plaque purified. The resulting cDNAs were subcloned into pBluescript vectors (Stratagene) for further analysis.

Cloning of rfp protein. Two partially complementary synthetic oligonucleotides (5'-TGGCCTCCGGGAGTGTGGCCGAGTGCC-3' and 5'-GGTGGTCTCCTGCTGCAGGCACTCGGCC-3') were designed based on the 5' sequence of the rfp cDNA reported (13). They were mixed and labeled with $[\alpha^{32}P]$ -ATP using the standard fill-in reaction of Klenow polymerase (12) and used to screen 10⁶ recombinant from the MOLT-4 cell cDNA library as described above. Two initial positives were detected and one of these, 2rl, was purified and the cDNA insert subcloned to pBluescript SK plasmid for further analysis. A rabbit anti-rfp reference serum was kindly provided by Dr. M. Takahashi, Aichi Cancer Center Research Institute, Nagoya, Japan.

Affinity-purified antibodies. λ Zap clone Cl phage plaques were induced to produce recombinant protein by overlaying with isopropyl β -thiogalactopyranoside (IPTG, Sigma Chemical Co., St. Louis, MO) saturated nitrocellulose filters and grown overnight at 37°C. These nitrocellulose filters were washed with PBS-T, incubated with diluted serum for 1 h, washed again with PBS-T before elution of bound antibodies with 0.1% BSA in 0.1 M phosphate buffer, pH 2.2. Affinity-purified antibodies were immediately neutralized by the addition of 1 M Tris-HCl, pH 8.8. The antibodies were concentrated with Centricon 100 microconcentrators (Amicon Corp., Danvers, MA).

Purification of recombinant protein. Plasmid pCl was transformed in Escherichia coli strain XL1-Blue (Stratagene). A 200-ml culture of the recombinant cells was grown to $OD_{600} = 0.6$ at 37°C and IPTG was added to a final concentration of 10 mM. The culture was allowed to resume growth overnight before harvesting by centrifugation. Purification of recombinant protein was performed as described by Adam et al. (14). The final pellet was extracted in 8 M urea for 15 min and the supernatant was stored at -70° C in aliquots.

ELISA. Standard protocol for ELISA was employed as described by Rubin (15). Purified recombinant proteins in 8 M urea were diluted 1,000-64,000-fold in PBS for coating Immulon 2 microtiter plates (Dynatech Laboratories, Inc., Alexandria, VA). Peroxidase-conjugated goat anti-human IgG + M reagents (Caltag Laboratories, So. San Francisco, CA) and the substrate 2,2'-azinobis(3-ethylbenzthiazoline sulfonic acid) (Boehringer Mannheim GmbH, Mannheim, FRG) were used as detecting agents.

Southern analysis. For determining the complexity of the gene encoding the 52-kD SS-A/Ro protein, human genomic DNA samples (10 μ g/lane) from HeLa cells and normal peripheral blood were digested with EcoRI or HindIII restriction enzyme to completion and analyzed essentially according to Southern (16) using labeled pCl cDNA insert.

In vitro RNA transcription and translation. The EcoRI λ gtll cDNA insert of 52FL was subcloned into pBluescript SK vector. The resulting plasmid p52FL with the cDNA insert in the same orientation as β -galactosidase was linearized by restriction enzyme ClaI at the 3' end of the insert. RNA was transcribed in vitro from the linearized plasmid using T3 RNA polymerase and was translated in vitro in a rabbit reticulocyte lysate (Promega, Madison, WI) in the presence of [³⁵S]-methionine (Tran³⁵S-label, 70% methionine and 15% cysteine, ICN Biochemicals) as described in the manufacturers' instructions. Products of in vitro translation were stored in aliquots for immunoprecipitation (17, 18) and two-dimensional gel analysis (19).

DNA sequencing and analysis. The sequence strategy was to combine data derived from restriction fragments subcloned in pBluescript vector and the use of synthetic oligonucleotides for primers. Plasmids were purified by the Chen and Seeburg method (20) and DNA sequencing was performed by the dideoxy chain termination method of Sanger et al. (21). DNA and protein sequences were analyzed by the Genetics Computer Group Sequence Analysis Software Package for VAX computers (22). Alignment of protein sequences was initially achieved with the GAP program that employed the algorithm of Needleman and Wunsch (23). Multiple sequence alignments were performed with CLUSTAL programs (24, 25).

Results

Cloning and characterization of 52-kD SS-A/Ro cDNAs. In the screening of 10⁶ recombinant phages from the λ Zap HepG2 cell cDNA library, clone Cl was selected by human serum Bo and was purified with three rounds of plaque purification. To show that Cl encoded the 52-kD SS-A/Ro protein, purification of affinity antibody was performed with Cl phage protein and a different human autoimmune serum Ca that recognized both SS-A/Ro and SS-B/La proteins in MOLT-4 cell extracts (Fig. 1, lane 2). Affinity purified antibody from the Cl phage protein alone (lane 4). This specific reactivity was consistent with our previous finding that human affinity purified anti-52-kD SS-A/Ro or SS-B/La proteins (4).

The λ Zap clone Cl was subcloned in vivo into pBluescript plasmid pCl. Recombinant protein derived from pCl grown in XL1-Blue bacteria had a molecular mass of $\simeq 46,000$ as detected by immunoblotting and was purified as described in Methods. The final preparation was a highly antigenic substrate for the detection of anti-52-kD SS-A/Ro antibody in an



Figure 1. Immunoblotting analysis of affinity purified antibodies from λ Zap Cl-derived fusion protein. Molt-4 cell extracts were separated on SDS-PAGE using a 20-cm 15% separating gel as described in Methods. After transfer to nitrocellulose filter, strips were cut and probed with different human antibodies: lane 1, a normal human serum; lane 2, human serum Ca with weak reactivity to the 60-kD SS-A/Ro protein and strong reactivities to both the 52-kD SS-A/ Ro and 47-kD SS-B/La proteins; lane 3, Ca serum absorbed with Cl phage proteins shows a drop in intensity of the 52-kD band but no effect on the reactivity to SS-B/La; lane 4, antibody from Ca serum affinity purified from Cl phage protein shows very strong reactivity with 52-kD

SS-A/Ro alone; lane 5, serum Bo, the original serum used in the cloning of Cl. The results show that affinity purified antibody from Cl recombinant protein recognized only the 52-kD SS-A/Ro protein.

ELISA format (Fig. 2 A). Even with the highest dilution of 1:64,000 for coating microtiter wells, an OD_{410} reading of 2 was obtained 1 h after the addition of detecting reagents. Fig. 2 B illustrates the specificity of the ELISA when a 1:30,000 dilution of the purified recombinant antigen was used for the detection of SS-A/Ro antibodies. Recombinant 60-kD SS-A/Ro and SS-B/La proteins derived from cDNA clones previously reported (17, 26) were purified similarly and are included in this figure for comparison. High OD readings for the 52-kD SS-A/Ro recombinant protein were only seen in sera with detectable antibodies by immunoblotting. Our data indicated that in most anti-SS-A/Ro sera tested, the reactivities as determined by ELISA were higher for the 52-kD than the 60-kD recombinant antigens. Besides the negative controls shown in Fig. 2 B, 20 normal human sera and standard prototype autoimmune sera were also negative including antibodies to PCNA, Sm, U1-RNP, ribosomal RNP, Jo-1, Scl-70, SL/Ki, and Ku specificities (data not shown).

The immunological data above suggested that pCl encoded at least one immunoreactive region of the 52-kD SS-A/Ro protein and the next step was to determine if the cDNA insert of pCl encoded the entire 52-kD protein. Plasmid pCl was digested with EcoRI restriction enzyme and a cDNA insert of 1.5 kb was detected after electrophoresis in a standard agarose gel. Northern blot analysis of the mRNA from MOLT-4 and HeLa cells using radiolabeled 1.5-kb fragment showed that there was a single band of ~ 1.9 kb. This indicated that pCl probably contained only a fraction of the complete cDNA for the 52-kD SS-A/Ro protein. To obtain a full-length cDNA, 10⁶ recombinant phages from the MOLT-4 cell cDNA library were screened with a radiolabeled DNA probe corresponding to the 5' sequence of pCl cDNA insert. Of the two clones selected, clone 52FL had the longest insert of ~ 1.9 kb and was sub-cloned into pBluescript plasmid p52FL for further analysis.

To determine if p52FL cDNA insert encoded the complete 52-kD protein, comparison was made between the in vitro transcription and translation products of p52FL and the 52-kD SS-A/Ro protein from MOLT-4 cells. Preliminary data showed that they comigrated in standard one-dimensional gel SDS-PAGE, and therefore two dimensional gels were used for more discriminating analysis (Fig. 3). In vitro [S³⁵]-methionine labeled translation products of p52FL RNA gave a major spot (arrow) and several minor spots that were observed only in an overexposed autoradiogram (Fig. 3 A). Immunoprecipitation of these translation products with human SS-A/Ro sera showed that the major 52-kD product was immunoreactive. Immunoblotting of MOLT-4 cell proteins with serum Ca detected multiple SS-B/La isospecies of different pI migrating at



Figure 2. Reactivity of purified recombinant protein in ELISA. Recombinant protein derived from pCl plasmid was purified as described in Methods. (A) Titration curve of the reactivities (OD_{410}) of a prototype SS-A/Ro serum Ge and a normal human serum with recombinant protein diluted from 1:1,000 to 1:64,000. (B) Reactivities of human sera with respect to different purified recombinant proteins. Recombinant 60-kD SS-A/Ro and 47-kD SS-B/La proteins were purified from cDNA clones previously reported (17, 26). Serum Ze is the CDC prototype anti-SS-B/La, sera Ge and Bo are mostly reactive with 52-kD SS-A/Ro, sera Ca, Wr, Be recognized both SS-A/Ro and SS-B/La, and serum We is the CDC prototype anti-Sm.



Figure 3. Autoradiograms of two-dimensional gels separating MOLT-4 cell proteins and in vitro translation products of p52FL RNA transcript. (A) Separation of [35S]-methionine labeled in vitro translation product of p52FL RNA. An overexposed autoradiogram of the dried gel is shown to indicate the major 52-kD spot (arrowhead) and other minor species. (B) The immunoblot of MOLT-4 cell proteins with serum Ca. showing strong reactivity with SS-B/La isospecies migrating at 47 kD and isoforms of the 60-kD SS-A/Ro proteins. In contrast the reactivity with 52kD SS-A/Ro protein is restricted to a major spot (arrowhead) with a pI of $\simeq 6.5$. The pI estimated by the **ISOELECTRIC program (22)** based on the protein sequence is 6.35. Arrows marked a and t

correspond to the positions of major cell proteins actin and tubulin, respectively. To determine unambiguously the gel migration positions of the 52-kD SS-A/Ro protein and the recombinant protein, a mixture of MOLT-4 cell proteins and the [35 S]-methionine labeled in vitro translation product of p52FL RNA was separated on a single gel and then transferred to nitrocellulose filter. (*C*) 4-d exposure of the labeled [35 S]-methionine on the nitrocellulose filter. (*D*) 12-h exposure with intensifying screen for the same nitrocellulose after immunoblotting with human serum Ge, which is specific for 52-kD SS-A/Ro protein, and [125 I]-protein A. The major spots (*arrowhead*) in *C* and *D* match to the same position showing that p52FL encoded the entire 52-kD protein of SS-A/Ro.

47 kD and several isoforms of the 60-kD SS-A/Ro proteins (Fig. 3 *B*). On the other hand, the 52-kD SS-A/Ro protein was restricted to a major spot with an estimated $pI \simeq 6.5$ (Fig. 3 *B*, *arrowhead*). The colocalization of the p52FL encoded protein and the 52-kD SS-A/Ro protein was determined unambiguously by running both samples in the same gel (Fig. 3, *C* and *D*). These independent results provided further confirmation that p52FL cDNA encoded the entire 52-kD protein of SS-A/Ro.

Southern analysis. Human genomic DNA samples from HeLa cells and normal peripheral blood were digested with EcoRI or HindIII restriction enzyme to completion. Either restriction site was not found in the complete 52FL sequence. Results of the genomic blot probed with Cl cDNA insert showed a major band of ≈ 5 kb and a minor band of ≈ 8 kb (HindIII) and ≈ 20 kb (EcoRI) (Fig. 4). These data indicated that there was probably only one and at most two genes for the 52-kD SS-A/Ro protein in the human genome.

Sequence analysis. The nucleotide sequences of the cDNA inserts of Cl and 52FL were determined by sequencing both DNA strands as outlined in Fig. 5. Clone Cl represented the 3' 80% of clone 52FL and there was no difference detected between Cl and 52FL clones in the corresponding regions. The DNA sequence of 52FL and the deduced amino acid sequence are shown in Fig. 6. There is only one large open reading frame composed of 475 amino acids corresponding to a calculated molecular mass of 54,082. The sequence flanking the first ATG codon is in close agreement with the eukaryotic translation

initiation consensus sequence (27). In the 5' noncoding region, there are three stop codons that are in-frame with the ATG start codon (Fig. 6). Two polyadenylation signal sequences are found at the end of the 3' noncoding sequence. The consensus sequence for nuclear localization signal has not been detected.

Special features of the 52-kD SS-A/Ro protein are summarized in Fig. 7. The interesting features are the amino-terminal zinc finger domains, a central domain with a leucine zipper motif, and a carboxy-terminal domain that is highly similar to human protein rfp. Fig. 8 shows the putative zinc-finger structures in the 52-kD protein. The arrangement of the Cys residues for the chelation of metal such as zinc has been found in many DNA/RNA-binding proteins (28, 29) and these zinc finger structures are generally thought to be responsible for the interaction with either DNA or RNA. The proposed finger structures in Fig. 8 are atypical as compared to most known finger protein sequences (28, 29). There is no published data at the present to indicate whether the SS-A/Ro protein binds zinc or other metal ions. However, a single zinc finger motif has also been described in the 60-kD SS-A/Ro protein (7, 17). A leucine zipper motif (residues 211-232) was detected in the center domain of the 52-kD protein. Leucine zippers were originally described in DNA-binding proteins (30) and now these motifs are known to participate in protein-protein interaction or dimer formation that is important for DNA-binding (31-33). This motif in 52-kD SS-A/Ro may be important for interaction with other proteins or for homo-dimer formation.

When the DNA and protein sequences were determined for



Figure 4. Southern blot analysis of human genomic DNA samples probed with Cl cDNA insert. Lane 1, HeLa cell DNA digested with EcoRI; lane 2, normal human peripheral blood DNA digested with EcoRI; lane 3, human peripheral blood DNA digested with HindIII restriction enzyme. The results suggest only one or two genes encoding the 52-kD SS-A/Ro protein.

the 52-kD protein, homology search with other known proteins was performed with the WORDSEARCH program available from the Genetics Computer Group Sequence Analysis Software Package (22). Three proteins were identified with high degree of homology to the 52-kD protein and they were the mouse rpt-1 protein (34), human ret transforming protein (35), and its related protein rfp (13). Fig. 9 A shows the alignments among the three protein sequences. Note that the least number of gaps has been introduced to optimize the alignment. There is over 30% identity and 50-60% similarity in sequences at the corresponding amino-terminal halves of these proteins. The cysteine residues of the zinc finger motifs are conserved among all three sequences. The 52-kD protein is more similar to either rfp or rpt-1 than rfp is to rpt-1 (Fig. 9 A). The data suggest that these three protein domains might be related as a subfamily of finger proteins. In addition, the carboxy-terminal halves of the 52-kD SS-A/Ro and rfp proteins are highly similar with > 50% of identity and 65% similarity (Fig. 9 B).

rfp cDNA isolation and analysis for antigenic reactivity. Because up to 50% identity in amino acid residues was found between the carboxy-terminal halves of the 52-kD SS-A/Ro and rfp proteins, it was of interest to determine whether (a) anti-52-kD SS-A/Ro autoantibodies could recognize rfp protein, and (b) rfp might be a member of SS-A/Ro proteins. A cDNA clone 2rl was isolated from the same MOLT-4 library and restriction enzyme analysis confirmed its identity to the published rfp cDNA clone (13). This was supported by DNA sequencing data which showed that 2rl was an authentic rfp protein cDNA with 104 extra nucleotides at the 5' end as compared to the original clone described by Takahashi et al. (13). The extra 5' sequence is TCCGCTCGGACGCGGCCACGT- TGTCTTGCGCGCTTTGCCCGCCTGGCCCTGGGAC-TCTGACCCTCGGCTACCCTTTCCTGCCCCACTAGCG-TGGCCGCGAGCCT. In vitro [S³⁵]-methionine labeled translation product of 2rl-encoded RNA was a protein of \sim 58 kD. None of the 10 different human anti-SS-A/Ro sera recognized this 58-kD protein in a standard immunoprecipitation assay. In immunoblotting, the rabbit antiserum to recombinant rfp protein recognized a 58-kD MOLT-4 cell protein that was distinct from the 52-kD and 60-kD SS-A/Ro proteins. These data indicated that, although there are significant sequence similarities between the 52-kD SS-A/Ro and rfp proteins, no immunological cross-reaction could be observed between these proteins using autoantibodies. It also reconfirms the general observation that human autoantibodies are highly specific for their antigenic targets and are generally noncross-reactive with related proteins (1).

Discussion

A cDNA encoding the complete 52-kD SS-A/Ro protein described by Ben-Chetrit et al. (4) was cloned from a human T-cell (MOLT-4) cDNA library. A partial clone was also obtained from a human liver (HepG2) cDNA library and there was no difference detected between their overlapping regions. The nucleotide and deduced amino acid sequences for the 52kD protein showed no homology with the corresponding sequences for the related 60-kD SS-A/Ro (7, 17) and 47-kD SS-B/La antigens (26). DNA hybridization experiments using the cDNAs from all three antigens above also showed no cross-hybridization (data not shown). Human SS-A/Ro autoantibodies recognized epitope(s) expressed on the recombinant proteins derived from the initial λ Zap phage and the subcloned pBluescript plasmid. The partially purified recombinant protein was a specific substrate for the sensitive detection of anti-SS-A/Ro antibodies in an ELISA format. In addition, in vitro translation product derived from cDNA clone could be immunoprecipitated by anti-SS-A/Ro sera and was found to comigrate with the cellular 52-kD protein as determined by immunoblotting and two-dimensional gel electrophoresis. Because data from mRNA blotting indicated that there was only one band of \simeq 1.9 kb in MOLT-4 cells, the 1.9-kb cDNA insert of p52FL probably represented the full-length mRNA. The presence of putative zinc fingers and a leucine zipper in the 52-kD protein



Figure 5. Schematic representation of the cDNA clones Cl and 52FL derived from HepG2 and MOLT-4 cell libraries, respectively. Coding regions are indicated by solid bars and the 5' and 3' untranslated regions are represented by thin lines. Thin and thick arrows represent DNA sequence derived from restriction fragment subcloning and sequencing with synthetic oligonucleotides, respectively. Restriction enzyme sites R, N, and S are EcoRI, NcoI, and SacI, respectively.

- 85	GGGA	GCI	GGC	TAC	CAG	CGT	TGA	GTT	GCC	CTG	TAA	AGC	CAA	AACO	ccc	TAA	AGG	TCT	CCA	CAC	CTGC	TGI	TTA	ACC	GC	ACAC	CTTG	ACA
1	ATG	GCT	TCA	GCA	GCA	CGC	TTG		ATG	ATG	TGG	GAG	GAG	GTC		TGC	сст	ATC	TGC	сто	GAC	ccc	TTC	GTC	GGAC	SCCI	GTG	AGC
1	M	<u>د ان ان</u>	. 10/1	Δ	Δ	R	T	т	M	M	w	E	E	v	т	С	P	T	C	L	D	P	F	v	Е	P	v	S
T	п	ñ	3	'n	'n	ĸ	1	•	1.1			5		•	-	Ŭ	-	-	Ũ	-	-	-	-					
05	ATC	~~~	ייייי				TTO	TCC			тсс	• • • • •	ידרי		CTT	000		сст	າດດດ	ccc		CTC	TGT	GCI	GTO	TGC	CGG	CAG
20	I	UNU F	,101 C	000	U U	c	F	- C	040	F	- C	T	,10.	0000	v	C	v	с. С	C	G	S	v	С	A	v	C	R	0
29	T	E	U	G	п	э	r	U	Q	E	U	Ŧ	3	Q	v	9	ĸ	0	0	0	5	•	Ŭ		•	Ŭ	••	*
1.00																	0000				• • • •	~	ATC	• A C C	CAC	CAC	ccc	ACA
169	CGC	111	.016	010	AAG	AAI	010	566		AAI	UGA	ICAG	1014	1000	AAU	AIG		AAU	AAU N			UCAA F		c		F	.000	D
57	R	F	L	L	ĸ	N	L	ĸ	P	N	ĸ	Q	L	A	N	м	v	N	N	L	K	E	T	3	Q	Ľ	~	ĸ
			_																-			~ • •	000			OTT	moo	TCC
253	GAG	GGC	ACA	CAG	GGG	GAA	CGG	TGT	GCA	GTG	CAI	GGA	GAC	JAGA	CIT	CAC	CIG	TIC	IGI	GAG	AAA	GAI	666	AAG			.160	100
85	Е	G	Т	Q	G	Е	R	С	Α	v	н	G	Е	R	L	н	L	F	C	E	ĸ	D	G	ĸ	A	L	C	w
337	GTA	TGI	GCC	CAG	TCT	CGG	AAA	CAC	CGT	GAC	CAC	CCCC	ATC	GTC	ССТ	CTT	GAG	GAG	GCT	GCA	CAG	GAG	TAC	CAG	GAG	AAG	CTC	CAG
113	v	С	Α	Q	S	R	ĸ	н	R	D	н	Α	М	v	Р	L	Е	Е	A	A	Q	Е	Y	Q	Е	ĸ	L	Q
421	GTG	GCA	TTA	GGG	GAA	CTG	AGA	AGA	AAG	CAG	GAG	TTG	GCI	GAG	AAG	TTG	GAA	GTG	GAA	ATT	GCA	ATA	AAG	AGA	GCA	GAC	TGG.	AAG
141	v	Α	L	G	Ε	L	R	R	ĸ	Q	Е	L	Α	Е	ĸ	L	Е	v	Е	Ι	A	I	ĸ	R	A	D	W	ĸ
	AAA																											
169	К	Т	v	Е	Т	Q	ĸ	S	R	I	н	Α	Е	F	V	Q	Q	ĸ	N	F	L	v	Е	Е	Е	Q	R	Q
589	CTG	CAG	GAG	CTG	GAG	AAG	GAT	GAG	AGG	GAG	CAG	CTG	AGA	ATC	CTG	GGG	GAG	AAA	GAG	GCC	AAG	CTG	GCC	CAG	CAG	AGC	CAG	GCC
	L																											
		`									`				~							^						
673	СТА	CAG	GAG	стс	ATC	TCA	GAG	CTA	GAT	CGA	AGG	TGC	CAC	AGC	TCA	GCA	CTG	GAA	CTG	CTG	CAG	GAG	GTG.	ATA	ATT	GTC	CTG	GAA
225	L	0	E	L	T	S	E	L	D	R	R	C	н	S	S	A	L	Е	L	L	0	Е	v	I	I	v	L	Е
	^							~																				
	ACC							^																				
757	AGG	AGT	GAG	тсс	TGG	AAC	CTG	ÂÂG	GAC	CTG	GAT	ATT	ACC	тст	CCA	GAA	CTC	AGG.	AGT	GTG	TGC	CAT	GTG	CCA	GGG	CTG	AAG	AAG
757	AGG R	AGT	GAG	тсс	TGG	AAC	CTG	ÂÂG	GAC	CTG	GAT	ATT	ACC	тст	CCA	GAA	CTC	AGG.	AGT	GTG	TGC	CAT	GTG	CCA	GGG	CTG	AAG	AAG
757 253	R	AGI S	GAG E	тсс s	TGG W	AAC N	CTG L	AAG K	GAC D	CTG L	GAT D	ATT I	ACC T	TCT S	CCA P	GAA E	CTC. L	AGG. R	AGT S	GTG V	TGC C	CAT H	GTG V	CCA P	GGG G	CTG. L	AAG/ K	AAG K
757 253 841	R ATG	AGI S CTG	GAG E AGG	TCC S ACA	TGG W TGT	AAC N GCA	CTG L GTC	AAG K CAC	GAC D ATC	CTG L ACT	GAT D CTG	ATT I GAT	ACC T CCA	TCT S	CCA P ACA	GAA E GCC	CTC. L	AGG. R CCG	AGT S TGG	GTG V CTG	TGC C ATA	CAT H CTT	GTG V TCA	CCA P GAA	GGG G GAT	CTG. L CGG.	AAGA K AGA(AAG K CAA
757 253 841	R	AGI S CTG	GAG E AGG	TCC S ACA	TGG W TGT	AAC N GCA	CTG L GTC	AAG K CAC	GAC D ATC	CTG L ACT	GAT D CTG	ATT I GAT	ACC T CCA	TCT S	CCA P ACA	GAA E GCC	CTC. L	AGG. R CCG	AGT S TGG	GTG V CTG	TGC C ATA	CAT H CTT	GTG V TCA	CCA P GAA	GGG G GAT	CTG. L CGG.	AAGA K AGA(AAG K CAA
757 253 841 281	R ATG M	AGT S CTG L	GAG E AGG R	TCC S ACA T	TGG W TGT C	AAC N GCA A	CTG L GTC V	AAG K CAC H	GAC D ATC I	CTG L ACT T	GAT D CTG L	GATT I GAT D	ACC T CCA P	GAC.	CCA P ACA T	GAA E GCC/ A	CTC. L AAT N	AGG. R CCG ⁴ P	AGT S TGG W	GTG V CTG L	TGC C ATA	CAT H CTT L	GTG V TCA S	CCA P GAA E	GGG G GAT D	CTG L CGG R	AAGA K AGA(R	AAG K CAA Q
757 253 841 281 925	R ATG M GTG	AGI S CTG L AGG	GAG E AGG R	TCC S ACA T GGA	TGG W TGT C GAC	AAC N GCA A	CTG L GTC V CAG	AAG K CAC H CAG	GAC D ATC I AGC	CTG L ACT T	GAT D CTG L CCT	GATT I GAT D	ACC T CCA P AAT	GAA	CCA P ACA T GAG	GAA E GCC/ A AGA?	CTC. L AAT N	AGG R CCG P GATA	AGT S TGG W AGT	GTG V CTG L	TGC C ATA I	CAT H CTT L	GTG V TCA S GTC	CCA P GAA E CTG	GGG G GAT D GGT	CTG L CGG R GCC	AAGA K AGA(R CAG(AAG K CAA Q CAC
757 253 841 281 925	R ATG M	AGI S CTG L AGG	GAG E AGG R	TCC S ACA T GGA	TGG W TGT C GAC	AAC N GCA A	CTG L GTC V CAG	AAG K CAC H CAG	GAC D ATC I AGC	CTG L ACT T	GAT D CTG L CCT	GATT I GAT D	ACC T CCA P AAT	GAA	CCA P ACA T GAG	GAA E GCC/ A AGA?	CTC. L AAT N	AGG R CCG P GATA	AGT S TGG W AGT	GTG V CTG L	TGC C ATA I	CAT H CTT L	GTG V TCA S GTC	CCA P GAA E CTG	GGG G GAT D GGT	CTG L CGG R GCC	AAGA K AGA(R CAG(AAG K CAA Q CAC
757 253 841 281 925 309	R ATG M GTG V	AGI S CTG L AGG R	GAG E AGG R CTT L	TCC S ACA T GGA G	TGG W TGT C GAC D	AAC N GCA A ACC T	CTG L GTC V CAG Q	AAG K CAC. H CAG. Q	GAC D ATC I AGC S	CTG L ACT T ATA I	GAT D CTG L CCT P	GATT I GAT D GGA G	ACC T CCA P AAT N	GAC. CGAA	CCA P ACA T GAG	GAA E GCC/ A AGA? R	CTC. L AAT N TTT F	AGG R CCG ⁴ P GATA D	AGT S TGG W AGT S	GTG V CTG L IAT Y	TGC C ATA I CCTA P	CAT H CTT L ATG	GTG V TCA S GTC V	CCA P GAA E CTG L	GGG G GAT D GGT G	CTG, L CGG, R GCC(A	AAGA K AGA R CAG Q	AAG K CAA Q CAC H
757 253 841 281 925 309 1009	R ATG M GTG V TTT	AGI S CTG L AGG R CAC	GAG E AGG R CTT L	TCC S ACA T GGA GGA	TGG. W TGT C GAC. D	AAC N GCA A ACC T CAT	CTG L GTC V CAG Q TAC	AAG K CAC. H CAG. Q TGG	GAC D ATC I AGC S GAG	CTG L ACT T ATA I GTA	GAT D CTG L CCT P GAT	GATT I GAT D GGA G GTG	ACC T CCA P AAT N ACA	GAC. D GAA E .GGA	CCA P ACA T GAG E	GAAG E GCC/ A AGA? R GAGG	CTC. L AAT N TTT F	AGG. R CCG [*] P GAT. D	AGT S TGG W AGT S GAC	GTG V CTG L IAT Y	TGC C ATA I CCT/ P	CAT H CTT L ATG M	GTG V TCA S CTC V TGC	CCA P GAA E CTG L	GGG G GAT D GGT G GAC	CTG, L CGG, R GCC(A	AAGA K AGA R CAG Q GTG	AAG K CAA Q CAC H CGC
757 253 841 281 925 309 1009	R ATG M GTG V	AGI S CTG L AGG R CAC	GAG E AGG R CTT L	TCC S ACA T GGA GGA	TGG. W TGT C GAC. D	AAC N GCA A ACC T CAT	CTG L GTC V CAG Q TAC	AAG K CAC. H CAG. Q TGG	GAC D ATC I AGC S GAG	CTG L ACT T ATA I GTA	GAT D CTG L CCT P GAT	GATT I GAT D GGA G GTG	ACC T CCA P AAT N ACA	GAC. D GAA E .GGA	CCA P ACA T GAG E	GAAG E GCC/ A AGA? R GAGG	CTC. L AAT N TTT F	AGG. R CCG [*] P GAT. D	AGT S TGG W AGT S GAC	GTG V CTG L IAT Y	TGC C ATA I CCT/ P	CAT H CTT L ATG M	GTG V TCA S CTC V TGC	CCA P GAA E CTG L	GGG G GAT D GGT G GAC	CTG, L CGG, R GCC(A	AAGA K AGA R CAG Q GTG	AAG K CAA Q CAC H CGC
757 253 841 281 925 309 1009 337	R ATG M GTG V TTT F	AGI S CTG L AGG R CAC H	GAG E AGG R CTT L TCT S	TCC S ACA T GGA G GGA G	TGG W TGT C GAC D AAA	AAC N GCA A ACC T CAT	CTG L GTC V CAG Q TAC Y	AAG K CAC H CAG Q TGG	GAC D ATC I AGC S GAG E	CTG L ACT T ATA I GTA	GAT D CTG L CCT P GAT D	GATT I GAT D GGA G GTG	ACC T CCA P AAT N ACA T	GAA GAA GAA GGAA G	CCA P ACA T GAG E AAG K	GAAG E GCC/ A AGA? R GAGG E	CTC. L AAT N TTT F GCC	AGG. R CCG ⁴ P GATA D IGGG	AGT S TGG W AGT S GACC D	GTG V CTG L IAT Y CTG L	TGC C ATA I CCT P GGT G	CAT H CTT L ATG M STC: V	GTG V TCA S CTC V C	CCA P GAA E CTG L CTG R	GGGG G GAT G GGT G GAC D	CTG L CGG R GCC A TCT S	AAGA K AGAG R CAGG Q GTGG V	AAG K CAA Q CAC H CGC R
757 253 841 281 925 309 1009 337	R ATG M GTG V TTT F AGG	AGI S CTG L AGG R CAC H	GAG E AGG R CTT L TCT S GGG	TCC S ACA T GGA G GGA G CAC	TGG W TGT C GAC D AAA K TTT	AAC N GCA A ACC T CAT H	CTG L GTC V CAG Q TAC Y	AAG K CAC. H CAG. Q TGG W	GAC D ATC I AGC S GAG E TCC	CTG L ACT T ATA I GTA V	GAT D CTG L CCT P GAT D	GAT D GGA G GTG V GGC	ACC T CCA P AAT N ACA T	GAA GAA CAA CAA CAA CAA C C C C C C C C	CCAC P ACAC T GAG E AAGO K	GAAG E GCC/ A AGA R GAGG E ATTT	CTC. L AATO N TTTO F GCCT A	AGG. R CCG ⁴ P GAT. D IGGG W	AGT S TGG W AGT S GACO D	GTG V CTG L TAT Y CTG L	TGC C ATAC I CCT P GGTC G	CAT H CTT L ATG M GTC V	GTG V TCA S GTC V FGC C	CCA P GAA E CTG L CTG L AGA R	GGGG GAT D GGT G GAC D GAG	CTG L CGG R GCC A FCT S GCT	AAGA K AGAC R CAGC Q GTGC V GGCA	AAG K CAA Q CAC H CGC R
757 253 841 281 925 309 1009 337	R ATG M GTG V TTT F	AGI S CTG L AGG R CAC H	GAG E AGG R CTT L TCT S GGG	TCC S ACA T GGA G GGA G CAC	TGG W TGT C GAC D AAA K TTT	AAC N GCA A ACC T CAT H	CTG L GTC V CAG Q TAC Y	AAG K CAC. H CAG. Q TGG W	GAC D ATC I AGC S GAG E TCC	CTG L ACT T ATA I GTA V	GAT D CTG L CCT P GAT D	GAT D GGA G GTG V GGC	ACC T CCA P AAT N ACA T	GAA GAA CAA CAA CAA CAA C C C C C C C C	CCAC P ACAC T GAG E AAGO K	GAAG E GCC/ A AGA R GAGG E ATTT	CTC. L AATO N TTTO F GCCT A	AGG. R CCG ⁴ P GAT. D IGGG W	AGT S TGG W AGT S GACO D	GTG V CTG L TAT Y CTG L	TGC C ATAC I CCT P GGTC G	CAT H CTT L ATG M GTC V	GTG V TCA S GTC V FGC C	CCA P GAA E CTG L CTG L AGA R	GGGG GAT D GGT G GAC D GAG	CTG L CGG R GCC A FCT S GCT	AAGA K AGAC R CAGC Q GTGC V GGCA	AAG K CAA Q CAC H CGC R
757 253 841 281 925 309 1009 337	R ATG M GTG V TTT F AGG	AGI S CTG L AGG R CAC H	GAG E AGG R CTT L TCT S GGG	TCC S ACA T GGA G GGA G CAC	TGG W TGT C GAC D AAA K TTT	AAC N GCA A ACC T CAT H	CTG L GTC V CAG Q TAC Y	AAG K CAC. H CAG. Q TGG W	GAC D ATC I AGC S GAG E TCC	CTG L ACT T ATA I GTA V	GAT D CTG L CCT P GAT D	GAT D GGA G GTG V GGC	ACC T CCA P AAT N ACA T	GAA GAA CAA CAA CAA CAA C C C C C C C C	CCAC P ACAC T GAG E AAGO K	GAAG E GCC/ A AGA R GAGG E	CTC. L AATO N TTTO F GCCT A	AGG. R CCG ⁴ P GAT. D IGGG W	AGT S TGG W AGT S GACO D	GTG V CTG L TAT Y CTG L	TGC C ATAC I CCT/ P GGTC G	CAT H CTT L ATG M GTC V	GTG V TCA S GTC V FGC C	CCA P GAA E CTG L CTG L AGA R	GGGG GAT D GGT G GAC D GAG	CTG L CGG R GCCC A FCTC S GCTC	AAGA K AGAC R CAGC Q GTGC V GGCA	AAG K CAA Q CAC H CGC R
757 253 841 281 925 309 1009 337	R ATG M GTG V TTT F AGG R	AGT S CTG L AGG R CAC H AAG K	GAG E AGG R CTT L TCT S GGG G	TCC S ACA T GGA G GGA G CACC H	TGG W TGT C GAC D AAAA K TTT F	AAC N GCA A ACC T CAT H L	CTG L GTC V CAG Q TAC Y CTT. L	AAG K CAC. H CAG. Q TGGG W AGT S	GAC D ATC. I AGC. S GAG E TCC. S	CTG L ACT T ATA I GTA V AAG	GAT D CTG L CCT P GAT D S	GATT I GGAT D GGGA G G G G G G G G G G G	ACC T CCA P AAT N ACA T TTC F	GGAA GGAA GGAA GGGA G TGGA W	CCAA P T GAGA E AAGG K ACAA T	GAAA E GCCA A AGA3 R GAG4 E ATT1 I	CTC. L AATO N TTTO F GCCC A IGG7 W	AGG. R CCCG [*] P GATA D IGGG W ITG7 L	AGT S TGG W AGT S GAC D D GAC D TGG A	GTG V CTG L TAT Y CTG L N	TGC C ATA I CCT/ P GGT G G G K	CAT H CTT L ATG M GTC V CAAJ Q	GTG V TCA S GTC V V C C XAA? K	CCA P GAA E CTG L CTG L AGA R R TATC Y	GGG G GAT D GGT G GAC D GAG C E	CTG. L CGGG, R GCCC A FCTC S GCTC A	AAGA K AGAG R CAGG Q GTGG V GGGCA G	AAG K CAA Q CAC H CAC R CAC R CAC T
757 253 841 281 925 309 1009 337 1093 365 1177	R ATG M GTG V TTT F AGG R	AGT S CTG L AGG R CAC H AAGG K CCC	CAGG E CAGG R CTTT L TCT S GGG G CAG	TCC S ACA T GGA G GGA G CAC H ACT	TGG W TGT C GAC D AAAA K TTT F	AAC N GCA A ACC T CAT H TTG L CTC	CTG L GTC V CAG Q TAC Y CTT. L CAC	AAG K CAC. H CAG. Q TGGG W AGT S	GAC D ATC. I AGC. S GAG E TCC. S CAG	CTG L ACT T ATA I GTA V AAG K GTG	GAT D CTG L CCT P GAT D S CCT	GATT I GGAT D GGGA G G G G G G G C C A	ACC T CCA P AAT N ACA T TTC F	GAAC GAAC D GGAA G GGA G TGGA W CAAG	CCAI P ACAI T GAGJ E AAGG K ACAJ T GTTC	GAAA E GCC2/A AGAA R GAGAA E ATTT I GGGGA	CTC. L AATO N TTTTO F GCCC A TGGT W	AGG. R CCG [°] P GATA D TGGG W TTGT L	AGT S TGG W AGT S GAC D TGG W CTGC	GTG V CTG L TATO Y CTGO L N SAC/	TGC C ATA I CCTA P GGTC G G AAAAC K	CAT H CTT L ATG M GTC V CAAJ Q GAG	GTG V TCA S GTC V V TGC C K S GCTC	CCA P GAA E CTG L CTG L A GAA R TATC Y GGC/	GGGG GAT D GGT G GAC D GAC E SAGG	CTG, L CGG, R SGCC(A SGCT(A STCT	AAGA K AGAA R CAGC Q GTGC V GGCA G	AAG K CAA Q CAC H CGC R CGC T T TC
757 253 841 281 925 309 1009 337 1093 365 1177	R ATG M GTG V TTT F AGG R TAC	AGT S CTG L AGG R CAC H AAGG K CCC	CAGG E CAGG R CTTT L TCT S GGG G CAG	TCC S ACA T GGA G GGA G CAC H ACT	TGG W TGT C GAC D AAAA K TTT F	AAC N GCA A ACC T CAT H TTG L CTC	CTG L GTC V CAG Q TAC Y CTT. L CAC	AAG K CAC. H CAG. Q TGGG W AGT S	GAC D ATC. I AGC. S GAG E TCC. S CAG	CTG L ACT T ATA I GTA V AAG K GTG	GAT D CTG L CCT P GAT D S CCT	GATT I GGAT D GGGA G G G G G G G C C A	ACC T CCA P AAT N ACA T TTC F	GAAC GAAC D GGAA G GGA G TGGA W CAAG	CCAI P ACAI T GAGJ E AAGG K ACAJ T GTTC	GAAA E GCC2/A AGAA R GAGAA E ATTT I GGGGA	CTC. L AATO N TTTTO F GCCC A TGGT W	AGG. R CCG [°] P GATA D TGGG W TTGT L	AGT S TGG W AGT S GAC D TGG W CTGC	GTG V CTG L TATO Y CTGO L N SAC/	TGC C ATA I CCT P GGT G G G C K	CAT H CTT L ATG M GTC V CAAJ Q GAG	GTG V TCA S GTC V V TGC C K S GCTC	CCA P GAA E CTG L CTG L A GAA R TATC Y GGC/	GGGG GAT D GGT G GAC D GAC E GAG C E	CTG, L CGG, R SGCC(A SGCT(A STCT	AAGA K AGAA R CAGC Q GTGC V GGCA G	AAG K CAA Q CAC H CGC R CGC T T TC
757 253 841 281 925 309 1009 337 1093 365 1177	R ATG M GTG V TTT F AGG R TAC Y	AGT S CTG L AGG R CAC H AAGG K CCC P	GAG E AGG R CTT L TCT S GGG G CAG. Q	TCC S ACA T GGA G GGA G CACC H ACT T	TGG W TGT C GAC D AAAA K TTT F CCCC P	AAC N GCA A A CCT H TTG L CTC L	CTG L GTC V CAG Q TAC Y CTT, L CAC H	AAG K CAC. H CAG. Q TGG W AGT S CTT(L	GAC D ATC. I AGC. S GAG E TCC. S CAG(Q	CTG L ACT T ATA I GTA V AAG K GTG V	GAT D CTG L CCT P GAT D AGT S S CCT P	GGAT D GGAT D GGGA G G G G G G G C C A P	ACC T CCA P AAT N ACA T TTC F IGC C	GGAC D GGAA G GGAA G TGGJ W CAAG Q	CCAI P ACAI T GAGJ E AAGG K T T GTTC V	GAAI E GCC/ A AGA: R GAGG E GGGA G	CTC. L AAAT F GCCC: A IGG7 W ATT7 I	AGG. R CCCG ^P GATA D IGGG W ITGT L TTCC F	AGT S TGGG W AGT S GACC D TGG U CTGC L	GTG V CTG, L TATO Y CTGO L SACI D	TGC C ATA I CCTA P GGT(G G AAAA(K K TAT(Y	CAT H CTT L ATG M GTC V CAAJ Q GAGC E	GTG V TCA S GTC V V C C K GCTC A	CCA P GAA E CTG L CTG L AGA R R TATC Y GGC Z	GGGG GAT D GGT G GAC D GAC SAGC E ATGC M	CTG, L CGG, R GCCC A FCTC V	AAGA K AGAG R CAGG Q GTGG V GGGCA G CCCT S	AAG K CAA Q CAC H CAC H CAC R CAC T T T C F
757 253 841 281 925 309 1009 337 1093 365 1177 393 1261	R ATG M GTG V TTT F AGG R TAC Y	AGT S CTG L AGG R CAC H AAG K CCC P	GAG E CAGG R CTT L TCT S GGG G G CAG Q ATC	TCC S ACA T GGA G GGA G CACC H ACT T	TGG. W TGT C GAC. D AAAA K TTT F CCCC P GACC	AAC N GCA A ACC T CAT H TTG L CTC L CAT	CTG L GTC V CAG Q TAC Y CTT L CAC H	AAG K CAC. H CAG. Q TGGG W AGT S CTTC L	GAC D ATC. I AGC. S GAG E TCC. S CAG(Q CTC.	CTG L ACT T ATA I GTA V AAG V STG V	GAT D CTG L CCT P GAT D S CCT P	ATT I GGAT D GGGA G G G G G G G G C C A Z P T C C Z	ACC T CCA P AAT N ACA T TTC F C C TTC	CAAC CAAC CAAC CAAC CAAC Q TCCCC	CCAI P ACAI T GAGI E AAGG K T T GTTC V GAA1	GAAA E GCC/A A AGAA R GAGAA E ATTT I I GGGGA G	CTC. L AATO N TTTTO F GCCC A U GCCC I GCCCI	AGG. R CCCG [*] P GATA D IGGG W TTG7 L TTCC F	AGT S TGGG W AGT S GACC D TGGA W CTGC L	GTG V CTG, L TATO Y CTGO L SACI D GGAO	TGC C ATAA I CCTA P GGTC G AAAAC K FATC Y	CAT H CTT L ATG M GTC V CAAJ Q GAGC E CTGC	GTG V TCA S GTC V C C C C C C C C C C C C C C C C C C	CCA P GAAA E CTG L AGAA R TATO Y GGCA G CCCI	GGGG G GAT D GGT G GGAC C D GGAC E SAGC E M TTCT	CTG, L CGG, R GCCC A FTCT S GCTC A GTCT V TTCA	AAGA K AAGAG R CAGG Q GGTGC V GGGCA G G CCCI S	AAG K CAA Q CAC H CAC R CGC T T CTC F
757 253 841 281 925 309 1009 337 1093 365 1177 393 1261	R ATG M GTG V TTT F AGG R TAC Y	AGT S CTG L AGG R CAC H AAG K CCC P	GAG E CAGG R CTT L TCT S GGG G G CAG Q ATC	TCC S ACA T GGA G GGA G CACC H ACT T	TGG. W TGT C GAC. D AAAA K TTT F CCCC P GACC	AAC N GCA A ACC T CAT H TTG L CTC L CAT	CTG L GTC V CAG Q TAC Y CTT L CAC H	AAG K CAC. H CAG. Q TGGG W AGT S CTTC L	GAC D ATC. I AGC. S GAG E TCC. S CAG(Q CTC.	CTG L ACT T ATA I GTA V AAG V STG V	GAT D CTG L CCT P GAT D S CCT P	ATT I GGAT D GGGA G G G G G G G G C C A Z P T C C Z	ACC T CCA P AAT N ACA T TTC F C C TTC	CAAC CAAC CAAC CAAC CAAC Q TCCCC	CCAI P ACAI T GAGI E AAGG K T T GTTC V GAA1	GAAA E GCC/A A AGAA R GAGAA E ATTT I I GGGGA G	CTC. L AATO N TTTTO F GCCC A U GCCC I GCCCI	AGG. R CCCG [*] P GATA D IGGG W TTG7 L TTCC F	AGT S TGG W AGT S GACC D TGGA W CTGC L	GTG V CTG, L TATO Y CTGO L SACI D GGAO	TGC C ATAA I CCTA P GGTC G AAAAC K FATC Y	CAT H CTT L ATG M GTC V CAAJ Q GAGC E CTGC	GTG V TCA S GTC V C C C C C C C C C C C C C C C C C C	CCA P GAAA E CTG L AGAA R TATO Y GGCA G CCCI	GGGG G GAT D GGT G GGAC C D GGAC E SAGC E M TTCT	CTG, L CGG, R GCCC A FTCT S GCTC A GTCT V TTCA	AAGA K AAGAG R CAGG Q GGTGC V GGGCA G G CCCI S	AAG K CAA Q CAC H CAC R CGC T T CTC F
757 253 841 281 925 309 1009 337 1093 365 1177 393 1261 421	R ATG M GTG V TTT F AGG R TAC Y TAC. Y	AGT S CTG L AGG R CAC H AAGG K CCC P AAC N	GAG E AGG R CTT L TCT S GGG G CAG Q ATC. I	TCC S ACA T GGA G GGA G CACC H ACT T	TGG. W TGT C GAC. D AAAA K TTT F C CCCC P GACC D	AAC N GCA A CAT H TTG L CTC L CATC H	CTG L GTC V CAG Q TAC Y CTT L CAC H H G G	AAG K CAC. H CAG. Q TGGG W AGT S CTTC L	GAC D ATC. I AGC. S GAGG E TCC. S CAGG Q CTC. L	CTG L ACT T ATA I GTA V AAGG K GTG V V AATC I	GAT D CTG L CCT P GAT D S CCT P CCT P	GGAT I GGAT GGGA G GGGC G G CCA S	ACC T CCA P AAT N ACA T TTC C F F	GGAA GGAA GGAA G GGAA G CAAG Q TCTC S	CCAA P ACAA T GAGA K AAAGO K T GTTC V GAAJ E	GAAA E GCC/ A AGAA R GAGA E ATTTI I GGGA G C	CTC. L AAT' N TTTC F GCCC A I GCCCI A	AGGA R CCCG ^P D GATA D F GGATA D T GGATA D T T GGATA D T T GGATA D T T GGATA D T GGATA D T GGATA D T GGATA D T GGATA D T GGATA D T G G T T D T G G G T D T G G G T D T G G G G	AGT S TGG W AGT S GAC(D CTGC L CTGC L ACAC	GTG V CTG L TAT V CTG L CTG L N SAC7 D GGAC G	TGC C ATAA I CCT/ P GGT(G AAAA(K FAT(Y CCTC P	CAT H CTT L ATGO M GTC V CAAM Q GAGO E CTGO L	GTG V TCAI S GTCI V C TGCI C AAA2 K S GCTC A C GGCC R	CCA P GAA E CTG L CTG R R FATC Y G G CCCT P	GGGG GAT D GGT G GAC D GAG E ATG M F	CTG, L CGGG, R GCCC A F CTCA F	AAGA K AGAG Q GTGC V GGCA G CCCI S S AGTC S	AAG K CAA Q CAC H CGC R ACC T T C F T C F
757 253 841 281 925 309 1009 337 1093 365 1177 393 1261 421 1345	R ATG M GTG V TTT F AGG R TACC Y GGT	AGT S CTG L AGG R CAC H AAG K CCCC P AAAC N	GAG E AGG R CTT L TCT S GGG G CAG. Q ATC. I	TCC S ACA T GGA G GGA G CACT H ACT T T GAT	TGG. W TGT C GAC. D AAAA K TTT F CCCC P GACC D GGAC	AAC N GCA T CAT H TTG L CTC L CAT H GGA	CTG L GTC V CAG Q TAC Y CTT L CAC H B G G	AAG K CAC. H CAG. Q TGGG W S CTTC L S CTTC S	GAC D ATC. I AGC. S GAG E TCC. S CAGC Q CTC. L	CTG L ACT T AATA I GTA V AAG K STG V V AATC I GCCC	GAT D CTG L CCT P GAT D S CCT P FACT Y	GATT I GATT D GGA G G G G G G G C C C T C C T C C T C C	ACC T CCA P AAT N ACA T TTC F IGC C	GAA GAA GGAA GGAA G GGAA G CAA Q CAA C CAA C S CTCT	CCAA P T GAGA E AAAGO K T T GAAJ E CGTC	GAAA E GCC/ A AGAA R GGGAA G GGGAA C C CCAC	CTC. L AAAT F GCCC A IGG7 W ATT1 I GCC7 A CTGA	AGG. R CCCG [*] P GAT. D GAT. D TTGC F TTCC F	AGT S TGG W AGT S GAC D D TGG L CTG C L CTG C T S	GTG V CTG L TAT Y CTG L SAC D GGAC G GGA1	TGC C ATAA I CCT/ P GGT(G AAAA K TAT(Y CCT(P CCAC	CAT H CTT L ATG M GTC V CAAA Q GAG C E CTG C L CAAG	GTG V TCAA S GTCA V C GTCA C C AAA7 K SGCTCA R SGGCA I	CCA P GAA E CTG L CTG R R G G G G G CCCA	GGGG GAT' D GGT' G GAC' D GAGC E ATGC M F	CTG. L CGG. R GCCC A F CTCA F GCTC C A F GACT	AAGA K AGAG Q GTGC V GGCA G CCTI S AGTC S CATT	AAAG K Q CAAC H CCC R CCC T T CCC F CCT P GA
757 253 841 281 925 309 1009 337 1093 365 1177 393 1261 421 1345	R ATG M GTG V TTT F AGG R TACC Y GGT	AGT S CTG L AGG R CAC H AAG K CCCC P AAAC N	GAG E AGG R CTT L TCT S GGG G CAG. Q ATC. I	TCC S ACA T GGA G GGA G CACT H ACT T T GAT	TGG. W TGT C GAC. D AAAA K TTT F CCCC P GACC D GGAC	AAC N GCA T CAT H TTG L CTC L CAT H GGA	CTG L GTC V CAG Q TAC Y CTT L CAC H B G G	AAG K CAC. H CAG. Q TGGG W S CTTC L S CTTC S	GAC D ATC. I AGC. S GAG E TCC. S CAGC Q CTC. L	CTG L ACT T AATA I GTA V AAG K STG V V AATC I GCCC	GAT D CTG L CCT P GAT D S CCT P FACT Y	GATT I GATT D GGA G G G G G G G C C C T C C T C C T C C	ACC T CCA P AAT N ACA T TTC F IGC C	GAA GAA GGAA GGAA G GGAA G TGGA Q CAA C CAA C S CTCT	CCAA P T GAGA E AAAGO K T T GAAJ E CGTC	GAAA E GCC/ A AGAA R GGGAA G GGGAA C C CCAC	CTC. L AAAT F GCCC A IGG7 W ATT1 I GCC7 A CTGA	AGG. R CCCG [*] P GAT. D GAT. D TTGC F TTCC F	AGT S TGG W AGT S GAC D D TGG L CTG C L CTG C T S	GTG V CTG L TAT Y CTG L SAC D GGAC G GGA1	TGC C ATAA I CCT/ P GGT(G AAAA K TAT(Y CCT(P CCAC	CAT H CTT L ATG M GTC V CAAA Q GAG C E CTG C L CAAG	GTG V TCAA S GTCA V C GTCA C C AAA7 K SGCTCA R SGGCA I	CCA P GAA E CTG L CTG R R G G G G G CCCA	GGGG GAT' D GGT' G GAC' D GAGC E ATGC M F	CTG. L CGG. R GCCC A F CTCA F GCTC C A F GACT	AAGA K AGAG Q GTGC V GGCA G CCTI S AGTC S CATT	AAG K CAA Q CAC H CGC R ACC T T C F T C F
757 253 841 281 925 309 1009 337 1093 365 1177 393 1261 421 1345 449	R ATG M GTG V TTT F AGG R TAC Y GGT G	AGI S CTG L AGG R CACC H AAGG K CCCC P AACC N TTC. F	GAG E AGG R CTT L TCT S GGG G CAG CAG I AATC N	TCC S ACA T GGA G GGA G CAC H T T ACT T T SATC	TGG. W TGT C GAC. D AAAA K TTT F CCCC P GACC D GACC G GGAC G	AAC N GCA A ACCC T CAT H CTCC L CTCC L CTCC H SCAA G	CTG L GTC V CAG Q TAC Y CTT L CAC H G G G AAAA K	AAG K CAC. H CAG. Q TGG W TGG W S CTTC L CTTC S S AACA N	GAC D ATC. I AGC. S GAG E TCC. S CAGC Q CTC. L CAGC C T	CTG L ACT T ATA I GTA V AAG V V AAG V I SCCC A	GAT D CTG L CCT P GAT D S CCT P TAC Y CCT C P	GATT I GGAT D GGGA G G G G G G G G C C A S C TAA L	ACC T CCA P AAT N ACA T T T C T T C T	GAAC GAAC D GAAA G GGAA G TGGJ W CAAA Q TCTC S CTCT L	CCAA P ACAA T GAGJ E AAGO K T GTTC C C C C C	GAAA E GCC2 A AGAA R GAGA E ATTTI I GGGGA G C C C C C C C C C C C C C C C C	CTC. L AAAT(N TTTC F GCCT A IGGT W ATTTI I GCCT A CTGA L	AGG. R CCG [°] P GATI D IGGG W FTG [°] F TTCC F	AGT S TGG W AGT S GACC D TGG L CTGC L TGC T T CTGC T	GTG V CTG L TAT Y CTG L CTG L SAC/ N SAC/ D SGAC G G G G G G G G G	TGC C ATAA I CCT/ P G G G AAAA K FATC P CCTC P CCAC S	CAT H CTT L ATG M GTC V CAAA Q GAGC E CTGC L CAAC Q	GTG V TCA S GTC V C GTC C V G G G G G G G G G G G G G G G G G	CCA P GAAA E CTG L AGAA R FATO Y GGCA G CCCA S	GGGG GAT D GGT G GAC E GAG G GAG M F F ACTC T	CTG. L CGG. R GCCC A F CTCT S GCTC A S TCT V TCA F S ACT D	AAGA K AGAG R CAGG Q GTGC V GGGCA G CCTT S AGTC S CATT Y	AAG K CAA Q CAC H CGC R T TC F TC F CT P GA * 475
757 253 841 281 925 309 337 1009 337 1093 365 1177 393 1261 421 1345 449 1429	R ATG M GTG V TTT F AGG R TAC Y GGT G GGT	AGI S CTG L AGG R CAC H AAG K CCC P AAC N TTC. F	GAG E AGG R CTT L TCT S GGG G G CAG Q ATC I N N TCT	TCC S ACA T GGA G GGA G CAC T T ACT T T SATC	TGG. W TGT C GAC. D AAAA K TTT F CCCC P GACC G GACC G GACC	AAC N GCA A CAT T CAT H TTG L CAT C H CATC H G GCA/ G	CTG. L GTC V CAG Q TAC Y CTT. L CAC H G G G K K GCC/	AAG K CAC. H CAG. Q TGGQ W AGT S CTTC L ICCCO S AACA N	GAC D ATC. I AGC. S GAG E TCC. S CAGC Q CTC. L ACAC T CTCC	CTG L ACT T ATA I GTA V AAG V AAG V SCCCA	GAT D CTG L CCT P GAT D AGT S CCT P FAC Y	GATT I GGAT D GGGA G G G G G G G G C C C A L G G G C C A L	ACC T CCA P AAT N ACA T T C T T C T T ACC C	GAAC GAAC CAAC GGAA G GGAA G CAAC Q TCTC C CTC7 L GCT1	CCAA P ACAA T GAGJ E AAGO K T GTTC C C C C C C C C C C C C C C C C	GAAA E GCCJ A GCCJ R GGCA C C C C C C C C C C C C C C C C C	CTC. L AATO N TTTO F GCCTA A TTTTI I GCCTA A CTGA L CCAC	AGG. R CCCG [*] P GATI D IGGG W F IGGG F F TTTA F AATA N CAAA	AGT S TGG4 W AGT S GAC0 D TGG4 W CTG0 L CTG0 T S ACA0 T S ACA0 T S ACA0 T	GTG V CTG L TATO Y CTGO L SGACI D SGACI G G G G G G G G G G G TGO	TGC C ATA I CCT/ P GGTC G AAAAC K FATC P CCTC S SCTC	CAT H CTT L ATG M GTC V CAAA Q GAGC E CTGC L CAAC Q CAAA Q CAAA Q CAT TTI	GTG V TCA S GTC V C GTC C A A A C G G C R C R C R C R C R C R C R C R C	CCA P GAAA E CTG L AGAA R FATC Y GGCA G G CCCA S S CCCA	GGGG GAT D GGT G GAC D GAC D GAC E M TTCI F A CTC T A TGA	CTG. L CGGG. R GCCCI A F CTCTC S GCTCI S CTCI V TTCA F S ACTI D ACTI	AAGA K AGAG Q GTGG V GGGCA G CCTG S CATT Y CCTG	AAAG K Q CAAC H CACC R CCC T T C C T C C T C C T C C T C
757 253 841 281 925 309 1009 337 1093 365 1177 393 1261 421 1345 449 1429 1513	R ATG M GTG V TTT F AGG R TAC Y GGT G GGT G GGT	AGI S CTG L AGG R CAC H AAAG K CCCC P AAAC N TTC. F CTT	GAGG E CTT L TTT S GGG G CAG. Q ATC. I N TTTG	TCC S ACA T GGA G GGA G CACC H ACT T T SAT(D CTG(CTC)	TGG W TGT C GAC D AAAA K TTT F GAC G G GGAC G GGAC G GGAC G GGAC	AAC N GCA T CAT H TTG L CTC CAT H G GGA G G ACT C CAGA	CTG L GTC V CAG Q TAC' Y CTT/ L CACC H SGC' K SGC/ K	AAG K CAC. H CAG. Q TGGQ W M TGGQ W M CTTC L ICCCC S AAC/A N ACTC CATC	GAC D ATC. I S GAG E S CAG Q CTCA L CTCA T CTCC CCGG	CTG L ACT T T ATA I GTA V V AAG V K GTG V V AAG V SGTG C CCA SATC	GAT D CTG L CCT P GAT D AGT S CCT P FAC Y CCTC P	GATT D GGAT D GGC G G G G G G G G G C C C TA L S G C C TA L S G C C A G C	ACC T CCA P AAT N ACA T T C F F C C T C C T C C T C C C C C C	CAAC CAAC CAAC CAAC CAAC CAAC CAAC CTCT L CGCTT CGAC	CCAA P ACAA T GAGA E AAAGO K T GTTC C C C C C C C C C C C C C C C C	GAAA E GCCJ A GGCGA E GGCGA C C C C C C C C C C C C C C C C C	CTC. L AATO N TTTC F GCC: A IGG? W ATTI I GCCI A CTGA CCCC CCCC CCCCCCCCCCCCCCCCCCCCCC	AGG. R CCG ^P D GAT. D IGGG W ITGC F ITTA F ATA N CAAA	AGT S TGG4 W AGT S GAC0 D TGG4 W CTG0 L CTG0 T S CTG0 L ACA0 T S ACA0 T	GTG V CTG L TAT Y CTG C CTG C G G G G G G G G G G G G G G	TGC C ATA I CCT/ P GGTC G AAAAC K FATC P CCTC S CCTC S CCTC S	CAT H CTT L ATG M GTC V CAAA Q CAAA Q SAGC E CTGC L CTGC L CAAC Q CTTI CAAC	GTG V TCA S GTC V V TGC C AAA7 K G G G G G G G G G G G G G G G G G G	CCA P GAAA E CTG L AGAA R FATC Y GGC/ P CCCA S CCCA S CCCA	GGGG GAT D GGT G GAC D GAC E SAGC E M TTCI F A CTC T A TGA	CTG. L CGG. R GCCC A F CTCA F CTCA F CCT	AAGA K AGAG Q GTGG Q GGGCA G GGCCS CATT Y CCTG GGCC	AAG K CAA Q CAC H CGC R CCC T CCC F GA * 475 AA CCC
757 253 841 281 925 309 337 1009 337 1093 365 1177 393 1261 421 1345 449 1429	R ATG M GTG V TTT F AGG R TAC Y GGT G TGGG CCAA	AGI S CTG L AGG R CAC H AAAG K CCCC P AAAC N TTC. F CTT CCT	GAG E AGG R CTT L TCT S GGG G CAG. Q AATC. I AATC. N TCTCTCTCTCTCTCTCTCCTCTCCCTTCCCCTCCCC	TCC S ACA T GGA G GGA G GCA C H T T SATC D CTGO IGTO	TGG. W TGT C GAC. D AAAA K TTT F CCCC P GACC D GGAC G GGAC G GGAC G GGAC	AAC N GCA' A CAT' H TTGO L CAT' H CTCO L CATO H G G ACTO CAGA	CTG L GTC V CAG Q TAC' Y CTT L CAC' H SGC' K SGCC K SGCC	AAGG K CAC. H CAG. Q TGGG W AGT S CTTC L CTTC L S CTTC L S AACZ N ACTC CAGI	GAC D ATC. I S GAGC E S CAGC Q CTCA C CAGC T C CCGC C GAA	CTG L ACT T T ATA I GTA V V AAG K GTG V V AAG X C CCA A CCCA SATC	GAT D CTG L CCT P GAT D D GAT D D CCT P P CCT C P CCT C P	GATT D GGAT D GGGA G G G G G G G G C C A C C TA/ L G G C C TA/ L G G C C TA/ C S C TA/ C S C TA/ C S C C C C S C C C C C C C C C C C C	ACC T CCA P AAT N ACA T T T C T T C T T C T C T T ACCC T	GAA GAA GGAA GGAA G GGAA W CAAA Q TCTC S CTC7 L GCT7 L GCT7 CGAC	CCAA P ACAA T GAGJ E AAGO K ACAA T GTTC C C C C C C C C C C C C C C C C	GAAA E GCCL A AAGA? R GGGGA G GGGGA C C C C C C C C C C C C C	CTC. L AAT F F GCCT A IGGT V ATTI I GCCT A CTGA L CCAC CCAC	AGG. R CCCG ^P D GAT. D ICGC W ITCC F ICTCC F ITTA F AATA N SAAA SGGA	AGT S TGG4 W AGT S GACC D TGG4 W CTGC L ACAC T ACCC AGT CAT	GTG V CTG L TAT Y CTG C C C G G G G G G G G G G G C C C C	TGC C ATA I CCT P G G G G AAAO K CCTC P CCAC S CCTC S CCTC CCTC CCTC CCTC CCTC	CAT H CTT L ATG M GTC V CAAA Q CAAA Q CAAA Q CAAA C CAAA C CAAA C CAAA C CAAA C C CAAA C	GTG V TCAI S GTCI V C C AAA2 C C AAA2 K G G C C AAA2 K G G C C C C C C C C C C C C C C C C C	CCA P GAA E CTG L CTG C CCG R FATC F CCCA S CCCA S CCCA S CCCA CCCA CCCA	GGGG GAT D GGT G GAC D GAGC E SAGC B GAGC T T ACTC T	CTG. L CGG. R GCCC A TCTC S GCTC A TCA F S ACT D ACT TCTTTT	AAGA K AGAG Q GTGG V GGCA G GGCA G CCCT S CATT Y CCTG CCCT TTC	AAAG K CAAA Q CAAC H CGC R CTC F CTC F GA * 475 AAA CC TG

Figure 6. Nucleotide sequence of 52FL cDNA insert and deduced amino acid sequence of the encoded 52-kD SS-A/Ro protein. Three inframe stop codons at the 5'untranslated region and two polyadenylation signal sequences are underlined. A leucine zipper motif is also underlined with the leucine residues marked with arrowheads. These sequences have been submitted to EMBL/Genbank under accession No. M35041.



Total sequence showing high degree of homology with human *rfp* protein

Sequence showing homology with human ret and mouse rpt-1 proteins

Figure 7. Diagrammatic summary of the features of the 52-kD SS-A/ Ro protein. The entire 52-kD protein sequence shows high degree of homology with human *rfp* protein. In addition, the amino-terminal half of the 52-kD protein is similar to the corresponding halves of human *ret* transforming protein and mouse *rpt*-1 gene regulatory protein (see Fig. 9).



Figure 8. Putative zinc fingers in the amino-terminal domain of 52kD SS-A/Ro protein. These finger structures are arranged based on the proposed finger structures in *rpt*-1 by Patarca et al. (34) and are further modified for the 52-kD protein. In the region encompassing residues 16-54, either one of two possible zinc fingers can be formed in the configurations depicted in A and B. Another region, residues 92-114 (C) could be the site of another zinc finger configuration (also see text).

<u>rfp</u> SS-A/Ro <u>rpt</u> -1	1 M 1 M 9	MASGSVAECLQQETTCPVCLQYFAEPMMLDCGHNICCACLARCWGTAETNVSCPQCRETFPQRHMRP MASAARLTMMWEEVTCPICLDPFVEPVSIECGHSFCQECISQV-GKGGGSV-CAVCRQRFLLKNLRP MAS-SVLEMIKEEVTCPICLELLKEPVSADCNHSFCRACITLNYESNRNTDGKGNCPVCRVPYPFGNLRP *** * . ***	67 65 69	Figure 9. Comparison of amino acid sequences generated from the CLUSTAL programs. (*) identical match and (\cdot) conservative substitution in
	66 N 70 N	NRHLANVTQLVKQLRTERPSGPGGEMGVCEKHREPLKLYCEEDQMPICVVCDRSREHRGHSVLPLEEAVE NRQLANMVNNLKEISQEAREGTQGERCAVHGERLHLFCEKDGKALCWVCAQSRKHRDHAMVPLEEAAQ NLHVANIVERLKGFKSIPEEEQKVNICAQHGEKLRLFCRKDMMVICWLCERSQEHRGHQTALIEEVDQ **** * * * * * * * * * * * * * * * *	137 133 137	amino acid residue, respectively. (A) Comparison of the amino- terminal halves of 52-kD SS-A/ Ro, human <i>rfp</i> , and mouse <i>rpt</i> -1
	134 H 138 H	GFKEQIQNQLDHLKRVKDLKKRRRAQGEQARAELLSLTQMEREKIVWEFEQLYHSLKEHEYRLLARLEEL EYQEKLQVALGELRRKQELAEKLEVEIAIKRADWKKTVETQKSRIHAEFVQQKNFLVEEEQRQLQELEKD EYKEKLQGALWKLMKKAKICDEWQDDLQLQRVDWENQIQINVENVQRQFKGLRDLLDSKENEELQKLKKE *.* .* .* .*	207 203 207	proteins. The sequence shown for <i>rfp</i> protein is identical in human <i>ret</i> transforming protein. The sequence for the 52-kD
	204 I 208 I	DLAIYNSINGAITQFSCNISHLSSLIAQLEEKQQQPTRELLQDIGDTLSRAERIRIPEPWI 268 EREQLRILGEKEAKLAQQSQALQELISELDRRCHSSALELLQEVIIVLERSESWNLKDLDI 264 KKEVMEKLEESENELEDQTELVRDLISDVEHHLELSTLEMLQGANCVLRRSQSLSLQQPQT 268	A	protein is more similar to <i>rfp</i> (33% identity and 54% similarity) and to <i>rpt</i> -1 (39% identity and 61% similarity) than the <i>rfp</i> protein is to the <i>rpt</i> -1 protein (29% identity and 52%
SS-A/Ro <u>rfp</u>	316	VHITLDPDTANPWLILSEDRRQVRLGDTQQSIPGNEERFDSYPMVLGAQHFHSGKHYWEVDVTGKEAWDL VDVTLDPDTAYPSLILSDNLRQVRYSYLQQDLPDNPERFNLFPCVLGSPCFIAGRHYWEVEVGDKAKWTI ******** * **** **** *** *** *** * *** * .*.*.*.*.*.*.*.*.*.*.*.*.*.	356 385	similarity). (B) Comparison of the carboxy-terminal halves of 52-kD SS-A/Ro and human rfp
	386 (GVCRDSVRRKGHFLLSSKSGFWTIWLWNKQKYEAGTYPQTPLHLQVPPCQVGIFLDYEAGMVSFYNITDH GVCEDSVCRKGGVTSAPQNGFWAVSLWYGKEYWALTSPMTALPLRTPLQRVGIFLDYDAGEVSFYNVTER *** *** ******** ** * * * *.** .******	426 455	proteins. > 50% of amino acid residues are identical, and there is a 65% similarity between the two sequences. Percent identity
		GSLIYSFSECAFTGPLRPFFSPGFNDGGKNTAPLTLCPLNIGSQGSTDY 475 -CHTFTFSHATFCGPVRPYFSLSYS-GGKSAAPLIICPMSGIDGFSGHVGNHGHSMETSP 513 **. * **.** ** *****.	В	and similarity are obtained from the GAP program (23).

sequence might suggest DNA and/or RNA binding activities although experimental evidence for the direct association of the 52-kD protein with DNA/RNA is still lacking. The molecular characteristics and cDNA sequence described in this report provides a precise definition for the 52-kD SS-A/Ro protein. It is necessary to mention that due to a sequencing error, an inframe carboxy-terminal region rich in Ser/Thr/Pro amino acids was reported to be present in the 52-kD clone (36). The error was detected and the corrected sequence is presented in this report.

Other SS-A/Ro protein sequences. There were three previous cDNA sequences reported for proteins of the SS-A/ Ro complex. Deutscher et al. (7) and Ben-Chetrit et al. (17) both described the cloning and sequence of cDNA for the 60kD SS-A/Ro protein. Their corresponding deduced protein sequences were identical except for short regions in their carboxy-termini. Both recombinant proteins were recognized by human autoantibodies in several different immunoassay formats (7, 17, unpublished data). The common region between the two deduced protein sequences contained a putative zinc finger and a RNA-binding protein consensus motif that could account for the direct interaction with hY-RNAs. Deutscher et al. (7) were able to reconstitute in vitro ribonucleoprotein complexes composed of hY1 RNA and recombinant 60-kD SS-A/ Ro protein. Because the two cDNAs were obtained from different human cells, the differences in sequence could be a result of differential expression of separate SS-A/Ro genes or differential mRNA splicing. A third sequence was reported initially by Lieu et al. (37) describing the amino-terminal 24 residues determined by standard protein sequencing of a purified 60-kD protein from Wil-2 cells. A synthetic peptide corresponding to this amino-terminal region was reported to be recognized by SS-A/Ro positive autoimmune sera (37); it was thus proposed that a major autoepitope resided within this amino-terminal region (37). McCauliffe et al. (38) from the same laboratory recently described a cDNA encoding this protein. The deduced amino acid sequence appears to be a human analogue of the rabbit and mouse calreticulin (calregulin), a high-affinity calcium binding protein present in the lumen of endoplasmic reticulum (39). It is unclear whether the protein of McCauliffe et al. (38) is another species of SS-A/Ro antigen because the reactivity of the recombinant protein with human anti-SS-A/Ro sera was not presented (38). The 52-kD SS-A/Ro protein described in this report is distinct from the cDNA sequence reported by McCauliffe and from other sequences reported to date.

Homology with ret, rfp, and rpt-1 proteins. In 1985, the activation of a transforming gene ret was first detected during transfection of NIH 3T3 cells with DNA from a human T lymphoma (40). The ret gene was cloned from transformed NIH cells by hybridization with human sequence probes and was shown to be a result of DNA rearrangement (40). Later, it was found that the ret gene was a fusion between two unlinked segments of human DNA encoding for an amino-terminal finger domain and a carboxy-terminal tyrosine kinase domain, respectively (35). This kinase domain was preceded by a hydrophobic transmembrane sequence and thus the ret protein was thought to be a cell surface receptor (35). A cellular homologue corresponding to the amino-terminal half of ret was cloned and named rfp (ret finger protein) and we show that the latter has a high degree of homology with the 52-kD SS-A/Ro protein (Fig. 9). A nuclear localization signal sequence was found in the rfp sequence, suggesting that it might be a nuclear protein. Its mRNA level was 20 times higher in testis than in liver and kidney and this observation led to a proposed functional role in male germ cell development (13). It was also noted that the

mRNA level was highest in 11.5-d mouse embryos as compared to other stages (13). The fact that rfp is part of a transforming gene ret raises the question whether the 52-kD protein might have similar transforming potential. The second protein that the 52-kD SS-A/Ro protein was found to share similarity with was rpt-1 (for regulatory protein of T-lymphocyte [1]). Rpt-1 was first described by Patarca et al. (34) as a protein selectively expressed by resting but not activated CD4+ inducer T cells. Rpt-1 is a 41-kD nuclear protein that down-regulates gene expression of IL-2 receptor α -chain gene and human immunodeficiency virus type 1 genes. Sequence similarities among the 52-kD SS-A/Ro, rfp, and rpt-1 proteins can be useful when the function of an index protein such as the 52-kD SS-A/Ro is unknown, but it is well acknowledged that such interpretations should be taken with caution. Nevertheless, it provides new avenues for future investigations into the possible functional roles of the 52-kD SS-A/Ro protein.

Because the report of Ben-Chetrit et al. (4) showing that SS-A/Ro autoantibody targets consisted of at least two components of 52 and 60 kD, the results of many studies based on the assumption that the 60-kD protein was the only SS-A/Ro autoantigen would have to be reevaluated. Buyon et al. (41) recently showed that in neonatal lupus syndrome with congenital heart block, all 20 mothers of permanently affected infants had antibodies to either SS-A/Ro or SS-B/La antigens. The predominant antibody response was to the 52-kD SS-A/Ro antigen (41). In another report by Ben-Chetrit et al. (42), it was shown that although most SS-A/Ro-positive autoimmune sera had reactivity to both the 52- and 60-kD antigens in immunoblotting, antibody to the 52-kD antigen without concomitant antibody to the 60-kD antigen was seen only in patients with primary Sjogren's syndrome whereas antibody to the 60-kD antigen without concomitant antibody to the 52-kD antigen was seen only in patients with systemic lupus erythematosus (42). The latter observation that there was a dissociation of immune responses to the two SS-A/Ro components might suggest different events stimulating the autoimmune process in these diseases. The current report on the sequence and molecular characteristics of the 52-kD protein establishes its identity and might be useful in elucidating its function and that of the multicomponent SS-A/Ro complex. Such information could contribute to understanding some features of this autoimmune reaction.

Acknowledgments

This is publication 6400-MEM from the Research Institute of Scripps Clinic. This work is supported by Grants AR32063 and AI10386 from the National Institutes of Health. Dr. Chan is a recipient of an Arthritis Foundation Investigator Award.

Note added in proof: Recently we have detected protein sequence similarity between the 52-kD SS-A/Ro protein and the 55-kD RAD18 gene product of *Saccharomyces cerevisiae* (Chanet et al. 1988. *Gene (Amst.).* 74:543–547). The region of similarity is restricted to NH₂-terminal 54 amino acid residues, and the Cys residues of the zinc finger domain are conserved.

References

1. Tan, E. M. 1989. Antinuclear antibodies: diagnostic markers for autoimmune diseases and probes for cell biology. *Adv. Immunol.* 44:93-151. 2. Buyon, J. P., and R. Winchester. 1990. Congenital complete heart block. A human model of passively acquired autoimmune injury. *Arthritis Rheum*. 33:609-614.

3. Provost, T. T., and M. Reichlin. 1988. Immunologic studies of cutaneous lupus erythematosus. J. Clin. Immunol. 8:223-233.

4. Ben-Chetrit, E., E. K. L. Chan, K. F. Sullivan, and E. M. Tan. 1988. A 52-kD protein is a novel component of the SS-A/Ro antigenic particle. J. Exp. Med. 167:1560-1571.

5. Wolin, S. L., and J. A. Steitz. 1984. The Ro small cytoplasmic ribonucleoprotein: identification of the antigenic protein and its binding site on the Ro RNAs. *Proc. Natl. Acad. Sci. USA*. 81:1996-2000.

6. Boire, G., and J. Craft. 1989. Biochemical and immunological heterogeneity of the Ro ribonucleoprotein particles. Analysis with sera specific for the Ro^{bYS} particles. J. Clin. Invest. 84:270–279.

7. Deutscher, S. L., J. B. Harley, and J. D. Keene. 1988. Molecular analysis of the 60-kDa human Ro ribonucleoprotein. *Proc. Natl. Acad. Sci. USA*. 85:9479– 9483.

8. Buyon, J. P., S. G. Slade, E. K. L. Chan, E. M. Tan, and R. Winchester. 1990. Effective separation of the 52kD SS-A/Ro polypeptide from the 48kD SS-B/La polypeptide by altering conditions of polyacrylamide gel electrophoresis. J. Immunol. Methods. 129:207-210.

9. Rader, M. D., C. O'Brien, Y. Liu, J. B. Harley, and M. Reichlin. 1989. Heterogeneity of the SS-A/Ro antigen: different molecular forms in lymphocytes and red blood cells. J. Clin. Invest. 83:1293-1298.

10. Laemmli, U. K. 1970. Cleavage of structural protein during the assembly of the head of bacteriophage T4. *Nature (Lond.).* 227:680-685.

11. Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets. Procedures and applications. *Proc. Natl. Acad. Sci. USA*. 76:4350-4354.

12. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* 132:6–13.

13. Takahashi, M., Y. Inaguma, H. Hiai, and F. Hirose. 1988. Developmentally regulated expression of a human "finger"-containing gene encoded by the 5' half of the ret transforming gene. *Mol. Cell. Biol.* 8:1853–1856.

14. Adam, S. A., T. Nakagawa, M. S. Swanson, T. K. Woodruff, and G. Dreyfuss. 1986. mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell Biol.* 6:2932–2943.

15. Rubin, R. L. 1986. Enzyme-linked immunosorbent assay for anti-DNA and anti-histone antibodies. *In* Manual of Clinical Laboratory Immunology. 3rd ed. N. R. Rose, H. Friedman, and J. L. Fahey, editors. American Society for Microbiology, Washington, DC. 744–748.

 Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

17. Ben-Chetrit, E., B. J. Gandy, E. M. Tan, and K. F. Sullivan. 1989. Isolation and characterization of a cDNA clone encoding the 60-kD component of the human SS-A/Ro ribonucleoprotein autoantigen. J. Clin. Invest. 83:1284–1292.

 Chan, E. K. L., A. M. Francoeur, and E. M. Tan. 1986. Epitopes, structural domains and asymmetry of amino acid residues in SS-B/La nuclear protein. *J. Immunol.* 136:3744–3749.

19. Garrels, J. I. 1983. Quantitative two-dimensional gel electrophoresis of proteins. *Methods Enzymol.* 100:411-423.

20. Chen, E. Y., and P. H. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA. 4:165-170.

21. Sanger, F. 1981. Determination of nucleotide sequence in DNA. Science (Wash. DC). 214:1205-1210.

22. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387-395.

23. Needleman, S. B., and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48:443–453.

24. Higgins, D. G., and P. M. Sharp. 1988. CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. *Gene (Amst.)*. 73:237-244.

25. Higgins, D. G., and P. M. Sharp. 1989. Fast and sensitive multiple sequence alignments on a microcomputer. *Comp. Appl. Biosci.* 5:151-153.

26. Chan, E. K. L., K. F. Sullivan, and E. M. Tan. 1989. Ribonucleoprotein SS-B/La belongs to a protein family with consensus sequences for RNA-binding. *Nucleic Acids Res.* 17:2233-2244.

27. Kozak, M. 1989. The scanning model for translation: an update. J. Cell Biol. 108:229-241.

28. Berg, J. M. 1986. Potential metal-binding domains in nucleic acid binding proteins. *Science (Wash. DC).* 232:485–487.

29. Klug, A., and D. Rhodes. 1987. Zinc fingers: a novel protein fold for nucleic acid recognition. Cold Spring Harbor Symp. Quant. Biol. 52:473-82.

30. Landschultz, W. H., P. F. Johnson, and S. L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science (Wash. DC).* 240:1759–1764.

31. Kouzarides, T., and E. Ziff. 1988. The role of the leucine zipper in the fos-jun interaction. *Nature (Lond.)*. 336:646–651.

32. Landschultz, W. H., P. F. Johnson, and S. J. McKnight. 1989. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science (Wash. DC).* 243:1681–1688.

33. Sassone-Corsi, P., L. J. Ransone, W. W. Lamph, and I. M. Verma. 1988. Direct interaction between fos and jun nuclear oncoproteins: role of the "leucine zipper" domain. *Nature (Lond.).* 336:692–695.

34. Patarca, R., J. Schwartz, R. P. Singh, Q.-T. Kong, E. Murphy, Y. Anderson, F.-Y. W. Sheng, P. Singh, K. A. Johnson, S. M. Guarnagia, T. Durfee, F. Blattner, and H. Cantor. 1988. rpt-1, an intracellular protein from helper/inducer T cells that regulates gene expression of interleukin 2 receptor and human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA*. 85:2733–2737.

35. Takahashi, M., and G. M. Cooper. 1987. ret transforming gene encodes a fusion protein homologous to tyrosine kinases. *Mol. Cell. Biol.* 7:1378-1385.

36. Chan, E. K. L., J. C. Hamel, C. L. Peebles, J. P. Buyon, and E. M. Tan. 1989. Protein domain of the 52kDa SS-A/Ro ribonucleoprotein particle is homologous to the N-terminal metal-finger domain of human ret/rfp and mouse rpt-1 proteins. J. Cell Biol. 109:14a. (Abstr.)

37. Lieu, T.-S., M. M. Newkirk, J. D. Capra, and R. D. Sontheimer. 1988.

Molecular characterization of human Ro/SS-A antigen. Amino terminal sequence of the protein moiety of human Ro/SS-A antigen and immunological activity of a corresponding synthetic peptide. J. Clin. Invest. 82:96-101.

38. McCauliffe, D. P., F. A. Lux, T.-S. Lieu, I. Sanz, J. Hanke, M. M. Newkirk, L. L. Bachinski, Y. Itoh, M. J. Siciliano, M. Reichlin, R. D. Sontheimer, and J. D. Capra. 1990. Molecular cloning, expression, and chromosome 19 localization of a human Ro/SS-A autoantigen. J. Clin. Invest. 85:1379–1391.

39. Lieu, T. S., D. P. McCauliffe, P. Volpe, B. H. Alderson-Lang, M. Michalak, J. D. Capra, and R. D. Sontheimer. 1990. Structural and functional homology between a human Ro/SS-A autoantigen and the calcium-binding protein, calreticulin. *Clin. Res.* 38:408. (Abstr.)

40. Takahashi, M., J. Ritz, and G. M. Cooper. 1985. Activation of a novel human transforming gene, ret, by DNA rearrangement. *Cell*. 42:581-588.

41. Buyon, J. P., E. Ben-Chetrit, S. Karp, R. A. S. Roubey, L. Pompeo, W. H. Reeves, E. M. Tan, and R. Winchester. 1989. Acquired congenital heart block. Pattern of maternal antibody response to biochemically defined antigens of the SS-A/Ro-SS-A/La system in neonatal lupus. J. Clin. Invest. 84:627-634.

42. Ben-Chetrit, E., R. I. Fox, and E. M. Tan. 1990. Dissociation of immune responses to the SS-A (Ro) 52-kd and 60-kd polypeptides in systemic lupus ery-thematosus and Sjogren's syndrome. *Arthritis Rheum.* 33:349–355.