

One Gene, Two Transcripts: Isolation of an Alternative Transcript Encoding for the Autoantigen La/SS-B from a cDNA Library of a Patient with Primary Sjögrens' Syndrome

By Helmut Tröster,* Thomas E. Metzger,* Imre Semsei,*
Martin Schwemmle,[§] Andreas Winterpacht,[‡] Bernhard Zabel,[‡]
and Michael Bachmann*

From the Institute für *Physiologische Chemie and [‡]Paediatric, Johannes-Gutenberg Universität, D-55099 Mainz; and [§]the Institute für Virologie, Albert-Ludwigs-Universität Freiburg, D-79104 Freiburg, Germany

Summary

A cDNA library was prepared from peripheral blood lymphocytes of an autoimmune patient with primary Sjögrens' syndrome. The cDNA library was screened with the patients own autoimmune serum being monospecific for the nuclear autoantigen La/SS-B. Thereby an alternative type of La mRNA was identified that differed from the known La mRNA due to an exchange of the exon 1. Sequencing of the genomic region between the exons 1 and 2 showed that the alternative 5'-end is a part of the intron. In addition, the presence of an alternative promoter site, which exists within the intron downstream of the exon 1, became evident. In consequence, the alternative La mRNA is the result of a promoter switching combined with an alternative splicing mechanism. In the intron, further transcription factor binding sites, including a NF- κ B element, were identified leading to the suggestion that the expression of the gene encoding for the nuclear autoantigen La/SS-B alters in dependence on disease conditions.

The basic mechanisms for development of autoimmunity, including the development of autoantibodies directed to nuclear antigens (ANAs)¹, in patients with systemic lupus erythematosus (SLE) or primary Sjögrens' syndrome (pSS) are not understood. It is assumed that unknown genomic prerequisites together with viruses may play a role during development of autoimmunity. Especially herpesviruses, including the Epstein Barr virus as well as exogenous or endogenous retroviruses, were discussed as possible causative agents of these diseases (1, 2). An involvement of retroviruses in pSS is supported by (a) an appearance of antibodies to retroviral proteins in patients with pSS and (b) the isolation of an A-type retroviral particle in lymphoblastoid cells exposed to homogenates of salivary tissue from patients with pSS (3). In addition, the human immunodeficiency virus (HIV), can produce a clinical syndrome and immunologic changes similar to those seen in pSS, including the development of ANAs (4). Nonetheless, even if these or other viruses are involved in autoimmunity in pSS and SLE patients, the mechanism by which viruses can lead to the development

of ANAs remains obscure, as these autoantibodies are at least partially the product of an autoantigen driven oligoclonal B cell response (5, 6).

One of the targets of an ANA system is the nuclear autoantigen La/SS-B (7). La protein has been suggested to be involved in transcription termination of RNA polymerase III and internal initiation of translation of at least poliovirus mRNA (8-10). Up to now more than 20 La cDNAs have been isolated by six independent groups and no dramatic differences had been observed (11-16). The transcription of this described type of La mRNA was suggested to occur from a GC-rich promoter region upstream of the exon 1 lacking a TATA-element, leading to the conclusion that La protein belongs to the housekeeping proteins (17).

In case of some protooncogenes, inflammatory mediators, signal transduction components, proteins involved in the immune response, transcription factors, growth factor receptors, and some other housekeeping proteins, in addition to the respective translatable mRNA, an alternative mRNA has been described. Usually the exon 1 is replaced in the respective alternative mRNA. The alternative 5'-terminus alters the properties of the mRNAs dramatically as it contains GC-rich regions and/or additional less functional AUGs upstream of the major open reading frame (18). Due to their unusual 5'-terminal structure, such mRNAs were hypothesized to ei-

¹ Abbreviations used in this paper: ANA, autoantibodies directed to nuclear antigen; pSS, primary Sjögrens' syndrome; SLE, systemic lupus erythematosus.

ther throttle the translation of the functional counterpart or represent intermediates that can be further processed depending on the tissue (18).

Here we describe the isolation and characterization of such an alternative type of mRNA derived from the gene encoding for the autoantigen La/SS-B.

Materials and Methods

Preparation of a cDNA Library from the PBL of a Patient with pSS. RNA was isolated from PBL of a patient with pSS (Ma) as basically described by Chirgwin et al. (19) using modifications according to Semsei and Cutler (20). Poly(A)⁺ mRNA was isolated from the RNA preparation using oligo(dT)-column chromatography (20). The cDNAs were synthesized using the ZAP-cDNA synthesis kit and ligated to the Uni-ZAPXR vector arms (Stratagene, Heidelberg, Germany, 21). The λ -library was packaged by Gigapack II Gold packaging extract (Stratagene), plated and amplified in the *Escherichia coli* cell lines PKL-F' and XL1-Blue, respectively. The complexity of the library was 10⁶ pfu.

Immunoscreening. Immunoscreening was performed using the patient's anti-La serum (10 ml). Before use, the serum was extensively preadsorbed to a Sepharose 4B column (5 ml), which had been covalently linked to a total extract obtained from *E. coli* (strain Y1090) at a concentration of 10 mg/ml gel. The first screening was followed by second and third screenings using the antibody solution left from the first screening step. The clones were checked in parallel with a cell culture supernatant of a hybridoma line producing a monoclonal anti-La antibody (termed La4B6, obtained by immunization of BALB/c mice with a recombinant human La protein; Tröster, H., T. Metzger, R. Klein, G. Pollak, I. Semsei, M. Schwemmle, G. Pruijn, W. van Venrooij, and M. Bachmann, unpublished results). The immunoscreening was performed following the protocol of Mierendorf et al. (22). A total amount of 4 × 10⁵ recombinant phages were screened and five positive La cDNAs were isolated with the patient's antiserum. In addition 12 cDNAs encoding for IgG heavy chains were coisolated, which was not surprising as the cDNA library was constructed from PBL and screened with anti-human IgG antibodies. The coisolated IgG heavy chain cDNAs were differentiated from the La cDNAs by (a) their different restriction enzyme analysis pattern and (b) sequencing. Isolation and storage of the clones was done according to Sambrook et al. (23). Phagemids (pBluescript SK(-)) were excised from λ phages during an *in vivo* excision procedure described by Stratagene using the filamentous helper phage strain R408 and the bacterial strain XL1-Blue as a host (23).

Construction of Deletion Clones. The two clones termed La19 and La23, which contained the largest inserts, were selected for sequencing. The other cDNAs were only sequenced from the 3'- and the 5'-end to characterize them as La cDNAs. DNA was prepared from small scale cultures (up to 12 ml) using the protocol of Holmes and Quigley (24) modified according to Sambrook et al. (23). To create deletion clones the method originally described by Henikoff (25) was used. The DNA to be deleted was cut with BstXI and EcoRI from which site deletions started.

Construction of a Genomic Subclone. For subcloning of the genomic fragment derived from the charon phage λ La2-1 (17), which was kindly provided by Dr. Keene (Duke University, Durham, NC), the "shotgun" cloning technique was used. 13 ng of λ La2-1 DNA was digested with EcoRI (5 U/ μ g DNA in 2× One-Phor-All-Buffer PLUS [Pharmacia LKB, Freiburg, Germany]). In earlier studies Chambers et al. (17) had shown that a digestion of genomic DNA with EcoRI resulted among other fragments

in a 4.6-kb fragment of the La gene. According to our calculations the molecular mass of this fragment might be about 4.8 kb. The reaction was stopped by the addition of EDTA to a final concentration of 17 mM followed by a heat treatment for 20 min at 68°C. The pBluescript SK(-) vector DNA (Stratagene) was digested with EcoRI (2.5 U/ μ g of DNA) at a concentration of 0.88 μ g DNA/ μ l in 2× One-Phor-All-Buffer PLUS. After dilution (1:4) and an inactivation step by heating to 68°C for 20 min, calf intestine phosphatase (CIP; Boehringer Mannheim, Mannheim, Germany) was added (0.3 U CIP/ μ g DNA) and incubated for 30 min at 37°C in the recommended buffer (10×: 0.5 M Tris/HCl, pH 8.5; 1 mM EDTA). The enzyme was inactivated by the addition of EDTA (16 mM final concentration) in combination with heating for 30 min to 75°C. The ligation of the EcoRI-fragments with the phosphatase-treated EcoRI-digested vector DNA took place in a 10 μ l volume at 16°C overnight. The reaction mixture consisted of 200 ng vector DNA, 30 ng of EcoRI-fragments in Tris/HCl, pH 7.5 containing 1 mM ATP, 1 mM dithiothreitol, and 5 mM MgCl₂. Then the ligation reaction was adjusted to a volume of 100 μ l with H₂O and added to 200 μ l of competent *E. coli* XL1-Blue cells for transformation according to Sambrook et al. (23). The bacteria were plated on ampicillin (150 μ g/ml) agar containing 2.5 mM isopropyl β -D-thiogalactoside (IPTG) and 0.4% (wt/vol) X-Gal in order to discriminate religated vector DNA (blue colonies) from colonies containing vectors with inserts (white colonies). Plasmid DNAs derived from white colonies were found to contain the desired 4.8-kb fragment of the La gene.

Southern Blotting. DNA was isolated as basically described by Chirgwin et al. (19) using minor modifications according to Semsei and Cutler (20). The DNA was digested with EcoRI and 13 μ g of DNA per lane were loaded on an 0.7% agarose gel. The gel was run for 16 h with 1.8 V/cm. Southern blotting was performed according to Sambrook et al. (23). After transfer onto a positively charged Nylon membrane (Boehringer Mannheim) the DNA was cross-linked by UV irradiation with 120 mJ using a Stratalinker 1800 (Stratagene) followed by a heat treatment for 60 min at 80°C. Hybridizations and labelings of the DNA were performed according to the nonradioactive detection system (Boehringer Mannheim). The signals were developed using anti-DIG antibodies covalently linked to either alkaline phosphatase and CSPD (TRopix, Serva Heidelberg, Germany) as substrate (see Fig. 2 A) or peroxidase (Boehringer Mannheim) and the enhanced chemiluminescence (ECL) detection system (Amersham-Buchler, Braunschweig, Germany) (see Fig. 2, B and C). Between the three hybridizations, membrane stripping was performed as follows. The filter was washed twice in 0.1% SDS at room temperature followed by a proteinase K treatment (1 mg/ml) in 0.1% SDS for 35 min at 55°C and 10 min at 65°C. Afterward the membrane was further stripped as described by Sambrook et al. (23). For this purpose the blot was washed with 50% formamide, 2× SSPE at room temperature and in the same buffer at 65°C for 45 min. The membrane was rinsed in 0.1% SDS, 0.1× SSC at room temperature, dried on 3MM-paper (Whatman International Ltd., Maidstone, UK) and stored at room temperature in the dark and dry.

PCR. PCR was performed using a TC9600 Cycler (Perkin-Elmer, Überlingen, Germany). The 50- μ l assay in 1× Taq buffer contained 2 U Taq polymerase (Boehringer Mannheim), 1.5 mM MgCl₂, 5% (vol/vol) DMSO, 200 mM of each dNTP, 20 pmol of each primer, and 100 ng of genomic DNA or 1 ng of plasmid DNA. Cycling was started by heating for 2 min to 94°C. 40 cycles followed, each consisting of 15 s at 94°C, 15 s at 55–60°C depending on the primer pairs used, and 15–60 s at 72°C depending on the length of the resulting fragments. Then the temperature

was held for 10 min at 72°C and cooled down to 4°C. 6 µl of the respective assay were taken, and 1.5 µl of sample buffer (40% sucrose, 0.25% bromophenol blue in H₂O) was added and applied to a 1.5–1.8% agarose (Biozym, Hameln, Germany) gel in TAE buffer (40 mM Tris-acetate, pH 8.0; 1 mM EDTA). For determination of the molecular weights of the PCR products 250 ng of the DNA molecular weight marker VI (Boehringer Mannheim) was run in parallel. The gels were run for 10 min at 5 V/cm and 120 min at 3 V/cm and stained with ethidium bromide. Photographs were taken with polaroid film 667.

DNA Sequence Analysis. The DNA for sequencing of the cDNAs was prepared by alkaline lysis miniprep method according to Sambrook et al. (23). The DNA samples were sequenced using the automatic fluorescent DNA sequencing system of the EMBL. The sequences were evaluated running the evaluation program COMGRA on a 486 PC system. DNA comparison and alignments were performed using the program package HUSAR of the DKFZ (Heidelberg, Germany).

The genomic sequences were prepared from genomic DNA isolated from different tissues including PBL of the patient (Ma; [8]) with pSS, an SLE patient (p62), and her baby born with AV block (p61), as well as several healthy donors, and the subclone derived from the genomic clone λLa2-1 (see above), the liver of a tumor patient (We), and the embryonic spleen. The genomic sequence between the promotor upstream of the exon 1 and the end of the exon 2 was determined for the subclone prepared from the genomic clone λLa2-1 (see above) as well as for the genomic DNA of the patient (Ma). For that purpose overlapping PCR fragments were prepared and sequenced using the Cycle Sequencing Kit from Pharmacia LKB and/or the PCR fragments were directly sequenced (see below). In the case of the GC-rich region between the exon 1 and the oligo(dT)-tail the PCR fragments prepared were not only directly sequenced but subcloned into pBluescript and at least three subclones were sequenced.

The following primers were used for sequencing of the genomic region between promoter 1 and the exon 2: P1 (GCACTTCACATGAATGGAGG); P2 (GAGTCGTTGCTGTTGCTGTTT); P3 (GCACAGGCTCACAAACAGCAAC); P4 (GCTGTTTGTGAGCCTGTGCGCGGC); P5 (GTTCTTGTAGATACTCTATGC); P6 (CGCTTTACTAGTGC GCGACTGCGCGTTTCC); P7 (AAGACGCAATGGGGATGAGG); P8 (ACCGCCTTCTAGTCTCACCGAA); P9 (ACAAGCCTTCGGTGAGACTAGA); P10 (CCCACCTCCCAGAACTTTACAG); P11 (GCACGGGTAACCGCCGGAGGGTTC); P12 (CTGAAACCTGATGTGAGCGATG); P13 (GCCAAATAGGTTTGCAGGGTG); P14 (ATGCTTCAGGAAGTGTACC); P15 (GTTAGTATGACCACATTCTCCC); P16 (GGACCCTTTGTAGCTTCTC); P17 (AGTTCAAGACCA GCCTGACC); P18 (CTTCACTTAGCTGGTCAAGC); P19 (CTCAACAAAGCAACACTTTGG); P20 (TCACCATTTTCAGCCATTGCGG); P21 (GATGACAGATTTTGGCCTCCAG).

Cycle sequencing was performed using the Cycle Sequencing Kit from Pharmacia LKB and the Cyler TC9600 (Perkin-Elmer). Four caps were prepared by adding to each 2 µl of one of the four termination mixtures (A, C, G, T) and stored at 4°C until use. Then master mixes were prepared consisting of 2 µl primer solutions containing up to 10 pmol primer, 5 µl of 5× sequencing buffer, 1 µl dNTP mix, 1 µl Tth (Thermus thermophilus) DNA polymerase (1 U/µl), 1 µl α-[³³P]dATP (3,000 Ci/mmol; NEN, Dreieich, Germany), and 2 µg of template DNA. 5 µl of the respective master mix were added to each of the four termination mixtures (A, C, G, T). Cycling was started by heating for 2 min to 95°C. Then 25 (to 50) cycles followed, each consisting of 30 s at 95°C, 36 s at 55°C, and 84 s at 72°C. Finally the reaction mix-

tures were held for 5 min at 72°C and cooled to 4°C until the addition of 3 µl of the stop solution.

The radioactivity labeled samples were separated on a sequencing gel using the sequencing system LKB 2101 MacroPhor (Pharmacia LKB).

For directly sequencing of the PCR fragments the T7-Sequase-Kit from Pharmacia LKB was used according to the following procedure. The PCR products were first separated on 3 to 4% NuSieve (Biozym) agarose gels and eluted using GENECLEAN II (Dianova, Hamburg, Germany). 4 µl of isolated DNA at a concentration of 10–20 µg/ml were added to 2 µl annealing buffer. Then 2 µl primer solutions containing 300 ng of primer were added, followed by 1.2 µl of NP-40 (5%) and 2.8 µl H₂O. The complete annealing mixture (12 µl) was briefly centrifuged using an Eppendorf centrifuge (Eppendorf, Hamburg, Germany) and rapidly heated to 95°C for 3 min. Then the annealing mixture was immediately transferred to liquid nitrogen. For labeling 6 µl of the labeling mixture (T7-Sequase-Kit; Pharmacia, LKB), which consisted of 3 µl Labeling-Mix, 0.5 µl α-[³⁵S]-dATP (1,000 Ci/mmol), 2 µl T7-polymerase (2 U/µl), and 0.5 µl NP-40 (5%), was added. 4 µl aliquots were used for each dNTP. The labeling reaction was started by adding and briefly centrifuging of the 4 µl labeling mixtures to the respective mixture consisting of 2 µl of one of the four short mixes (A, C, G, T) and 0.5 µl NP-40 (5%). Labeling was performed for 5 min at 37°C. Then 2 µl of a chase mixture consisting of 2 µl dNTPs (2.5 mM), 2 µl NP-40 (5%), 2 µl NaCl (500 mM), and 14 µl H₂O was added, briefly centrifuged and incubated for 5 min at 37°C. Finally, the reaction was stopped by adding 5 µl stop solution and briefly centrifuging.

5'-RACE. For an analysis of the 5'-ends of the La RNAs we chose the 5'-RACE system supplied by GIBCO BRL (Eggenstein, Germany). It includes a first strand cDNA synthesis and an PCR amplification step. During reverse transcription, 40 U of RNase inhibitor (Boehringer Mannheim) was added to each sample. The total RNA samples used were isolated from either (adult) liver of the tumor patient (We) with liver metastasis (1.4 µg), human embryonic (20 week) spleen (5 µg), PBL of the patient (Ma) with pSS (0.3 µg), PBL of a (adult) control person (0.6 µg). The reactions and the following 5'-RACE steps were performed according to the instructions of the supplier. For the first strand synthesis a primer locating within the exon 3 was used (TGTCCCGTGCAAATTTGAAGTCGCC). Amplification was performed using the primers P20 and P12 in combination with the Anchor Primer of the 5'-RACE kit (GIBCO BRL, Eggenstein, Germany).

Results

Based on the suggestion that the development of autoantibodies to La protein in patients with SLE and pSS is at least partially driven by the autoantigen itself, one might expect an altered expression of the La gene in patients in dependence on disease conditions. Therefore, we looked for alternative forms of La mRNAs in tissue obtained from a La positive pSS patient (Ma) (8).

Isolation of Alternative La cDNAs. For that purpose a cDNA library was prepared from the patient's PBLs. This cDNA library was screened with the patient's own autoimmune serum. Thereby five La cDNAs were isolated (Fig. 1). Three of them had derived from shortened La mRNAs starting within or downstream of the exon 2, as became already evidence from restriction enzyme fragment analysis (Fig. 1) and sequencing from the 5'- and 3'-end (data not shown). The

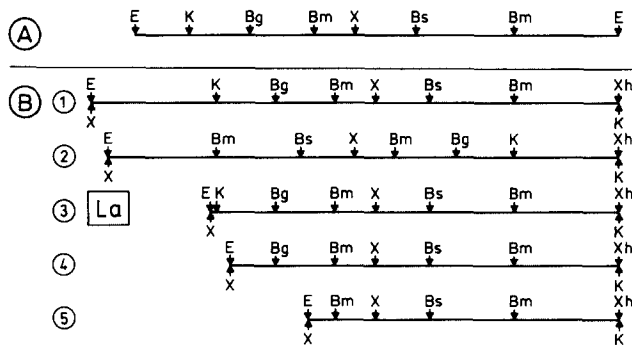


Figure 1. Characterization of the La cDNAs isolated from the cDNA library of the patient with pSS by immunoscreening using restriction enzyme fragment analysis. The screening of the patient's cDNA library with her own autoimmune serum resulted in the isolation of five cDNAs (B; 1-5) that had been derived from La mRNAs. These La cDNAs showed a restriction enzyme fragment pattern similar to the pattern of a La cDNA (A) described by Chan et al. (13). The clones with the two longest La inserts were termed as La23 (1) and La19 (2). The insert in the clone La19 has the opposite orientation as in the other La clones.

residual two La cDNAs (Fig. 1, clones 1 and 2, which were termed as La23 and La19) were sequenced and their sequence was given to the EMBL Data Library (EMBL accession number X69804). Both La cDNAs were found to start upstream of the exon 2, but displayed the following three unusual properties: (a) they started at their 5' terminus with an oligo(dT)-stretch; (b) as already evident from the restriction enzyme fragment analysis (Fig. 1), the La cDNA insert was inverted in the clone La19, probably due to the presence of the oligo(dT)-stretch, and; (c) upstream of exon 2, exactly at the splice site, the complete exon 1 was replaced by an unknown 5'-terminal sequence, termed exon 1' (Fig. 2). Downstream of the exon 2, the sequence was identical to the described La sequences (Fig. 2).

Southern Blot Analysis. To rule out a cloning artifact or the isolation of a splicing intermediate the genomic origin of the exon 1' sequence was determined. By performing Southern blot experiments with genomic DNA of healthy individuals and the patient (Fig. 3) the exon 1' was located on a 4.8-kb EcoRI fragment (Fig. 3 A), which was described as containing the exons 1 and 2, as well as the intron between them (17). In agreement, the same fragment also hybridized to a probe being specific for the exon 1 sequence (Fig. 3 B) and was confirmed to be a part of the La gene, as it also hybridized to a probe containing the complete La cDNA sequence (Fig. 3 C). From this experiment it became also evident that the exon 1' sequence is not a patient specific sequence (Fig. 3, A-C, lane c) but exists in healthy donors as well (Fig. 3, A-C, lanes a and b).

The Genomic Origin of the Alternative La cDNAs. The next steps were to isolate the 4.8-kb fragment containing the exon 1' sequence and to develop a sequencing scheme that allowed us to compare the sequences from healthy donors and patients. For that purpose the 4.8-kb fragment was subcloned from the genomic clone λ La2-1 (17) into pBluescript SK(-) and sequenced starting upstream of the proposed promoter of exon 1 going downstream through the intron into the

```
(I) [La19/23]          TTTT.....TTTTTTTTTTTTTTT
1 ACCTCCACCCGCTCTAGTCTCACCGAAGGCTGTGGCCATCTCTGTAAA 50
51 GTTCTGGGGAGGTGGGCACGGGTAAACGCCGGAGGOTCTGTGGAGCC 100
101 TATGGGACTGACTGGAGGCCAAGAAGCTTTGGGCCGAGATGTACACCAG 150
(II) 1 .....CCGGCGGCGCTGGGAGGTGGA 21
151 CGTCACCTCTTTGGTCTGTTAGGGCCTTCATCGCTCACATCAGGTTTCAG 200
22 GTCGTGCTGTTGCTGTTTGTGAGCCCTGTGGGC.....GGCTTCTGTG 65
201 TGTGAACGGGAAACATAGCCGTAATGGCTGAAAAATGGTGATAATGAAA 250
66 GGCCGGAACCTTAAAGATAGCCGTAATGGCTGAAAAATGGTGATAATGAAA 115
251 AGATGGCTGC.....//.....AACTAATAAAATATATACTATATGAAA 1750
116 AGATGGCTGC.....//.....AACTAATAAAATATATACTATATGAAA 1615
```

Figure 2. Characterization of the alternative form of La mRNA. A comparison of the sequence derived from the La cDNAs La19 and La23 (I; EMBL accession number X69804) with the human La cDNA described by Chan et al. (13) (II; EMBL accession number X13697) was performed. Upstream of the acceptor splice site at exon 2 (nt 216; [↓]) the sequences differed, whereas the sequence downstream of the exon 1' is identical to the sequence described by Chan et al. (13). The La cDNAs La19 and La23 started with an oligo(dT)-tail of different length (La19, 31 (dT)-residues; La23, 52 (dT)-residues). At the genomic level the length of the (dT)-tail was determined with 23 (dT)-residues (see also Fig. 4 B).

exon 2 using a series of primers as schematically summarized in Fig. 4 A. The resulting complete genomic sequence was given to the EMBL Data Library (EMBL accession number Z35127) and is partially shown in Fig. 4 B.

We identified the exon 1' sequence, including its oligo(dT)-tail, 325 nt downstream of the exon 1 donor splice site. From this result we concluded that (a) the exon 1' sequence is a part of the intron between exon 1 and 2; (b) the alternative La cDNAs La19 and La23 must have been derived from processed hnRNAs; and (c) the 5'-terminal oligo(dT)-tail of La19 and La23 was not the result of a cloning artifact.

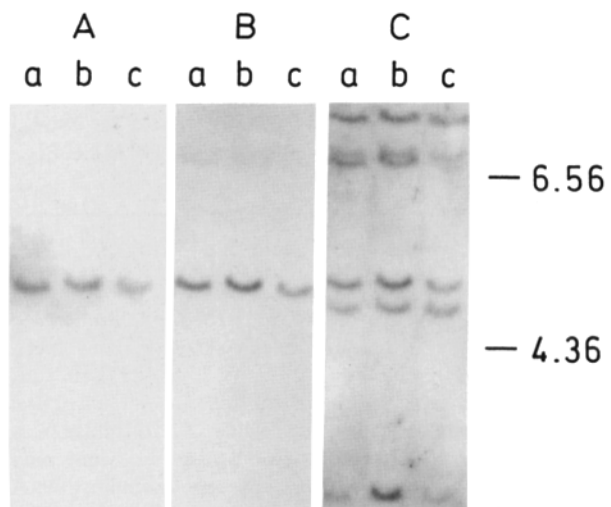
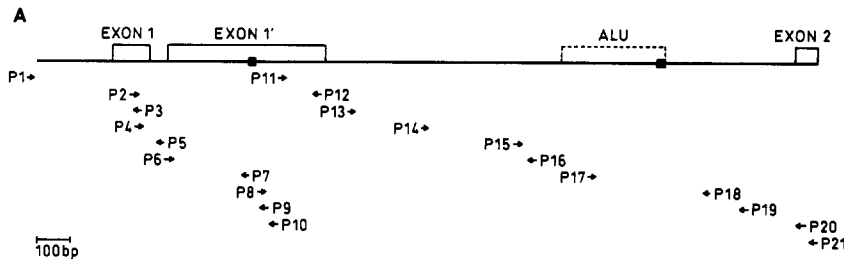


Figure 3. Southern blot experiment of the La gene. Genomic DNAs obtained from liver of a tumor patient (We, lane a), from PBL of either a healthy donor (lane b), or from the patient (Ma) with pSS (lane c) were analyzed. The same blot was hybridized three times with probes specific for either exon 1' (A) or exon 1 (B) or the complete La gene (C).



B

```

5' 10      20      30      40      50
   GGGCTGGGGT ACAGACGCCG TCCAGAAATA CTTAAAATA AATGTTTATA
   60      70      80      90     100
   CTAACACAAA CTAGCAGCCA TTCCCCTACA ACCTGGAGTA CTTTTTAACC
   110     120     130     140     150
   CAATCTGTTT GGGATTGAGG GCGGGACTCC AGGGTCCCTG GGAGGCCCCG
                               (i) =>>>>
   160     170     180     190     200
   GCAGCCGCGC ATGCTCAGAG CACAGGGACC GCTCCACCTC GTCCGTGGCC
   210     220     230     240     250
   CTGCCACCC  AGGCCGCAAG AGCTGCCGGG ACGGTCCCCA TCTTCTTGGA
                               (ii) =>>>>
   260     270     280     290     300
   GCGCTTTAGG CTGGCCGGCG GCGCTGGGAG GTGGAGTCGT TGCTGTTGCT
   310     320     330     340     350
   GTTTGTGAGC CTGTGGCGCG GCTTCTGTGG GCCGGAACCT TAAAGGTGAG
   360     370     380     390     400
   TAACTCTCGG TGGCTAATGA GAAGCTACAA AAGGGTCCGC TTTGCTGGTG
                               (ii) =>>>>
   410     420     430     440     450
   CGCGACTGCG CGTTTCCGTT CTTACTGCT TCTCGGGCG TCCGTCGCC
                               (ii) =>>>>
   460     470     480     490     500
   CCTCGCCTTC CTCACGTTCA GGCCGAGCGC CGCGTGGGCC GCAGCCGGAG
   510     520     530     540     550
   GCGGCGGAGC CGAGGGGGCG CGGGCGGCGG CGAGGCTGCC CCCGCGAGGA
   560     570     580     590     600
   ATGCGGGATC CTGGGGTTC CCGCACGTGG CTGGTGCCCA GGCAGTCCG
   610     620     630     640     650
   CGGTTGCTTA GAGACGGGAC TGTGGCTGCC CTCATCCCCA TTGCGTCTTT
   660     670     680     690     700
   TTTTTTTTTT TTTTTTTTTT ACCTCCACCG CCTTCTAGTC TCACCGAAGG
   710     720     730     740     750
   CTTGTGGCCA TCTCTGTAAA GTTTCTGGGG AGGTGGGCAC GGGGTAAACG
   760     770     780     790     800
   CCGGAGGGTT CTGTGGAGCC TATGGGACTG ACTGGAGGCC AAAGAACTTT
   810     820     830     840     850
   GGGCCGCGAG TGTACACCAG CGTCACTCTT TTGGTCTGTT AGGGCCTTCA
   860     870     880     890     900
   TCGCTCACAT CAGGTTTCAG TGTGAAACGG GAAAACGTGG GTAATATTTT
   910     920
   ATTTACAGTC GTACATTCCTG .....//.... CGGGATATTA ATACTTTGTA
   2270     2280     2290     2300     2310
   AACATTATCT TTCTATTCTT GAGTTTTATA TAAAAAATA CTTTATGCTA
   2320     2330     2340     2350     2360
   ATTTGGGAAA TTTTACAGAT AGCCGCAATG GCTGAAAATG GTGATAATGA
   2370     2380     2390     2400
   AAAGATGGCT GCCCTGGAGG CCAAAATCTG TCATCAAAAT GAG 3'

```

Figure 4. The promoter regions, the exon 1, the intron containing the exon 1', and exon 2 of the La/SS-B gene of DNA of the patient (Ma) and the genomic clone λ La2-1. (A) The genomic region starting upstream of the promoter of exon 1 and ending downstream of the exon 2 was determined using the 21 primers (P1 to P21) (see Materials and Methods) as schematically shown. The complete sequence is available from the EMBL Data Library (EMBL accession number Z35127). (B) Sequencing started 118 nt upstream of the promoter region (i) =>>>> described by Chambers et al. (17); nt 240 transcription start site (⊞) of exon 1; nt 240-345 exon 1; nt 345 (⊥) donor splice site of exon 1; (ii) =>>>> putative alternative TFIID binding site (TACAAA; see also Fig. 5); nt 413 transcription start site (⊞) of exon 1' as determined with 5'-RACE; nt 413-886 exon 1'; nt 648-886 5'-end of the La cDNAs La19 and La23; nt 886 (⊥) donor splice site of exon 1'; nt 921-2240 of the intron containing at nt 1604-1904 an Alu repeat are not shown (...//...); nt 2328 acceptor splice site (⊥) of exon 2; nt 2329-2403 exon 2; nt 2403 (⊥) donor splice site of exon 2.

The same sequencing scheme was then used to sequence the genomic DNA of the patient (Ma). The sequence of the genomic DNA of the patient (Ma) differed from λ La2-1 at nt 2003 with respect to a conversion of the nt T to C. Moreover, in one allele of the patients' DNA a deletion of the two nts (CA (1266/1267) was found. However this deletion and the mutation is not common to other patients, as they were not found in the DNA of a female SLE patient or her baby born with AV block. A further point mutation (C to A) was found in the sequence of the tumor patient (We) with liver metastasis at nt 495.

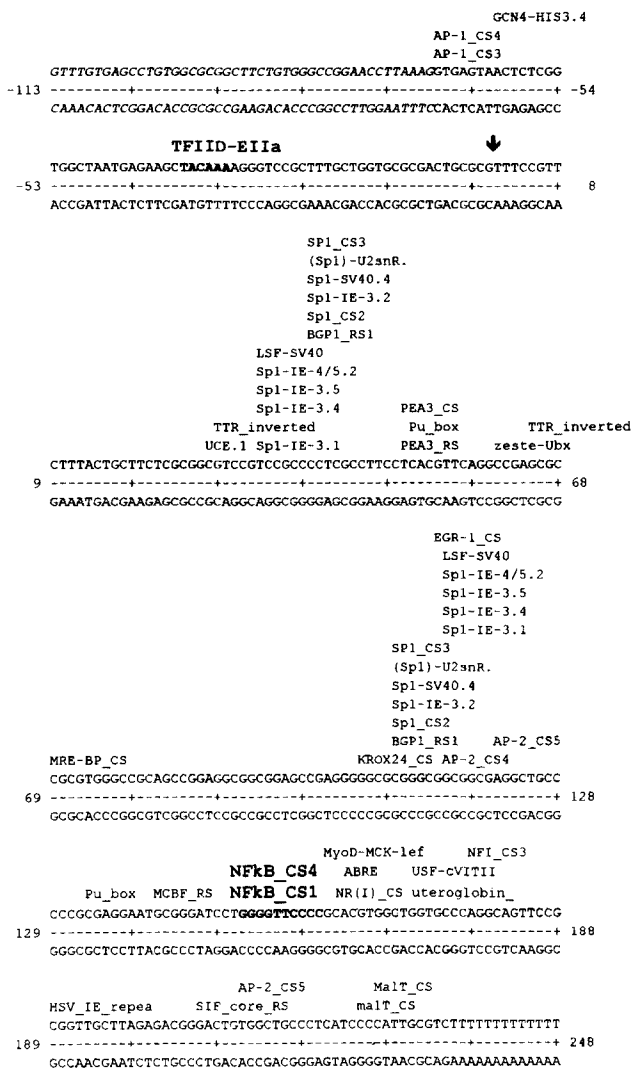


Figure 5. Consensus sequences for transcription factors. The genomic sequence between the 3'-end of exon 1 (*italics*) and the oligo(dT)-tail, which is the start of the La cDNAs La19 and La23, was analyzed for transcription factor consensus sequences using the program Husar (DKFZ). In agreement with the 5'-RACE result, a TFIID binding site exists -37 nt upstream of the transcription start site (\downarrow nt 1). In addition, the sequence contains a series of further putative transcription factor binding sites, including a NF- κ B element. It should be mentioned that two heat shock protein (hsp) 70 transcription factor consensus sequences are located about 270 nt upstream of the TACAAA site, which is already upstream of the promoter for the exon 1. Moreover, a true TATA box locates 50 nt further upstream (not shown).

Determination of the Alternative Transcription Initiation Site. As the oligo(dT)-tail was found to be a true part of the alternative La mRNA, it appeared rather unlikely that it should be the transcription initiation site of the La mRNA. Therefore, in a further approach we looked for the transcription initiation site of the alternative La mRNA. Using the 5'-RACE technique, the transcription start was determined to locate 235 nt upstream of the oligo(dT)-stretch, which is 68 nt downstream of the exon 1 but still within the intron between exon 1 and exon 2 (Fig. 4 B). By consequence, the La gene might contain two promotor sites. One upstream of the exon 1 and an alternative promotor within the intron downstream of the exon 1'. Indeed searching for hypothetical transcription factor binding sites (Fig. 5) a hypothetical TFIID binding site was identified to locate -37 nt upstream of the determined transcriptional start site (Figs. 4 and 5).

A Mechanism Leading to the Alternative La mRNA Form. In summary, we conclude that the alternative La mRNA is transcribed from an alternative promoter within the intron and represents the result of an alternative splicing event. The proposed mechanism leading to the two forms of La mRNAs is schematically summarized in Fig. 6 as follows: when the promoter upstream of exon 1 is used, the complete intron between exons 1 and 2 including the exon 1' sequence locating within the intron is removed during splicing. If the promoter upstream of exon 1' but downstream of the exon 1 is used, then (a) the exon 1 is not transcribed and (b) an alternative donor splice site within the intron is spliced to the same acceptor splice site at the exon 2. At the mRNA level this results in a complete exchange of the exon 1 by the exon 1'.

Properties of the Resulting Alternative La mRNA. One major difference between the two types of La mRNAs is based on the existence of three ATGs in the exon 1'. As shown in Fig. 7, there are two stop codons in the reading frame of the first ATG. The second ATG is in the +1 and the third ATG in the -1 frame. Interestingly, upstream of the first ATG there are two GC-rich regions.

Further Properties of the Analyzed Genomic Region of the La Gene. Based on the presence of the oligo(dT)-tail and of a further homopolymeric sequence, an oligo(dA)-tail of 26 (A)-residues (nt 1905-1930; EMBL accession number Z35127), which locates downstream of an Alu repeat (nt 1604-1904), we asked whether the alternative 5' terminus represents a part of a pseudogene. Comparison with the EMBL Data Library resulted in the identification of a sequence element of 89 nt starting 153 nt upstream of the oligo(dT)-tail, which has homologies in the range of 65% to the reversed sequence of the avian myelocytomatosis provirus and pseudorabies virus (Fig. 8). The region of homology to the avian myelocytomatosis virus locates within the v-myc region of the gag-myc fusion protein.

Discussion

Frequently sera of patients with pSS and SLE contain autoantibodies. One of the targets of the ANAs is the nuclear autoantigen La/SS-B (7). Different mechanisms had been discussed for the development of the autoantibodies directed to

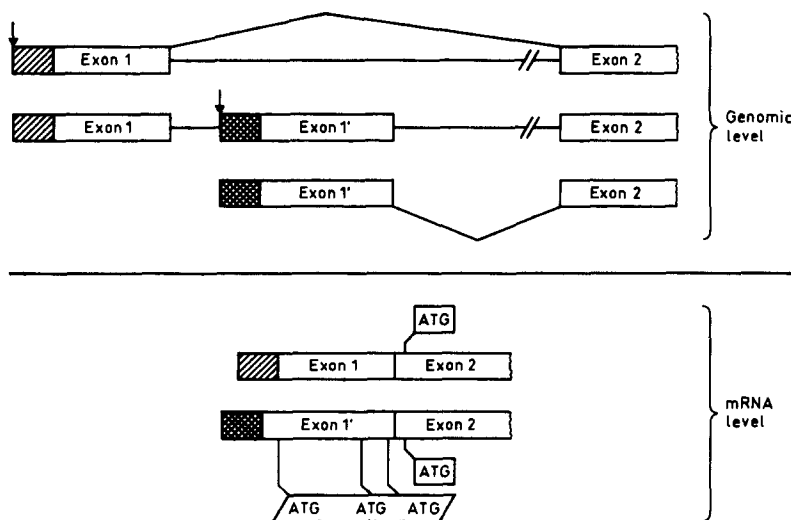


Figure 6. Origin of the two different forms of La mRNAs. The La gene contains two alternative transcription start sites (\downarrow). In the case of the exon 1 containing La mRNA the transcription start was originally determined with primer extension (17). This site was confirmed by the use of the 5'-RACE technique. The same technique was used to determine the initiation site of the exon 1' type La mRNA. It was found to locate in the intron downstream of the exon 1. By consequence, the two forms are expressed as follows: when the promoter upstream of exon 1 is used, the complete intron between exon 1 and exon 2, including the exon 1' sequence locating within the intron, is removed during splicing. If the promoter upstream of exon 1' but downstream of the exon 1 is used, (a) the exon 1 is not transcribed and (b) an alternative donor splice site within the intron is spliced to the same acceptor splice site at the exon 2. Thereby the exon 1 is replaced by the exon 1'. One major difference between the two types of La mRNAs is based on the fact that the translation initiation site of the exon 1 type La mRNA locates in the exon 2, since the exon 1 does not contain an ATG sequence. In contrast, the exon 1' contains three further ATGs.

```

10      20      30      40      50
5' TTTCGGTTC T TACTGCTTC TCGCGGCGTC CGTCCGCCCC TCGCCTTCCT
60      70      80      90      100
CACGTT CAGG CCGAGCGCCG CGTGGGCGCG AGCCGGAGGC GCGGGAGCCG
110     120     130     140     150
AGGGGGCGCG GCGGCGGCGG AGGCTGCCCC CGCGAGGAAT GCGGGATCCT
0
160     170     180     190     200
GGGGTTCCCC GCACGTGGCT GGTGCCCAGG CAGTTCGCGG GTTGCTTAGA
0 G V P R T W L V P R Q F R G A L E
210     220     230     240     250
GACGGGACTG TGGCTGCCCT CATCCCCATT GCGTCTTTT TTTTTTTTT
0 T G L W L P S S P L R L F F F F F
260     270     280     290     300
TTTTTTTTAC TCACCACGCC TTCTAGTCTC ACCGAAGGCT TGTGGCCATC
0 F F T S T A F * S H R R L V A I
310     320     330     340     350
TCTGTAAGT TTCTGGGAG GTGGCACGG GGTAAACGCC GGAGGGTTC
0 S V K F L G R W A R G K R R R V L
360     370     380     390     400
GTGGAGCCTA TGGGACTGAC TGGAGGCCAA AGAACTTTGG GCCGCGAGAT
0 W S L W D * L E A K E L W A A D V
-1 M G L T G G Q R T L G R R C
+1
410     420     430     440     450
TACACCAGCG TCACTCTTTT GGTCTGTTAG GGCCTTCATC GCTCACATCA
0 H Q R H S F G L L G P S S L T S
-1 Y T S V T L L V C *
+1 T P A S L F W S V R A F I A H I R
460     470     480     490     500
GGTTTCAGTG TGAACGGGA AAACATAGCC GCAATGGCTG AAAATGGTGA
0 G F S V K R E N I A A M A E N G D
+1 F Q C E T G K H S R N G * *
510     520     530     540     550
TAATGAAAAG ATGGCTGCC TGGAGGCCAA AATCTGTCAT CAAATTGAGT
0 N E K M A A L E A K I C H Q I E Y

```

Figure 7. The alternative 5' terminus contains three ATGs. All the hypothetical reading frames starting within the exon 1' are italics and underlined. Stop codons are indicated by (*). Upstream of the first ATG the exon 1' sequence contains two GC-rich regions (*underlined*). The first reading frame contains two stop codons. The second ATG is in the +1 frame, the third ATG in the -1 frame. The () marks the splice junction

the nuclear antigen La/SS-B including either a molecular mimicry with (retro)viral epitopes or an antigen driven mechanism (5, 6).

In case of an autoantigen driven mechanism alterations in patients with respect to the expression of the autoantigen should occur. However, such pathophysiological alterations had not been described up to now. Therefore we started to analyze the expression of the La gene searching for alternative transcripts overexpressed in patients.

For that purpose a cDNA library was made from PBL of the patient (Ma; 8) with pSS and the library was screened with her own autoimmune serum. Thereby a novel form of La mRNA was identified. The two isolated alternative La cDNAs differed from the hitherto known La cDNA sequence at the 5' terminus. Both alternative La cDNAs started with an unusual 5'-oligo(dT)-tail and lacked the exon 1 sequence. Instead of the exon 1 an alternative 5'-terminal sequence which was termed as exon 1', was found. Using the technique of Southern blotting, an EcoRI fragment was identified to contain the exon 1' in healthy donors as well as in patients. Therefore, we concluded that the exon 1' is not a patient-specific insert. The 4.8-kb fragment containing the exon 1' was subcloned from a genomic λ clone into pBluescript and sequenced. Thereby the exon 1' including its oligo(dT)-tail was found to be a part of the intron between exons 1 and 2.

As the oligo(dT) was also found within the genomic sequence, the presence of the oligo(dT)-tail at the 5' terminus of the cDNAs is not the result of a cloning artifact, although it must be mentioned that the length found for the genomic oligo(dT)-tail (23 [dT]-residues) was smaller than the length of the oligo(dT)-tails of the cDNAs (31 [dT]-residues in the

between exon 1' and exon 2. The proposed NH₂-terminal aa sequence of the La protein (13), if translation starts at the first ATG in the exon 2 (*bold*; nt 484).

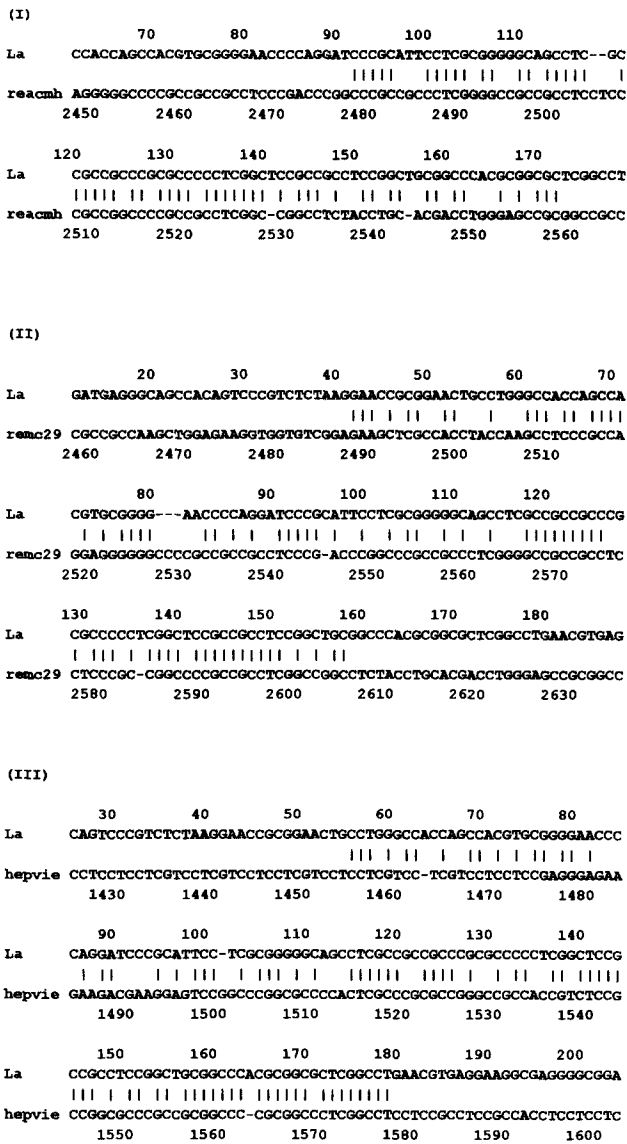


Figure 8. Sequence homologies of the 5'-end upstream of the oligo(dT)-tail. Comparisons of the reversed sequence upstream of the oligo(dT)-tail with the EMBL DNA Data Library revealed homologies of (a) 67.5% to a 83 bp overlap within the *v-myc* region of the avian myelocytomatosis HBI provirus (*reacmh*; EMBL accession number M11784); (b) 60.8% homology to a 120 bp overlap encoding for the *gag-myc* fusion protein of the avian myelocytomatosis virus of the same region (*remc*; EMBL accession numbers V01174 and J02247); or (c) 64% homology to a 125-bp overlap of an immediate-early gene of pseudorabies virus (*hepvie*; EMBL accession number X15120).

case of La19 and 52 in the case of La23). In fact these differences may have resulted during the cloning procedure.

By comparing the La cDNA sequences with the genomic sequence all the used splice sites became evident, including the exon 1 intron donor splice site, the alternative exon 1' intron donor splice site and the common intron exon 2 acceptor splice site. As the genomic sequence contains a 1443-nt-long intron between the 3'-end of the exon 1' and the 5'-start of the exon 2, which is absent in the sequence of the La cDNAs La19 and La23, it is clear that these La cDNAs

must have been derived from alternatively spliced La mRNAs.

Using the 5'-RACE technique the transcription initiation site of the alternative La mRNA was determined to locate 68 nt downstream of the exon 1 donor splice site. According to this result, the transcription of the alternative La mRNA starts within the intron between exon 1 and exon 1'. Therefore, an alternative promoter might locate upstream of this alternative transcription initiation site. Indeed, searching for putative transcription factor binding sites resulted in the identification of a putative TFIID binding site at the expected position (Fig. 5). It is interesting to note that in addition to the TFIID binding site a series of further consensus sequences for other transcription factors, which could be involved in regulation of a differential expression of the two types of La mRNAs, exists in the intron downstream of the exon 1 as summarized in Fig. 5. These transcription factor binding sites include common transcription factors such as SP1 but also of the AP1 and AP2 family. Of most interest is the presence of a NF- κ B element 218 nt downstream of the exon 1, as NF- κ B is known to be activated by UV irradiation, dsRNA, acute phase proteins, including some interleukins, interferons, tumor necrosis factor, and retroviruses (26). Moreover, NF- κ B elements also exist in retroviruses and are assumed to be involved in activation of retroviruses, viral latency, and development of autoimmunity. The NF- κ B element in the La gene (5'-GGGGTTCCCC) is very similar to the NF- κ B binding site in the HIV virus (5'-GGGGAT-TCCCC), which has been described to cause symptoms similar to pSS (4).

Taken together, the conditions known to activate NF- κ B are the conditions known to cause and induce acute disease in SLE patients. Therefore, it is tempting to speculate that here could be a first hint for a link between the development of disease as well as inducers of acute disease and an alternative expression of the gene encoding for the autoantigen La/SS-B.

In summary our genomic data show that the mechanism leading to the alternative form of La mRNA includes a promoter switching and an alternative splicing event.

At present the function of the alternative La mRNA remains obscure, as the complete exon 1' as determined with the 5'-RACE experiment does not only contain the unusual oligo(dT)-tail, but also two GC-rich regions and three ATGs. However, there are two stop codons in the reading frame of the first ATG. Moreover, the second and the third ATG are not in the correct reading frame (Fig. 7). Similar looking alternative transcripts had been described for other housekeeping proteins. In these cases it had been suggested that the alternative transcripts either are not translated into protein but should throttle the translation or represent stored intermediates for a further tissue-dependent 5'-terminal processing (18).

By consequence, future studies are required to clarify the physiological function of the alternative La mRNA. Moreover, the behavior under pathophysiological conditions must be elucidated. For example it is well established that active retroviruses take use of read through of stop codons and ribosomal frame shiftings during translation of their proteins.

So it is tempting to speculate that similar events could occur during translation of the alternative La mRNA in retroviral infected cells leading to NH₂ terminally altered La proteins,

which could finally induce the autoimmune reactions against the La protein.

We thank Dr. Kenan and Dr. Keene (Duke University, Durham, NC) for providing us with the charon phage λ La2-1 and Dr. Mayet (I. Med. Universität, Mainz, Germany) for the blood sample of the patient (Ma).

This investigation was supported by a grant from the Deutsche Forschungsgemeinschaft (Ba1145/1-2). M. Bachmann is a recipient of a C3 Hermann-and-Lilly Schilling professorship granted by the "Stifterverband für die Deutsche Wissenschaft." I. Semsei is a recipient of a fellowship donated by Alexander von Humboldt-Stiftung.

Address correspondence to Dr. Michael Bachmann, Institut für Physiologische Chemie, Johannes-Gutenberg Universität, Duesbergweg 6, 55099 Mainz, Germany.

Received for publication 26 May 1994 and in revised form 16 August 1994.

References

1. Fox, R.I., M. Luppi, P. Pisa, and H.-L. Kang. 1992. Potential role of Epstein-Barr virus in Sjögren's syndrome and rheumatoid arthritis. *J. Rheumatol.* 19:18.
2. Talal, N., M.J. Dauphinee, H. Dang, S.S. Alexander, D.J. Hart, and R.F. Garry. 1990. Detection of serum antibodies to retroviral proteins in patients with primary Sjögren's syndrome (autoimmune exocrinopathy). *Arthritis Rheum.* 33:774.
3. Moutsopoulos, H.M., and P. Youinou. 1991. New developments in Sjögren's syndrome. *Curr. Opin. Rheumatol.* 3:815.
4. Talal, N. 1991. AIDS and Sjögren's syndrome. *Bull. Rheum. Dis.* 40:6.
5. Meilof, J.F., K.M. Hebeda, J. de-Jong, and R.J. Smeenk. 1992. Analysis of heavy and light chain use of lupus-associated anti-La/SS-B and anti-Sm autoantibodies reveals two distinct underlying immunoregulatory mechanisms. *Res. Immunol.* 143:711.
6. McNeilage, L.J., K. Umaphysivam, E. Macmillan, A. Guidolin, S. Whittingham, and T. Gordon. 1992. Definition of a discontinuous immunodominant epitope at the NH₂ terminus of the La/SS-B ribonucleoprotein autoantigen. *J. Clin. Invest.* 89:1652.
7. Tan, E.M. 1989. Antinuclear antibodies: diagnostic markers for autoimmune diseases and probes for cell biology. *Adv. Immunol.* 44:93.
8. Bachmann, M., K. Pfeifer, H.C. Schröder, and W.E.G. Müller. 1990. Characterization of the autoantigen La as a nucleic acid dependent ATPase/dATPase with melting properties. *Cell.* 60:85.
9. Gottlieb, E., and J.A. Steitz. 1989. Function of the mammalian La protein: evidence for its activity in transcription termination by RNA polymerase III. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:851.
10. Meerovitch, K., V.V. Svitkin, H.S. Lee, F. Leibkowitz, D. Kenan, E.K.L. Chan, V.I. Agol, J.D. Keene, and N.J. Sonenberg. 1993. La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulo lysate. *J. Virol.* 67:3798.
11. Chambers, J.C., and J.D. Keene. 1985. Isolation and analysis of cDNA clones expressing the human lupus La antigen. *Proc. Natl. Acad. Sci. USA.* 82:2115.
12. Rauh, A.J.G., H. Hornig, and R. Lührmann. 1988. At least three distinct B cell epitopes reside in the C-terminal half of La protein, as determined by a recombinant DNA approach. *Eur. J. Immunol.* 18:2049.
13. 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.
14. Sturgess, A.D., M.G. Peterson, L.J. McNeilage, S. Wittingham, and R.L. Coppel. 1988. Characteristics and epitope mapping of a cloned human autoantigen La. *J. Immunol.* 140:3212.
15. Kohsaka, H., K. Yamamoto, H. Fujii, H. Miura, and N. Miyasaka. 1990. Fine epitope mapping of the human SS-B/La protein. Identification of a distinct autoepitope homologous to a viral Gag polyprotein. *J. Clin. Invest.* 85:1566.
16. Puijn, G.J.M., R.L. Slobbe, and W.J. van Venrooij. 1991. Analysis of protein-RNA interactions within Ro ribonucleoprotein complexes. *Nucleic Acids Res.* 19:5173.
17. Chambers, J.C., D. Kenan, B.J. Martin, and J.D. Keene. 1988. Genomic structure and amino acid sequence domains of human La autoantigen. *J. Biol. Chem.* 263:18043.
18. Kozak, M. 1991. An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* 115:887.
19. Chirgwin, J., A. Przybyla, R. McDonald, and W. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294.
20. Semsel, I., S. Ma, and R.G. Cutler. 1989. Tissue and age specific expression of the myc proto-oncogene family throughout the life span of C57BL/6J mouse strain. *Oncogene.* 4:465.
21. Short, J.M., J.M. Fernandez, J.A. Sorge, and W. Huse. 1988. Lambda ZAP: a bacteriophage Lambda expression vector with in vivo excision properties. *Nucleic Acids Res.* 16:7583.
22. Mierendorf, R.C., C. Percy, and R.A. Young. 1987. Gene isolation by screening Lambda gt11 libraries with antibodies. *Methods Enzymol.* 152:458.
23. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, NY. I:25; II:59; IX:44.
24. Holmes, D.S., and B. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114:193.
25. Henikoff, S. 1987. Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods Enzymol.* 11:156.
26. Schreck, R., K. Albermann, and P.A. Baeuerle. 1992. Nuclear factor kappa B: an oxidative stress-responsive transcription factor of eukaryotic cells (a review). *Free Radical Research Communications.* 17:221.