Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1777–1781, March 1990
Biochemistry

Isolation and characterization of cDNA encoding the 80-kDa subunit protein of the human autoantigen Ku (p70/p80) recognized by autoantibodies from patients with scleroderma–polymyositis overlap syndrome

(DNA-binding protein/molecular cloning/autoimmune disease/"leucine zipper")

TSUNEYO MIMORI*†, YASUO OHOSONE*, NOBUAKI HAMA*, AKIRA SUWA*, MASASHI AKIZUKI*, MITSUO HOMMA*, ANDREW J. GRIFFITH‡, AND JOHN A. HARDIN‡

*Department of Medicine, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160, Japan; and ‡Department of Medicine, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510

Communicated by Osamu Hayaishi, November 27, 1989

ABSTRACT **Anti-Ku (p70/p80) autoantibodies in patients with scleroderma–polymyositis overlap syndrome recognize a 70-kDa/80-kDa protein heterodimer which binds to terminal regions of double-stranded DNA. In the present study, we isolated full-length cDNAs that encode the 80-kDa Ku subunit. Initial screening of a human spleen cDNA library with anti-Ku antibodies yielded a cDNA of 1.0 kilobase (kb) (termed K71) encoding a portion of the 80-kDa Ku polypeptide (identification based on immunological criteria). In RNA blots, this cDNA hybridized with two mRNAs of 3.4 and 2.6 kb. In rescreening of a cDNA library constructed from simian virus 40-transformed human fibroblast mRNA with the K71 cDNA as a hybridization probe, three positive clones were isolated, and that bearing the longest insert (termed Ku80-6) was selected for further characterization. *In vitro* transcription and translation experiments produced an immunoprecipitable polypeptide which comigrated with the 80-kDa Ku subunit. The Ku80-6 cDNA proved to be 3304 nucleotides in length, with an additional poly(A) tail, closely approximating the size of the larger mRNA. It contains a single long open reading frame encoding 732 amino acids (M_r = 82,713). The putative polypeptide has a high content of acidic amino acids and a region with periodic repeat of leucine in every seventh position which may form the "leucine zipper" structure. In genomic DNA blots, probes derived from the opposite ends of cDNA Ku80-6 hybridized with several nonoverlapping restriction fragments from human leukocyte DNA, indicating that the gene encoding the 80-kDa Ku polypeptide is divided into several exons by intervening sequences.**

Various autoantibodies to cellular constituents are produced in patients with autoimmune diseases. Autoantibodies to the Ku antigen were originally recognized in Japanese patients with scleroderma–polymyositis overlap syndrome ("Ku" is derived from the prototype patient's name) (1). Subsequently, immunoprecipitation studies demonstrated that the Ku autoantigen consists of two polypeptides of 70 and 80-kDa [hence Ku is also referred to as p70/p80 (2)] which form a heterodimer and bind to terminal regions of double-stranded DNA (2–5). An exchange of autoimmune sera among laboratories has demonstrated that p66/p86 (Ki) antigen (5) and the 86–70-kDa protein complex (6) are identical with the Ku (p70/p80) protein. Although the precise function of the Ku antigen is still unknown, a role in DNA repair or transposition has been proposed because of its DNA-terminal binding activity (4).

In the present study, we have cloned two full-length cDNAs which encode the 80-kDa Ku subunit. Sequence analysis§ demonstrates that this subunit shows no homology with the 70-kDa component (p70) which was cloned recently by Reeves and Sthoeger (7). Since patients usually make autoantibodies that are focused against both of the Ku subunits (3) even though they are unrelated, it appears that relatively intact Ku heterodimer–DNA complexes are actively involved in autoimmune responses.

MATERIALS AND METHODS

Sera. Sera containing antibodies to Ku (p70/p80) (MO, KK, and TH) were used as probes for screening and characterizing cDNA clones that expressed the Ku epitopes. Specificities for antibodies were established by Ouchterlony immunodiffusion, immunoblotting, and immunoprecipitation assays (1, 3).

Screening of cDNA Libraries. A phage λgt11 cDNA library constructed from mRNA of human spleen (Clontech) was screened with an anti-Ku serum as described by Young and Davis (8). Positive plaques were subcloned by repeating sequentially until all progeny plaques were recognized by the antiserum. To isolate longer cDNAs encoding the Ku autoantigen, a pCD cDNA library constructed from mRNA of simian virus 40-transformed human fibroblasts was screened as described by Okayama and Berg (9) (gift from Joan A. Steitz, Yale University), using the 1.0-kilobase (kb)-long K71 cDNA (encoding the 80-kDa Ku as described in *Results*) as a probe for colony hybridization. The K71 insert cDNA was labeled with [^{32}P]dCTP (Amersham) by the random priming method (10).

Affinity Purification of Antibodies and Immunoblotting. For the characterization of cDNA clones, "elution blot" experiments were performed. Isolated recombinant phages were plated on a Luria–Bertani agar plate (9-cm diameter) with *Escherichia coli* strain Y1090 at 1×10^4 plaque-forming units per plate. After a 3-hr incubation at 42°C, a nitrocellulose filter (82-mm diameter; BA85, Schleicher & Schuell) saturated with 10 mM isopropyl β-D-thiogalactopyranoside (IPTG; Boehringer Mannheim) was overlaid on the agar for 3 hr at 37°C to induce the expression of β-galactosidase fusion proteins. After blocking with 1% skim milk solution in 10 mM

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

†To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M30938)

sodium phosphate/150 mM NaCl (PBS, pH 7.5) for 1 hr at room temperature, the filter was incubated with a diluted anti-Ku serum. After three washes with PBS containing 0.05% Triton X-100, filter-bound antibodies were eluted by 0.5 M glycine·HCl at pH 2.8, neutralized immediately by adding Tris base, and used as probes for immunoblotting against purified Ku antigen from HeLa cells. Affinity-column purification of the Ku antigen from HeLa cell extracts and immunoblot analysis were described previously (3).

DNA Sequencing. cDNA inserts isolated from recombinant phages (*Eco*RI fragments) and pCD plasmid (*Bam*HI fragments) were digested with several restriction enzymes (*Hae* III, *Alu* I, *Pvu* II, and *Pst* I; Takara Shuzo, Kyoto) and the resulting DNA fragments were ligated into the polylinker regions of M13mp18 or M13mp19 replicative form DNA. Nucleotide sequences were determined by using the dideoxy chain termination method of Sanger *et al.* (11), employing Klenow enzyme (Takara Shuzo) or modified T7 DNA polymerase (Sequenase; United States Biochemical) (12). Sequences were compared and aligned by using a computer program for DNA analysis (MicroGenie; Beckman). To avoid compression in G+C-rich sequences, some sequencing reactions were performed with dITP alternating with dGTP.

DNA and RNA Blot Analysis. In DNA blots, high-molecular-weight genomic DNA was extracted from human peripheral blood leukocytes according to Gross-Bellard *et al.* (13). Ten micrograms of DNA was digested with each of several restriction enzymes for 2 hr at 37°C, fractionated on 1% agarose gel electrophoresis, and transferred to a nitrocellulose membrane. For a hybridization probe, the Ku80-6 cDNA insert was digested with *Kpn* I to remove its poly(A) tail, and the 3.3-kb fragment was purified from an agarose gel slice by using a Geneclean kit (Bio 101, La Jolla, CA). In some experiments, the same probe was further cut with *Pst* I to form 5′ and 3′ fragments, separated and purified from an agarose gel. In RNA blots, 20 μg of cytoplasmic RNA from HeLa cells was fractionated on a 1.0% agarose gel containing 2.2 M formaldehyde and transferred to a nylon membrane (GeneScreen*Plus*; NEN). In both DNA and RNA blots, hybridization was performed in a buffer containing ^{32}P-labeled cDNA probe (heat-denatured, 5×10^5 cpm/ml), heat-denatured salmon testis DNA (100 μg/ml), 50% (vol/vol) deionized formamide, 1 M NaCl, 10% dextran sulfate, and 1% $NaDodSO_4$ at 42°C for 12–16 hr. The membranes were serially washed two times each for 30 min at 65°C with a solution containing 0.3 M NaCl/0.03 M sodium citrate and 1% $NaDodSO_4$ followed by a wash with 15 mM NaCl/1.5 mM sodium citrate, and then exposed to x-ray films.

***In Vitro* Transcription, *in Vitro* Translation, and Immunoprecipitation.** The Ku80-11 cDNA (2.4-kb *Bam*HI fragment) was inserted into the *Bam*HI site of pBluescript II SK+ DNA (Stratagene). The resulting plasmid was linearized with *Eco*RI or *Xba* I and transcribed by using T7 (for *Eco*RI-cut) or T3 (*Xba* I-cut) RNA polymerase. The expressed RNAs were translated in a rabbit reticulocyte lysate system (Amersham) in the presence of [^{35}S]methionine (Amersham). Immunoprecipitation was performed as described previously (3), and the precipitates were analyzed by $NaDodSO_4$/polyacrylamide gel electrophoresis and autoradiography.

RESULTS

Isolation of cDNA Encoding Ku Polypeptide. Using a patient anti-Ku serum as a probe, we screened 1×10^6 clones of the human spleen cDNA library in phage λgt11 and isolated 3 positive clones, termed K14, K68, and K71. These clones were recognized by all (10 of 10 for K14 and K68) or nearly all (9 of 10 for K71, see below) anti-Ku sera when their fusion proteins were induced by isopropyl thiogalactoside. None of the clones were bound by individual sera containing anti-

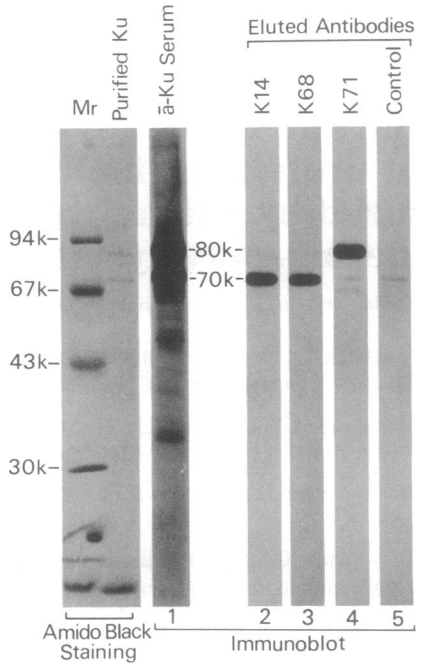

FIG. 1. Characterization of antigenic specificities of isolated clones by elution blot. Affinity-purified antibodies eluted from membrane-bound fusion proteins that had reacted with an anti-Ku serum were used to probe the HeLa cell Ku antigen in immunoblotting. Lanes: 1, a whole anti-Ku serum (MO); 2, antibodies eluted from the phage K14; 3, antibodies eluted from the phage K68; 4, antibodies eluted from the phage K71; 5, antibodies eluted from a wild-type λgt11 phage; Mr, molecular weight markers (94k = 94,000, etc.).

U1RNP, anti-Sm, anti-La/SSB, and anti-Ro/SSA autoantibodies or by a normal control serum. Restriction enzyme *Eco*RI digestion of the recombinant phage DNA demonstrated 0.8-, 1.8-, and 1.0-kb cDNA inserts in K14, K68, and K71 clones, respectively. In immunoblotting, an anti-Ku serum (patient MO; but no normal control serum) known to have antibodies to both 70- and 80-kDa Ku subunits recognized the 130-, 150-, and 130-kDa fusion proteins in K14, K68, and K71 phage lysates, respectively. In contrast, an anti-Ku serum specific for the 70-kDa subunit (patient TH) gave positive results with only the K14 and K68 phages. These results suggested that the K71 clone corresponds to the 80-kDa Ku subunit (data not shown).

To identify the clones with greater certainty, affinity-purified antibodies eluted from membrane-bound fusion proteins were used to probe purified Ku proteins in immuno-

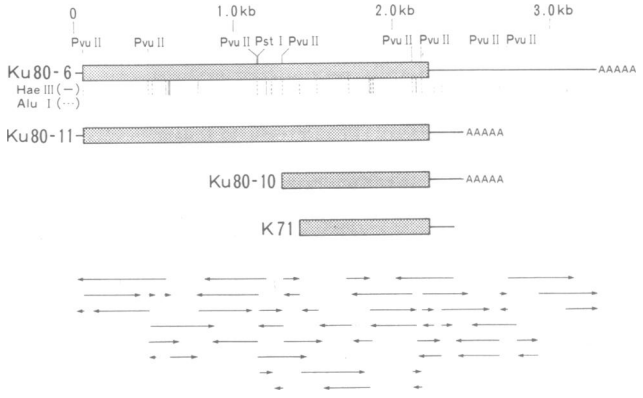

FIG. 2. Restriction map of cDNAs encoding the 80-kDa Ku polypeptide and strategy for determining their sequences. Arrowheads show the direction of sequencing.

blots. Fig. 1 shows results of such experiments. A starting whole anti-Ku serum (MO) contained antibodies which bound both Ku components (lane 1). Antibodies from this serum purified by elution from the K14 (lane 2) and K68 (lane 3) fusion proteins bound selectively to the 70-kDa Ku protein, whereas such antibodies recovered from the K71 fusion protein recognized only the 80-kDa Ku protein (lane 4). As a control, antibodies eluted from an indifferent clone (wild-type λgt11) did not react with any of the purified Ku polypeptides (lane 5). Therefore, the K14 and K68 cDNA inserts must correspond to the 70-kDa Ku polypeptide, while the K71 clone encodes the 80-kDa Ku component.

RNA Blot Analysis. The K68 and K71 cDNA inserts were used to probe HeLa cell mRNA in RNA blots. The K68 probe (encoding the 70-kDa subunit) hybridized to a single mRNA of 2.4 kb, whereas the K71 cDNA (encoding the 80-kDa

```
1            CGACCAAAGCGCCTGAGGACCGGCAAC ATG GTG CGG TCG GGG AAT AAG GCA GCT GTT GTG CTG TGT ATG GAC GTG GGC TTT ACC ATG AGT  90
[1]                                      Met Val Arg Ser Gly Asn Lys Ala Ala Val Val Leu Cys Met Asp Val Gly Phe Thr Met Ser [21]

91    AAC TCC ATT CCT GGT ATA GAA TCC CCA TTT GAA CAA GCA AAG AAG GTG ATA ACC ATG TTT GTA CAG CGA CAG GTG TTT GCT GAG AAC AAG  180
[22]  Asn Ser Ile Pro Gly Ile Glu Ser Pro Phe Glu Gln Ala Lys Lys Val Ile Thr Met Phe Val Gln Arg Gln Val Phe Ala Glu Asn Lys [51]

181   GAT GAG ATT GCT TTA GTC CTG TTT GGT ACA GAT GGC ACT GAC AAT CCC CTT TCT GGT GGG GAT CAG TAT CAG AAC ATC ACA GTG CAC AGA  271
[52]  Asp Glu Ile Ala Leu Val Leu Phe Gly Thr Asp Gly Thr Asp Asn Pro Leu Ser Gly Gly Asp Gln Tyr Gln Asn Ile Thr Val His Arg [81]

271   CAT CTG ATG CTA CCA GAT TTT GAT TTG CTG GAG GAC ATT GAA AGC AAA ATC CAA CCA GGT TCT CAA CAG GCT GAC TTC CTG GAT GCA CTA  360
[82]  His Leu Met Leu Pro Asp Phe Asp Leu Leu Glu Asp Ile Glu Ser Lys Ile Gln Pro Gly Ser Gln Gln Ala Asp Phe Leu Asp Ala Leu [111]

361   ATC GTG AGC ATG GAT GTG ATT CAA CAT GAA ACA ATA GGA AAG AAG TTT GAG AAG AGG CAT ATT GAA ATA TTC ACT GAC CTC AGC AGC CGA  450
[112] Ile Val Ser Met Asp Val Ile Gln His Glu Thr Ile Gly Lys Lys Phe Glu Lys Arg His Ile Glu Ile Phe Thr Asp Leu Ser Ser Arg [141]
                                                                                                                         *
451   TTC AGC AAA AGT CAG CTG GAT ATT ATA ATT CAT AGC TTG AAG AAA TGT GAC ATC TCC CTG CAA TTC TTC TTG CCT TTC TCA CTT GGC AAG  540
[142] Phe Ser Lys Ser Gln Leu Asp Ile Ile Ile His Ser Leu Lys Lys Cys Asp Ile Ser Leu Gln Phe Phe Leu Pro Phe Ser Leu Gly Lys [171]
                       *                           *                           *                           *
541   GAA GAT GGA AGT GGG GAC AGA GGA GAT GGC CCC TTT CGC TTA GGT GGC CAT GGG CCT TCC TTT CCA CTA AAA GGA ATT ACC GAA CAG CAA  630
[172] Glu Asp Gly Ser Gly Asp Arg Gly Asp Gly Pro Phe Arg Leu Gly Gly His Gly Pro Ser Phe Pro Leu Lys Gly Ile Thr Glu Gln Gln [201]
                  *
631   AAA GAA GGT CTT GAG ATA GTG AAA ATG GTG ATG ATA TCT TTA GAA GGT GAA GAT GGG TTG GAT GAA ATT TAT TCA TTC AGT GAG AGT CTG  720
[202] Lys Glu Gly Leu Glu Ile Val Lys Met Val Met Ile Ser Leu Glu Gly Glu Asp Gly Leu Asp Glu Ile Tyr Ser Phe Ser Glu Ser Leu [231]

721   AGA AAA CTG TGC GTC TTC AAG AAA ATT GAG AGG CAT TCC ATT CAC TGG CCC TGC CGA CTG ACC ATT GGC TCC AAT TTG TCT ATA AGG ATT  810
[232] Arg Lys Leu Cys Val Phe Lys Lys Ile Glu Arg His Ser Ile His Trp Pro Cys Arg Leu Thr Ile Gly Ser Asn Leu Ser Ile Arg Ile [261]

811   GCA GCC TAT AAA TCG ATT CTA CAG GAG AGA GTT AAA AAG ACT TGG ACA GTT GTG GAT GCA AAA ACC CTA AAA AAA GAA GAT ATA CAA AAA  900
[262] Ala Ala Tyr Lys Ser Ile Leu Gln Glu Arg Val Lys Lys Thr Trp Thr Val Val Asp Ala Lys Thr Leu Lys Lys Glu Asp Ile Gln Lys [291]

901   GAA ACA GTT TAT TGC TTA AAT GAT GAT GAT GAA ACT GAA GTT TTA AAA GAG GAT ATT ATT CAA GGG TTC CGC TAT GGA AGT GAT ATA GTT  990
[292] Glu Thr Val Tyr Cys Leu Asn Asp Asp Asp Glu Thr Glu Val Leu Lys Glu Asp Ile Ile Gln Gly Phe Arg Tyr Gly Ser Asp Ile Val [321]

991   CCT TTC TCT AAA GTG GAT GAG GAA CAA ATG AAA TAT AAA TCG GAG GGG AAG TGC TTC TCT GTT TTG GGA TTT TGT AAA TCT TCT CAG GTT  1080
[322] Pro Phe Ser Lys Val Asp Glu Glu Gln Met Lys Tyr Lys Ser Glu Gly Lys Cys Phe Ser Val Leu Gly Phe Cys Lys Ser Ser Gln Val [351]

1081  CAG AGA AGA TTC TTC ATG GGA AAT CAA GTT CTA AAG GTC TTT GCA GCA AGA GAT GAT GAG GCA GCT GCA GTT GCA CTT TCC TCC CTG ATT  1170
[352] Gln Arg Arg Phe Phe Met Gly Asn Gln Val Leu Lys Val Phe Ala Ala Arg Asp Asp Glu Ala Ala Ala Val Ala Leu Ser Ser Leu Ile [381]

1171  CAT GCT TTG GAT GAC TTA GAC ATG GTG GCC ATA GTT CGA TAT GCT TAT GAC AAA AGA GCT AAT CCT CAA GTC GGC GTG GCT TTT CCT CAT  1260
[382] His Ala Leu Asp Asp Leu Asp Met Val Ala Ile Val Arg Tyr Ala Tyr Asp Lys Arg Ala Asn Pro Gln Val Gly Val Ala Phe Pro His [411]

1261  ATC AAG CAT AAC TAT GAG TGT TTA GTG TAT GTG CAG CTG CCT TTC ATG GAA GAC TTG CGG CAA TAC ATG TTT TCA TCC TTG AAA AAC AGT  1350
[412] Ile Lys His Asn Tyr Glu Cys Leu Val Tyr Val Gln Leu Pro Phe Met Glu Asp Leu Arg Gln Tyr Met Phe Ser Ser Leu Lys Asn Ser [441]

1351  AAG AAA TAT GCT CCC ACC GAG GCA CAG TTG AAT GCT GTT GAT GCT TTG ATT GAC TCC ATG AGC TTG GCA AAG AAA GAT GAG AAG ACA GAC  1440
[442] Lys Lys Tyr Ala Pro Thr Glu Ala Gln Leu Asn Ala Val Asp Ala Leu Ile Asp Ser Met Ser Leu Ala Lys Lys Asp Glu Lys Thr Asp [471]

1441  ACC CTT GAA GAC TTG TTT CCA ACC ACC AAA ATC CCA AAT CCT CGA TTT CAG AGA TTA TTT CAG TGT CTG CTG CAC AGA GCT TTA CAT CCC  1530
[472] Thr Leu Glu Asp Leu Phe Pro Thr Thr Lys Ile Pro Asn Pro Arg Phe Gln Arg Leu Phe Gln Cys Leu Leu His Arg Ala Leu His Pro [501]

1531  CGG GAG CCT CTA CCC CCA ATT CAG CAG CAT ATT TGG AAT ATG CTG AAT CCT CCC GCT GAG GTG ACA ACA AAA AGT CAG ATT CCT CTC TCT  1620
[502] Arg Glu Pro Leu Pro Pro Ile Gln Gln His Ile Trp Asn Met Leu Asn Pro Pro Ala Glu Val Thr Thr Lys Ser Gln Ile Pro Leu Ser [531]

1621  AAA ATA AAG ACC CTT TTT CCT CTG ATT GAA GCC AAG AAA AAG GAT CAA GTG ACT GCT CAG GAA ATT TTC CAA GAC AAC CAT GAA GAT GGA  1710
[532] Lys Ile Lys Thr Leu Phe Pro Leu Ile Glu Ala Lys Lys Lys Asp Gln Val Thr Ala Gln Glu Ile Phe Gln Asp Asn His Glu Asp Gly [561]

1711  CCT ACA GCT AAA AAA TTA AAG ACT GAG CAA GGG GGA GCC CAC TTC AGC GTC TCC AGT CTG GCT GAA GGC AGT GTC ACC TCT GTT GGA AGT  1800
[562] Pro Thr Ala Lys Lys Leu Lys Thr Glu Gln Gly Gly Ala His Phe Ser Val Ser Ser Leu Ala Glu Gly Ser Val Thr Ser Val Gly Ser [591]

1801  GTG AAT CCT GCT GAA AAC TTC CGT GTT CTA GTG AAA CAG AAG AAG GCC AGC TTT GAG GAA GCG AGT AAC CAG CTC ATA AAT CAC ATC GAA  1890
[592] Val Asn Pro Ala Glu Asn Phe Arg Val Leu Val Lys Gln Lys Lys Ala Ser Phe Glu Glu Ala Ser Asn Gln Leu Ile Asn His Ile Glu [621]

1891  CAG TTT TTG GAT ACT AAT GAA ACA CCG TAT TTT ATG AAG AGC ATA GAC TGC ATC CGA GCC TTC CGG GAA GAA GCC ATT AAG TTT TCA GAA  1980
[622] Gln Phe Leu Asp Thr Asn Glu Thr Pro Tyr Phe Met Lys Ser Ile Asp Cys Ile Arg Ala Phe Arg Glu Glu Ala Ile Lys Phe Ser Glu [651]

1981  GAG CAG CGC TTT AAC AAC TTC CTG AAA GCC CTT CAA GAG AAA GTG GAA ATT AAA CAA TTA AAT CAT TTC TGG GAA ATT GTT GTC CAG GAT  2070
[652] Glu Gln Arg Phe Asn Asn Phe Leu Lys Ala Leu Gln Glu Lys Val Glu Ile Lys Gln Leu Asn His Phe Trp Glu Ile Val Val Gln Asp [681]

2071  GGA ATT ACT CTG ATC ACC AAA GAG GAA GCC TCT GGA AGT TCT GTC ACA GCT GAG GAA GCC AAA AAG TTT CTG GCC CCC AAA GAC AAA CCA  2160
[682] Gly Ile Thr Leu Ile Thr Lys Glu Glu Ala Ser Gly Ser Ser Val Thr Ala Glu Glu Ala Lys Lys Phe Leu Ala Pro Lys Asp Lys Pro [711]

2161  AGT GGA GAC ACA GCA GCT GTA TTT GAA GAA GGT GGT GAT GTG GAC GAT TTA TTG GAC ATG ATA TAG GTCGTGGATGTATGGGGAATCTAAGAGAGCT  2257
[712] Ser Gly Asp Thr Ala Ala Val Phe Glu Glu Gly Gly Asp Val Asp Asp Leu Leu Asp Met Ile ***                                [732]

2258  GCCATCGCTGTGATGCTGGGAGTTCTAACAAAACAAGTTGGATGCGGCCATTCAAGGGGAGCCAAAATCTCAAGAAATTCCCAGCAGGTTACCTGGAGGCGGATCATCTAATTCTCTGT  2376
                                                                        ▼
2377  GGAATGAATACACACATATATATTACAAGGGATAATTTAGACCCCATACAAGTTTATAAAGAGTCATTGTTATTTTCTGGTTGGTGTATTATTTTTTCTGTGGTCTTACTGATCTTTGT  2495

2496  ATATTACATACATGCTTTGAAGTTTCTGGAAAGTAGATCTTTTCTTGACCTAGTATATCAGTGACAGTTGCAGCCCTTGTGATGTGATTAGTGTCTCATGTGGAACCATGGCATGGTTA  2614

2615  TTGATGAGTTTCTTAACCCTTTCCAGAGTCCTCCTTTGCCTGATCCTCCAACAGCTGTCACAACTTGTGTTGAGCAAGCAGTAGCATTTGCTTCCTCCCAACAAGCAGCTGGGTTAGGA  2733

2734  AAACCATGGGTAAGGACGGACTCACTTCTCTTTTTAGTTGAGGCCTTCTAGTTACCACATTACTCTGCCTCTGTATATAGGTGGTTTTCTTTAAGTGGGGTGGGAAGGGGAGCACAATT  2852

2853  TCCCTTCATACTCCTTTTAAGCAGTGAGTTATGGTGGTGGTCTCATGAAGAAAAGACCTTTTGGCCCAATCTCTGCCATATCAGTGAACCTTTAGAAACTCAAAAACTGAGAAATTTAC  2971

2972  TTCAGTAGTTAGAATTATATCACTTCACTGTTCTCTACTTGCAAGCCTCAAAGAGAGAAAGTTTCGTTATATTAAAACACTTAGGTAACTTTTCGGTCTTTCCCATTTCTACCTAAGTC  3090

3091  AGCTTTCATCTTTGTGGATGGTGTCTCCTTTACTAAATAAGAAAATAACAAAGCCCTTATTCTCTTTTTTTCTTGTCCTCATTCTTGCCTTGAGTTCCAGTTCCTCTTTGGTGTACAGA  3209

3210  CTTCTTGGTACCCAGTCACCTCTGTCTTCAGCACCCTCATAAGTCGTCACTAATACACAGTTTTGTACATGTAACATTAAAGGCATAAATGACTC(A)100                     3304
```

FIG. 3. Nucleotide and deduced amino acid sequence of the Ku80-6 cDNA. Ku80-11 has an identical open reading frame but shorter 5′- and 3′-untranslated regions (see text). * indicates a periodic repeat of leucine and/or serine in every seventh amino acid residue. Underlines indicate polyadenylation signals. ▼ is a site where the poly(A) tail of Ku80-10 and Ku80-11 starts.

subunit) bound to two mRNA species of 2.6 and 3.4 kb (data not shown). These results indicate separate mRNAs for the two Ku subunits. In addition, it was apparent that the cloned cDNAs were smaller than their mRNAs and thus represented only partial-length sequences.

Isolation and Characterization of Full-Length cDNAs Encoding the 80-kDa Ku Polypeptide. Using the K71 insert cDNA as a hybridization probe, we screened the pCD cDNA library constructed from mRNA of simian virus 40-transformed human fibroblasts. Among 5×10^4 colonies of transformed cells, 3 positive clones were isolated and referred to as Ku80-6, Ku80-10, and Ku80-11. Of these clones, Ku80-6 had the longest insert cDNA (*Bam*HI fragment), 3.4 kb, identical in size with the larger mRNA which hybridized to the K71 cDNA (see above).

We next determined the nucleotide sequences of all cDNAs isolated from the pCD and λgt11 libraries. The restriction map and sequencing strategy for the Ku80-6 cDNA are shown in Fig. 2, along with a comparison of the shorter Ku80-10, Ku80-11, and K71 cDNAs. The Ku80-6 cDNA proved to be 3304 nucleotides in length, with an additional poly(A) tail of approximately 100 nucleotides. Its primary nucleotide and deduced amino acid sequence are shown in Fig. 3. The only difference between the Ku80-6 and Ku80-11 cDNAs was that the latter lacked two nucleotides at its 5′ end and the 856 nucleotides at its 3′ end. The nucleotide sequences of cDNA Ku80-6 and Ku80-11 contain a single long open reading frame of 2196 base pairs (bp) (nucleotides 28–2223 of Ku80-6) that encodes 732 amino acids. The predicted molecular weight for the encoded polypeptide is 82,713, which closely approximates the apparent molecular weight of the 80-kDa Ku subunit estimated by $NaDodSO_4$/polyacrylamide gel electrophoresis.

Several features characterize the Ku80-6 cDNA sequence. The putative polypeptide has a high content of acidic amino acids (14.9 mol % aspartate and glutamate), resulting in a low pI (6.03). A periodic repetition of leucine residues (or replacements with serine) in every seventh position occurs at amino acids 140–175, thus forming a sequence consistent with the "leucine zipper" structure (14, 15). This region is followed by a cluster of basic amino acids (amino acids 120–131), as recognized in other proteins with the leucine zipper (16). The open reading frame is preceded by a 5′ untranslated region of 27 bp. The sequence GGCAACATGG (nucleotides 22–31) surrounds the proposed initiator methionine codon and closely approximates the potential ribosomal binding sequence, GCCACCATGG (17). The TAG stop codon (positions 2224–2226) is followed by a long 3′ noncoding region of 1081 bp. This region includes four putative polyadenylylation signals at nucleotides 2431–2436 (TATAAA), 3125–3130 (AAATAA), 3133–3138 (AAATAA), and 3285–3290 (ATTAAA) (18). A poly(A) tail starts at nucleotide 3305 in the Ku80-6 cDNA and at the corresponding upstream sites (nucleotide 2449 of the Ku80-6 sequence) in the Ku80-10 and Ku80-11 cDNAs (see Fig. 3 and *Discussion*). In searches of both DNA (GenBank; version 57) and protein (National Biomedical Research Foundation; version 17) data banks, we found no homologies between the Ku80-6 cDNA and any known sequences at either the nucleotide or the amino acid level.

***In Vitro* Expression of the Ku80-11 cDNA.** To confirm the identity and full length of the isolated cDNA, the Ku80-11 cDNA was inserted into a pBluescript expression vector and mRNAs transcribed *in vitro* were used in a rabbit reticulocyte system for *in vitro* translation. As shown in Fig. 4, the polypeptide translated from the sense mRNA was immunoprecipitated with an anti-Ku serum (lane 2) and comigrated with the 80-kDa subunit of HeLa cell Ku protein (lane 1). Translation of the antisense mRNA did not yield any immunoprecipitable polypeptide (lane 3), and a normal human serum did not precipitate the translated polypeptide from the sense mRNA (lane 4).

DNA Blot Analysis of Genomic DNA. To analyze the genomic structure of the Ku gene, human genomic DNA was analyzed in DNA blots. In initial studies, human leukocyte DNA was cleaved with *Eco*RI and *Hin*dIII and probed with the Ku80-6 cDNA from which the poly(A) tail was removed

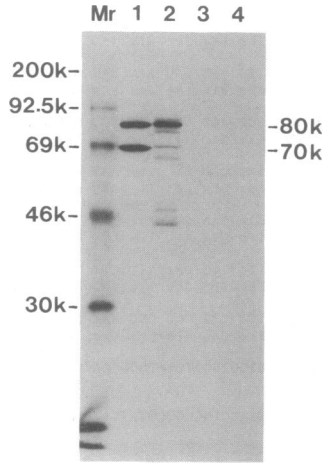

FIG. 4. Immunoprecipitation of an *in vitro* translated polypeptide from the clone Ku80-11. The Ku80-11 cDNA insert was subcloned in pBluescript plasmid DNA, transcribed *in vitro* with T3 or T7 RNA polymerase, and translated in a rabbit reticulocyte lysate system. Lanes: Mr, molecular weight markers; 1, the Ku polypeptides immunoprecipitated from [^{35}S]methionine-labeled HeLa cell extracts; 2, sense mRNA (T3 RNA polymerase transcript) was translated *in vitro* and the ^{35}S-radiolabeled polypeptide was immunoprecipitated with an anti-Ku serum; 3, antisense mRNA (T7 RNA polymerase transcript) was treated with the same procedure as for lane 2; 4, the polypeptide of lane 2 was immunoprecipitated with a normal human serum.

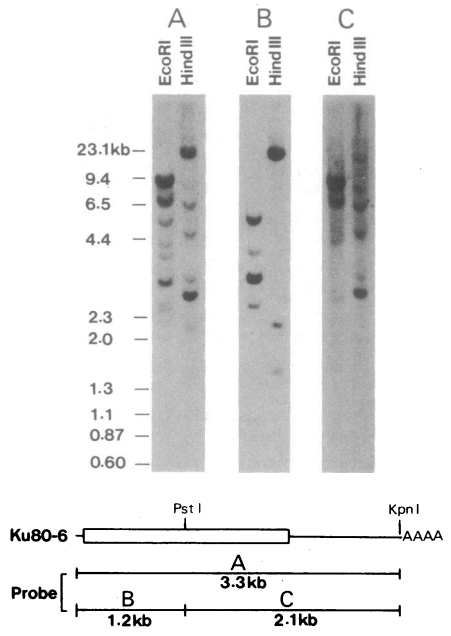

FIG. 5. DNA blot analysis of human genomic DNA hybridized with the Ku80-6 cDNA. Human leukocyte DNA was digested with *Eco*RI or *Hin*dIII, fractionated on a 1% agarose gel, and transferred to nitrocellulose. DNA blots were probed with Ku80-6 cDNA (3.4-kb *Bam*HI fragment) from which the poly(A) tail was removed by cutting with *Kpn* I (blot A), 5′ *Pst* I fragment (1.2 kb) (blot B), or 3′ *Pst* I fragment (2.1 kb) (blot C).

by cutting with *Kpn* I (probe A in Fig. 5). Seven or eight different DNA bands were recognized in each digest examined with this probe (blot A). This finding suggests several possibilities; the presence of multiple genes encoding the 80-kDa Ku protein, pseudogenes, or multiple introns within a single gene. To distinguish among these possibilities, the Ku80-6 cDNA (probe A) was cut with *Pst* I to form a 5′ fragment of 1.2 kb (probe B) and a 3′ fragment of 2.2 kb (probe C). As shown in blots B and C of Fig. 5, each of these probes hybridized with several nonoverlapping DNA bands. In repeated studies these digestion patterns were always reproducible even when digestion times were extended. Since the Ku80-6 cDNA contains no *Eco*RI or *Hin*dIII sites, these results indicate the presence of multiple introns within the gene for the 80-kDa Ku subunit.

DISCUSSION

We have derived a cDNA referred to as Ku80-6 that contains a full-length sequence for the 80-kDa subunit of the Ku autoantigen. The identification of this clone is based on the observation that expression of a corresponding partial-length cDNA termed K71 in phage λgt11 leads to the generation of a fusion protein that is recognized by all anti-Ku sera that contain antibodies to this Ku component. Moreover, antibodies eluted from phage expressing this fusion protein selectively recognize the 80-kDa Ku subunit. For several reasons, we conclude that cDNA Ku80-6 and Ku80-11 are full-length clones: (*i*) their sizes are nearly identical with each of two mRNAs; (*ii*) the molecular weight of the putative polypeptide encoded by its longest open reading frame is 82,713, which agrees with the apparent molecular weight of the 80-kDa Ku polypeptide determined by $NaDodSO_4$/polyacrylamide gel electrophoresis (2, 3, 5, 6); (*iii*) the initiator methionine codon is surrounded by the sequence GGCAACATGG, which closely approximates the ideal ribosomal binding sequence GCCACCATGG (17); and (*iv*) *in vitro* expression study using the Ku80-11 cDNA yielded an immunoprecipitable polypeptide which comigrates with the 80-kDa Ku protein.

Recently Yaneva *et al.* (19) described a cDNA for the larger Ku protein subunit which is identical within its open reading frame to the sequence reported here. In addition, they found that their cDNA hybridizes to two mRNAs which closely approximate the 3.4- and 2.6-kb mRNAs detected in our RNA blots. In comparison with the sequence of Yaneva *et al.* (19), our longest sequence (Ku80-6) lacks 6 nucleotides at its 5′ end but contains an additional 258 nucleotides of the 3′ noncoding sequence and a poly(A) tail. At positions 1599 and 2677 we find guanines in place of adenines, but the open reading frames encode identical amino acid sequences. Thus, the present study confirms the previously reported cDNA sequence and extends this work by demonstrating that the cDNA product is immunologically and electrophoretically identical to the 80-kDa Ku subunit.

The present work also sheds new light on how the two mRNAs for the Ku polypeptide are related. First it should be noted that they closely approximate the Ku80-6 and Ku80-11 cDNAs in size, and it is tempting to speculate that these cDNAs represent the full length of both messages. These two cDNAs have identical open reading frames and encode identical amino acid sequences. They differ only because the Ku80-11 cDNA lacks 2 nucleotides at its extreme 5′ end and the 856 nucleotides which immediately precede its poly(A) tail. The noncoding region of the Ku80-6 cDNA contains four putative polyadenylylation signals, but only one of these is found in the noncoding region of cDNA Ku80-11. The simplest explanation is that both of these cDNAs are derived from a common gene and differ because alternative polyadenylylation signals are utilized for initiation of the poly(A) tails. Finally, the DNA blots prepared with probes derived from the opposite ends of cDNA Ku80-6 demonstrate hybridization to nonoverlapping restriction fragments derived from human genomic DNA. Thus, the gene encoding the 80-kDa Ku polypeptide is likely to be divided into several portions by intervening sequences. This result does not exclude the possible presence of multiple genes or pseudogenes.

The complete nucleotide and amino acid sequence of a full-length cDNA encoding the 70-kDa Ku polypeptide (p70) was recently described by Reeves and Sthoeger (7). Their studies indicated a similarity between the 70-kDa Ku polypeptide and *myc* gene family products. The corresponding region contains a periodic repeat of leucine and serine in every seventh position, thus forming a leucine zipper (14, 15). The putative polypeptide encoded by cDNA Ku80-6 also contains a leucine and/or serine repeat at every seventh position between amino acid residues 140 and 175. This region is followed by a cluster of basic amino acids which might constitute a DNA-binding site (16). Therefore, it seems reasonable to hypothesize that the two Ku protein subunits interact at these sites. This idea is consistent with the earlier observation that the two Ku polypeptides form a stable heterodimer complex that cannot dissociate easily (2, 3, 5, 6).

Most patient sera which contain anti-Ku autoantibodies recognize both Ku subunits in immunoblots (3, 5), and immunologic cross-reactivity of the two polypeptides has been reported (5). However, the primary sequence of cDNA Ku80-6 shows no homology to that predicted from the cDNAs for the 70-kDa Ku subunit (p70) (7), and it seems clear that the two polypeptides are products of different genes. It is also unlikely that the two proteins share cross-reactive epitopes. Rather, both Ku subunits are usually targeted for autoimmune responses simultaneously in a focused manner.

We thank Dr. Joan A. Steitz for the kind gift of the human fibroblast cDNA library. We are also grateful to her and Drs. Joe Craft, Mark Mamula, and Peter Blier for helpful discussion and review of this manuscript. This work was supported by grants-in-aid for scientific research from the Ministry of Education, Science and Culture of the Japanese Government (62570298 and 0157369), grants from the Foundation for Intractable Diseases, the Uehara Memorial Foundation, and the Japan Rheumatism Foundation, and a grant from Keio University. Work in the United States was supported in part by grants from the National Institutes of Health (AR32549 and AM07107), the Arthritis Foundation and its Connecticut chapter, the Connecticut chapter of the Lupus Foundation, and the Scleroderma Society of New York.

1. Mimori, T., Akizuki, M., Yamagata, H., Inada, S., Yoshida, S. & Homma, M. (1981) *J. Clin. Invest.* **68**, 611–620.
2. Reeves, W. H. (1985) *J. Exp. Med.* **161**, 18–39.
3. Mimori, T., Hardin, J. A. & Steitz, J. A. (1986) *J. Biol. Chem.* **261**, 2274–2278.
4. Mimori, T. & Hardin, J. A. (1986) *J. Biol. Chem.* **261**, 10375–10379.
5. Francoeur, A.-M., Peebles, C. L., Gompper, P. T. & Tan, E. M. (1986) *J. Immunol.* **136**, 1648–1653.
6. Yaneva, M. & Busch, H. (1986) *Biochemistry* **25**, 5057–5063.
7. Reeves, W. H. & Sthoeger, Z. M. (1989) *J. Biol. Chem.* **264**, 5047–5052.
8. Young, R. & Davis, R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194–1198.
9. Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2**, 161–170.
10. Feinberg, A. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
12. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
13. Gross-Bellard, M., Oudet, P. & Chambon, P. (1973) *Eur. J. Biochem.* **36**, 32–38.
14. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
15. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1989) *Science* **243**, 1681–1688.
16. Gentz, R., Rauscher, F. J., Abate, C. & Curran, T. (1989) *Science* **243**, 1695–1699.
17. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
18. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41**, 349–359.
19. Yaneva, M., Wen, J., Ayala, A. & Cook, R. (1989) *J. Biol. Chem.* **264**, 13407–13411.