Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 7317–7321, October 1988
Immunology

Primary structure of the human M2 mitochondrial autoantigen of primary biliary cirrhosis: Dihydrolipoamide acetyltransferase

(autoantibodies/molecular cloning)

ROSS L. COPPEL*†, L. JANE MCNEILAGE*, CHARLES D. SURH‡, JUDY VAN DE WATER‡, TERRY W. SPITHILL*, SENGA WHITTINGHAM*, AND M. ERIC GERSHWIN‡

*The Walter and Eliza Hall Institute of Medical Research, Royal Parade, Parkville Victoria 3050, Australia; and ‡Division of Rheumatology/Allergy and Clinical Immunology, University of California at Davis, Davis, CA 95616

Communicated by G. J. V. Nossal, June 2, 1988

ABSTRACT **Primary biliary cirrhosis is a chronic, destructive autoimmune liver disease of humans. Patient sera are characterized by a high frequency (>95%) of autoantibodies to a M_r 70,000 mitochondrial antigen, a component of the M2 antigen complex. We have identified a human cDNA clone encoding the complete amino acid sequence of this autoantigen. The predicted structure has significant similarity with the dihydrolipoamide acetyltransferase (EC 2.3.1.12) of the *Escherichia coli* pyruvate dehydrogenase multienzyme complex. The human sequence preserves the Glu-Thr-Asp-Lys-Ala motif of the lipoyl-binding site and has two potential binding sites. Expressed fragments of the cDNA react strongly with sera from patients with primary biliary cirrhosis but not with sera from patients with autoimmune chronic active hepatitis or sera from healthy subjects.**

Primary biliary cirrhosis (PBC) is an enigmatic chronic autoimmune liver disease characterized by inflammation and obliteration of the intrahepatic bile ducts (1) and the presence of high-titer autoantibodies to mitochondria (2). These autoantibodies react with multiple targets within mitochondria designated M1–M9 (3), but PBC-specific responses are directed only to the M2, M4, and M8 proteins. Of these, anti-M2 is the most common specificity in PBC sera, with >95% of patients having these antibodies. Patients with other clinical conditions may have anti-mitochondrial antibodies as demonstrated by immunofluorescence, but these tend to be directed to components of the mitochondria other than M2.

The M2 complex was originally described as a trypsin-sensitive protein on the inner mitochondrial membrane (4). Reactivity of anti-M2 antibodies with submitochondrial fractions suggested that M2 comprised the F_1 portion of the H^+-ATPase (4, 5) or the adenine nucleotide translocator (6). However, recent studies have failed to demonstrate reactivity of anti-M2 antibodies with any of the isolated major subunits of the F_1 ATPase (7, 8), and anti-adenine nucleotide translocator reactivity has been distinguished from anti-M2 reactivity (9) and is not specific for PBC (10). Recently the M2 complex has been shown to comprise three distinct inner mitochondrial membrane proteins: two minor autoantigens of M_r 39,000 and 45,000 and a dominant autoantigen of M_r 70,000 (7).

We have previously identified a cDNA from a rat liver library that encoded a portion of the M_r 70,000 component of the M2 complex (11). By using this sequence as a probe, we have isolated a clone encoding the complete structure of the human M_r 70,000 polypeptide.§ This sequence shows 86% identity at the amino acid level with the rat M_r 70,000 M2 sequence and shows significant similarity to *Escherichia coli* dihydrolipoamide acetyltransferase (E2p; EC 2.3.1.12), a component of the inner mitochondrial membrane multienzyme complex, the pyruvate dehydrogenase (PDH) complex.

MATERIALS AND METHODS

Cloning Methods. The human placental cDNA library in λgt11 was obtained from Clontech (Palo Alto, CA). The insert of the rat M2 cDNA clone pRMIT was nick-translated and used to screen the plated library by plaque hybridization (12). Positive plaques were passaged to homogeneity, and DNA was prepared as described (12). The phage DNA was digested with *Eco*RI, and the inserts were separated by agarose gel electrophoresis and purified by using the glass milk procedure (Bio 101 Pty., La Jolla, CA). Inserts were ligated separately into phage M13 (13), and the nucleotide sequence was determined by the dideoxy method (14). Each fragment was sequenced completely in both orientations by using synthetic oligonucleotides to prime the reaction.

Expressing Clones. Fragments of the pHUMIT sequence were ligated into appropriate plasmid vectors to ensure in-frame translation of the insert (15–17). Liquid cultures were induced to maximize expression (11, 15), and lysates of expressing clones were subjected to $NaDodSO_4$/PAGE and immunoblotting as described (11).

DNA and RNA Blot Analysis. Southern blot analysis of human total genomic DNA was performed as described (18). Poly(A)$^+$ RNA was prepared from total cellular RNA from several human cell lines and was analyzed by blot hybridization as described (18). Washes were performed in 2× SSC (1× SSC = 0.15 M NaCl/0.015 sodium citrate, pH 7) at 65°C.

Sera. Five anti-mitochondrial autoantibody positive sera were obtained from patients who met established criteria for PBC. These sera gave the characteristic anti-mitochondrial autoantibody pattern of immunofluorescence and reacted with the M_r 70,000 M2 antigen on immunoblots of human placental mitochondria. Sera from five patients with autoimmune chronic active hepatitis were used as controls. All sera were obtained from patients and normal subjects giving informed consent.

RESULTS

Nucleotide Sequence of the M2 Autoantigen. Screening of a human placental library with the rat liver M2 cDNA clone pRMIT identified four hybridizing clones. The longest of these, with a total insert length of 2540 base pairs, was designated pHUMIT and was selected for study. DNA was

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.

Abbreviations: E2p, dihydrolipoamide acetyltransferase; PBC, primary biliary cirrhosis; PDH, pyruvate dehydrogenase.

†To whom reprint requests should be addressed.

§The sequence reported in this paper is being deposited in the EMBL/GenBank data base (IntelliGenetics, Mountain View, CA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03866).

prepared from the pHUMIT clone and was digested with *Eco*RI. Three fragments of sizes 1.5, 0.8, and 0.3 kilobases (kb) were released, which is consistent with the presence of internal *Eco*RI sites in the sequence. The complete nucleotide sequence of the three fragments was determined. Two of the fragments overlapped the previously determined rat sequence, which allowed determination of the correct arrangement of these fragments. The third fragment (≈0.8 kb) was assigned to the 5′ region of the clone on the basis of restriction mapping studies of the 2540-base-pair insert (data not shown). The complete nucleotide sequence and deduced amino acid sequence are shown in Fig. 1.

A long open reading frame commencing at position 636 and terminating at 2477 predicts a protein of 614 amino acids with a M_r of 65,640. There are two in-frame methionines at the beginning of the open reading frame. The first of these precedes a sequence of 18 residues that contains two arginines and two serines and a number of hydrophobic residues. If residues 1–18 are plotted on a helical wheel projection beginning with methionine at position 1, they can be arranged to form an amphipathic α-helix, a structure that has been described as typical of the signal sequence of proteins transported into mitochondria (19). In addition, residues 1–18 lack aspartate, glutamate, valine, and isoleucine, another feature typical of mitochondrial targeting sequences (8). We have previously shown that the rat M2 sequence was encoded by nuclear DNA (11), and therefore such a signal sequence would be expected to be present at the amino terminus of the M2 protein to allow entry into the mitochondria. In contrast, the residues following the second in-frame methionine (residues 18–36) do not form an amphipathic α-helical structure of appropriate composition. We therefore believe that the methionine at nucleotides 636–638 is the first residue of the translated M2 protein. The encoded polypeptide is rich in proline (10.3%) and alanine (9.5%), and there is a region of internally repeated sequence (nucleotides 897–1043 and 1278–1424). There is no sequence element with the characteristics of a classical anchor sequence. The long 5′ untranslated sequence is unusual; however, a stretch of similar length has been reported for another mitochondrial protein, PDH (20).

pHUMIT Is Similar to *E. coli* E2p. Comparison of the amino acid sequence of the predicted pHUMIT polypeptide with sequences in the protein sequence data bases [NBRF (ver-

```
GAATTCCGCGCCAGGGCCGACCGTTCTTCCTGTACTACGCTTCCCACCACACTCACTACCCTCAGTTCAGTGGACAAAGCTTCACCAAGCGCTCAGGCCGTGGGCCATTTGGGGACTCCT
      10        20        30        40        50        60        70        80        90       100       110       120
TGATGGAGCTGGATGGAGCTGTAGGGGCCTTGATGACAACTGTGGGGGACCTCGGTCTGCTGGAAGAGACACTAGTCATCTTCACTGCAGATAACGGTCCTGAGTTGATGCGCATGTCCA
     130       140       150       160       170       180       190       200       210       220       230       240
ATGGCGGCTGCTCTGGCCTCTTGAGATGTGGAAAAGGAACAACTTTTGAAGGTGGCGTCCGAGAGCCTGCCTTGGTCTACTGGCCAGGTCACATTACTCCTGGTGTAACCCATGAGCTGG
     250       260       270       280       290       300       310       320       330       340       350       360
CCAGCTCTCTGGACCTGCTGCCCACCCTGGCAGCCCTGACCGGGGTCCGCTGCCCAACGTCACCTTGGATGGTGTTGACATCAGCCCCTTGCTGCTAGGCACAGGCAAGAGCCCACGGAA
     370       380       390       400       410       420       430       440       450       460       470       480
GTCTGTCTTCTTCTACCCGCCCTACCCAGACGAGATCCATGGGGTCTTTGCTGTTCGGAATGGGAAATACAAGGCTCATTTCTTCACCCAGGGCTCCGCCCACAGTGACACCACTTCAGA
     490       500       510       520       530       540       550       560       570       580       590       600
                                 M   S   P   H   C   S   T   T   Y   L   R   T   L   G   R   T   T   M   F   W   K   T   T   E   G   R   D   G
TCCTGCCTGTCATGCTGCCAACCGTCTGACGGCTCATGAGCCCCCACTGCTCTACGACTTATCTCAGGACCCTGGGGAGAACTACAATGTTTTGGAAAACCACAGAGGGAAGAGATGGGA
     610       620       630       640       650       660       670       680       690       700       710       720
K   M   A   V   Q   E   F   S   E   F   G   L   L   L   Q   L   L   G   S   P   G   R   R   Y   Y   S   L   P   P   H   Q   K   V   P   L   P   S   L   S   P
AAATGGCTGTCCAGGAGTTTTCGGAATTCGGGTTACTGCTGCAGCTTTTGGGGTCGCCCGGCCGCCGCTATTACAGTCTTCCCCCGCATCAGAAGGTTCCATTGCCTTCTCTTTCCCCCA
     730       740       750       760       770       780       790       800       810       820       830       840
T   M   Q   A   G   T   I   A   R   W   E   K   K   E   G   D   K   I   N   E   G   D   L   I   A   E   V   E   T   D   K   A   T   V   G   F   E   S   L   E
CAATGCAGGCAGGCACCATAGCCCGTTGGGAAAAAAAAAGAGGGGGACAAAATCAATGAAGGTGACCTAATTGCAGAGGTTGAAACTGATAAAGCCACTGTTGGATTTGAGAGCCTGGAGG
     850       860       870       880       890       900       910       920       930       940       950       960
E   C   Y   M   A   K   I   L   V   A   E   G   T   R   D   V   P   I   G   A   I   I   C   I   T   V   G   K   P   E   D   I   E   A   F   K   N   Y   T   L
AGTGTTATATGGCAAAGATACTTGTTGCTGAAGGTACCAGGGATGTTCCCATCGGAGCGATCATCTGTATCACAGTTGGCAAGCCTGAGGATATTGAGGCCTTTAAAAATTATACACTGG
     970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080
D   S   S   A   A   P   T   P   Q   A   A   P   A   P   T   P   A   A   T   A   S   P   P   T   P   S   A   Q   A   P   G   S   S   Y   P   P   H   M   Q   V
ATTCCTCAGCAGCACCTACCCCACAAGCGGCCCCAGCACCAACCCCTGCTGCCACTGCTTCGCCACCTACACCTTCTGCTCAGGCTCCTGGTAGCTCATATCCCCCTCACATGCAGGTAC
     1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
L   L   P   A   L   S   P   T   M   T   M   G   T   V   Q   R   W   E   K   K   V   G   E   K   L   S   E   G   D   L   L   A   E   I   E   T   D   K   A   T
TTCTTCCTGCCCTCTCTCCCACCATGACCATGGGCACAGTTCAGAGATGGGAAAAAAAAGTGGGTGAGAAGCTAAGTGAAGGAGACTTACTGGCAGAGATAGAAACTGACAAAGCCACTA
     1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320
I   G   F   E   V   Q   E   E   G   Y   L   A   K   I   L   V   P   E   G   T   R   D   V   P   L   G   T   P   L   C   I   I   V   E   K   E   A   D   I   S
TAGGTTTTGAAGTACAGGAAGAAGGTTATCTGGCAAAAATCCTGGTCCCTGAAGGCACAAGAGATGTCCCTCTAGGAACCCCACTCTGTATCATTGTAGAAAAAGAGGCAGATATATCAG
     1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440
A   F   A   D   Y   R   P   T   E   V   T   D   L   K   P   Q   V   P   P   P   T   P   P   P   V   A   A   V   P   P   T   P   Q   P   L   A   P   T   P   S
CATTTGCTGACTATAGGCCAACCGAAGTAACAGATTTAAAACCACAAGTGCCACCACCTACCCCACCCCCGGTGGCCGCTGTTCCTCCAACTCCCCAGCCTTTAGCTCCTACACCTTCGA
     1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560
T   P   C   P   A   T   P   A   G   P   K   G   R   V   F   V   D   P   L   A   K   K   L   A   V   E   K   G   I   D   L   T   Q   V   K   G   T   G   P   D
CACCCTGCCCAGCTACTCCTGCTGGACCAAAGGGAAGGGTGTTTGTTGACCCTCTTGCAAAGAAGTTGGCAGTAGAGAAAGGGATTGATCTTACACAAGTAAAAGGGACAGGACCAGATG
     1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680
G   R   I   T   K   K   D   I   D   S   F   V   P   S   K   V   A   P   A   P   A   A   V   V   P   P   T   G   P   G   M   A   P   V   P   T   G   V   F   T
GTAGAATCACCAAGAAGGATATCGACTCTTTTGTGCCTAGTAAAGTTGCTCCTGCTCCGGCAGCTGTTGTGCCTCCCACAGGTCCTGGAATGGCACCAGTTCCTACAGGTGTCTTCACAG
     1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
D   I   P   I   S   N   I   R   R   V   I   A   Q   R   L   M   Q   S   K   Q   T   I   P   H   Y   Y   L   L   S   C   K   Y   G   E   V   L   L   V   R   K
ATATCCCAATCAGCAACATTCGTCGGGTTATTGCACAGCGATTAATGCAATCAAAGCAAACCATACCTCATTATTACCTTCTATCGTGTAAATATGGAGAAGTTTTGTTGGTACGGAAAG
     1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920
E   L   N   K   I   L   E   G   R   S   K   I   S   V   N   D   F   I   I   K   A   S   A   L   A   C   L   K   V   P   E   A   N   S   S   W   M   D   T   V
AACTTAATAAGATATTAGAAGGGAGAAGCAAAATTTCTGTCAATGACTTCATCATAAAAGCTTCAGCTTTGGCATGTTTAAAAGTTCCCGAAGCAAATTCTTCTTGGATGGACACAGTTA
     1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040
I   R   Q   N   H   V   V   D   V   S   V   A   V   S   T   P   A   G   L   I   T   P   I   V   F   N   A   H   I   K   G   V   E   T   I   A   N   D   V   V
TAAGACAAAATCATGTTGTTGATGTCAGTGTTGCGGTCAGTACTCCTGCAGGACTCATCACACCTATTGTGTTTAATGCACATATAAAAGGAGTGGAAACCATTGCTAATGATGTTGTTT
     2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160
S   L   A   T   K   A   R   E   G   K   L   Q   P   H   E   F   Q   G   G   T   F   T   I   S   N   L   G   M   F   G   I   K   N   F   S   A   I   I   N   P
CTTTAGCAACCAAAGCAAGAGAGGGTAAACTACAGCCACATGAATTCCAGGGTGGCACTTTTACGATCTCCAATTTAGGAATGTTTGGAATTAAGAATTTCTCTGCTATTATTAACCCAC
     2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280
P   Q   A   C   I   L   A   I   G   A   S   E   D   K   L   V   P   A   D   N   E   K   G   F   D   V   A   S   M   M   S   V   T   L   S   C   D   H   R   V
CTCAAGCATGTATTTTGGCAATTGGTGCTTCAGAGGATAAACTGGTCCCTGCAGATAATGAAAAAGGGTTTGATGTGGCTAGCATGATGTCTGTTACACTCAGTTGTGATCACCGGGTGG
     2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
V   D   G   A   V   G   A   Q   W   L   A   E   F   R   K   Y   L   E   K   P   I   T   M   L   L   *
TGGATGGAGCAGTTGGAGCCCAGTGGCTTGCTGAGTTTAGAAAGTACCTTGAAAAACCTATCACTATGTTGTTGTAACTAACTCAAGAATTTCTAAACTCTCCCAGGTCACACTGATTCA
     2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520
TTCTTAACAAGCCCGAATTC
     2530      2540
```

FIG. 1. Nucleotide sequence and predicted amino acid sequence of the pHUMIT cDNA. Internal *Eco*RI sites are underlined.

sion 14), GenBank (protein translation of version 50 of GenBank nucleic acid), SwissProt (version 5), and PSD/Kyoto (March 1986)] showed three homologous sequences including pRMIT, the partial sequence for the rat M_r 70,000 mitochondrial antigen and the probe used to detect the pHUMIT sequence. The sequences are 85% identical at the nucleotide level and 86% identical at the amino acid level (Fig. 2). The other computer matches are two related sequences, *E. coli* E2p and *E. coli* dihydrolipoamide succinyltransferase (21, 22). The E2p sequence shows 26% identity over a 550-amino-acid overlap, whereas the succinyltransferase shows a similar level of identity over a 293-amino-acid overlap. The comparison of pHUMIT, pRMIT, and *E. coli* E2p is shown in Fig. 2. The sequences are arranged so as to align the lipoyl-binding domains, and the binding sites from several species are compared in Table 1. Two lines of evidence support the proposal, derived from the sequence similarity, that M2 is the enzyme E2p. First, $NaDodSO_4$/PAGE analysis shows that mammalian E2p has a M_r of 70,000, the same as the observed size of the M2 autoantigen. In contrast, dihydrolipoamide succinyltransferase has a M_r of 43,600 (22). Second, in other experiments we have shown that anti-M2 antibodies from PBC patients react with the PDH enzyme complex containing E2p, purified from human mitochondria (C.D.S., J.V.d.W, M.E.G., and R.L.C., unpublished observations).

Expression of pHUMIT and Reaction with PBC Sera. Fragments of the pHUMIT sequence were ligated into expression vectors and assayed for reactivity with sera from patients with PBC. Fragment 1 (nucleotides 745–2206) was

Table 1. Lipoyl-binding regions of E2p from several species

Source	Residues	Sequence
E. coli E2p	37–52	VEGDKASMEVPSPQAG
E. coli E2p	140–155	VEGDKASMEVPAPFAG
E. coli E2p	241–256	VEGDKASMEVPAPFAG
Bovine		VETDKATVGF
Human	95–109	VETDKATVGFESLEEC
Human	200–215	IETDKATIGFEVQEEG
Rat	83–98	IETDKATIGFEVQEEG

The *E. coli*, bovine, and rat sequences are from refs. 20, 22, and 11, respectively. Amino acids are indicated by the single-letter code.

ligated into a derivative of pUR292 (17), and lysates of expressing clones were electrophoresed in $NaDodSO_4$/PAGE, transferred to nitrocellulose, and probed with a number of sera. Sera from PBC patients but not from patients with autoimmune chronic active hepatitis reacted with the expressed fusion protein (Fig. 3). Similar results were obtained after ligation of fragment 1 into a pUC19 derivative (16) followed by screening of lysates of expressing clones with patient sera (data not shown). A second fragment (nucleotides 1095–2207) was ligated into pGEX (15), a glutathione transferase fusion vector, and the expressed protein was assayed for antigenicity. This fusion protein also reacted strongly and specifically with PBC sera (data not shown), showing that autoantigenic epitopes are present in this region of sequence.

DNA and RNA Blot Analysis. Human genomic DNA was probed with a subclone of pHUMIT (nucleotides 1095–2207)

```
pHuMIT               MSPHCSTTYLRTLGRTTMFWKTTEGRDGKMAVQEFSEFGLLLQLLGSPGRRYYSLPPHQKVPLPSLSPTMQAGTIARWEKKE....  82
                           *  *   *             *  *  *                              *      *      ****
E2p       MAIEIKVPDIGADEVEITEILVKVGDKVEAEQSLITVEGDKASMEVPSPQAGIVKEIKVSVGDKTQTGALIMIFDSADGAADAAPAQAEEKKEAAPA  97

pRMIT                                                                                           GPEAFKNYTLD  11
                                                                                                  *********
PHuMIT    ..............................GDKINEGDLIAEVETDKATVGFESLEECYMAKILVAEGTRDVPIGAIICITVGKPEDIEAFKNYTLD  149
                                        ***         ** ***              * *  *   *  *            *  *
E2p       AAPAAAAAKDVNVPDIGSDEVEVTEILVKVGDKVEAEQSLITVEGDKASMEVPAPFAGTVKEIKVNVGDK.VSTG..........SLIMVFEVAGEA  183

pRMIT     SATAAT.QAAPAPAAAPAAAPAAPSASAPGSSYPVHMQIVLPALSPTMTMGTVQRWEKKVGEKLSEGDLLAEIETDKATIGFEVQEEGYLAKIL.VP  106
          *  * * ******  *  * *  *** ******* ***  ****************************************************** **
pHuMIT    SSAAPTPQAAPAPTPAATASPPTPSAQAPGSSYPPHMQVLLPALSPTMTMGTVQRWEKKVGEKLSEGDLLAEIETDKATIGFEVQEEGYLAKIL.VP  245
            ***      **  ** * *      * *        *  *          *     *** *      *   * ***         *   * * *
E2p       GAAAPAAKQEAAPA.AAPA.P......AAGVKE.....VNVPDI..GGDEVEVTEVMVKVGDKVAAEQSLITVEGDKASMEVPAPFAGVV.KELKVN  264

pRMIT     EGTRDVPLGTPLCIIVEKQEDIAAFADYRPTEVTSLKPQAPPPVPPPVAAVPPIPQPLAPTPS....AAPAGPKGRVFVSPLAKKLAAEKGIDLTQV  199
          ******************  ** *********** **** *** ********* *********    * ********** ******* *********
pHuMIT    EGTRDVPLGTPLCIIVEKEADISAFADYRPTEVTDLKPQVPPPTPPPVAAVPPTPQPLAPTPSTPCPATPAGPKGRVFVDPLAKKLAVEKGIDLTQV  342
           *  *  *    *   *                    *       *  **    *   **          *     *   **   ** * *  *  *
E2p       VGDKVKTGSLIMIFEVE.............GAAPAAAPAKQEAAAPAPAAKAEAP.AAAPAAKAEGKSEFAENDAYVHATPLIRRLAREFGVNLAKV  347

pRMIT     KGTGPEGRIIKKDIDSFV..PTKAAPAAAAAAPP.GPRVAP...TPAGVFIDI...PISNIRRVIAQRLMQSKQTIPHY.YLSVDVNMGEVLLVRKE  286
          ***** *** ********  * * *** **  ** **  **    * *** **   *********************** **      *********
pHuMIT    KGTGPDGRITKKDIDSFV..PSKVAPAPAAVVPPTGPGMAP...VPTGVFTDI...PISNIRRVIAQRLMQSKQTIPHY.YLLSCKY.GEVLLVRKE  429
          ****  ***   *         * * *  *      *** *        *  *       *       *      ***           *    **
E2p       KGTGRKGRILREDVQAYVKEAIKRAEAAPAATGGGIPGMLPWPKVDFSKFGEIEEVELGRIQKISGANLSRNWVMIPHVTHFDKTDI.TELEAFRKQ  443

pRMIT     LNKMLEGK...GKISVNDFIIKASALACLKVPEANSSWMDTVIR..QNHVVDVSVAVSTPAGLITPIVFNAHIKGLETIASDVVSLASKAREGKLQP  378
          *** ***     ********************************  ***************************** **** ****** *********
pHuMIT    LNKILEGR...SKISVNDFIIKASALACLKVPEANSSWMDTVIR..QNHVVDVSVAVSTPAGLITPIVFNAHIKGVETIANDVVSLATKAREGKLQP  521
           *     *    **    ** ** * *    *  ***      *          *** ** **  *       **             *** ***
E2p       QNEEAAKRKLDVKITPVVFIMKAVAAALEQMPRFNSSLSEDGQRLTLKKYINIGVAVDTPNGLVVPVFKDVNKKGIIELSRELMTISKKARDGKLTA  540

pRMIT     HEFQGGTFTISNLGMFGIKNFSAIINPPQACILAIGASEDKLIPADNEKGFDVASVMSVTLSCDHRVVDGAVGAQWLA  457
          ****************************************** ************ **********************
pHuMIT    HEFQGGTFTISNLGMFGIKNFSAIINPPQACILAIGASEDKLVPADNEKGFDVASMMSVTLSCDHRVVDGAVGAQWLAEFRKYLEKPITMLL  613
           * *** ****  *  *   *  * * *   **  * *     *  * * *    *    ** **** *** **         *   *  *
E2p       GEMQGGCFTISSIGGLGTTHFAPIVNAPEVAIL..GVSKSAMEPVWNGKEFVPRLMLPISLSFDHRVIDGADGARFITIINNTLSD.IRRLVM  630
```

FIG. 2. Comparison of the amino acid sequences of the *E. coli* E2p, pHUMIT, and pRMIT. Spaces indicated by dots have been inserted to align the lipoyl-binding domains, and identities are indicated by asterisks. Predicted lipoic acid-binding sites are in boldface. The pRMIT sequence (11) has been corrected by deletion of two incorrect nucleotides (1311–1312) from the published sequence.

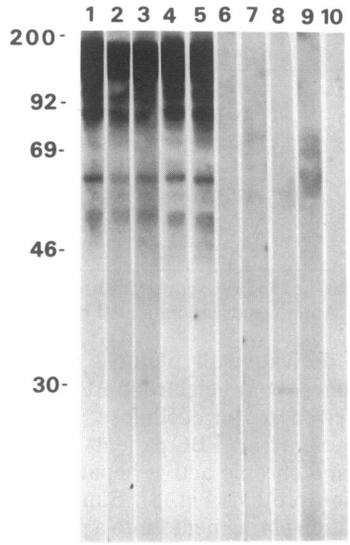

FIG. 3. PBC sera recognize determinants in the expressed pHUMIT sequence. Lysates of pHUMIT-transformed *E. coli* were fractionated by $NaDodSO_4$/PAGE and were immunoblotted. PBC sera (lanes 1–5) but not sera from patients with autoimmune chronic active hepatitis (lanes 6–10) react with a polypeptide of a M_r of ≈175,000 and with multiple breakdown products. Sizes (in $M_r \times 10^{-3}$) are indicated.

to determine the complexity of DNA sequences homologous to the pHUMIT sequence. As shown in Fig. 4*A*, restriction enzymes that do (*Hin*dIII) or do not (*Eco*RI and *Bam*HI) cut the probe sequence generate 3–7 bands in Southern blots. This is consistent with the presence of one or more genes that may contain introns. Hybridization of the probe to poly(A)$^+$ RNA from three human cell lines detected several RNA species of 1.8–3.2 kb (Fig. 4*B*), the most prominent of which

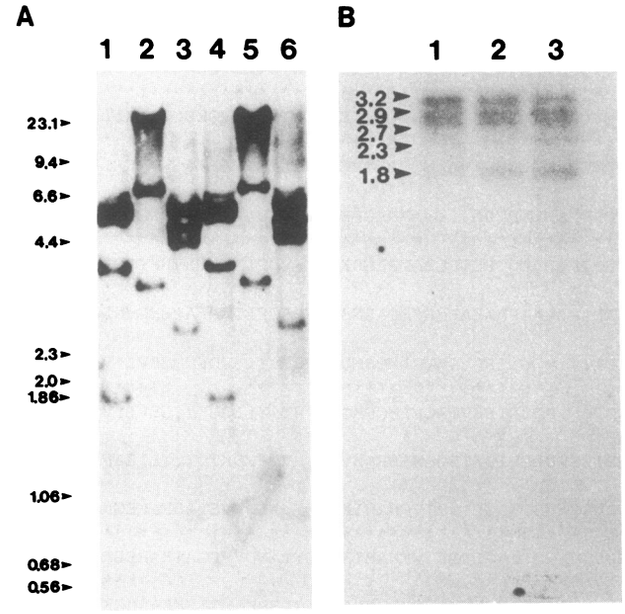

FIG. 4. (*A*) Southern analysis of human genomic DNA from the Joy cell line (lanes 1–3) or human peripheral blood lymphocytes (lanes 4–6) digested with *Eco*RI (lanes 1 and 4), *Bam*HI (lanes 2 and 5), or *Hin*dIII (lanes 3 and 6). Sizes (in kb) are indicated. (*B*) There are five mRNA species homologous to pHUMIT in human cells. Poly(A)$^+$ RNA from the human cell lines Joy (lane 1), J1 (lane 2), and K562 (lane 3) was analyzed by RNA gel blotting. The probe for both experiments corresponded to nucleotides 1095–2207 of the pHUMIT sequence. Sizes (in kb) are indicated.

were the 2.7-, 2.9-, and 3.2-kb species. These multiple RNAs could arise by transcription of multiple copies of one or more related genes, by read through of polyadenylylation signals, or by an alternate splicing mechanism, but further analysis is needed to resolve this. The degree of sequence similarity between *E. coli* E2p and the gene for dihydrolipoamide succinyltransferase suggests that the DNA and RNA gel blots could have detected the sequences of both genes (21, 22).

DISCUSSION

This study describes the isolation and characterization of a human cDNA clone encoding the M_r 70,000 M2 mitochondrial autoantigen and identifies the autoantigen as E2p. This enzyme is a component of the PDH complex, which consists of the dehydrogenase (E1), the acetyltransferase (E2), and lipoamide dehydrogenase (E3). Comparison of the human sequence with the *E. coli* sequence shows that identity is greatest in the carboxyl-terminal regions of the protein and is much less at the amino-terminal region, presumably reflecting the presence of sequence elements that direct the protein to different locations in the two organisms. Support for the identity of the M2 antigen with E2p is provided by the demonstration that anti-M2 reactivity is directed to the M_r 70,000 component of the purified PDH complex.

The PDH complex catalyzes the conversion of pyruvate to acetyl-CoA in the mitochondria, and E2p, the structural core of the complex, contains covalently bound lipid cofactors. Lipoic acid is bound to the ε-amino group of the lysine contained within the motif Val-Glu-Gly-Asp-Lys-Ala (V E G D K A), and the *E. coli* E2p has three binding sites for lipoic acid. The corresponding human enzyme has only two regions of homologous sequence and presumably has only two lipoic acid binding sites, as has been suggested for the bovine E2p (23). The first binding site in the human enzyme is unusual in that it has a cysteine instead of the glycine normally found 11 amino acid residues carboxyl terminal to the lipoic acid binding lysine (Table 1). Identity between the human and *E. coli* E2p is greatest in the carboxyl-terminal region, the binding site for the other components of the PDH complex (21).

We have previously shown that antibodies to M2 are directed in part to a 20-amino-acid sequence present in the rat sequence (24). This sequence is completely conserved in the human sequence and includes one of the predicted lipoic acid binding sites of E2p. This raises the possibility that the autoantibodies may be capable of interfering with enzymic activity, and it will be of interest to assess whether the degree of inhibition correlates with the severity of clinical disease. The establishment of the identity of the M2 autoantigen will allow the formulation of experiments to test the role of this protein in the aetiology and pathogenesis of PBC.

It is still unclear how an internal antigen may be the target of immune effector mechanisms. It has been demonstrated in other systems that an internal protein in virally infected cells may induce a cytotoxic lymphocyte response (25). Similarly, in PBC, viral infection of biliary epithelium may result in altered synthesis of E2p, subsequent degradation, and presentation to lymphocytes in association with class 1 major histocompatibility complex molecules. The protein sequence of the M2 antigen will allow the mapping of the epitopes of E2p recognized by both T and B cells of PBC patients. The expressed human protein should also prove useful in formulating rapid, sensitive, and specific diagnostic tests for the presence of autoantibodies, in the same way as has been demonstrated for the La and centromere autoantigens (26, 27).

Note Added in Proof. To check on the unusual finding of a cysteine residue located 11 residues carboxyl terminal to the lysine in the first lipoic acid-binding site, we have sequenced multiple M13 clones, and

the results confirm the presence of a cysteine at this position in the derived protein sequence. We have been unable to isolate an independent cDNA clone that includes this region of the sequence.

We thank Dr. I. K. Hariharan for the Joy cell DNA and RNA; Dr. A. Kyne for help with computer analysis; and Fiona Smith, Jeanette West, and Gaetano Naselli for technical assistance. We thank Prof. J. R. Guest for assistance in the correction of the pRMIT sequence. This work was supported by the Australian National Health and Medical Research Council, National Institutes of Health Grant DK 39588, the John D. and Catherine T. MacArthur Foundation, and the Reid Memorial Laboratory.

1. Kaplan, M. M. (1987) *Intern. Med.* **32,** 359–370.
2. Berg, P. A. & Klein, R. (1985) *Mol. Aspects Med.* **8,** 235–247.
3. Baum, H. & Berg, P. A. (1981) *Semin. Liver Dis.* **1,** 309–321.
4. Berg, P. A., Klein, R., Lindenborn-Fotinos, J. & Klöppel, W. (1982) *Lancet* **ii,** 1423–1426.
5. Lindenborn-Fotinos, J., Sayers, T. J. & Berg, P .A. (1982) *Clin. Exp. Immunol.* **54,** 648–654.
6. Schultheiss, H. P., Berg, P. & Klingenberg, M. (1984) *Clin. Exp. Immunol.* **50,** 267–274.
7. Frazer, I. H., Mackay, I. R., Jordan, T. W., Whittingham, S. & Marzuki, S. (1985) *J. Immunol.* **135,** 1739–1745.
8. Lindenborn-Fotinos, J., Baum, H. & Berg, P. A. (1985) *Hepatology* **5,** 763–769.
9. Mendel-Hartvig, I., Nelson, B. D., Lööf, L. & Tötterman, T. H. (1985) *Clin. Exp. Immunol.* **62,** 371–379.
10. Mendel-Hartvig, I., Frostell, A. & Nelson, B. D. (1986) *Clin. Exp. Immunol.* **66,** 399–405.
11. Gershwin, M. E., Mackay, I. R., Sturgess, A. & Coppel, R. (1987) *J. Immunol.* **138,** 3525–3531.
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
14. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
15. Smith, D. B. & Johnson, K. S. (1988) *Gene,* **67,** 31–40.
16. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
17. Rüther, V. & Müller-Hill, B. (1983) *EMBO J.* **2,** 1791–1794.
18. Spithill, T. W. & Samaras, N. (1987) *Mol. Biochem. Parasitol.* **24,** 23–37.
19. vón Heijne, G. (1986) *EMBO J.* **5,** 1335–1342.
20. DeMeirleir, L., Mackay, N., Wah, A. M. L. H. & Robinson, B. H. (1988) *J. Biol. Chem.* **263,** 1991–1995.
21. Stephens, P. E., Darlison, M. G., Lewis, H. M. & Guest, J. R. (1983) *Eur. J. Biochem.* **133,** 481–489.
22. Spencer, M. C., Darlison, M. G., Stephens, P. E., Duckenfield, I. K. & Guest, J. R. (1984) *Eur. J. Biochem.* **141,** 361–374.
23. Bradford, A. P., Howell, S., Aitken, A., James, L. A. & Yeaman, J. J. (1987) *Biochem. J.* **245,** 919–922.
24. Van de Water, J., Gershwin, M. E., Leung, P., Ansari, A. & Coppel, R. L. (1988) *J. Exp. Med.* **167,** 1791–1799.
25. Townsend, A. R. (1987) *Immunol. Res.* **6,** 80–100.
26. Whittingham, S., Naselli, G., McNeilage, L. J., Coppel, R. L. & Sturgess, A. D. (1987) *Lancet* **i,** 1–3.
27. Earnshaw, W. C., Machlin, P. S., Bordwell, B. J., Rothfield, N. F. & Cleveland, D. W. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4979–4983.