Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 258–262, January 1991
Biochemistry

Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1

(protein degradation)

PATRICIA M. HANDLEY*†‡, MIKE MUECKLER*, NED R. SIEGEL§, AARON CIECHANOVER¶, AND ALAN L. SCHWARTZ†

Departments of *Cell Biology, †Pharmacology, and †Pediatrics, Washington University School of Medicine, and §Monsanto Corporation, St. Louis, MO; and ¶Department of Biochemistry, Technion-Medical School, Haifa, Israel

Communicated by Emil R. Unanue, October 8, 1990

ABSTRACT **The ubiquitin-activating enzyme E1 catalyzes the first step in ubiquitin conjugation. We have cloned and sequenced the cDNA for human E1. This clone predicts a protein of 110,450 Da. Cys-194 lies within a region of identity to active-site Cys-88 of the ubiquitin carrier protein E2, suggesting a potential role for this region in enzymatic function of this protein. In addition, Cys-454 lies within a region of identity to the thiol ester consensus sequence of several proteins involved in thioester formation. Tissue distribution reveals a single 3.5-kilobase E1 message ubiquitous among tissues and cell lines.**

Ubiquitination is a covalent modification of various cellular proteins. This posttranslational modification exerts a wide range of effects within all eukaryotic cells. The most thoroughly characterized conjugation event occurs in the ubiquitin-dependent degradative pathway (1). In addition, stable ubiquitin adducts can be found intracellularly—for example, those involving histones and arthrin (1–3). Furthermore, several cell-surface molecules have been found to be modified by ubiquitin (i.e., platelet-derived growth factor and growth hormone receptors) (4, 5).

The first reaction in the ubiquitin conjugation system is the activation of ubiquitin to a high-energy intermediate. This reaction is catalyzed by E1, the ubiquitin-activating enzyme. The reaction takes place in two steps during which ubiquitin first forms an adenylate intermediate before being transferred to a thiol site (6, 7). The ubiquitin is then transferred to one of the ubiquitin carrier proteins (E2s) to generate a similar thioester linkage. At this point, ubiquitin can either be linked directly to a target protein or conjugated to proteins destined for degradation via the ubiquitin–protein ligases (E3s). Thus, E1 provides the initial activated form of ubiquitin necessary for any conjugation reaction to take place. In this capacity, E1 plays a key regulatory role in any process affected by ubiquitin modification.

The ubiquitin-activating enzyme catalyzes the activation of ubiquitin in the following steps.

$$(i)\ \mathrm{E1} + \mathrm{Ub} + \mathrm{ATP} \overset{Mg^{2+}}{\rightleftharpoons} \mathrm{E1}_{\mathrm{AMP\sim Ub}} + \mathrm{PP_i}$$

$$(ii)\ \mathrm{E1}_{\mathrm{AMP\sim Ub}} \rightleftharpoons \mathrm{E1}^{\mathrm{s\sim Ub}} + \mathrm{AMP}$$

$$(iii)\ \mathrm{E1}^{\mathrm{s\sim Ub}} + \mathrm{Ub} + \mathrm{ATP} \rightleftharpoons \mathrm{E1}^{\mathrm{s\sim Ub}}_{\mathrm{AMP\sim Ub}} + \mathrm{PP_i},$$

where Ub = ubiquitin, s = thioester, and PP_i = pyrophosphate. There are two active sites within the E1 molecule, allowing it to accommodate two ubiquitin moieties at a time, with a new ubiquitin forming an adenylate intermediate as the previous one is transferred to the thiol site (6). The molecular mass of purified E1 as determined by gel filtration is ≈210 kDa. However, on SDS/PAGE the apparent molecular mass decreases to ≈110 kDa, a finding that suggests E1 may be dimeric in structure.

Recently, two mammalian cell-cycle-arrest mutants have been characterized that appear to have a single temperature-sensitive defect in E1 (8–10). The more thoroughly characterized cell line (ts85, which is derived from the parental mouse mammary carcinoma FM3A cell) exhibits a variety of phenotypic abnormalities at the nonpermissive temperature (1, 8, 9). These include a marked inhibition in the degradation of short-lived and abnormal proteins. The ts85 cells also display a decrease in ubiquitination of histones at the nonpermissive temperature, as well as apparent defects in DNA synthesis and nucleoside transport (1). Thus, E1 appears to be directly or indirectly involved in these processes.

We report the amino acid sequence of human E1, deduced from the cDNA clone.‖ Availability of the clone will allow future studies that will provide better understanding of the structure and function of E1.

MATERIALS AND METHODS

Purification of Human E1 and Anti-E1 Antibodies. E1 was purified to homogeneity from human erythrocytes by covalent affinity chromatography on immobilized ubiquitin (7). E1 was bound covalently to the ubiquitin on the affinity support with ATP and Mg^{2+} and then washed. E1 was specifically eluted by reversing its enzymatic reaction with the addition of pyrophosphate (see Introduction). The enzyme was >95% pure as judged by SDS/PAGE with Coomassie blue stain (Fig. 1) and silver stain (data not shown). The purified enzyme was assayed for ubiquitin-dependent pyrophosphate–ATP exchange activity (7) as well as E1-dependent ubiquitin conjugation activity (11) (data not shown). To obtain anti-E1 antibody, rabbits were immunized with repeated injections of the purified enzyme in Freund's adjuvant. The antiserum was purified on immobilized E1. Immunoblot analysis of E1 was carried out by electroblotting samples onto nitrocellulose after separation on 10% SDS/PAGE. E1 was detected with the affinity-purified antibody and iodinated protein A.

Amino Acid Sequencing and Analysis. Seventy micrograms of E1 was purified using 10% SDS/PAGE, blotted onto 0.45 μm nitrocellulose (Schleicher & Schuell), and stained with Ponceau S (Sigma). The excised 110-kDa bands were then incubated in polyvinylpyrrolidone 40 and digested in an *in situ* fashion (12) using HPLC-purified, L-1-tosylamido-2-

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 at: Box 8116, Department of Pediatrics, Division of Hematology-Oncology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110.

‖The sequence reported in this paper has been deposited in the GenBank data base (accession no. M58028).

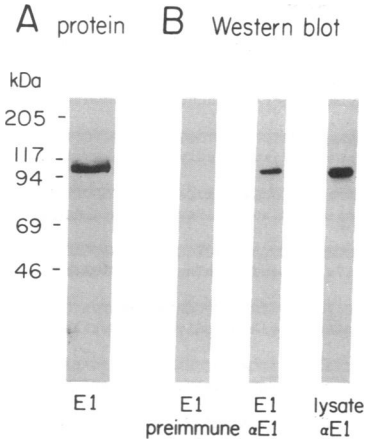

FIG. 1. Purified human E1 and rabbit anti-E1 antibody. (*A*) Silver stain of 10% SDS/PAGE analysis of human E1 (0.5 μg) purified from erythrocytes. Molecular mass markers are at left. (*B*) Immunoblot analysis of purified human E1 (0.5 μg) and erythrocyte lysate (7) (50 μg) as described. αE1, anti-E1 antibody.

phenylethyl chloromethyl ketone (TPCK) trypsin (Sigma). Digestions were done overnight at 37°C in 100 mM Tris·HCl, pH 8.2/acetonitrile (95:5, vol/vol) using an enzyme/substrate ratio of 1:20 (wt/wt). The acidified digest was chromatographed on an Aquapore Bu-300 (Brownlee Lab) reversed-phase HPLC column (2.1 × 100 mm) using 70% acetonitrile/0.1% trifluoroacetic acid. Sequence information was derived from the reversed-phase HPLC fractions by using automated Edman degradation chemistry performed by an Applied Biosystems model 470A gas-phase sequencer equipped with on-line identification of the phenylthiohydantoins (13).

For amino acid compositional analysis, 12 μg of E1 was transferred to an evacuated sealed glass tube and hydrolyzed using 6 M HCl at 110°C for 24 hr. Compositional amino acid data were collected from the hydrolysate by using postcolumn ninhydrin-based analysis and a Beckman 6300 autoanalyzer (14). The ratio of amino acids obtained was ultimately compared to the cDNA-deduced protein sequence-derived data by using the program of Cornish–Bowden (15).

Library Screening. The cDNA encoding E1 was isolated from a human placental cDNA library in λ-zap, which was provided by Jonathan Gitlin (Washington University) (16). The library was screened for the expression product E1 with the anti-E1 antibody and iodinated second antibody. Positive clones were excision-rescued by means of the standard protocol of Stratagene. The largest clone was estimated at 3.5 kilobases (kb) and was prepared for sequencing.

Sequencing. The 3.5-kb E1 clone was sequenced using sequence-specific primers and the Sanger method. The shotgun approach (17) was taken to obtain sequence data for the full coding region as follows. Ten micrograms of the E1 insert was end-repaired using Klenow and DNA polymerase and then blunt ligated end to end, forming large concatamers. These products were then sonicated; the resulting fragments were size-selected (0.5–1 kb), and the DNA was purified from the agarose with the glass milk procedure (18), end repaired, and blunt ligated into the *Sma* I site of the M13mp19 sequencing vector. After transformation of the resulting ligation mixture into *Escherichia coli* Xl1-Blue cells, phage minipreps (19) were done to purify single-stranded DNA for sequencing. The data was analyzed and compiled using the DNASTAR software package, (Dnastar, Madison, WI).

Transcription and Translation. cDNA was subcloned into pGEM7zf(+), linearized with *Hin*dIII, and transcribed by using SP6 RNA polymerase. The mRNA was translated with reticulocyte lysate (Promega), according to the manufacturer's instructions. The translated product was immunoprecipitated by using the rabbit anti-human E1 immune serum (see above). Normal rabbit serum was used in a preimmune control.

Tissue Distribution. mRNAs were prepared from human tissues and various cell lines according to standardized procedures (21). Ten micrograms of each sample was analyzed via agarose gel electrophoresis in the presence of formaldehyde and transferred to nylon membrane (22). The blot was probed with a random prime-labeled insert (23) from the 3.5-kb E1 clone.

RESULTS

Human E1 and Anti-E1 Antibody. E1 purified from human erythrocytes via immobilized ubiquitin affinity chromatography yielded a single polypeptide of ≈110 kDa after SDS/PAGE (Fig. 1*A*). Polyclonal rabbit antibody to this protein recognizes a single protein species of ≈110 kDa in erythrocyte lysates (Fig. 1*B*).

cDNA for Human E1. The 3.5-kb cDNA clone contains a 287-base 3′ untranslated region, complete with a polyadenylylation signal (Fig. 2). A 251-base 5′ untranslated region is also found in this clone. This region contains one stop codon and is G+C rich. Primer-extension analysis demonstrated an additional 8 nucleotides at the extreme 5′ end of the E1 message (data not shown). Thus, this 3.5-kb clone includes an essentially complete nucleotide sequence of the human E1 mRNA.

Identification of the Clone. The 3.5-kb cDNA clone identified with the anti-E1 antibody contains a 2991-base-pair (bp) open reading frame encoding a protein of 996-amino acid residues (Fig. 2). This predicts a polypeptide with a molecular size of 110,450 Da.

The following experiments confirm the identity of the clone in addition to identification by the antibody. (*i*) Four peptides from a tryptic digest of the purified protein were sequenced and found identical to corresponding sequences in the clone (underlined in Fig. 2). (*ii*) Amino acid composition of the purified protein (Table 1) was found very similar to the theoretical amino acid composition as deduced from the cDNA. The $S\ \delta n/N$ statistic for compositional comparison equals 0.346; values <0.42 indicate >95% certainty of identity (15). (*iii*) *In vitro* translation of the mRNA transcribed from the clone yielded a major product of ≈110 kDa (Fig. 3) (see *Discussion*). This major product was also immunoprecipitable by the anti-human E1 antibody used to identify the clone (Fig. 3).

Tissue Distribution. Northern (RNA) blot analysis of several human tissues and cell lines reveals a single message, ≈3.5 kb in length. This band appeared in all samples examined (Fig. 4). In addition, mRNA from Chinese hamster ovary cells reveals a product of the same size.

DISCUSSION

We have cloned and sequenced the cDNA-encoding human E1. The clone was found to be unique in that there was no significant homology to any known sequence in either Genbank/European Molecular Biology Laboratory nucleotide release 64.0 or the National Biomedical Research Foundation protein data banks release 25.0.

The second active site within the E1 molecule involves transfer of the ubiquitin moiety from the adenylate site to a cysteine residue as a high-energy thioester (1, 6). Our clone contains 19 cysteine residues. Recently, Sung *et al.* reported that Cys-88, within one of the yeast ubiquitin carrier proteins RAD6, is the active site of this enzyme (24). Ubiquitin carrier proteins also bind ubiquitin as a high-energy thioester intermediate (1). Therefore, homology might be found between the region containing the single cysteine that binds ubiquitin

```
   1  GAGAAGGCGGCAGCGGCGATTCTAGACGGCCCAGGCGGCGGGGAGGAGGAGAAGGAGGAGGGTGGCGGCCGGGCTTGGCTTCGGCTCCTTGAGGAGTTGGCGGCGGC  107

 108  GCGACCCGGGGAACCGGCATTGATGTCCAGCTCGCCGCTGTCCAAGAAACAGTCGCGTGTCCGGGCCTGATCCAAAGCCGGGTTCTAACTGCTCCCCTGCCCAGTCCG  215

 216  TGTTGTCCGAAGTGCCCTCGGTGCCAACCAACGGAAATGGCCAAGAACGGCAGTGAAGCAGACATAGACGAGGGCCTTTACTCCCGGCAGCTGTATGTGTTGGGCCAT  323
   1                                  MetAlaLysAsnGlySerGluAlaAspIleAspGluGlyLeuTyrSerArgGlnLeuTyrValLeuGlyHis   24

 324  GAGGCAATGAAGCGGCTCCAGACATCCAGTGTCCTGGTATCAGGCCTGCGGGGGCCTGGGCGTGGAGATCGCTAAGAACATCATCCTTGGTGGGGTCAAGGCTGTTACC  431
  25  GluAlaMetLysArgLeuGlnThrSerSerValLeuValSerGlyLeuArgGlyLeuGlyValGluIleAlaLysAsnIleIleLeuGlyGlyValLysAlaValThr   60

 432  CTACATGACCAGGGCACTGCCCAGTGGGCTGATCTTTCCTCCCAGTTCTACCTGCGGGAGGAGGACATCGGTAAAAACCGGGCCGAGGTATCACAGCCCCGCCTCGCT  539
  61  LeuHisAspGlnGlyThrAlaGlnTrpAlaAspLeuSerSerGlnPheTyrLeuArgGluGluAspIleGlyLysAsnArgAlaGluValSerGlnProArgLeuAla   96

 540  GAGCTCAACAGCTATGTGCCTGTCACTGCCTACACTGGACCCCTCGTTGAGGACTTCCTTAGTGGTTTCCAGGTGGTGGTGCTCACCAACACCCCCCTGGAGGACCAG  647
  97  GluLeuAsnSerTyrValProValThrAlaTyrThrGlyProLeuValGluAspPheLeuSerGlyPheGlnValValValLeuThrAsnThrProLeuGluAspGln  132

 648  CTGCGAGTGGGTGAGTTCTGTCACAACCGTGGCATCAAGCTGGTGGTGGCAGACACGCGGGGCCTGTTTGGGCAGCTCTTCTGTGACTTTGGAGAGGAAATGATCCTC  755
 133  LeuArgValGlyGluPheCysHisAsnArgGlyIleLysLeuValValAlaAspThrArgGlyLeuPheGlyGlnLeuPheCysAspPheGlyGluGluMetIleLeu  168

 756  ACAGATTCCAATGGGGAGCAGCCACTCAGTGCTATGGTTTCTATGGTTACCAAGGACAACCCCGGTGTGGTTACCTGCCTGGATGAGGCCCGACACGGTTTGAGAGCG  863
 169  ThrAspSerAsnGlyGluGlnProLeuSerAlaMetValSerMetValThrLysAspAsnProGlyValValThrCysLeuAspGluAlaArgHisGlyLeuArgAla  204

 864  GGGACTTTGTCTCCTTTTCAGAAGTACAGGGCATTGGTTGAACTCAACGGAAATCAGCCCATGGAGATCAAAGTCCTGGGTCCTTATACCTTTAGCATCTGTGACACC  971
 205  GlyThrLeuSerProPheGlnLysTyrArgAlaLeuValGluLeuAsnGlyAsnGlnProMetGluIleLysValLeuGlyProTyrThrPheSerIleCysAspThr  240

 972  TCCAACTTCTCCGACTACATCCGTGGAGGCATCGTCAGTCAGGTCAAAGTACCTAAGAAGATTAGCTTTAAATCCTTGGTGGCCTCACTGGCAGAACCTGACTTTGTG  1079
 241  SerAsnPheSerAspTyrIleArgGlyGlyIleValSerGlnValLysValProLysLysIleSerPheLysSerLeuValAlaSerLeuAlaGluProAspPheVal  276

1080  GTGACGGACTTCGCCAAGTTTTCTCGCCCTGCCCAGCTGCACATTGGCTTCCAGGCCCTGCACCAGTTCTGTGCTCAGCATGGCCGGCCACCTCGGCCCCGCAATGAG  1187
 277  ValThrAspPheAlaLysPheSerArgProAlaGlnLeuHisIleGlyPheGlnAlaLeuHisGlnPheCysAlaGlnHisGlyArgProProArgProArgAsnGlu  312

1188  GAGGATGCAGCAGAACTGGTAGCCTTAGCACAGGCTGTGAATGCTCGAGCCCTGCCAGCAGTGCAGCAAAATAACCTGGACGAGGACCTCATCCGGAAGCTGGCATAT  1295
 313  GluAspAlaAlaGluLeuValAlaLeuAlaGlnAlaValAsnAlaArgAlaLeuProAlaValGlnGlnAsnAsnLeuAspGluAspLeuIleArgLysLeuAlaTyr  348

1296  GTGGCTGCTGGGGATCTGGCACCCATAAACGCCTTCATTGGGGGCCTGGCTGCCCAGGAAGTCATGAAGGCCTGCTCCGGGAAGTTCATGCCCATCATGCAGTGGCTA  1403
 349  ValAlaAlaGlyAspLeuAlaProIleAsnAlaPheIleGlyGlyLeuAlaAlaGlnGluValMetLysAlaCysSerGlyLysPheMetProIleMetGlnTrpLeu  384

1404  TACTTTGATGCCCTTGAGTGTCTCCCTGAGGACAAAGAGGTCCTCACAGAGGACAAGTGCCTCCAGCGCCAGAACCGTTATGACGGGCAAGTGGCTGTGTTTGGCTCA  1511
 385  TyrPheAspAlaLeuGluCysLeuProGluAspLysGluValLeuThrGluAspLysCysLeuGlnArgGlnAsnArgTyrAspGlyGlnValAlaValPheGlySer  420

1512  GACCTGCAAGAGAAGCTGGGCAAGCAGAAGTATTTCCTGGTGGGTGCGGGGGCCATTGGCTGTGAGCTGCTCAAGAACTTTGCCATGATTGGGCTGGGCTGCGGGGAG  1619
 421  AspLeuGlnGluLysLeuGlyLysGlnLysTyrPheLeuValGlyAlaGlyAlaIleGlyCysGluLeuLeuLysAsnPheAlaMetIleGlyLeuGlyCysGlyGlu  456

1620  GGTGGAGAAATCATCGTTACAGACATGGACACCATTGAGAAGTCAAATCTGAATCGACAGTTTCTTTTCCGGCCCTGGGATGTCACGAAGTTAAAGTCTGACACGGCT  1727
 457  GlyGlyGluIleIleValThrAspMetAspThrIleGluLysSerAsnLeuAsnArgGlnPheLeuPheArgProTrpAspValThrLysLeuLysSerAspThrAla  492

1728  GCTGCAGCTGTGCGCCAAATGAATCCACATATCCGGGTGACAAGCCACCAGAACCGTGTGGGTCCTGACACGGAGCGCATCTATGATGACGATTTTTTCCAAAACCTA  1835
 493  AlaAlaAlaValArgGlnMetAsnProHisIleArgValThrSerHisGlnAsnArgValGlyProAspThrGluArgIleTyrAspAspAspPhePheGlnAsnLeu  528

1836  GATGGCGTGGCCAATGCCCTGGACAACGTGGATGCCCGCATGTACATGGACCGCCGCTGTGTCTACTACCGGAAGCCACTGCTGGAGTCAGGCACACTGGGCACCAAA  1943
 529  AspGlyValAlaAsnAlaLeuAspAsnValAspAlaArgMetTyrMetAspArgArgCysValTyrTyrArgLysProLeuLeuGluSerGlyThrLeuGlyThrLys  564

1944  GGCAATGTGCAGGTGGTGATCCCCTTCCTGACAGAGTCGTACAGTTCCAGCCAGGACCCACCTGAGAAGTCCATCCCCATCTGTACCCTGAAGAACTTCCCTAATGCC  2051
 565  GlyAsnValGlnValValIleProPheLeuThrGluSerTyrSerSerSerGlnAspProProGluLysSerIleProIleCysThrLeuLysAsnPheProAsnAla  600

2052  ATCGAGCACACCCTGCAGTGGGCTCGGGATGAGTTTGAAGGCCTCTTCAAGCAGCCAGCAGAAAATGTCAACCAGTACCTCACAGACCCCAAGTTTGTGGAGCGAACA  2159
 601  IleGluHisThrLeuGlnTrpAlaArgAspGluPheGluGlyLeuPheLysGlnProAlaGluAsnValAsnGlnTyrLeuThrAspProLysPheValGluArgThr  636

2160  CTGCGGCTGGCAGGCACTCAGCCCTTGGAGGTGCTGGAGGCTGTGCAGCCGAGCCTGGTGCTGCAGCGACCACAGACCTGGGCTGACTGCGTGACCTGGGCCTGCCAC  2267
 637  LeuArgLeuAlaGlyThrGlnProLeuGluValLeuGluAlaValGlnProSerLeuValLeuGlnArgProGlnThrTrpAlaAspCysValThrTrpAlaCysHis  672

2268  CACTGGCACACCCAGTACTCGAACAACATCCGGCAGCTGCTGCACAACTTCCCTCCTGACCAGCTCACAAGCTCAGGAGCGCCGTTCTGGTCTGGGCCCAAACGCTGT  2375
 673  HisTrpHisThrGlnTyrSerAsnAsnIleArgGlnLeuLeuHisAsnPheProProAspGlnLeuThrSerSerGlyAlaProPheTrpSerGlyProLysArgCys  708

2376  CCACACCCGCTCACCTTTGATGTCAACAATCCCCTGCATCTGGACTATGTGATGGCTGCTGCCAACCTGTTTGCCCAGACCTACGGGCTGACAGGCTCTCAGGACCGA  2483
 709  ProHisProLeuThrPheAspValAsnAsnProLeuHisLeuAspTyrValMetAlaAlaAlaAsnLeuPheAlaGlnThrTyrGlyLeuThrGlySerGlnAspArg  744

2484  GCTGCTGTGGCCACATTCCTGCAGTCTGTGCAGGTCCCCGAATTCACCCCCAAGTCTGGCGTCAAGATCCATGTTTCTGACCAGGAGCTGCAGAGCGCCAATGCCTCT  2591
 745  AlaAlaValAlaThrPheLeuGlnSerValGlnValProGluPheThrProLysSerGlyValLysIleHisValSerAspGlnGluLeuGlnSerAlaAsnAlaSer  780

2592  GTTGATGACAGTCGTCTAGAGGAGCTCAAAGCCACTCTGCCCAGCCCAGACAAGCTCCCTGGATTCAAGATGTACCCCATTGACTTTGAGAAGGATGATGACAGCAAC  2699
 781  ValAspAspSerArgLeuGluGluLeuLysAlaThrLeuProSerProAspLysLeuProGlyPheLysMetTyrProIleAspPheGluLysAspAspAspSerAsn  816

2700  TTTCATATGGATTTCATCGTGGCTGCATCCAACCTCCGGGCAGAAAACTATGACATTCCTTCTGCAGACCGGCACAAGAGCAAGCTGATTGCAGGGAAGATCATCCCA  2807
 817  PheHisMetAspPheIleValAlaAlaSerAsnLeuArgAlaGluAsnTyrAspIleProSerAlaAspArgHisLysSerLysLeuIleAlaGlyLysIleIlePro  852

2808  GCCATTGCCACGACCACAGCAGCCGTGGTTGGCCTTGTGTGTCTGGAACTGTACAAGGTTGTGCAGGGGCACCGACAGCTTGACTCCTACAAGAATGGTTTCCTCAAC  2915
 853  AlaIleAlaThrThrThrAlaAlaValValGlyLeuValCysLeuGluLeuTyrLysValValGlnGlyHisArgGlnLeuAspSerTyrLysAsnGlyPheLeuAsn  888

2916  TTGGCCCTGCCTTTCTTTGGTTTCTCTGAACCCCTTGCCGCACCACGTCACCAGTACTATAACCAAGAGTGGACATTGTGGGATCGCTTTGAGGTACAAGGGCTGCAG  3023
 889  LeuAlaLeuProPhePheGlyPheSerGluProLeuAlaAlaProArgHisGlnTyrTyrAsnGlnGluTrpThrLeuTrpAspArgPheGluValGlnGlyLeuGln  924

3024  CCTAATGGTGAGGAGATGACCCTCAAACAGTTCCTCGACTATTTTAAGACAGAGCACAAATTAGAGATCACCATGCCTGTCCCAGGGCGTGTCCATGCTCTATTCCTT  3131
 925  ProAsnGlyGluGluMetThrLeuLysGlnPheLeuAspTyrPheLysThrGluHisLysLeuGluIleThrMetProValProGlyArgValHisAlaLeuPheLeu  960

3132  CTTCATGCCAGCTGCCAAGCTCAAGGAACGGTTGGATCAGCCGATGACAGAGATTGTGAGCCGTGTGTCGAAGCGAAAGCTGGGCCGCCACGTGCGGGCGCTGGTGCT  3239
 961  LeuHisAlaSerCysGlnAlaGlnGlyThrValGlySerAlaAspAspArgAspCysGluProCysValGluAlaLysAlaGlyProProArgAlaGlyAlaGlyAla  996

3240  TGAGCTGTGCTGTAACGACGAGAGCGGCGAGGATGTCGAGGTTCCCTATGTCCGATACACCATCCGCTGACCCCGTCTGCTCCTCTAGGCTGGCCCCTTGTCCACCCC  3347
 997  End

3348  TCTCCACACCCCTTCCAGCCCAGGGTTCCCATTTGGCTTCTGGCAGTGGCCCAACTAGCCAAGTCTGGTGTTCCCTCATCATCCCCCTACCTGAACCCCTCTTGCCAC  3455

3456  TGCCTTCTACCTTGTTTGAAACCTGAATCCTAATAAAGAATTAATAACTCCCAAAAAAAAAAAAAAAAAAAA  3527
```

FIG. 2. Nucleotide and deduced amino acid sequence of human E1 cDNA. The sequence is presented with numbering for both nucleotides and amino acids. Coding region begins at nucleotide 252 and terminates at 3242. Four amino acid sequences obtained from the purified protein are underlined. Positions of the two potential consensus sequences for thiol reactivity containing Cys-194 and Cys-454 are boxed.

Table 1. Comparison of the E1 cDNA derived amino acid sequence with amino acid composition data from the purified protein

Amino acid	Amino acids, no. Experimental	Theoretical	Experimental/ theoretical, %
Aspartic acid and asparagine	102.18	106	96.4
Threonine	48.26	48	100.5
Serine	60.38	54	111.8
Glutamic acid and glutamine	99.52	117	85.1
Proline	54.81	57	96.2
Glycine	0.00	0	N/A
Alanine	76.90	88	87.4
Cysteine	0.00	0	N/A
Valine	73.71	73	101.0
Methionine	22.16	19	116.6
Isoleucine	38.55	38	101.4
Leucine	109.28	103	106.1
Tyrosine	29.43	28	101.5
Phenylalanine	51.13	49	104.4
Histidine	22.76	24	94.8
Lysine	53.21	48	110.9
Arginine	53.30	47	113.4
Total	895.89	899	

Sixty picomoles of hydrolyzed E1 were analyzed. Tryptophan residues were destroyed during analysis. Neither glycine nor cysteine were included in the calculation due to a high experimental value for the former caused by contamination and an artificially low value for the latter due to poor yield from the carboxymethylation. The theoretical molecular weight used in the calculation was, therefore, 102,812, and the yield/residue was ≈55 pmol. The $S\ \delta n/N$ value (0.0346) was used to compare theoretical and experimental compositions (15). NA, not analyzed.

in E2 and one of the cysteine-containing regions in the E1 clone. Indeed, Cys-192 of human E1 is followed by leucine and then aspartic acid, which is identical to the E2 thiol active site. In addition, Cys-454 is located within a region that is strikingly homologous to the thioester consensus region of several proteins (25). Complement proteins C3 and C4, α_2-macroglobulin, and pregnancy zone protein all contain a Pro-Xaa-Gly-Cys-Gly-Glu-Glu-Xaa-Met sequence. The E1 sequence contains the core Gly-Cys-Gly-Glu sequence at

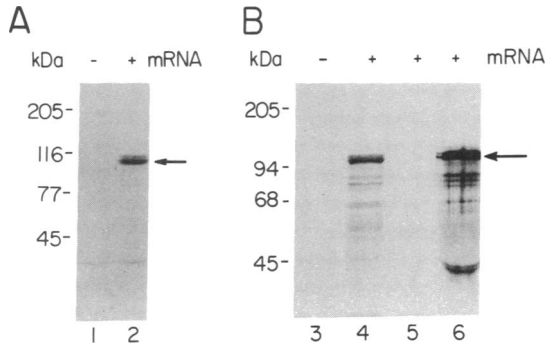

FIG. 3. *In vitro* translation of mRNA transcribed from E1 cDNA. (*A*) Reticulocyte lysate was incubated as described without (lane 1) or with (lane 2) E1 mRNA. Aliquots of the total translation mixture were subjected to SDS/PAGE and autoradiographed. Arrow, specific translation products of human E1 mRNA at ≈110 kDa. Molecular mass markers are indicated at left. (*B*) E1 was immunoprecipitated from a similar translation mixture by using the rabbit anti-human E1 antibody. Lanes: 3, total translation mixture minus E1 mRNA; 4, total translation mixture plus E1 mRNA; 5, immunoprecipitate with preimmune serum; 6, immunoprecipitate with anti-E1 serum. Arrow, major translation product of E1 mRNA and corresponding immunoprecipitated species in lanes 4 and 6.

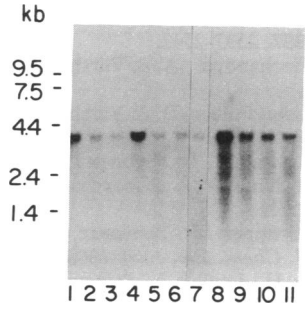

FIG. 4. Tissue distribution of E1 mRNA. Approximately 10 μg of total RNA was run on agarose gel, transferred, and probed as described; tissues are from normal human sources. All cell lines were grown with serum and harvested at confluency for RNA isolation. Markers appear at left. Lanes: 1, placenta; 2, liver; 3, skeletal muscle; 4, lung; 5, distal colon; 6, small bowel; 7, stomach; 8, HeLa human cervical carcinoma cells; 9, BeWo human choriocarcinoma cells; 10, HepG2 human hepatoma cells; 11, CHO Chinese hamster ovary cells.

residues 453–456. It is of interest that residue 457 is glycine in E1. In contrast, the thioester consensus sequence contains a second glutamic acid in this position that is necessary for formation of the *internal* thioester bond characteristic of these four proteins. A glutamic residue in position 457 of E1 would not allow formation of an *external* thioester bond between E1 and ubiquitin. These two cysteine residues are the most likely candidates for the thioester active site of E1.

It should be noted that various preparations (8, 11) of E1 contain a minor band a few kDa greater than the major (≈110 kDa) band (Fig. 1; ref. 8). It has been suggested that this minor form is a ubiquitinated form of E1 (8). The translated E1 message yields similar products, in the appropriate size range between 109 and 114 kDa, both of which are recognized by the antibody (Fig. 3). Premature termination of translation, an additional in-frame initiation site, or ubiquitination of the translated product are possible explanations for this finding. The single 3.5-kb band seen on Northern analysis suggests a single E1 message; therefore, the precise nature of the minor form of E1 remains to be determined.

Availability of a full-length E1 clone will allow genetic analysis of the role of E1 in cell-cycle progression as well as other pleiotropic effects seen in the E1 mutant cell ts85.

We dedicate this paper to the memory of Dr. Howard A. Schneiderman (deceased December 5, 1990). The authors acknowledge the generous contributions of Marilyn Strube, Julie Trausch, Burton Wice, and Dr. Jonathan Gitlin. This study was supported by National Institutes of Health (GM38284, DK38495, and GM07805), United States–Israel Binational Science Foundation, Israel Cancer Research Fund, and Monsanto Corporation.

Note Added in Proof. Following submission of the present manuscript, Hatfield *et al.* (26) reported the cloning of wheat ubiquitin-activating enzyme.

1. Reichsteiner, M., ed. (1988) *Ubiquitin* (Plenum, New York).
2. Hunt, L. T. & Dayhoff, M. O. (1977) *Biochem. Biophys. Res. Commun.* **74**, 650–655.
3. Ball, E., Karlik, C. C., Beall, C. J., Saville, D. L., Sparrow, J. C., Bullard, B. & Fyrberg, E. A. (1987) *Cell* **51**, 221–228.
4. Yarden, Y., Escobedo, J. A., Kuang, W.-J., Yang-Feng, T. L., Daniel, T. O., Tremble, P. M., Chen, E. Y., Ando, M. E., Harkins, R. N., Francke, U., Fried, V. A., Ulrich, A. & Williams, L. T. (1986) *Nature (London)* **323**, 226–232.
5. Spencer, S. A., Hammonds, R. G., Henzel, W. J., Rodriguez, H., Waters, M. J. & Wood, W. I. (1988) *J. Biol. Chem.* **263**, 7862–7867.
6. Haas, A. L. & Rose, I. A. (1982) *J. Biol. Chem.* **257**, 10329–10337.

7. Ciechanover, A., Elias, S., Heller, H. & Hershko, A. (1982) *J. Biol. Chem.* **257,** 2537–2542.
8. Finely, D., Ciechanover, A. & Varshavsky, A. (1984) *Cell* **37,** 43–56.
9. Ciechanover, A., Finely, D. & Varshavsky, A. (1984) *Cell* **37,** 57–66.
10. Kulka, R. G., Raboy, B., Schuster, R., Parag, H. A., Diamond, G., Ciechanover, A. & Marcus, M. (1988) *J. Biol. Chem.* **263,** 15726–15731.
11. Mayer, A., Gropper, R., Schwartz, A. L., Ciechanover, A. (1989) *J. Biol. Chem.* **264,** 2060–2068.
12. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. H. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6970–6974.
13. Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J. & Hood, L. E. (1983) *Methods Enzymol.* **91,** 399–413.
14. Moore, S. & Stein, W. H. (1963) *Methods Enzymol.* **6,** 819–831.
15. Cornish-Bowden, A. (1980) *Anal. Biochem.* **105,** 233–238.
16. Flemming, R. E. & Gitlin, J. D. (1990) *J. Biol. Chem.* **265,** 7701–7707.
17. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 13.14–13.33.
18. Davis, L. G., Dibner, M. D. & Battey, J. F. (1986) *Basic Methodsin Molecular Biology* (Elsevier, New York), pp. 123–125.
19. Miller, H. (1987) *Methods Enzymol.* **152,** 145–172.
20. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
21. MacDonald, R. J., Swift, G. H., Przybyla, A. E. & Chirguin, J. M. (1987) *Methods Enzymol.* **152,** 219–227.
22. Fourney, R. M., Miyakoshi, G., Day, R. S., III, & Paterson, M. C. (1988) *BRL Focus* **10.1,** 5–6.
23. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
24. Sung, P., Prakash, S. & Prakash, L. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2695–2699.
25. Hostetter, M. K. & Gordon, D. L. (1987) *Rev. Infect. Dis.* **9,** 97–109.
26. Hatfield, P. M., Callis, J. & Vierstra, R. D. (1990) *J. Biol. Chem.* **265,** 15813–15817.