JOURNAL OF VIROLOGY, Apr. 1996, p. 2658–2663 Vol. 70, No. 4
0022-538X/96/$04.00+0

Genetic Organization, Size, and Complete Sequence of Early Region 3 Genes of Human Adenovirus Type 41

HUNG-YUEH YEH,[1] NORMAN PIENIAZEK,[2] DANUTA PIENIAZEK,[3] AND RONALD B. LUFTIG[1]*

Department of Microbiology, Immunology and Parasitology and Stanley S. Scott Cancer Center, Louisiana State University Medical Center, New Orleans, Louisiana 70112,[1] and Parasitic Diseases Branch[2] and AIDS Program,[3] Centers for Disease Control and Prevention, Atlanta, Georgia 30333

Received 9 October 1995/Accepted 18 December 1995

The complete nucleotide and predicted amino acid sequences for open reading frames (ORFs) of the human adenovirus type 41 (Ad41) early region 3 (E3) gene have been determined. The sequence of the Ad41 E3 gene (map units 74 to 83.9) consists of 3,373 nucleotides and has one TATA box and two polyadenylation signals (AATAAA). Analysis of the nucleotide sequence reveals that the E3 gene can encode six ORFs, designated RL1 to RL6. These are all expressed at the mRNA level, as determined by reverse transcription-PCR analysis of Ad41-infected cell RNA. When compared with known E3 sequences of most other human adenoviruses deposited in GenBank, the sequences of RL1 to RL3 were found to be unique to subgroup F adenoviruses (Ad40 and Ad41). They encode putative proteins of 173 amino acids (19.4 kDa) and 276 amino acids (31.6 kDa) in one reading frame as well as a 59-amino-acid (6.7 kDa) protein in an overlapping reading frame. RL4 encodes a 90-amino-acid protein (10.1 kDa) with 40% homology to the Ad2 E3 10.4-kDa protein, which induces degradation of the epidermal growth factor receptor and functions together with the Ad2 E3 14.5-kDa protein to protect mouse cell lines against lysis. RL5 encodes a protein of 107 amino acid residues (12.3 kDa) and is analogous to the Ad2 E3 14.5-kDa protein. RL6 codes for a protein of 122 amino acids (14.7 kDa) that is analogous to the Ad2 14.7-kDa protein, which functions to protect Ad-infected cells from tumor necrosis factor-induced cytolysis. This finding of three unique (RL1 to RL3) E3 gene ORFs may explain why subgroup F adenoviruses differ substantially from other human adenoviruses in their host range; i.e., they replicate predominantly in the host's gastrointestinal rather than respiratory tract. A recent phylogenetic study that compared subgroup F Ad40 DNA sequences with representatives of subgroups B (Ad3), C (Ad2), and E (Ad4) reached a similar conclusion about the uniqueness of RL1 and RL2.

Human adenoviruses (Ad) are nonenveloped, icosahedral, linear double-stranded DNA viruses (24). So far, 49 serotypes, classified into six subgenera (A to F), have been identified on the basis of their biophysical, biochemical, and biological properties (11, 18, 49, 51). Subgenus F (also called enteric or fastidious Ad) consists of serotypes 40 and 41 (Ad40 and Ad41) and causes acute infantile gastroenteritis, accounting for up to 17% of all infant diarrheas (1, 4, 5, 33, 45, 48, 50). Like other diarrhea-associated viruses, both Ad40 and Ad41 do not replicate efficiently in many human cell culture systems (see reference 46 for a review), which commonly permit propagation of other Ads. Another distinctive characteristic of these viruses is that Ad40 and Ad41 are the only human Ads which contain two (long and short) fibers instead of one (long) fiber (27–29, 38, 39). In Ad40 and Ad41 infection, both fibers are expressed at late times (48 h) in infected cells (27, 57).

At the genome level, all human adenoviruses are subdivided into two regions: an early one, which is expressed before the beginning of viral DNA replication, and a late one, which is transcribed after the initiation of DNA replication (24). The early region contains four transcription units, E1, E2, E3, and E4 (24). E3 is nonessential for Ad replication in tissue culture but is conserved to some degree among all Ad serotypes, suggesting that E3 is needed for other functions, such as host-virus interactions and viral pathogenesis (55). Comparative studies of Ad2 and Ad5 DNA sequences have demonstrated that their E3 genes contain nine open reading frames (ORFs), with seven proteins having already been identified in infected cells (55). The following host-virus defense functions can be ascribed to these proteins: (i) inhibition of cytolysis by tumor necrosis factor (E3 14.7-kDa protein or a combination of 14.5- and 10.4-kDa proteins) (13–15, 31, 44, 47), (ii) interference with cell surface expression of major histocompatibility complex class I protein (E3 gp19) (6, 25, 36), (iii) function as a signal for membrane insertion (E3 6.7-kDa protein) (52–54), and (iv) induction of endosome-mediated internalization and degradation of the epidermal growth factor receptor (E3 10.4- and 13.7-kDa proteins) (7, 20, 21, 31, 41, 44). These results support the model that E3 gene products play a pivotal role in permitting the virus to overcome host defense mechanisms and persist inside infected hosts (30).

Previously, we have determined that Ad41 is unique in that downstream of its E3 genes, there are two different genes encoding the Ad41 fiber (57); one is short and the other is long, with both fiber gene products incorporated into virions (27, 57). In this paper, we report on the identification of six ORFs for the Ad41 E3 gene which can potentially encode proteins with expected molecular masses of 19.4, 31.6, 6.7, 10.1, 12.3, and 14.0 kDa (Fig. 1).

Overall Ad41 E3 sequence analysis. The Ad41 TAK strain used initially in our studies was a gift from Jan C. deJong (Laboratory of Virology, Bilthoven, The Netherlands) (11) and was adapted by blind passage seven times in HEp-2 cells in order to provide high virus yields (37, 56, 57). Viral DNA was purified from Ad41-infected HEp-2 cell monolayers by the Hirt method (19) and digested with *Eco*RI. The *Eco*RI B

* Corresponding author. Mailing address: Department of Microbiology, Immunology and Parasitology and Stanley S. Scott Cancer Center, Louisiana State University Medical Center, 1901 Perdido St., New Orleans, LA 70112-1393. Phone: (504) 568-4063. Fax: (504) 568-2918. Electronic mail address: hyeh@nomvs.lsumc.edu.

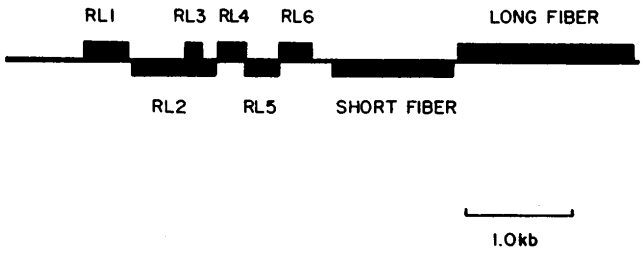

FIG. 1. Locations of ORFs in the E3 and fiber genes of Ad41 (TAK strain). RL1 to RL6 designate the proteins (5′-to-3′ direction) encoded by the E3 region genes.

```
(A) RL1
   1  MKICVLFCVL SLTSSLRTSP TTVGSLRQLQ DSTKGTHQTL YFSESTTSIA LNCSCRNQLV
  61  QWRANRQFCK LFWDALIVQG NNSLCNNCTA TTLTLTPPFV PGPYLCIGTG RGPSCFNRWT
 121  LQKENLTTTT LLPLTTYTFS QKKIYFLPII ALLAFVCVIT ANYILIFNLD NFY

(B) RL2
   1  MLLFLLCLLF CSAYAAVPEK TLNNLVRVYA LVGTNLSLDS MKTPQIDELT SLSWIKQEDN
  61  PNKNLQSFFF IGQKLCEVTK DKITVFNYYP LEFSCANVTL YLYNLKTDDS GLYNGKAHTK
 121  ELEHNTYVRL YVIDIPPPKC DITSRYLGIQ ATGEDYCLIE INCTNSKYPA VVKFNGRQSN
 181  FYHYVSENGN KELPNFYETH ITVNGTHKSF HFNYPFNDLC QTTSALQYND NVQVVLILLI
 241  VVGLIIISAS LILLYCHRKK IKAEVQHQPV HICLEK

(C) RL3
   1  MAGKATSTIM LAKTETKNFQ IFMKHTSLLM VPTKAFTLIT LLTTFVKQPA LYNIMTMSR
```

FIG. 2. Predicted amino acid sequences for RL1 (A), RL2 (B), and RL3 (C), indicated in the one-letter amino acid code.

fragment (map units 74 to 92) was separated from other DNA fragments by agarose electrophoresis, purified by using Geneclean (Bio 101, La Jolla, Calif.), and then subcloned into the *Eco*RI site of pBluescript II SK(+) (Stratagene, La Jolla, Calif.) by standard recombinant DNA techniques (40). Preliminary sequencing was performed by the method of Deininger (10). The B fragment was then briefly sheared; the ends were filled by incubation with T4 DNA polymerase, and then the fragment was cloned into the *Sma*I site of an M13 mp18 phage vector (10). M13 clones were sequenced by using a Sequenase kit (United States Biochemical, Cleveland, Ohio). After both the 5′ and 3′ ends of the Ad41 E3 gene were located by homology to the Ad5 E3 gene (8), both strands of the Ad41 E3 gene were sequenced by using custom oligonucleotide primers, which were chemically synthesized by standard procedures at the Louisiana State University Medical Center Core Laboratory Facilities. DNA sequences were analyzed by using an IBI computer program.

The complete sequence of the Ad41 E3 region (*Eco*RI-*Esp*I fragment; map units 74 to 83.9) consisted of 3,373 nucleotides. This sequence encodes six complete ORFs, designated as RL1 to RL6. The E3 DNA sequence has one TATA box located between residues 248 and 251. Two different polyadenylation signals (AATAAA) were found between residues 2033 and 2038 (at the end of the RL2 ORF) and residues 3022 and 3027 (downstream of the RL6 ORF), suggesting that as in subgroup C Ads (8), the E3 gene can essentially be divided into two major regions (E3a and E3b). Table 1 compiles base composition and codon usage of the Ad41 E3 region DNA. Highest usage involved T, with codons TTT, CTT, ACT, and AAA each being used at least 3% of the time. Also, a comparison of the complete nucleotide sequence for the Ad41 E3 region with that of Ad40 (9), which likewise contains 3,373 nucleotides, shows that the two E3 genes are highly homologous (about 96.2%) but clearly different.

Ad41 E3 RL1. The Ad41 E3 RL1 ORF is the first in the E3a region. It extends from bp 683 to 1204 and encodes a putative protein of 173 amino acids with a molecular mass of 19.4 kDa (Fig. 2). Using a BLAST (basic local alignment search tool) computer program (2), we examined whether this predicted protein is homologous to Ad E3 proteins from other Ad subgroups whose sequences have been deposited in GenBank. It was found that although the 19.4-kDa protein corresponds in location to the E3 gp19 protein of Ad2 and Ad5 (subgroup C Ads), its sequences are unique to subgroup F Ads. In addition, comparison of this 19.4-kDa Ad41 protein with the 264-amino-acid protein of Ad12 E3 shows about 30% homology, although the former is 92 amino acids shorter than the latter. This E3 sequence similarity between a subgroup F virus (Ad40) and a subgroup A virus (Ad12) had been noted by Bailey and Mautner (3). The Ad40 and Ad41 RL1 sequences are 99% homol-

TABLE 1. Base composition and codon usage of Ad41 E3 region DNA

Nucleotide	No.	%	Nucleotide	No.	%	Nucleotide	No.	%	Nucleotide	No.	%
A	852	25.3	T	996	29.5	G	600	17.8	C	925	27.4
ATT	21	1.8	TTT	44	3.9	GTT	20	1.7	CTT	39	3.4
ATC	17	1.5	TTC	28	2.4	GTC	14	1.2	CTC	23	2.0
ATA	7	0.6	TTA	27	2.4	GTA	8	0.7	CTA	10	0.8
ATG	10	0.8	TTG	19	1.6	GTG	13	1.1	CTG	28	2.4
ACT	37	3.2	TCT	19	1.6	GCT	24	2.1	CCT	23	2.0
ACC	32	2.8	TCC	30	2.6	GCC	22	1.9	CCC	20	1.7
ACA	10	0.8	TCA	15	1.3	GCA	7	0.6	CCA	22	1.9
ACG	14	1.2	TCG	5	0.4	GCG	11	0.9	CCG	13	1.1
AAT	21	1.8	TAT	26	2.3	GAT	11	0.9	CAT	14	1.2
AAC	23	2.0	TAC	30	2.6	GAC	21	1.8	CAC	16	1.4
AAA	35	3.1	TAA	12	1.0	GAA	27	2.4	CAA	28	2.4
AAG	20	1.7	TAG	4	0.3	GAG	10	0.8	CAG	21	1.8
AGT	10	0.8	TGT	20	1.7	GGT	8	0.7	CGT	6	0.5
AGC	17	1.5	TGC	24	2.1	GGC	10	0.8	CGC	11	0.9
AGA	8	0.7	TGA	13	1.1	GGA	7	0.6	CGA	2	0.1
AGG	13	1.1	TGG	11	0.9	GGG	7	0.6	CGG	7	0.6

TABLE 2. Amino acid sequences and hydrophobicity values of the signal anchor domains of E3 6.7 kDa proteins from Ad3, Ad5, and Ad40/Ad41

Virus	Sequence of signal anchor domain[a]	Hydrophobicity value[b]
Ad3	NH_2-W-V-V-A-G-F-V-T-L-G-V-V-A-G-G-L-V-L-I-L-C-COOH	−26.6
Ad5	NH_2-G-I-G-V-G-V-I-L-T-L-V-I-L-F-I-L-I-L-A-L-L-COOH	−29.5
Ad41/Ad40	NH_2-T-S-L-L-M-V-P-T-K-A-F-T-L-I-T-L-L-T-T-F-V-COOH	−19.7

[a] References to Ad sequences: Ad3 (42), Ad5 (8), Ad40 (9), and Ad41 (GenBank accession number M85254).
[b] Calculated by the method of Hopp and Wood (23).

ogous; amino acid 144 is isoleucine in Ad41 and asparagine in Ad40. We further analyzed the sequence of this protein for potential glycosylation sites and found four, at amino acid positions 52, 81, 87, and 125. Also, six possible phosphorylation sites exist at positions 15, 25, 32, 54, 109, and 140. Further, based on hydropathy plot analysis, the 19.4-kDa protein is predicted to be a membrane-associated glycoprotein, like Ad2 gp19. However, as noted earlier by us (33a) and others (3), it is not homologous to the gp19 of subgroup B, C, and E Ads. Whether Ad41, like Ad12 (12a), uses E1A instead of E3 genes to down-regulate the major histocompatibility complex class I molecules at the level of transcription is not known. Experiments to study RL1 functions are under way.

Ad41 E3 RL2. The Ad41 E3 RL2 ORF extends from bp 1207 to 2034 and encodes a 276-amino-acid protein with an expected molecular mass of 31.6 kDa (Fig. 2). The BLAST computer program search for homology to other human Ad E3 proteins revealed that the sequence of this protein was unique (compared with sequences of Ad2, Ad3, Ad4, and Ad5 proteins) and again 99% homologous to the Ad40 sequence; amino acids at positions 192 and 264 are glutamic acid in Ad41 and lysine in Ad40. Further, as noted for Ad40 (3), comparison of this 31.6-kDa protein with the 268-amino-acid protein of Ad12 E3 shows about 34% homology (3). This finding suggests that RL1 and RL2 may be involved in viral processes involving the host gastrointestinal tract, since both Ad41 and Ad12 cause diarrheas in humans. The 31.6-kDa protein has four potential glycosylation sites (at positions 35, 97, 162, and 204), five potential N-myristoylation sites (at positions of 33, 72, 111, 148, and 205), and a domain (residues 234 to 254) with high hydrophobicity, suggesting, as with RL1, that RL2 may be a membrane protein.

Ad41 E3 RL3. The Ad41 E3 RL3 ORF, extending from bp 1730 to 1909, is near the carboxy end of E3a and in a different reading frame from RL2; it consists of 59 amino acids with a molecular mass of 6.7 kDa (Fig. 2). Interestingly, the Ad2 (subgroup C) 6.7-kDa protein is also located at the carboxy terminus of the Ad2 E3a genes (54) and is homologous to the similarly located Ad3 16-kDa protein (16). Further, analysis of the Ad2 E3 6.7-kDa protein, which consists of 61 amino acids, showed that it is a membrane protein with a signal anchor domain at residues 15 to 35 and is homologous to other Asn-linked type III integral membrane glycoproteins localized in the endoplasmic reticulum (52, 53). When we examined the Ad41 RL3 protein by the method of Hopp and Wood (23), we found that it also had a hydrophobic transmembrane segment spanning residues 26 to 46, although the degree of hydrophobicity of this segment is less than the corresponding segment of the Ad2 E3 6.7-kDa protein (Table 2). This finding suggests that the Ad41 RL3 protein, although different in sequence, may have a function similar to that of the Ad2 E3 6.7-kDa protein, viz., directing glycoproteins to the endoplasmic reticulum. This putative protein is not mentioned in previous sequence comparison studies for Ad40 or Ad41 (3, 9, 46), possibly because it overlaps the RL2 reading frame (9). In addition to sequence identification of the Ad41 RL3 gene, we have also demonstrated below that RNA from Ad41-infected cells, when analyzed by reverse transcription (RT)-PCR, amplifies a segment with the expected size of the RL3 gene.

Other Ad41 E3 ORFs. Other Ad41 E3 ORFs are located in the E3b region and show similarity with E3b genes of most

```
                                  Lumenal
          <- Signal   Peptide -><- Domain -><- Transmembrane
Ad41      MVTPLLLLVCLPIIY-ASTTFAAVSHLDTDCLPALLTYLIFTSVCCTAIC   49
Ad40      MVTPLLLLVCLPIIY-ASTTFAAVSHLDTDCLPALLTYLIFTSVCCTAIC   49
Ad12      MVTVLLIFLCLPVIF-SSSTFAAVSDLDPECLAPFAVYLIFTFVTATCVC   49
Ad5       MIPRVFILLTLVALFCACSTLAAVSHIEVDCIPAFTVYLLYGFVTLTLIC   50
Ad2       MIPRVLILLTLVALFCACSTLAAVAHIEVDCIPPFTVYLLYGFVTLILIC   50
Ad3       MIPRNFFFTILICPFNVCATFTAVATASPDCIGPFASYALFAFVTCICVC   50
Ad7       MIPRNFFFTILICAFNVCATFTAVATASPDCIGPFASYALFAFVTCICVC   50

           Domain -><-    Cytoplasmic   Domain    ->
Ad41      SIATFFVAIFQTADYLYVRVAYYRHHPQYRNHEVAALLCL-S  90
Ad40      SIATFFVAIFQTADYLYVRVAYYRHHPQYRNHEVATLLCL-S  90
Ad12      SIITLLITSLQFFDYYYVRIVYRRHHPRYQNPQIAALLQLQP  91
Ad5       SLITVVIAFIQCIDWVCVRFAYLRHHPQYRDRTIAELLRIL-  91
Ad2       SLVTVVIAFIQFIDWVCVRIAYLRHHPQYRDRTIADLLRIL-  91
Ad3       SIVCLVINFFQLVDWIFVRIAYLRHHPEYRNQNVAALLRLI-  91
Ad7       SIVCLVINFFQLVDWIFVRIAYLRHHPEYRNQNVAALLRLI-  91
```

FIG. 3. Amino acid comparison of the predicted protein sequence for Ad41 RL4 with sequences of the E3 10.4-kDa proteins obtained from various human Ad serotypes. Amino acids conserved among Ad serotypes are in boldface. Domains are also indicated. References to sequences for Ad serotypes are as follows: Ad2 (17), Ad3 (42), Ad5 (8), Ad7 (22), Ad12 (43), Ad40 (9), and Ad41 (GenBank accession number M85254).

```
          <-Signal Peptide->
Ad41      MKV--PLLCLILLHKVLAN--------------CHLHRPTEFLRCYSTET   34
Ad40      MKV--PLLCLILLHKVLAN---------------CHLHRPTEFLRCYSTET   34
Ad12      MKT--ALVLFFMLIPVWAS--------------SCQLHKPWNFLDCYTKET   35
Ad2       MK--RSVIFVLLIFCALPVLCSQTSAPPKRHISCRFTQIWNIPSCYNKQS   48
Ad5       MK--FTVTF-LLIICTLSAFCSPTSKP-QRHISCRFTRIWNIPSCYNEKS   46
Ad3       MQAMLPVILILLLPCI-PLASTATRATPEQLRKCKFQQPWSFLDCYHEKS   49
Ad7       MQAMLPVILILLLPCI-ALASTATRATPEQLRKCKFQQPWSFLDCYHEKS   49
Ad11      MQAILPIFLLLLLPYA-VSTPAAYSTPPEHLRKCKFQQPWSFLACYREKS   49

           <-    Transmembrane Domain     ->
Ad41      SSF--WLYSIIFILIFFATFLGLQIYGCLHLGWMHPPN-NLPRFP----G   77
Ad40      SSF--WLYSIIFILIFFATFLGLQIYGCLHLGWMHPPN-NLPRFP----G   77
Ad12      NYIG-WVYGIMSGLVFVSSVVSLQLYARLNFSWNKYTD-DLPEYPNPQDD  83
Ad2       DLSEAWLYAIISVMVFCSTIFALAIYPYLDIGWNAIDAMNHPTFPVPAVI  98
Ad5       DLSEAWLYAIISVMVFCSTILALAIYPYLDIGWKRIDAMNHPTFPAPAML  96
Ad3       DFPTYWIVIVGIINILSCTFFSITIYPTFNFGWNSPNALGYPQEPD-EHI  98
Ad7       DFPTYWIVIVGIINILSCTFFSITIYPTFNFGWNSPNALGYPQEPD-EHI  98
Ad11      EIPPNLIMIAGIINIICCTIISFLIYPLFDFDWNAPNAHDHPQDPE-EHI  98

                                Hydrophobic
                              <-C terminus->
Ad41      FLLQPPPP------PPAPVQRAPSVISYFHLNSEDV    107
Ad40      FLLQPPPP------PPAPVQRAPSVISYFHLNSEDV    107
Ad12      LPLNIVFP------EPP---RPPSVVSYFKFTGEDD    110
Ad2       PLQQVIA----PINQPRPPSPTPTEISYFNLTGGDD    130
Ad5       PLQQVVAGGFVPANQPRPTSPTPTFISYFNLTGGDD    132
Ad3       PLQHIQQPLALVQYENEPQPSLPPAISYFNLTGGDD    134
Ad7       PLQHIQQPLALVEYENEPQPSLPPAISYFNLTGGDD    134
Ad11      PLQHIQQPLALVEYENEPQPSLLPAISYFNLTGGDD    134
```

FIG. 4. Amino acid comparison of the predicted Ad41 RL5 protein sequence with sequences of the E3 14.5-kDa proteins from various human Ad serotypes. Amino acids conserved among Ad serotypes are in boldface. The border amino acids of each domain are in italics and double underlined. References to sequences for Ad serotypes are as follows: Ad2 (17), Ad3 (42), Ad5 (8), Ad7 (22), Ad11 (35), Ad12 (43), Ad40 (9), and Ad41 (GenBank accession number M85254).

```
Ad12    MIEPD----------LEIDGRITEQRLLTDRARRRQQDQKNKELIDLQTV    40
Ad41    MSD-Q----------LEIDGQRTEQLILA----RRKLKQQNQELFNLQAL    35
Ad40    MSD-Q----------LEIDGQCTEQLILA----RRKLKQQNQELFNLQAL    35
Ad3     MTDPIATSSTAAKELLDMDGRASEQRLIQLRIRQQQ-ERAVKELRDAIGI    49
Ad7     MTEILTTSNSAE-DLLDMDGRVSEQRLAQLRIRQQQ-ERVTKELRDVIQI    48
Ad11    MTETLTTSNSAE-DLLDMDGRVSEQRLAQLRIRQQQ-ERAAKELRDVIQI    48
Ad5     MTDTLD---------LEMDGIITEQRLLERRRAAAEQQRMNQELQDMVNL    41
Ad2     MTESLD---------LELDGINTEQRLLERRKAASERERLKQEVEDMVNL    41

Ad12    HQCKKGLFCLVKQATLRYESLPGKEHQLCYTLPTQRQTFTAMVGSVPIKV    90
Ad41    HQCKKGLFCLVKQAELCYD-VTQQGHELSYTLNKQRQSFMTMVGVKPIKV    84
Ad40    HQCKKGLFCLVKQAELCYD-VTQQGHELSYTLNKQRQSFMTMVGVKPIKV    84
Ad3     HQCKKGIFCLVKQSKISYE-ITATDHRLSYELGPQRQKFTCMVGINPIVI    98
Ad7     HQCKKGIFCLVKQAKISYE-ITATDHRLSYELGPQRQKFTCMVGINPIVI    97
Ad11    HQCKKGIFCLVKQAKISYE-ITATDHRLSYELGPQRQKFTCMVGINPIVI    97
Ad5     HQCKRGIFCLVKQAKVTYDS-NTTGHRLSYKLPTKRQKLVVMVGEKPITI    90
Ad2     HQCKRGIFCVVKQAKLTYEK-TTTGNRLSYKLPTQRQKLVLMVGEKPITV    90

Ad12    SQQAGEQEGSIRCLCDNPECLYTLIKTLCGLRNLLPMN    128
Ad41    TQQSGPVEGSILCQCTNSECMYTMVKTLCGLRELLPFN    122
Ad40    TQQSGPVEGSILCQCTNPECMYTMVKTLCGLRELLPFN    122
Ad3     TQQSGDTKGCIQCSCDSTECIYTLLKTLCGLRDLLPMN    136
Ad7     TQQSGDTKGCIHCSCDSIECTYTLLKTLCGLRDLLPMN    135
Ad11    TQQSGDTKGCIHCSCDSIECTYTLLKTLCGLRDLLPMN    135
Ad5     TQHSVETEGCIHSPCQGPEDLCTLIKTLCGLKDLIPFN    128
Ad2     TQHSAETEGCLHFPYQGPEDLCTLIKTMCGIRDLIPFN    128
```

FIG. 5. Amino acid comparison of the predicted Ad41 RL6 protein sequence with sequences of the E3 14.7-kDa proteins from other Ad serotypes. Amino acids conserved among Ad serotypes are in boldface. Italic segments indicate a conserved domain. The four leucine residues of a leucine zipper motif found in Ad12 are underlined. References to sequences for Ad serotypes are as follows: Ad2 (17), Ad3 (42), Ad5 (8), Ad7 (22), Ad11 (35), Ad12 (43), Ad40 (9), and Ad41 (GenBank accession number M85254).

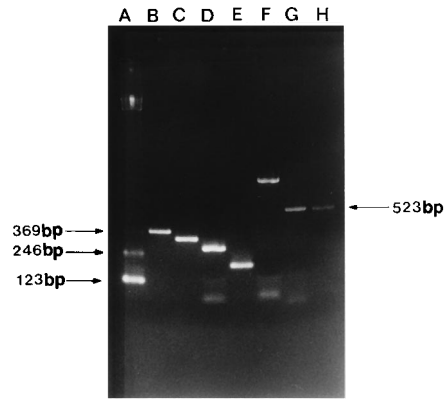

FIG. 6. Analysis of ethidium bromide-stained RT-PCR products of Ad41 E3 region genes. Total RNA from Ad41 infected HEp-2 cell monolayers was isolated by using a Tripure solution (Boehringer Mannheim) and reverse transcribed into cDNA by using a preamplification kit (GIBCO-BRL), and the cDNAs were amplified by using a PCR kit (Roche Molecular Division). The amount of DNA applied on each lane was 10 μl per well. Lanes: A, 123-bp multiple ladders as a molecular weight standard; B, RL6; C, RL5; D, RL4; E, RL3; F, RL2; G, RL1; H, 523-bp mRNA from the preamplification kit (GIBCO-BRL) as an internal positive control. Estimated positions of markers are shown on the left.

other Ads. The Ad41 E3 RL4 ORF, regarded as the first complete ORF in E3b, extends from bp 2056 to 2325 and encodes a protein of 90 amino acid residues with a putative molecular mass of 10.1 kDa. Comparison of amino acid sequences with those of the E3 10.4-kDa protein from other Ad serotypes shows that the expected Ad40 and Ad41 proteins exhibit 43% homology to the Ad2 and Ad5 E3 10.4-kDa proteins (Fig. 3), which are reported to induce an endosome-mediated internalization and degradation of the epidermal growth factor receptor (7). The Ad41 and Ad40 E3 10.4-kDa proteins are identical except for amino acid position 85 (alanine in Ad41 and threonine in Ad40). The Ad41 E3 RL4 protein, like its Ad counterparts, can be divided into a signal segment (residues 1 to 21), lumenal domain (residues 22 to 33), transmembrane domain (residues 34 to 59) and a relatively conserved (about 42%) carboxy terminus (residues 60 to 90) (Fig. 3).

The Ad41 E3 RL5 ORF extends from bp 2325 to 2648 and encodes a putative protein of 107 amino acids with a molecular mass of 12.3 kDa. Comparison of this protein with the E3 14.5-kDa proteins from other human Ad serotypes shows that the Ad41 RL5 protein has about 30% homology with the Ad2/Ad5 E3 14.5-kDa protein (Fig. 4). Both proteins exhibit a signal segment (residues 1 to 16) at the amino terminus, a transmembrane segment (residues 36 to 66), and a hydrophobic carboxy terminus (Fig. 4).

Finally, the Ad41 E3 RL6 ORF extends from bp 2641 to 3009 and encodes a putative protein of 122 amino acids with a molecular mass of 14.7 kDa which is analogous to the Ad2/Ad5 E3 14.7-kDa protein (Fig. 5). At both the DNA and amino acid sequence levels, RL6 is highly conserved (80 to 98%) within its own Ad subgroup but less so (45 to 60%) between different subgroups (data not shown). When analyzed further, the RL6-encoded protein was found to have certain highly conserved domains (particularly between amino acid residues 35 and 53) among all Ads, indicating that may be involved in a common function (Fig. 5). Also, the Ad41 E3 RL6 protein has fewer charged amino acids (arginine, aspartic acid, glutamic acid, and lysine) (20.6%) than the Ad2 14.7-kDa protein (28.9%). Third, it is of interest that the Ad12 14.7-kDa protein contains a putative leucine zipper motif (32, 34) between residues 104 and 125 (Fig. 5), suggesting that it may also act as a regulatory protein.

RT-PCR analysis of Ad41 E3 RL1 to RL6 ORFs. To determine whether the E3 region ORFs of RL1 to RL6 encode genes that are expressed inside Ad41-infected cells, we undertook a set of experiments using RT-PCR. HEp-2 cell monolayers in T-25 tissue culture flasks (Corning Glassware, Corning, N.Y.) were infected with Ad41 in the presence of cycloheximide and cytosine arabinoside as previously described (44). Total RNA was harvested at various times postinfection by using a Tripure reagent (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) according to the manufacturer's instructions. Concentration of total RNA was determined spectrophotometrically at a wavelength of 260 nm by using a DU-70 spectrophotometer (Beckman). In RT reactions, a preamplification kit purchased from GIBCO-BRL (Gaithersburg,

TABLE 3. Primers used in RT-PCR amplification of Ad41 E3 genes

Gene	Sense strand	Antisense strand
RL1	5′-CTCACCGGTATGAAGATCTGTGTTC-3′	5′-CTCTTAGTAAAAATTATCAAGATTG-3′
RL2	5′-CTCCTCGAGATGCTGCTGTTTTTACTTTGCCTTC-3′	5′-CTCTGATCATTATTTTTCTAAACAAATATGCACTGG-3′
RL3	5′-CTCCTCGAGATGGCAGGCAAAGCAACTTCTACCA-3′	5′-CTCTGATCACTACCTGGACATTGTCATTATATTG-3′
RL4	5′-CTCACCGGTATGGTAACTCCTCTTCTCCTGCTTG-3′	5′-TCATGACAGGCACAGAAGGGCGGCCACC-3′
RL5	5′-CTCACCGGTATGAAAGTTCCTCTTCTCTGTCTTATCC-3′	5′-TCAGACATCTTCAGAGTTAAGATG-3′
RL6	5′-CTCACCGGTATGTCTGACCAACTAGAAATCGACGG-3′	5′-TTAATTAAAGGGGAGAAGTTCCCTGA-3′

TABLE 4. Comparison of estimated lengths for Ad41 E3 RL genes obtained by RT-PCR with those calculated from the DNA ORF sequence

Gene	R_f on gel	Length (bp)	
		Estimated from agarose gel	Calculated from DNA sequence
RL1	0.588	523	522
RL2	0.471	830	826
RL3	0.794	194	180
RL4	0.735	269	273
RL5	0.691	330	324
RL6	0.662	365	369

Md.) was used to generate cDNA libraries. Total RNA treated with DNase prior to RT and samples with no reverse transcriptase added to the RT reaction were used as controls. In PCR amplification, the procedure of Kawasaki (26) was used as described in detail elsewhere (57). Primers synthesized for RT-PCR are listed in Table 3. Treatment of total RNA with DNase prior to RT-PCR was carried out to ensure that the mRNAs rather than residual viral DNA were amplified. As seen in Fig. 6, all six Ad41 E3 region genes are transcribed inside infected cells. The expected and calculated lengths of each gene are virtually identical (Table 4). Samples without reverse transcriptase added did not show any bands (data not shown), indicating that the genes (Fig. 6) were amplified from viral mRNAs and not from residual viral DNA.

We conclude that RL4 to RL6 appear to generally be conserved among other human Ad subgroups, such as B, C, and E. In contrast, RL1 and RL2, located in the E3a region, appear to be unique to subgroup F adenoviruses, just as the E3 20.5-kDa protein is unique to subgroup B Ads (Ad3, Ad7, Ad11, and Ad35) (12, 22, 35, 42). Both Ad41 E3 RL1 and RL2 proteins have characteristics of membrane proteins. RL3 encodes a 6.7-kDa protein which shows homology to other Ad subgroup proteins of this size. It has not previously been reported for subgroup F Ads and may have been missed because of an overlap in reading frame with RL2.

In agreement with the phylogenetic study for Ad40 (3), our sequence analysis of Ad41 shows that subgroups F and A Ads share a great deal of homology for E3a proteins. This finding is in contrast to a comparison with Ad subgroups B, C, and E. Thus, it suggests a distinctive role for these proteins in Ad gastrointestinal versus respiratory tract infections. Exact functions for the two unique E3a gene products (RL1 and RL2) in Ad41 pathogenesis will be of particular interest to determine in the future.

Nucleotide sequence accession number. The sequence of the Ad41 E3 region has been deposited in GenBank under accession number is M85254.

(A preliminary version of these observations was presented at the June 1990 American Society for Biochemistry and Molecular Biology/American Association of Immunologists meeting, New Orleans, La.)

We thank S. B. Mayo for excellent technical assistance and D. Jett for editing the manuscript.

This work was partially supported by a grant from the Stanley S. Scott Cancer Center of the LSU Medical Center and by LSU Medical Center Institutional Funds.

REFERENCES

1. **Albert, M. J.** 1986. Enteric adenoviruses. Arch. Virol. **88:**1–17.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Bailey, A., and V. Mautner.** 1994. Phylogenetic relationships among adenovirus serotypes. Virology **205:**438–452.
4. **Brandt, C. D., H. W. Kim, W. J. Rodriquez, J. O. Arrobio, B. C. Jeffries, E. P. Stallings, C. Lewis, A. J. Miles, M. K. Gardner, and R. H. Parrot.** 1985. Adenoviruses and pediatric gastroenteritis. J. Infect. Dis. **151:**437–443.
5. **Brandt, C. D., H. W. Kim, R. H. Yolken, A. Z. Kapikian, J. O. Arrobio, W. J. Rodriquez, R. G. Wyatt, R. M. Chanock, and R. H. Parott.** 1979. Comparative epidemiology of two rotavirus serotypes and other viral agents associated with pediatric gastroenteritis. Am. J. Epidemiol. **110:**243–254.
6. **Burgert, H.-G., and S. Kvist.** 1985. An adenovirus type 2 glycoprotein blocks cell surface expression of human histocompatibility class I antigens. Cell **41:** 987–997.
7. **Carlin, C. R., A. E. Tollefson, H. A. Brady, B. L. Hoffman, and W. S. M. Wold.** 1989. Epidermal growth factor receptor is down-regulated by a 10,400 MW protein encoded by the E3b region of adenovirus. Cell **57:**135–144.
8. **Cladaras, C., and W. S. M. Wold.** 1985. DNA sequence of the early E3 transcription unit of adenovirus 5. Virology **140:**28–43.
9. **Davison, A. J., E. A. R. Telford, M. S. Watson, K. McBride, and V. Mautner.** 1993. The DNA sequence of adenovirus type 40. J. Mol. Biol. **234:**1308–1316.
10. **Deininger, P. L.** 1983. Approaches to rapid DNA sequence analysis. Anal. Biochem. **135:**247–263.
11. **deJong, J. C., R. Wigand, A. H. Kidd, G. Wadell, J. G. Kapsenberg, C. J. Muzerie, A. G. Wermenbol, and R.-G. Firtslaff.** 1983. Candidate adenovirus 40 and 41: fastidious adenoviruses from human infant stool. J. Med. Virol. **11:**215–231.
12. **Flomenberg, P. R., M. Chen, and M. S. Horwitz.** 1988. Sequence and genetic organization of adenovirus type 35 early region 3. J. Virol. **62:**4431–4437.

12a. **Friedman, D. J., and R. P. Ricciardi.** 1988. Adenovirus type 12 E1A gene represses accumulation of MHC class I mRNAs at the level of transcription. Virology **165:**303–305.

13. **Gooding, L. R., L. W. Elmore, A. E. Tollefson, H. A. Brady, and W. S. M. Wold.** 1988. A 14.7 MW protein from the E3 region of adenovirus inhibits cytolysis by tumor necrosis factor. Cell **53:**341–346.
14. **Gooding, L. R., T. S. Ranheim, A. E. Tollefson, L. Aquino, P. Duerksen-Hughes, T. M. Horton, and W. S. M. Wold.** 1991. The 10,400- and 14,500-dalton proteins encoded by region E3 of adenovirus function together to protect many but not all mouse cell lines against lysis by tumor necrosis factor. J. Virol. **65:**4114–4123.
15. **Gooding, L. R., I. O. Sofola, A. E. Tollefson, P. Duerksen-Hughes, and W. S. M. Wold.** 1990. The adenovirus E3-14.7K protein is a general inhibitor of tumor necrosis factor-mediated cytolysis. J. Immunol. **145:**3080–3086.
16. **Hawkins, L. K., J. Wilson-Rawls, and W. S. M. Wold.** 1995. Region E3 of subgroup B human adenoviruses encodes a 16-kilodalton membrane protein that may be a distant analog of the E3-6.7K protein of subgroup C adenoviruses. J. Virol. **69:**4292–4298.
17. **Herisse, J., and F. Galibert.** 1981. Nucleotide sequence of the EcoRI E fragment of adenovirus 2 genome. Nucleic Acids Res. **9:**1229–1240.
18. **Hierholzer, J. C., R. Wigand, L. J. Anderson, T. Adrian, and J. W. M. Gold.** 1988. Adenoviruses from patients with AIDS: a plethora of serotypes and a description of five new serotypes of subgroup D (type 43-47). J. Infect. Dis. **158:**804–813.
19. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26:**365–369.
20. **Hoffman, P., P. Rajakumar, B. Hoffman, R. Heuertz, W. S. M. Wold, and C. R. Carlin.** 1992. Evidence for intracellular down-regulation of the epidermal growth factor (EGF) receptor during adenovirus infection by an EFG-independent mechanism. J. Virol. **66:**197–203.
21. **Hoffman, P., M. B. Yaffe, B. L. Hoffman, S. P. Yei, W. S. M. Wold, and C. R. Carlin.** 1992. Characterization of the adenovirus E3 protein that down-regulates the epidermal growth factor. J. Biol. Chem. **267:**13480–13487.
22. **Hong, J. S., K. G. Mullis, and J. A. Engler.** 1988. Characterization of the early region 3 and fiber genes of Ad7. Virology **167:**545–553.
23. **Hopp, T. P., and K. R. Wood.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78:**3824–3828.
24. **Horwitz, M. S.** 1990. Adenoviruses, p. 1723–1740. *In* B. N. Field and D. M. Knipe (ed.), Virology. Raven Press, New York.
25. **Kampe, O., D. Bellgrau, V. Hammerling, P. Lind, S. Paabo, L. Severinsson, and P. A. Peterson.** 1983. Complex-formation of class I transplantation antigens and a viral glycoprotein. J. Biol. Chem. **258:**10594–10598.
26. **Kawasaki, E. S.** 1990. Amplification of RNA, p. 21–27. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols. Academic Press, San Diego, Calif.
27. **Kidd, A. H., J. Chroboczek, S. Cusack, and R. W. H. Ruigrok.** 1993. Adenovirus type 40 virions contain two distinct fibers. Virology **192:**73–84.
28. **Kidd, A. H., and M. J. Erasmus.** 1989. Sequence characterization of the adenovirus 40 fiber gene. Virology **172:**134–144.
29. **Kidd, A. H., M. J. Erasmus, and C. T. Tiemessen.** 1990. Fiber sequence heterogeneity in subgroup F adenoviruses. Virology **179:**139–150.
30. **Korner, H., U. Fritzsche, and H.-G. Burgert.** 1992. Tumor necrosis factor α stimulates expression of adenovirus early region 3 proteins: implication for viral persistence. Proc. Natl. Acad. Sci. USA **89:**11857–11861.
31. **Krajcsi, P., A. E. Tollefson, and W. S. M. Wold.** 1992. The E3-14.5K integral

membrane protein of adenovirus that is required for down-regulation of the EGF receptor and for prevention of TNF cytolysis is O-glycosylated but not N-glycosylated. Virology **188:**570–579.
32. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240:**1759–1764.
33. **Luftig, R. B., D. Pieniazek, and N. J. Pieniazek.** 1990. Update on viral pathogenesis. ASM News **56:**366–368.
33a. **Luftig, R. B., N. J. Pieniazek, S. B. Slemenda, D. Pieniazek, and J. Velarde.** 1990. DNA sequence of the early E3 transcription unit of the human enteric adenovirus type 41 (Tak), abstr. 624. *In* Abstracts of the ASBMB/AAI meeting.
34. **McKnight, S. L.** 1991. Molecular zippers in gene regulation. Sci. Am. **264:** 54–64.
35. **Mei, Y. F., and G. Wadell.** 1992. The nucleotide sequence of adenovirus type 11 early 3 region: comparison of genome type Ad11p and Ad11a. Virology **191:**125–133.
36. **Paabo, S., F. Weber, O. Kampe, W. Schaffner, and P. A. Peterson.** 1983. Association between transplantation antigens and a viral membrane protein synthesized from a mammalian expression vector. Cell **33:**445–453.
37. **Pieniazek, D., N. J. Pieniazek, D. Macejak, J. Coward, M. Rayfield, and R. B. Luftig.** 1990. Differential growth of human enteric adenovirus 41 (Tak) in continuous cell lines. Virology **174:**239–249.
38. **Pieniazek, N. J., S. B. Slemenda, D. Pieniazek, J. Velarde, and R. B. Luftig.** 1989. Sequence of human enteric adenovirus type 41 Tak fiber. Nucleic Acids Res. **17:**9474.
39. **Pieniazek, N. J., S. B. Slemenda, D. Pieniazek, J. Velarde, and R. B. Luftig.** 1990. Human enteric adenovirus type 41 Tak contains a second fiber protein gene. Nucleic Acids Res. **18:**1901.
40. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
41. **Scaria, A., A. E. Tollefson, S. K. Saha, and W. S. M. Wold.** 1992. The E3-11.6K protein of adenovirus is an Asn-glycosylated integral membrane protein that localizes to the nuclear membrane. Virology **191:**743–753.
42. **Signäs, C., G. Akusjärvi, and U. Pettersson.** 1986. Region E3 of human adenoviruses: differences between the oncogenic adenovirus-3 and the non-oncogenic adenovirus-2. Gene **50:**173–184.
43. **Sprengel, J., B. Schmitz, D. Heuss-Neitzel, C. Zock, and W. Doerfler.** 1994. Nucleotide sequence of human adenovirus type 12 DNA: comparative functional analysis. J. Virol. **68:**379–389.
44. **Stewart, A. R., A. E. Tollefson, P. Krajcsi, S.-P. Yei, and W. S. M. Wold.** 1995. The adenovirus E3 10.4K and 14.5K proteins, which function to prevent cytolysis by tumor necrosis factor and to down-regulate the epidermal growth factor receptor, are localized in the plasma membrane. J. Virol. **69:**172–181.
45. **Svensson, L., I. Uhnoo, and G. Wadell.** 1990. Enteric adenoviruses of men, p. 97–114. *In* L. J. Saif and K. W. Theil (ed.), Viral diarrheas of man and animals. CRC Press, Inc., Boca Raton, Fla.
46. **Tiemessen, C. T., and A. H. Kidd.** 1995. The subgroup F adenoviruses. J. Gen. Virol. **76:**481–497.
47. **Tollefson, A. E., A. R. Stewart, S. Yei, S. K. Saha, and W. S. M. Wold.** 1991. The 10,400- and 14,500-dalton proteins encoded by region E3 of adenovirus form a complex and function together to down-regulate the epidermal growth factor receptor. J. Virol. **65:**3095–3105.
48. **Uhnoo, I., L. Svensson, and G. Wadell.** 1990. Enteric adenoviruses. Bailliere's Clin. Gastroenterol. **4:**627–642.
49. **Wadell, G., A. Allard, M. Evander, and Q.-C. Li.** 1986. Genetic variability and evolution of adenoviruses. Chem. Scr. **26B:**325–335.
50. **Wadell, G., A. Allard, M. Johansson, L. Svensson, and I. Uhnoo.** 1994. Enteric adenoviruses, p. 519–547. *In* A. Z. Kapikian (ed.), Viral infections of the gastrointestinal tract. Marcel Dekker, New York.
51. **Wigand, R., T. H. Adrian, and F. Bricoutt.** 1987. A new human adenovirus of subgroup D: candidate adenovirus type 42. Arch. Virol. **94:**283–286.
52. **Wilson-Rawls, J., S. L. Deutscher, and W. S. M. Wold.** 1994. The signal-anchor domain of adenovirus E3-6.7K, a type III integral membrane protein, can direct adenovirus E3-gp19K, a type I integral membrane protein, into the membrane of the endoplasmic reticulum. Virology **201:**66–76.
53. **Wilson-Rawls, J., S. K. Saha, P. Krajcsi, A. E. Tollefson, L. R. Gooding, and W. S. M. Wold.** 1990. A 6700 MW membrane protein is encoded by region E3 of adenovirus type 2. Virology **178:**204–212.
54. **Wilson-Rawls, J., and W. S. M. Wold.** 1993. The E3-6.7K protein of adenovirus is an Asn-linked integral membrane glycoprotein localized in the endoplasmic reticulum. Virology **195:**6–15.
55. **Wold, W. S. M., and L. R. Gooding.** 1991. Region E3 of adenovirus: a cassette of genes involved in host immunosurveillance and virus-cell interactions. Virology **184:**1–8.
56. **Yeh, H.-Y., and R. B. Luftig.** 1991. Human enteric adenovirus type 41 binding to HEp-2 cells is specific. Virology **183:**410–414.
57. **Yeh, H.-Y., N. Pieniazek, D. Pieniazek, H. Gelderblom, and R. B. Luftig.** 1994. Human adenovirus type 41 contains two fibers. Virus Res. **33:**179–198.