DNA Sequence of the Adenovirus Type 41 Hexon Gene and Predicted Structure of the Protein

By CELIA I. A. TOOGOOD AND RONALD T. HAY*

Department of Biochemistry and Microbiology, University of St Andrews, St Andrews, Fife KY16 9AL, U.K.

(Accepted 25 May 1988)

SUMMARY

The gene for the major capsid protein (hexon) of human adenovirus type 41 (Ad41) has been isolated and the complete DNA sequence determined. Comparison of the predicted amino acid sequence with hexons from human Ad2 and Ad5 and bovine adenovirus type 3 reveals regions of high homology at the N and C termini separated by a central region of low homology. Fitting of the Ad41 hexon sequence to the known three-dimensional structure of the Ad2 hexon demonstrates that both hexons have a common architecture. Regions of the hexon which in the trimer constitute the pseudohexagonal base are highly conserved, with the major amino acid changes concentrated in the domains forming the triangular towers which represent the surface of the capsid. Changes in the Ad41 towers therefore permit the virus to present a unique surface to the environment while conservation of residues in the base maintains the integrity of hexon–hexon contacts. A striking difference is the absence in the Ad41 sequence of 32 amino acids which are present in the Ad2 sequence. In Ad2 this region is highly charged and may be responsible for pH-induced conformational changes within the virus capsid. The DNA sequence in the region surrounding the Ad41 hexon gene was also determined and revealed an open reading frame which appeared to code for the homologue of the Ad2-coded endoprotease. Comparison of the predicted amino acid sequences of the Ad41 and Ad2 proteins revealed a high degree of homology suggesting that this protein may have an important role in the infectious cycle of the virus.

INTRODUCTION

Adenovirus (Ad) types 40 and 41 were first identified by electron microscopy of stool samples of infants suffering from acute gastroenteritis (Flewett et al., 1975). The involvement of Ad41 in the aetiology of diarrhoeal disease has been suggested by a number of studies (Whitelaw et al., 1977; Johansson et al., 1980; Brandt et al., 1984). At the acute stage of infection enteric adenoviruses are excreted in large quantities of up to $10^{11}$ particles/g of stool (Gary et al., 1979; Retter et al., 1979). Enteric adenoviruses have been visualized within the cells of the small intestine (Whitelaw et al., 1977) and it is therefore possible that multiplication is within the gastro-intestinal tract. Adenovirus types 40 and 41 can be distinguished from the other 39 human serotypes by serology, restriction enzyme cleavage analysis, polypeptide profiles and tissue culture growth characteristics (Wadell et al., 1980; Takiff et al., 1981; Uhnoo et al., 1983; Kidd, 1984). Although enteric adenoviruses are extremely difficult to propagate in conventional cell lines (de Jong et al., 1983) most strains can be successfully propagated in 293 cells which are human embryonic kidney cells transformed by Ad5 DNA (Graham et al., 1977).

Human adenoviruses have a common architecture and genome organization. Virions are non-enveloped icosahedrons containing a single copy of linear dsDNA of approx. 35000 base pairs.
and at least 11 virus-encoded polypeptides (reviewed by Ginsberg, 1984). The capsid consists of 252 morphological units or capsomers (Horne et al., 1959) which are arranged to form the 20 triangular facets of the icosahedron. At the 12 vertices the capsomers have five neighbours and are known as pentons, whereas the remaining capsomers have six neighbours and are known as hexons (Ginsberg et al., 1966). Pentons consist of a trimeric fibre complexed to a pentameric penton base (van Oostrum & Burnett, 1985). Hexons consist of three identical polypeptide chains which combine to give the pseudohexagonal shape (Grutter & Franklin, 1974). The three-dimensional structure of the Ad2 hexon has been determined by X-ray crystallography (Roberts et al., 1986) and has revealed a protein of great complexity (Fig. 6). Each hexon trimer has a triangular top superimposed on a pseudohexagonal base. The base is composed of three copies of each of the two pedestal domains (P1 and P2) which are very similar and share the same eight-stranded β-barrel structure. The triangular top consists of three tower domains (T) each of which is composed of three intertwined loops (l₁, l₂, and l₃) rising from the pedestal domains. Remarkably, the three loops which form each tower domain arise from different subunits and the intertwining of the individual polypeptide chains confers great stability on the hexon structure.

The hexons have been shown to carry type, group, intrasubgroup and intersubgroup antigenic determinants (Norrby, 1969; Norrby & Wadell, 1969). Type-specific determinants have been demonstrated on the surface of the virion (Norrby, 1969; Wilcox & Mautner, 1976) whereas most group-specific determinants appear to be internal (Norrby, 1969). Sera raised against purified Ad5 hexon antigen are capable of neutralizing virus infectivity of the homologous type 5 virus and the closely related type 1 virus (Kjellén & Pereira, 1968). Variation in the surface properties of the adenovirus virion is likely to be a consequence of evolutionarily driven changes in the genes coding for the surface proteins. Since the hexon is the major component of the virion we have determined the DNA sequence of the Ad41 hexon gene and have compared its predicted amino acid sequence to that of human Ad2 and Ad5 (Kinloch et al., 1984) and bovine adenovirus type 3 (BAV3) (Hu et al., 1984). Fitting of the Ad41 amino acid sequence to the three-dimensional model of the Ad2 hexon indicates that although the surface regions of the hexons are variable the internal regions are highly conserved.

**METHODS**

*Cells and virus.* The 293 cell line (Graham et al., 1977) was grown in Glasgow-modified Eagle's medium containing 10% newborn bovine serum. Adenovirus type 41 (strain Tak) was propagated in 293 cells and purified as described (Mautner & Wilcox, 1974). Virion DNA was extracted by the method of Pettersson & Sambrook (1973).

*Identification and isolation of the Ad41 hexon gene.* HindIII-cleaved Ad41 DNA was ligated into pUC13 and transformed into JM83 by standard procedures (Maniatis et al., 1982). Plasmids containing the hexon gene were identified by colony hybridization using a nick-translated (Rigby et al., 1977) 420 bp Smal–XhoI fragment derived from the 5’ end of the Ad2 hexon gene (Mautner & Boursnell, 1983). pCT3 was isolated from a positively hybridizing colony and was shown to contain the 5’ region of the hexon gene. An overlapping fragment containing the remainder of the hexon gene was identified using a BglII–HindIII fragment (Fig. 1) to probe Southern blots (Southern, 1975; Hay et al., 1984) of restriction enzyme-digested Ad41 DNA. Adenovirus type 41 DNA was therefore digested with EcoRI and BglII, the fragments ligated into BamHI–EcoRI-cleaved pUC13 and transformed colonies were probed with the BglII–HindIII fragment described above. pCT6 was thus isolated and contained the remainder of the hexon gene on a 2.7 kb fragment. Plasmid DNA was prepared by CsCl/ethidium bromide centrifugation as described (Hay et al., 1984).

*DNA sequencing.* Plasmid DNA was sonicated, treated with T4 DNA polymerase in the presence of dNTPs to generate blunt ends and size-fractionated on an agarose gel. DNA between 400 and 800 bp was isolated and ligated into Smal-cut dephosphorylated M13mp8 RF. Recombinant phage containing the hexon gene were detected by plaque hybridization using nick-translated plasmid inserts containing the Ad41 hexon gene. Phage DNA was sequenced by the dideoxy procedure (Sanger et al., 1977, 1980) using buffer gradient gels (Biggin et al., 1983). The complete sequence was assembled using the Staden DB suite of computer programs (Staden, 1982a).

*Materials.* Klenow polymerase was purified as described (Joyce & Grindley, 1983). Restriction enzymes, T4 DNA ligase, T4 DNA polymerase and bacterial alkaline phosphatase were obtained commercially from a variety of sources and used as specified by the manufacturers. Radioactive isotopes were obtained from Amersham.
The complete nucleotide sequence of the cloned HindIII and BgIII/EcoRI fragments was determined by shotgun dideoxy sequencing and is displayed in Fig. 2. An open reading frame of 2778 nucleotides encodes the hexon polypeptide of 924 amino acids with a predicted $M_r$ of 103844 which is smaller than the 109077 $M_r$ Ad2 hexon polypeptide (Akusjärvi et al., 1984). This difference in $M_r$ endows the Ad41 polypeptide with an increased electrophoretic mobility in SDS–polyacrylamide gels relative to its Ad2 counterpart (data not shown). Comparison of the Ad41 hexon amino acid sequence with that of Ad2 is shown in Fig. 3. Deletions and insertions were accommodated to align the two sequences maximally. The overall amino acid identity between the two hexon species including insertions and deletions is 77% and of the different amino acids more than one-third of the changes are conservative. Although this represents a high degree of homology between viruses in different subgroups the amino acid changes are not distributed randomly throughout the sequence, but are concentrated in two regions of the polypeptide spanning amino acids 134 to 296 and 409 to 490 (Fig. 3). In contrast the N and C termini of the Ad2 and Ad41 hexons are extremely well conserved, with only two non-conservative and three conservative changes in the first 100 amino acids. Regions 134 to 296 and 409 to 490 represent 49% and 58% homology respectively while regions 1 to 134 and 491 to 924 represent over 90% homology. This distribution of changes is similar to that observed between the hexons of Ad2 and Ad5, which are both subgroup C adenoviruses (Kinloch et al., 1984). The
Structure of Ad41 hexon

The most striking difference between Ad2 and Ad41 is the absence in the latter of 32 amino acids present in Ad2. This region (139 to 170) in Ad2 is extremely rich in acidic amino acids and contains 15 glutamic acid residues. Comparisons of the Ad41 with the Ad2 sequence and the Ad41 with the BAV3 hexon sequence (Hu et al., 1984) were performed by Diagon plots (Fig. 4). These plots stress the overall similarity in the organization of the hexon polypeptides in different human and animal adenoviruses. A comparison of the region following amino acid 124 in human Ad2, 5 and 41 and BAV3 is made in Fig. 5. The N-terminal sequence of all four hexons is extremely well conserved but diverges sharply after amino acid 134. At this point in Ad5 there is a small deletion with respect to Ad2 which is followed by a region containing a high proportion of changed amino acids. At residue 139, there is a deletion in Ad41 that removes the acidic region discussed above. The hexon from BAV3 is also deleted with respect to Ad2, although in

Fig. 2. Sequence of the Ad41 hexon gene and flanking regions. Translation of the hexon open reading frame is shown in single-letter amino acid code. The amino acid is shown above the middle base of the codon.
Fig. 3. Comparison between amino acid sequences of Ad41 and Ad2 hexons. Identical amino acids are highlighted by asterisks. Dots represent blank characters which have been inserted into the sequence to obtain maximal alignment.

this case the deletion occurs after residue 144 but still removes the highly acidic region present in Ad2.

**Predicted three-dimensional structure of the Ad41 hexon**

The similarity between the Ad2 and Ad41 amino acid sequences has allowed us to fit the Ad41 hexon to the three-dimensional model of the Ad2 hexon (Roberts *et al.*, 1986). Displayed in Fig. 6 is the Ad2 hexon structure with the changes present in Ad41 superimposed upon it. It is immediately apparent that the differences between the hexons are largely confined to the surface of the molecule, with the base being highly conserved. The obvious exception to this is the absence in Ad41 of the highly acidic region present in Ad2 (amino acids 139 to 170) which stretches from the surface of the molecule down into the β-sheet of domain P1. Structural data (Roberts *et al.*, 1986) indicate that of the 16 amino acid residues in the D strand five are found on...
Fig. 4. Comparison, by Diagon plot analysis (Staden, 1982b) of the complete Ad41 hexon amino acid sequence with that of Ad2 (a) and BAV3 (b). The proportional algorithm was applied using a span of 11 and a score of 132.

Ad2  APKGAPNSEQEQTEDSGRAVEEEDDDEEEEEDQNARQATKK
Ad41  ...T...P...KDNN-------------------------------.
Ad5    .......P......~DEAAAT..LEINL...D.N.D.VD...~E...~Q.
BAV3  ...S...NTQFR.ANHGP--------------------QIAQ.SYV....

Fig. 5. Amino acid comparison of human Ad2, 5, 41 and BAV3 in the highly acidic region that is present in Ad2. Dots indicate identical amino acids and dashes represent blank characters that have been inserted to obtain maximal alignment.

the upper external surface of the trimer and therefore would not be involved in hexon–hexon contact. Thus it appears that in Ad2 the 11 residues from 130 to 140 are important for the integrity of the P1 domain. The amino acids that constitute the lower region of the D strand of Ad2 were compared with those in the Ad41 hexon which would be expected to occupy an equivalent position. Four amino acids are identical and three represent conservative changes. Molecular modelling has indicated that these amino acids in Ad41 could form a β-sheet that would maintain the integrity of the P1 domain (R. Murali & R. M. Burnett, personal communication). As mentioned previously the β-sheet which constitutes the D strand of the Ad2 hexon extends above the P1 domain to the surface of the trimer. To determine whether the corresponding region in the Ad41 hexon had the potential to form a β-sheet we utilized the secondary structure prediction program of the University of Wisconsin Genetics Computer Group Sequence Analysis Software Package. Using the method of Robson–Garnier (Garnier et al., 1978) this program predicted that the region 141 to 145 in the Ad41 hexon was indeed likely
Fig. 6. Fitting of the Ad41 amino acid sequence to the three-dimensional model of the Ad2 hexon (Roberts et al., 1986). Non-conservative changes scoring less than 11 on a score matrix (Dayhoff, 1969) are represented by closed circles, regions not present in Ad41 are represented by cross-hatching and regions present in Ad41 but not Ad2 are represented by stars.
Structure of Ad41 hexon

Fig. 7. Comparison of the amino acid sequence from 1 to 198 of the Ad2 and Ad41 predicted protease genes. Identical amino acids are highlighted by asterisks.

to form a β-sheet. However, these predictions are in no way definitive and it is worth noting that the program failed to predict many of the β-sheets that have been shown to exist in the Ad2 hexon (Roberts et al., 1986). Furthermore, a number of β-sheets were predicted which are not apparent in the crystal structure. The function of the highly acidic region in the Ad2 hexon is not known although it could be involved in pH-dependent conformational changes (Roberts et al., 1986). Extreme variation in the amino acids which constitute this region among the different adenovirus hexons indicates that it could represent the type-specific determinants of the viruses and be responsible for the adaptation of specific virus types to a particular environment. One striking feature of the Ad2 and Ad41 comparison is the conservation of proline residues. Of the 61 prolines present in the type 41 hexon 53 are conserved with respect to Ad2. These residues are clearly important in chain folding and many of the β-strands in the structure are terminated by proline. The overall impression from the fitting of the Ad41 sequence to the Ad2 structure is that while the towers of the two molecules, which constitute the surfaces of the virions, are quite different the bases of the molecules, which are involved in hexon-hexon contact, are essentially unchanged.

Sequence features in the region surrounding the hexon gene

Inspection of the DNA sequence downstream from the hexon gene reveals an ATG initiator codon at position 3007 followed by an uninterrupted open reading frame which extends to the end of the sequence we have determined (Fig. 2). In Ad2 the corresponding region of the genome codes for a polypeptide of Mr 23000. Furthermore the lesion responsible for the temperature-sensitive phenotype of the mutant H2ts1 has been mapped to this region of the genome (Yeh-Kai et al., 1983). At the non-permissive temperature this mutant is defective in polypeptide processing and it is thought that the Mr 23000 polypeptide is a virus-encoded endoproteinase. Comparison of the Ad41 predicted amino acid sequence with that of the Ad2 protease gene indicates a high degree of homology (Fig. 7). Of the 198 amino acids compared 155 are identical, with many of the changes being conservative. The degree of homology observed between the Ad2 and Ad41 protease sequences may well be a reflection of the requirement of this protein to interact specifically with cleave structurally similar viral proteins.

Upstream from the hexon gene in Ad2 is the termination codon (position 18751) of the gene for the structural protein pVI. Comparison of the region upstream of the Ad41 hexon gene with the corresponding region in Ad2 by Diagon analysis (Fig. 8) reveals a region of homology corresponding to the pVI gene which is separated from the Ad41 hexon gene by a non-homologous region of 97 bp. It thus appears that in this region the genome organization of Ad2
and Ad41 is very similar with conservation of protein coding regions and divergence of non-coding regions.

We would like to thank R. M. Burnett and R. Murali at Columbia University for the molecular modelling data, and R. E. Randall and W. C. Russell for critically reading the manuscript. We also thank the secretarial staff at St Andrews for typing and Bill Blyth for photography. This work was supported by a grant from the Wellcome Trust.

REFERENCES


Fig. 8. Comparison by Diagon plot analysis (Staden, 1982b) of the upstream regions of the Ad2 (nucleotides 18628 to 19027) and the Ad41 (Fig. 2, 1–400) hexon genes. The parameters applied were a span of 13 and a score of 9. T denotes the termination codon of the polypeptide VI gene. I denotes the initiation codon of the hexon gene.


*(Received 31 March 1988)*