Analysis of the hexon gene sequence of bovine adenovirus type 4 provides further support for a new adenovirus genus (Atadenovirus)

Ádám Dán,1,2 Zsolt Ruzsics,3 W. C. Russell,3 Mária Benkő1 and Balázs Harrach1

1Veterinary Medical Research Institute, Hungarian Academy of Sciences, Budapest, PO Box 18, H-1581 Hungary
2Central Veterinary Institute, Budapest, PO Box 2, H-1581 Hungary
3School of Biomedical Sciences, University of St Andrews, Fife KY16 9AL, UK

The putative hexon gene of bovine adenovirus type 4 (BAV-4), encoding 910 amino acid residues, has been identified and sequenced. A characteristic codon usage biased towards the use of AT-rich triplets was observed. Comparative analysis with other hexon sequences detected a high level of amino acid identity in the regions corresponding to the pedestals of the hexon. Substitutions, insertions and deletions were identified mainly in the variable regions forming the loops which are exposed on the outer surface of the virion. In these variable regions, BAV-4 shared similarity only with egg drop syndrome (EDS) virus and ovine adenovirus isolate 287 (OAV287). The close relationship of these viruses was also demonstrated by phylogenetic analysis of the hexon gene. In addition to the two groups of the Mastadenovirus and Aviadenovirus genera, a third cluster appeared comprising BAV-4, OAV287 and EDS virus.

The family Adenoviridae is divided into the genera Mastadenovirus and Aviadenovirus comprising adenoviruses isolated from mammals or birds, respectively. The separation of the two genera was partially based on the lack of immunologically cross-reactive proteins between members of the two groups (Norrby et al., 1976).

The genus Mastadenovirus presently contains 10 bovine adenovirus (BAV) serotypes. Because of the lack of the genus-specific complement-fixing antigen and some obvious biological differences observed in certain serotypes, subdivision of BAVs was introduced (Bartha, 1969). BAVs which were similar to human and other mammalian adenoviruses were classified into subgroup 1, while the slow-growing, non-cross-reacting BAV serotypes (4–8) were separated into subgroup 2.

After DNA analysis of the genome, remarkable genetic differences between the members of the two subgroups were reported (Hu et al., 1984a; Benkő et al., 1988, 1990). More recently, based primarily on phylogenetic analysis performed on protease gene sequences, a proposal was made for the establishment of a new adenovirus genus for these BAV serotypes (i.e. members of subgroup 2) along with egg drop syndrome (EDS) virus and ovine adenovirus isolate 287 (OAV287) (Harrach et al., 1997; Harrach & Benkő, 1998).

In addition to the complete genome sequence of OAV287 (Vrati et al., 1996) and EDS virus (Hess et al., 1997), DNA sequences are available from different subgroup 1 BAVs (Hu et al., 1984b; Cai et al., 1990a; Mittal et al., 1992; Elgadi et al., 1993; Esford & Haj-Ahmad, 1994; Salmon & Haj-Ahmad, 1994; Fitzgerald et al., 1997). From subgroup 2 BAVs, however, only the protease gene (Cai et al., 1990b) and the major late promoter sequence (Song et al., 1996) of BAV-7 have been published so far. The information concerning the phylogenetic relatedness between subgroup 1 and subgroup 2 BAVs, and between human and animal adenoviruses, is therefore deficient (Bailey & Mautner, 1994; Harrach et al., 1997).

To enhance the reliability of comparisons aiming to clarify the taxonomic place and relationships of BAVs, we decided to sequence the complete genome of BAV-4, a typical representative of subgroup 2. In the present paper, the hexon gene sequence of BAV-4 is described. Since the hexon is the major structural component of the adenovirus capsid, containing type, subgenus- and genus-specific antigenic determinants (Adám et al., 1996; Norrby & Wadell, 1969; Wilcox & Mautner, 1976), it has been widely investigated. Hexon gene sequences from numerous human and animal adenovirus types have been published; thus a very comprehensive amino acid sequence alignment could be assembled for the examination of genus-specific characteristics. The results of the phylogenetic analysis and the putative three-dimensional structure of the BAV-4 hexon protein are presented.
Fig. 1. Nucleotide and predicted amino acid sequence of the BAV-4 genome between map units 45 ± 7 and 57 ± 0 containing the C-terminal (284 nt) part of the pVI protein, the complete hexon gene and the N-terminal (188 nt) part of the protease gene. The putative recognition site of the second protease cleavage on the pVI protein is underlined, and the hypothetical cleavage site is marked by an arrow. The putative cleavage yields an 11-residue-long protease cofactor (pVIc).
The genome of the reference strain (THT/62) of BAV-4 had previously been cloned and mapped (Benkô et al., 1990). From the viral insert of clone pBAV402, a HindIII–BamHI fragment was subcloned into plasmid pMOB, a vector designed specifically for use in a transposon insertion system. Using a TN1000 kit (according to the instructions of the manufacturer, Gold Biotecnology) the transposon was randomly introduced into the viral fragment. By restriction enzyme digestion of the positive clones, the insertion site of the transposon was determined, and a nested set of clones was selected for the generation of overlapping sequence data. Sequencing was performed at the two ends of the viral insert (using T3 and T7 primers), and from the two ends of the transposon (using the G186 and G187 primers complementary to the transposon sequence, and supplied with the kit). Compared to the published physical maps of BAV-4 (Benkô et al., 1990), an additional HindIII site at map position 46±4 yielding an additional small fragment (N) was found. Similarly, an additional XbaI site (at map position 53±9) and fragment (F) could be identified. These internal fragments were subcloned into plasmid pBluescript (pBS) SK (Stratagene), and sequenced with T7 and T3 primers. The PRISM Ready Reaction Dye Deoxy Cycle sequencing protocol (Perkin-Elmer) and an ABI 373A automated DNA sequencer (Applied Biosystems) were used. The nucleotide sequences were read using the Applied Biosystems 373A DNA Sequencer Data Analysis Program and assembled by the program package Lasergene (DNASTAR). The coded proteins were identified using the BLAST search program (Altschul et al., 1990).

The three-dimensional structure of BAV-4 was predicted using Swiss-Model (Peitsch, 1996). Adenovirus hexon sequences for multiple alignments were retrieved from the GenBank and EMBL databases. Multiple alignment of the amino acid sequences was performed with the MultAlin computer program (Corpet, 1988) using a blosum62 comparison table with gap weight 12 and gap length weight 2. Since only homologous residues can be used in the phylogenetic calculations, the highly variable regions were removed as described by Harrach & Benkô (1998), who also indicate other tips and possible pitfalls of the method. Phylogenetic trees were constructed with programs included in PHYLP, version 3.572c (Felsenstein, 1989).

The C-terminal region of the pVI protein, the entire hexon gene, and the 5’ end of the protease gene of BAV-4 are presented in Fig. 1. The hypothetical second protease cleavage site of pVI could be identified. The amino acid sequence of the putative protease cofactor (pVIc) was similar to that of OAV287 and thus corresponded to the consensus suggested by Vrati et al. (1996). The putative BAV-4 hexon gene consists of 2733 nucleotides encoding 910 amino acid residues. Its location on the BAV-4 genome was estimated to be between map units 46±7 and 56±4 at the conventional location of the adenoviral hexon gene. Interestingly, however, the BAV-4 hexon gene overlapped the protease gene by four nucleotides.

In the adenoviruses examined so far, there is generally a short distance (ranging from 7 to 37 nucleotides) between the stop codon of the hexon gene and the start codon of the protease, with the exception of BAV-7, EDS virus and OAV287 (Harrach et al., 1997). At the N-terminal part of the hexon gene, the distance between the hypothetical stop codon of the gene for the pVI protein and the start codon of the hexon in both BAV-4 and OAV287 was only 19 nucleotides, and was 21 in the EDS virus, compared to the 84 nucleotides present in human adenovirus type 2 (Ad2). These observed overlaps or shorter intergenic distances might partially account for the smaller overall genome size of the subgroup 2 BAVs (Hu et al., 1984a; Benkô et al., 1988). The AT content (64±8%) of the hexon coding region of BAV-4 was closer to that of OAV287 (64±0%) and EDS virus (57±6%) than the other sequenced adenoviral hexons, in which this value ranges between 39 and 52%.

The hexon protein consists of three identical polypeptides (van Oostrum et al., 1987). The three-dimensional structure of the Ad2 hexon had earlier been determined by X-ray crystallography (Roberts et al., 1986), and revealed a protein with a dense pedestal base composed of two eight-stranded, antiparallel beta barrels (P1 and P2) and a triangular top formed from three loops (l1, l2 and l3) which are exposed on the outer surface of the virion. Accordingly, in the primary amino acid sequences there are several highly conserved regions (corresponding to the beta sheets of P1 and P2) interrupted by the loop sequences which contain variable and hypervariable regions (Toogood & Hay, 1988; Weber et al., 1994; Sheppard et al., 1995; Crawford-Miksza & Schnurr, 1996; Reubel & Studdert, 1997a). In the l3 region, which is not part of the exposed hexon surface, the degree of amino acid conservation is higher than in the other loops. The three-dimensional structure of the BAV-4 hexon could be modelled on the basis of the Ad2 structure (Roberts et al., 1986), and is presented in Fig. 2.

The alignment of the deduced amino acid sequence of the BAV-4 hexon compared with that of representative adenoviruses from which full-length hexon sequences are available is shown in Fig. 3(a). Altogether 22 sequences were included in the phylogenetic analysis, but from human adenovirus serotypes only Ad2 and Ad40, and from the two sequenced canine (CAV) and equine (EAV) adenoviruses only CAV-1 and EAV-1, are presented in the figure. The highly acidic region (amino acid residues 139–170) previously found in Ad2 but not present in any other adenovirus examined (Weber et al., 1994) was also missing from the BAV-4 hexon.

In Fig. 3(a), the hexon sequences examined are arranged according to their genus affiliation. BAV-4, OAV287 and EDS virus, the candidate members of the proposed third adenovirus genus, are placed in the middle. In addition to the most conservative regions of the hexon retained in all three groups (underlined residues), a large number of identical amino acids shared only by the members of one or two of the groups
could be identified (amino acid residues printed in bold). However, no sequence motifs were identified which were conserved throughout the different adenovirus types of any individual host species and could have been considered as host-specific determinants. Thus, for example, no common patterns were recognized between BAV-3 and BAV-4, or between EDS virus and fowl adenovirus (FAV) type 1 or 10, although these viruses share distant evolutionary but common host origins. No further evidence was found for the hypothesis that \( l_2 \) is a host species-specific region (Crawford-Miksza \\& Schnurr, 1996; Vrati et al., 1996; Reubel \\& Studdert, 1997a, b).

Interestingly, there were some instances where all the sequences except those from one virus were identical. For example at the end of \( l_4 \), before the beginning of the P2 region,
BAV-4 hexon analysis: a new adenovirus genus

Fig. 3. For legend see page 1459.
there is a very conserved stretch of amino acid sequence which was present in every type but BAV-3. The published nucleic acid sequence of the BAV-3 hexon at this critical point indicates a suspected frame shift possibly caused by the compression of four consecutive C bases. The questionable conserved amino acid sequence is in fact present in BAV-3, but in another reading frame. Several similar situations could be observed in the hexon sequences of EAV-2 and FAV-10. It would be interesting to repeat the sequencing or read the original gels again in order to confirm the amino acid sequence.

The result of the phylogenetic (distance matrix) analysis performed with the same set of data (22 hexon sequences) is shown in Fig. 3(b). In addition to the clusters of the genera *Mastadenovirus* and *Aviadenovirus*, a third, well separated cluster appeared, which contained BAV-4, OAV287 and EDS virus. Phylogenetic analysis of these sequences thus provides persuasive data about the phylogenetic relations, but unfortunately a limitation of the approach is the poor availability of DNA sequences representing different genome regions from many different adenovirus types. The unrooted tree presented here is, however, almost identical with those obtained earlier by analysis of the protease (Harrach et al., 1997) and DNA polymerase (Harrach & Benko, 1998) sequences. Similar tree topology (and maximal bootstrap values for the proposed *Atadenovirus* cluster) were generated using maximum parsimony analysis, or when the DNA sequence alignment was analysed with either of the two programs (data not shown but available on request).

The present results confirm our earlier findings concerning the distinctiveness of subgroup 2 BAVs, and support the proposal for the establishment of a third genus within the family *Adenoviridae*. In addition to BAV-4 and its close relatives...
BAV-4 hexon analysis: a new adenovirus genus

Fig. 3. (a) Multiple alignment of the amino acid sequences of representative adenovirus hexons including BAV-4. Amino acids conserved in every type examined are underlined, while the amino acids shared by all members of one or two groups only are in bold. The positions where each group, corresponding to the three proposed genera, contains a different (but identical within-group) amino acid, are emphasized by italics in addition to bold. The locations of the loops and pedestals are marked according to the Ad2 sequence. The aligned sequences (and their abbreviations) are: human adenovirus types 2 (H2) and 40 (H40), equine adenovirus type 1 (E1), canine adenovirus type 1 (C1), porcine adenovirus type 3 (P3), BAV-3 (B3), murine adenovirus type 1 (M1), BAV-4 (B4), OAV287 (O), EDS virus (EDS), fowl adenovirus types 1 (F1) and 10 (F10). (b) Phylogenetic tree of 22 adenovirus hexon sequences showing the homology of BAV-4 (B4), OAV287 (O287) and EDS virus, and their distinctiveness from all other sequenced mastadenoviruses and aviadenoviruses. The length of the branches indicates the phylogenetic distance between the different viruses. The tree was generated by distance matrix analysis (PROTDIST, using the Dayhoff PAM 001 scoring matrix, followed by FITCH, applying Global search option). The high statistical significance of the tree topology is shown by the bootstrap values obtained by analysis of 100 randomly re-sampled data sets from the aligned sequences. The bootstrap values of human adenoviruses are not shown. The aligned sequences and their accession numbers are: human adenovirus (marked by serotype number only) type 2 (2), J01917; 3, X76549; 4, X84646; 5, M73260; 7, X76551; 12, X73487; 16, X74662; 40, X51782; 41, X51783; 48, U20821; B3, K01264; B4, AF036092; C1, U55001; C2, U77082; E1, L79955; E2, L80007; EDS, Y09598; F1, U46993; F10, U26221; M1, U57336; OAV287, U40837; P3, U34592. The edited alignment (in PHYLIP format) of the 22 hexon sequences used as infile for the phylogenetic analysis is available at http://www.vmri.hu/~harrach.

(BAV serotypes 5–8), OAV isolate 287 and EDS virus should be classified into this new taxon. The descriptive genus name Atadenovirus, referring to the high genomic AT content of the candidate members, has been proposed (Harrach & Benkó, 1998).

Zs.R. has spent a 6 month fellowship at the University of St Andrews, Scotland, provided by the EU. A.D. is an enrolled PhD student at the Eötvös Loránd University, Budapest. This work was supported by the following grants of the National Research Fund of Hungary: OTKA T016882, T022405, A312 and C043.

References


Received 5 December 1997; Accepted 18 February 1998