DNA sequencing and analysis of the right-hand part of the genome of the unique bovine adenovirus type 10

Krisztina Ursu,1,2 Bala´zs Harrach,1 Katalin Matiz1,3 and Mária Benko˝1

1Veterinary Medical Research Institute, Hungarian Academy of Sciences, PO Box 18, H-1581 Budapest, Hungary
2Central Veterinary Institute, PO Box 2, H-1581 Budapest, Hungary
3Veterinary Institute of Debrecen, PO Box 51, H-4002 Debrecen, Hungary

The prototype strain of bovine adenovirus (BAdV) type 10 and four additional isolates that were indistinguishable in serum-neutralization tests have been shown to have remarkable variation in their genome size and restriction maps. In the present study, more than 40 % of the DNA sequence of the BAdV-10 isolate with the longest genome was determined. A biased base composition resulting in low (<41 %) GC content was noticed. Analysis of the genes of the DNA-binding protein, 100K, 33K, pVIII and fibre proteins, as well as early regions E3 and E4, which are encoded by the genome fragment examined, confirmed that BAdV-10 is different from the other known BAdV types regarding its phylogenetic distance and the organization of its exceptionally short E3 region, apparently containing only two genes. A comparative analysis of the E3 and E4 regions of BAdV-10 with various animal adenoviruses revealed interesting features accounting for the very short genome of BAdV-10. In the examined BAdV-10 isolate, duplicated sequences were localized in and around the fibre gene. Since BAdV-10 appears to be pathogenic to cattle and is genetically distant from the other BAdVs, we suggest that BAdV-10 is not a genuine bovine virus, but has recently switched host and is now undergoing an adaptation process in its new host. In accordance with this hypothesis, the remarkable predominance of AT-rich codons along with the variable fibre gene might be signs of adaptation.

INTRODUCTION

The prevalence of adenovirus infection in cattle is reflected by the high number of serotypes identified so far. Officially, ten bovine adenovirus (BAdV) types are recognized (Benkő et al., 2000), but at least one additional candidate (isolate Rus) has already been partially characterized (Zakharchuk et al., 1993; Élő, 2002). Thus, after the human and fowl adenoviruses (HAdVs and FAdVs), BAdVs comprise the third largest group of adenoviruses originating from one host species. Interestingly, however, the genetic distance among BAdVs is remarkable and they have now been classified into two separate genera. Several BAdV types (namely types 4–8) having peculiar genome organization (Vrati et al., 1996a) are members of the recently established genus Atadenovirus (Benkő & Harrach, 1998; Both, 2002a). The conventional BAdV types (1–3 and 9) that belong to the genus Mastadenovirus do not constitute a uniform group either and seem to be representatives of different adenovirus species (Benkő et al., 2000). DNA hybridization and phylogenetic studies have shown that BAdV-2 most likely shares a common origin with a number of ovine adenovirus types (Barbezange et al., 2000; Benkő, 2000; Rusvai et al., 2000), while BAdV-9 is most closely related to members of the species Human adenovirus C (Benkő et al., 2000). BAdV-1 and -3 also occupy distinct places on phylogenetic trees calculated using various protein (hexon, DNA polymerase, protease) sequences. The reason for the existence of BAdVs of seemingly miscellaneous evolutionary origin, compared with human adenoviruses that form a monophyletic group, has yet to be elucidated.

The most recently described BAdV, BAdV-10 (Horner et al., 1989), is particularly interesting because it is apparently capable of causing a well-defined disease in cattle characterized by acute severe fibrinous enterocolitis of sporadic occurrence, and every isolate of BAdV-10 isolated to date has originated from such a fatal case (Horner et al., 1980; Smyth et al., 1986; Adair et al., 1996; Lehmkuhl et al., 1999). Another uncommon characteristic of this virus is the variation in the genome size of the different isolates (Benkő et al., 1995). BAdV-10 shares no detectable DNA homology with other known BAdV types and therefore a type-specific in situ
DNA hybridization method could be developed (Smyth et al., 1996). Physical and genetic mapping and phylogenetic analyses of BAdV-10 strain Belfast1, clarifying its taxonomic place in the genus Mastadenovirus, have been published (Matiz et al., 1996, 1998). In the present study, the right-hand part of the genome of strain Belfast1 was sequenced and analysed. Fine mapping of this region allowed localization of the fibre gene as the site responsible for the interstrain genomic size variation. The analysis of other genes of structural or non-structural proteins encoded by this part of the genome confirmed the uniqueness of BAdV-10 among other BAdVs and further justifies its classification as a separate bovine adenovirus species (Benkő et al., 2000).

We also performed comparative analysis of early regions E3 and E4 of BAdV-10. Each putative gene identified in these regions was compared with the corresponding region from a large number of human and animal adenovirus isolates for homology. This analysis revealed that the E3 and E4 regions have a very variable and characteristic gene content in the different AdVs and seem to be very simple in BAdV-10.

**METHODS**

**Virus propagation and DNA purification.** Physical maps of the four BAdV-10 strains isolated in Northern Ireland (Adair et al., 1996) and labelled as Belfast1 to -4 in our laboratory have previously been compared with those of the prototype strain Ruakura (Benkő et al., 1995). In the present study, strain Belfast1 (isolate 85-1183), possessing the longest genome, was studied. The virus was propagated in low-passage-number bovine testicular cell cultures. The viral DNA was purified by ethanol precipitation after phenol/chloroform extraction from virions concentrated in an ultracentrifuge as described previously (Benkő et al., 1988).

**Molecular cloning and DNA sequencing.** From the previously cloned Clal fragments of the Belfast1 genome (Matiz et al., 1998), subclones were prepared by cleavage with Apal, BamHI and HindIII restriction enzymes followed by ligation in pBluescript SK phagemid (Stratagene). The terminal fragments of the genome were cloned using alkaline treatment to remove possible traces of the terminal protein remaining at the 5' ends (Zakharchuk et al., 1993). In the clones encompassing the right-hand part of the genome, nested deletions were prepared with exonuclease III treatment and religation (ExoIII/SI Deletion Kit; MBI Fermentase). For DNA sequencing, the selected clones were purified on Qiagen columns (Plasmid Mini kit). Sequencing was performed using a PRISM ready reaction dideoxy cycle sequencing protocol (Perkin-Elmer) on an ABI 373A automated DNA sequencer (Applied Biosystems) or using a dS DNA Cycle Sequencing System kit (Gibco BRL) on an ALFExpress DNA Sequencer and Fragment Analysis System (Pharmacia Biotech) with T7 and T3 primers.

**DNA sequence analysis.** To identify the encoded proteins, the program BLASTX (Gish & States, 1993) was applied using GenBank at the National Center for Biotechnology Information or our own dedicated adenovirus database at GeneFarm (http://www.vmri.hu/blast.htm). Modelling of the fibre protein subunit was done by manual alignment of the characteristic amino acid motifs arranged as suggested by van Raaij et al. (1999).

The content of early regions E3 and E4 was compared with 17 and 18 different adenovirus types, respectively. Homology between the different open reading frames (ORFs) was identified by BLAST search. Besides the well-characterized genes, additional ORFs and unidentified reading frames (URFs) were also analysed for homology.

**RESULTS**

**Organization of the right-hand half of the BAdV-10 genome**

The right-hand half of the BAdV-10 genome (11 939 bp) was sequenced and found to contain the genes encoding the protease, 100K, 33K, pVIII, 12.5K and fibre proteins on the r strand from left to right and the genes of the 34K and DBP on the l strand. Besides the readily identifiable genes, two novel ORFs were found in the E3 and E4 regions (URF1 and URF2, respectively; Fig. 1), which showed no obvious homology to any known

![Fig. 1. Physical and genetic maps of the right-hand part of the genome of BAdV-10. Empty arrows indicate the repeated, most likely nonsense portion of the fibre gene and the end of the 34K protein of the E4 region. The extension of the supposedly reiterated sequence is marked with vertical lines; however, it was not possible to discern which one of the two identical sections was the original and which the repetition.](image-url)
adenovirus genes. The overall gene arrangement of this part of the genome corresponded to that found in mastadenoviruses. However, several peculiar characteristics, such as the very short E3 and E4 regions and an extremely long fibre gene, were also noted. Moreover, a short reiteration of 1223 bp (consisting of the knob region of the fibre gene and the end of the gene for the 34K protein on the complementary strand) was found (Fig. 1). The existence and size of this reiteration was also confirmed by PCR on full viral DNA (data not shown). The GC content of the determined nucleotide sequence was found to be 40·81%, which is extremely low for a mastadenovirus.

The sizes of the sequenced BAdV-10 genes are shown in Table 1 and compared with their counterparts in BAdV-3. The small size of the majority of these genes and the short intergenic regions account for the compact genome of BAdV-10. However, the fibre gene and the inverted terminal repeats (ITRs) of BAdV-10, which are 3431 and 369 bp, respectively, are the longest such sequences known from any AdVs to date.

The fibre gene

Putative models for the three main parts (tail, shaft and knob) of the predicted fibre amino acid sequences of BAdV types 3, 4 and 10 are presented in Fig. 2. The tail domain of the BAdV-10 fibre was predicted to consist of 38 aa with recognizable homology to other AdV tail sequences. The fibre shaft region of the examined BAdV-10 strain contained 54 repetitive turns, whereas in BAdV-3, which possesses the longest hitherto known fibre shaft of 23 nm, 47 turns have been found (Ruigrok et al., 1994). The fibre gene of the other BAdV-10 strains (Ruakura and Belfast2 to -4) seemed to be shorter with some variation in the shaft region (data not shown). Alignment of the sequence of the fibre knob region from BAdV types 3, 4 and 10 failed to identify amino acid motifs that could be considered characteristic for bovine AdVs.

Table 1. Comparison of the size of genes encompassed by the right-hand genome part of two bovine adenovirus types

<table>
<thead>
<tr>
<th>Protein</th>
<th>BAdV-3</th>
<th>BAdV-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protease</td>
<td>616</td>
<td>591</td>
</tr>
<tr>
<td>DNA-binding protein</td>
<td>1289</td>
<td>1307</td>
</tr>
<tr>
<td>100K protein</td>
<td>2552</td>
<td>1967</td>
</tr>
<tr>
<td>33K protein</td>
<td>824</td>
<td>536</td>
</tr>
<tr>
<td>Protein VIII precursor</td>
<td>650</td>
<td>596</td>
</tr>
<tr>
<td>E3 region</td>
<td>1230</td>
<td>585</td>
</tr>
<tr>
<td>Fibre protein</td>
<td>2930</td>
<td>3431</td>
</tr>
<tr>
<td>Reiteration</td>
<td>–</td>
<td>1223</td>
</tr>
<tr>
<td>E4 region</td>
<td>2916</td>
<td>1194</td>
</tr>
<tr>
<td>ITR</td>
<td>195</td>
<td>369</td>
</tr>
<tr>
<td>Complete sequence from</td>
<td>13901</td>
<td>11939</td>
</tr>
<tr>
<td>protease to the end</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparison of early regions E3 and E4

The ORFs identified in the E3 and E4 regions of different AdVs are presented in Fig. 3(a, b). It was remarkable that both regions of BAdV-10 proved to be the shortest compared with other AdVs.

Striking differences were found in the length and complexity of the E3 regions (Fig. 3a). Primate AdVs contain six to ten genes, whereas most of the non-primate animal AdVs encode only two or three putative proteins. Porcine adenovirus type 4 (PAdV-4) from the species PAdV-B is an exception with six short ORFs. There was no single ORF identified that had homologues in every mastadenovirus. Homologues of the gene encoding a protein named 12·5K in species HAdV-C were identified in the largest number in the AdV types examined. Moreover, the E3 region sensu stricto apparently exists in mastadenoviruses and siadenoviruses only (Davison et al., 2003b).

Adjacent to the E3 region but on the complementary (l) strand, an additional ORF, termed the U exon (Davison et al., 1993), was found in most AdVs. The U exon was not identified in BAdV-10 and is also missing from MAdV-1. Several bovine and PAdVs also seem to lack the U exon.

The E4 region proved to be more conserved (Fig. 3b) and its content was almost identical in the different primate AdV species. Among mastadenoviruses, BAdV-10 had the simplest E4 unit with only two putative genes.

DISCUSSION

BAdV-10 is an interesting subject because it is not closely related to any other BAdVs and is capable of provoking severe disease in cattle. Fatalities caused by BAdV-10 have been revealed by retrospective in situ DNA hybridization examination of archival histological samples originating from different countries including Canada, Great Britain and The Netherlands (Smyth et al., 1999). The presence of BAdV-10 has also been demonstrated in cattle in the USA (Lehmkuhl et al., 1998). BAdV-10 has only been isolated on six occasions (Horner et al., 1980; Adair et al., 1996; Lehmkuhl et al., 1999) and the available strains differed from each other in their genome size.

Because of the difficulties encountered during isolation and propagation, BAdV-10 was originally described by Horner et al. (1989) as a virus probably belonging to a peculiar group of BAdVs (so-called subgroup 2; Bartha, 1969), which were later moved into the novel genus Atadenovirus. The name of the new genus reflects the high AT content found in the genomic DNA of its initial members. This genus contains five out of the ten recognized BAdVs, including the candidate BAdV-11 (strain Rus) (Elö et al., 2003). Adenoviruses with typical genome organization of, and close phylogenetic relations to, atadenoviruses have recently been identified in a number of different reptilian hosts including snakes (Benkö et al., 2002; Farkas et al., 2002;
Marschang et al., 2003) and different lizard species (Wellehan et al., 2003). Interestingly, however, the genomes of these reptilian viruses seem to have an equilibrated base composition. We have hypothesized that atadenoviruses might represent an adenovirus lineage that has co-evolved with reptiles and that their occurrence now in ruminants and birds might be the result of host switches (Benkő & Harrach, 2003). The biased base composition of the genome is perhaps the sign of an adaptation process. The part of the genome of BAdV-10 sequenced to date also showed a low GC content. On the other hand, phylogenetic analyses and the presence of the gene for the V protein have indicated that BAdV-10 is indeed a mastadenovirus (Matiz et al., 1998). The results of the present study further support this

Fig. 2. Comparison of the predicted structure of the fibre protein of different BAdV types representing two genera. BAdV-3 and BAdV-10 are mastadenoviruses and BAdV-4 is an atadenovirus.

<table>
<thead>
<tr>
<th>BAdV-3</th>
<th>BAdV-10</th>
<th>BAdV-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHRKQVRQNLVYYVRRAFRHMTFED</td>
<td>MHRKQVRQNLVYYVRRAFRHMTFED</td>
<td>MHRKQVRQNLVYYVRRAFRHMTFED</td>
</tr>
<tr>
<td>MMRKQVRQNLVYYVRRAFRHMTFED</td>
<td>MMRKQVRQNLVYYVRRAFRHMTFED</td>
<td>MMRKQVRQNLVYYVRRAFRHMTFED</td>
</tr>
<tr>
<td>MHRKQVRQNLVYYVRRAFRHMTFED</td>
<td>MHRKQVRQNLVYYVRRAFRHMTFED</td>
<td>MHRKQVRQNLVYYVRRAFRHMTFED</td>
</tr>
<tr>
<td>MHRKQVRQNLVYYVRRAFRHMTFED</td>
<td>MHRKQVRQNLVYYVRRAFRHMTFED</td>
<td>MHRKQVRQNLVYYVRRAFRHMTFED</td>
</tr>
</tbody>
</table>
E3 is one of the best-studied transcription units of the mastadenovirus genome. By definition, it is an early region, but transcripts are produced even at late times of infection (Mahr & Gooding, 1999; Russell, 2000). In the different HAdVs, a varying number of proteins are encoded here, the functions of which have been partially elucidated and seem to lie mainly in helping the virus in the evasion of the host’s immune system (Wold et al., 1984; Wold & Gooding, 1991). In species HAdV-C, the existence of at least seven E3-encoded proteins has been confirmed (Sparer et al., 1996). It is also well known that E3 is not essential for in vitro virus replication and can easily be deleted and replaced by foreign genes that are expressed (Graham & Prevec, 1992). The E3 region has attracted much attention for its suitability as an insertion site for potential gene-delivery vectors, and it has been sequenced and analysed in many different animal AdVs, which have been studied for feasibility as a foreign gene expression system. The lack of a real homologue of the E3 region in non-mastadenoviruses has also been discovered during manipulation of the genome of ovine adenovirus type 7 (Vrati et al., 1995) and the CELO virus (Michou et al., 1999), the best-studied representatives of at- and aviadenoviruses, respectively. At the right end of the genome of atadenoviruses, a large number of as yet poorly characterized genes and ORFs has been described, which are unique for the genus and have been named the E3 region only because of their in vitro non-essential nature (Vrati et al., 1996a, b). In siadenoviruses, there is one large ORF situated between the genes for pVIII and the fibre protein, but it is not homologous to any other known E3 genes (Davison et al., 2000). In Fig. 3(a), the homologies presented are only those that were revealed with the use of the BLASTX search program, although additional homologies among ORFs in the E3 region have recently been described by Davison et al. (2003a) on the basis of manual sequence comparisons.

AdVs originating from primates have been found to have the longest and most complex E3 regions, whereas the E3 region of the mouse AdV (MAdV) is among the shortest, containing a single ORF (Meissner et al., 1997; Raviprakash et al., 1989). Considering the evolutionary distance between rodents and primates, representing the most ancient and most modern lineages of mammals, respectively, it is tempting to speculate that adenovirus evolution has been accompanied by the gradual sophistication of the E3 region. In BAdV-10, this region consists of only two ORFs and is even shorter than that in MAdV-1. The interpretation of this finding could be that BAdV-10 has originated from the most primitive rodents or that its E3 region is truncated; in both cases this assumes that cattle are not the original host of BAdV-10. In support of this theory is the observation that partial or complete deletion of the E3 region of HAdV-5 results in an increased pathogenicity in an animal model system (Ginsberg et al., 1989; Ginsberg & Prince, 1994). BAdV-10 also seems to be linked to a specific pathology.

Comparison of the E4 transcription units showed the separation of the four officially accepted AdV genera (Mayo, 2002), and BAdV-10 was again clearly among the mastadenoviruses. Homologues of the gene named 34K in HAdVs exist in every mastadenovirus studied to date. This gene, adjacent to a seemingly type-specific URF, is also conserved in the E4 region of BAdV-10, which possesses the shortest and simplest such region of all mastadenoviruses.

Interestingly, homologues of the 34K gene occur also in atadenoviruses, even in duplicate (Both, 2002b). Gene duplication in the E4 region of atadenoviruses seems to be common. It has been demonstrated in HAdV-C species that the 34K polypeptide of E4 forms a complex with the 55K protein of the E1B transcription unit and that this complex selectively influences the accumulation of mRNA in the cytoplasm by inhibiting the transport of the cellular and facilitating that of the viral mRNA (Shenk, 2001). It is noteworthy that the 55K protein is the only E1 gene that has homologues in atadenoviruses.

More interestingly, in aviadenoviruses, an additional stretch of DNA accounting for the exceptionally large genome size (42–45 kbp) can be found at the right end, which encompass a large number of genes and ORFs. Only a few of these have been studied in detail (Lehrmann & Cotten, 1999; Chiocca et al., 2002; Wick et al., 2003). In the members of the genus Aviadenovirus, the only gene (a putative dUTPase) that showed homology to E4-like genes could be identified at the left-hand side of the genome (Chiocca et al., 1996; Ojkic et al., 2002). dUTPase-like genes are found in the E4 region of certain mastadenoviruses (Weiss et al., 1997) but not in BAdV-10. In siadenoviruses, two unique small ORFs in a tail-to-tail position occupy the end of the genome before the right ITR. The function of these putative genes is as yet unknown and they might be considered to be the E4 region of siadenoviruses because of their localization.

Our results have shown that variations in the length and maps of the DNA of different BAdV-10 isolates described previously (Benkő et al., 1995) are primarily due to sequence changes occurring in and around the fibre gene. The fibre protein is known to play a crucial role in virion attachment to the cells, and several membrane proteins have been characterized as cellular receptors for certain HAdV types (Arnberg et al., 2002; Segerman et al., 2003). Possible association of the length of the fibre shaft with infectivity (Shayakhmetov & Lieber, 2000; Ambriovic-Ristov et al., 2003) or pathogenicity (Pallister & Prince, 1994) of different AdVs has been demonstrated. The presence of the reiterated sequence between the end of the fibre gene and the E4 region (Fig. 1) implies that actual variations or rearrangements take place in this genome segment. The phylogenetic distance of BAdV-10 from other bovine mastadenoviruses, the poor replication ability, the low genomic GC content and the variable fibre gene indicate that BAdV-10 might
Fig. 3. Comparison of the content of the E3 (a) and E4 (b) regions of different human and animal adenovirus types each representing a different species. Genes marked with the same colour were found to have homology with each other. Grey arrows indicate ORFs that were found only in one virus species. The ORF numbering and gene names are adopted from the original publications and database entries at GenBank. The homologues identified by us but not described in the original publications are in inverted commas. Abbreviation of the hosts are: B, bovine; C, canine; D, duck; F, fowl; Fr, frog; H, human; O, ovine; P, porcine; S, simian; T, turkey; GenBank accession numbers: BAdV-1, AF038868; BAdV-3, AF030154; BAdV-10, AF036092; BAdV-10, AF027599; CAdV-1, U55001; CAdV-2, U77082; SAdV-21, NC_004001; SAdV-25, AF394196.
be an example of an intrageneric host switch. Genomic characterization and comparison of further BAdV-10 isolates will help elucidate the origin of these viruses.

ACKNOWLEDGEMENTS

We wish to gratefully acknowledge the contribution made by Drs Joan A. Smyth and Brian M. Adair (Veterinary Sciences Division, Department of Agriculture and Rural Development for Northern Ireland, Belfast) by transferring isolates of BAdV-10 to our laboratory. The excellent technical assistance provided by Mrs Erika Molnár Szállné is much appreciated. This work was supported by grants provided by the Hungarian Scientific Research Fund (OTKA T043422 and A312) and the Hungarian Prime Minister’s Office (MEH 46761/2003).

REFERENCES


