A naturally occurring human adenovirus type 7 variant with a 1743 bp deletion in the E3 cassette

Juliana C. Marinheiro,1 Tatiana G. dos Santos,1 Joselma Siqueira-Silva,1 Xiaoyan Lu,2 Daniela Carvalho,1 Ataide A. da Camara,3 Eurico Arruda,3 Karla Arruda,3 Dean D. Erdman2 and Charlotte M. Härö1

Correspondence
Charlotte M. Härö
chmharsi@icb.usp.br

1Department of Microbiology, Virology Laboratories, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil
2Centers for Disease Control and Prevention, Gastroenteritis and Respiratory Viruses Laboratory Branch, Atlanta, GA, USA
3Department of Cell and Molecular Biology, School of Medicine, University of São Paulo, Ribeirão Preto, Brazil

Received 4 February 2011
Accepted 13 June 2011

Human adenovirus type 7 (HAdV-7) is an important cause of acute respiratory disease (ARD). Different genomic variants of HAdV-7 have been described, designated 7a–7l. In a previous study to investigate risk factors for ARD and wheezing, nasopharyngeal samples were collected from 90 ill children seeking medical attention in Ribeirão Preto, São Paulo, Brazil. HAdVs were identified in 31 samples and were characterized by serum neutralization and genome restriction analysis. Eleven HAdVs were identified as being HAdV-7, five of which were classified as being of genome type 7p (Gomen). Six other HAdV-7 isolates gave new restriction profiles with all enzymes used and were classified as being a new genomic variant, 7m. These isolates were further characterized by sequencing. The hexon and fiber genes of the 7m variant were nearly identical to the prototype, 7p. However, nucleotide sequences from the E3 cassette revealed a 1743 bp deletion affecting the 16.1K, 19K, 20.1K and 20.5K ORFs.

Human adenoviruses (HAdVs) belong to the family Adenoviridae, genus Mastadenovirus. They are non-enveloped, icosahedral viruses with dsDNA genomes that are classified into seven species (A–G) based on their nucleotide homology, haemagglutination, biochemical and biological properties. There are currently 56 recognized serotypes of HAdV that are associated with a wide range of important human diseases, including respiratory tract illnesses, conjunctivitis, cystitis and gastroenteritis (Benkö & Harrach, 2003; Jones et al., 2007).

Species B HAdV serotypes 3 and 7 are frequently associated with recurring community-wide outbreaks of acute respiratory disease (ARD) with occasional cases of severe disease and death being reported (Erdman et al., 2002; Gray et al., 2007; Kajon et al., 1990; Kim et al., 2000; Moura et al., 2007a, b). These types, along with HAdV-14 and HAdV-21 of species B and HAdV-4 of species E, are also responsible for outbreaks of ARD at military-recruit training centres (Metzgar et al., 2005, 2007). In addition to classical serotyping, genome restriction analysis and targeted gene sequencing have been used extensively to characterize the molecular epidemiology of these viruses (Kajon & Erdman, 2007). Currently 13 genomic variants of HAdV-7 are recognized, designated 7a–7l. The prototype, 7p (Gomen), was the first strain of serotype 7 identified.

Early region 3 (E3) of the HAdV genome encodes several proteins, many of which have been implicated in modulating the host immune response. The size and composition of the E3 region varies between HAdV species. Some of the E3 genes are present in all species (10.4K, 14.4K and 14.7K) or in most of them (12.5K and 19K). Other genes appear to be species specific, such as 20.1K and 20.5K of species B HAdV and 6.7K of species C HAdV (Burgert & Blusch, 2000). Most, if not all, E3 genes are not necessary for viral replication in cell culture; however, they have been thought to be essential for virus survival in vivo (Burgert & Blusch, 2000; Horwitz, 2001, 2004).

In a previous study of the aetiology of ARD and wheezing in children <12 years of age who sought medical care in Ribeirão Preto city in São Paulo State (Câmara et al., 2004), HAdVs were detected in 31 of 90 patients (34.4 %). Respiratory specimens were collected according to the ethical standards of the Committee on Ethics on Human A supplementary table of primer sequences is available with the online version of this paper.

The GenBank/EMBL/DDBJ accession number for the nucleotide sequence of the truncated and fused 16.1K/20.5K ORF of the variant HAdV-7m is HQ542295.
Experimentation of the Hospital, protocol HCRP no. 5.672/98. The present study was approved by the Institute of Biomedical Sciences of the USP (protocol no. 739/06).

Genome restriction analysis identified 11 isolates as being HAdV-7. These could be further differentiated into two distinct genome types: five showed a restriction profile similar to HAdV-7p and six showed a novel, previously undescribed, restriction profile which we designated 7m (Azar et al., 1998; Choi et al., 2006; Erdman et al., 2002; Kajon et al., 1996; Mistchenko et al., 1998; Wadell et al., 1985). BamHI and HindIII electrophoretic profiles of this new variant are presented in Fig. 1. Genome-restriction maps localized the predicted changes to the E3 cassette, revealing a large deletion of approximately 1700 nt. In this study, we further characterize this virus by targeted sequencing and speculate on its potential for enhanced pathogenesis.

HAdV hexon hypervariable regions 1–6 (Lu & Erdman, 2006), the complete fiber gene (Xu et al., 2000) and almost the complete E3 region were sequenced. The 7m region was amplified in four PCRs: the first reaction amplified a 3069 nt section comprising the ORFs 12.5K, 16.1K, 19K, 20.1K, 20.5K and 7.7K; this was followed by two nested-PCRs to amplify, separately, ORFs 12.5K–19K and ORFs 19K–20.5K, and a third reaction to amplify ORF 7.7K (Kajon et al., 2005). Primer sequences are provided in Supplementary Table S1 (available in JGV Online).

Sequencing reactions were performed in both directions by using a ABI Prism BigDye Terminator v3.1 Cycle Sequencing kit on an Applied Biosystems 3130xl Genetic Analyzer (Applied Biosystems). Sequences were assembled and edited using SEQUENCER version 3.1.1 (Gene Codes). The DNA sequence and the predicted amino acid sequences were aligned with published sequences of HAdV-7p and -3p (GenBank accession nos AY594255 and AY599834, respectively) by using CLUSTAL W, implemented in the BioEdit version 7.0.9.0 software.

PCR amplicons obtained for the hexon hypervariable regions 1–6 and full fiber genes of all HAdV-7 isolates were of the expected size. Sequences obtained for these regions were identical with each other and were identical to the sequence published for the reference HAdV-7p strain Gomen (GenBank accession no. AY594255). For the E3 cassette, PCR amplicons were of the expected size (~3069 bp) for the 7p genome types, but were substantially smaller (~1300 bp) for the 7m variants (Fig. 2a).

Sequences obtained for this region identified a 1743 bp deletion between nt 27 909 and 29 652 of the reference HAdV-7p genome, encompassing part of ORF 16.1K, all of ORFs 19K and 20.1K, and part of ORF 20.5K. The first 252 nt of ORF 16.1K and the terminal 100 nt of ORF 20.5K were retained, resulting in a new ORF of 352 nt (GenBank accession no. HQ542295) (Fig. 2b). The predicted translation of this ORF yields a protein composed of the 84 N-terminal amino acids of protein 16.1K extended by another 17 aa, and distinct from those known for the 20.5K protein. (Fig. 2c). BLAST analysis of this new ORF only matched the N-terminal region of protein 16.1K. The C-terminal 17 aa did not match with any known protein. Analysis of this region in other reading frames did not reveal any putative ORF. Sequences obtained from HAdV-7m DNA extracted directly from the clinical specimens were identical to the isolates, indicating that this deletion did not result from passage of the virus in cell culture. The same deletion was present in all six HAdV-7m-positive samples. The four ORFs to the right of the deletion (7.7K, 10.4K, 14.5K and 14.7K) were preserved in the HAdV-7m variants and were homologous to those of HAdV-7p.

The HAdV-7m variant was recovered from six epidemiologically unrelated immune-competent children <10 years of age who presented with wheezing. The samples were collected between September 1999 and March 2000, suggesting sustained community-wide circulation. Table 1 presents the limited demographic, clinical and laboratory data available from children infected with HAdV genome types 7p, 7m and other HAdV serotypes. According to these data, the use of accessory respiratory muscles was required by patients infected with HAdV-7m, suggesting severe respiratory distress. Unfortunately, given the few cases identified in this study, possible differences in pathogenicity among these viruses cannot be determined. As with the other patients in this study, these children attended the public University Hospital and were from low-income families with life activities restricted to the city or the neighbouring area. The predominance of these viruses in Ribeirão Preto, a moderate-size city of 977 446
inhabitants, located 330 km from the state capital, São Paulo city, may reflect their temporal or geographical restriction to this area. Whether HAdV-7m will persist or spread beyond its current geographical location remains to be seen. The study was concluded in 2000 and no other samples were collected after that.

HAdV-7 is recognized as being among the most pathogenic serotypes, causing severe and sometimes fatal respiratory illnesses (Gray et al., 2007). Berge et al. (1955) first reported the isolation of HAdV-7p in association with an outbreak of ARD among new military recruits. Designated strain Gomen, the genome restriction map of this virus was first published in 1980. The virus was later identified as HAdV-7p (GenBank accession no. HQ542295).

**Fig. 2.** (a) PCR amplification products using primers directed to part of the E3 region of human adenoviruses. 7p Lanes, adenovirus prototype Gomen (HAdV-7p); 7m lanes, adenovirus variant HAdV-7m; C−, negative control; M, molecular mass marker (100 bp DNA ladder; Biotools). Electrophoresis was carried out on an ethidium bromide stained, 1.0% agarose gel. (b) Nucleotide sequence of the truncated and fused 16.1K/20.5K ORF of the variant HAdV-7m (GenBank accession no. HQ542295). (c) Alignment of the amino acid sequence of the 16.1K protein of the prototype strain HAdV-7p with the predicted sequence of the new variant HAdV-7m. Transmembrane domains are underlined. There is a putative sorting motif in the cytoplasmic tail of HAdV-7p (aa 132–136, YLVIL) and a putative prenylation motif near the cytoplasmic terminus of the HAdV-7m (aa 99–101, CLLLL).
first characterized by Wadell & Varsanyi (1978) during an investigation of outbreaks of severe respiratory illness in Europe in the 1970s. This study revealed a new genome type, HAdV-7b, which has subsequently spread worldwide. The predominant type of HAdV-7 circulating throughout Brazil from the early 1980s through to the mid-1990s. It was detected in São Paulo city and Rio de Janeiro (Gomes et al., 1989; Hársi et al., 1995; Kajon et al., 1999; Moraes et al., 1997) and it was itself replaced, in 1995, by HAdV-7h, as documented from respiratory samples collected in São Paulo city (Moura et al., 2007a; Tanaka et al., 2000). Interestingly, neither HAdV-7b nor -7h were identified in our study. Instead, all of the patients with documented HAdV-7 were infected with HAdV-7p, which has only been sporadically detected since its original discovery in 1955 (Bailey & Richmond, 1986; Erdman et al., 2004). The HAdV-7m variant retains the first 252 nt of the 16.1K ORF, corresponding to 57% of the protein (Fig. 2b, c). The N-terminal sequence of the 16.1K protein is preserved, along with the signal sequence and part of the hydrophilic luminal region. The C-terminal 62 aa were lost and replaced by a new sequence of 17 aa. Analyses using the TMpred program (Hofmann & Stoffel, 1993) revealed that this new sequence is hydrophobic and may be a new transmembrane domain with a putative prenylation motif at the terminus (CLLL), as revealed by PSORT II (Horton & Nakai 1997). This novel, modified 16.1K protein is probably anchored in the membrane via its C-terminal region but has lost its sorting function. The C-terminal region of the native type 16.1K protein has a distant similarity to the 6.7K protein of species C HAdVs, which is a transmembrane protein that localizes to the endoplasmic reticulum (Hawkins et al., 1995). It collaborates with the receptor internalization and degradation (RID) complex (E310.4K/14.5K) in the degradation of tumour necrosis factor-related apoptosis-inducing ligand receptor 2 (TRAIL-R2) and the inhibition of apoptosis (Benedict, et al., 2001). Although some similarity between 16.1K and 6.7K proteins is mentioned by Hawkins et al. (1995), the functions of 16.1K in HAdV species B infection remain unknown.

The deletion observed in the HAdV-7m variant resulted in the complete loss of ORFs 19K, 20.1K and 20.5K. The 19K protein is an integral membrane glycoprotein that binds MHC class I antigens in the endoplasmic reticulum, thus blocking CTL recognition (Burgert & Blusch, 2000; Horwitz, 2001, 2004). McSharry et al. (2008) describe how this protein also protects against NK cell attack by down-regulating MHC class I chain-related proteins A and B (MICA and MICB) ligands. With the exception of species A and F, all HAdV species express a 19K-like protein (Burgert & Blusch, 2000). Studies of E3 19K function have been conducted with species C adenoviruses by using animal

![Table 1. Comparative demographic, clinical and laboratory data from children with HAdV infection](image)

The values shown in the table are the number of samples with the percentage of the total number shown in parentheses.

<table>
<thead>
<tr>
<th>Patient characteristic</th>
<th>HAdV-7p (n=5)</th>
<th>HAdV-7m (n=6)</th>
<th>HAdV-other (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age &lt;2 years</td>
<td>5 (100)</td>
<td>1 (33.3)</td>
<td>13 (65)</td>
</tr>
<tr>
<td>Age ≥2 years</td>
<td></td>
<td>4 (66.6)</td>
<td>7 (35)</td>
</tr>
<tr>
<td>Male sex</td>
<td>3 (60)</td>
<td>4 (66.6)</td>
<td>12 (60)</td>
</tr>
<tr>
<td>Hospitalized</td>
<td>1 (20)</td>
<td>2 (33)</td>
<td>10 (50)</td>
</tr>
<tr>
<td>Use of accessory respira</td>
<td>1 (20)</td>
<td>5 (83)</td>
<td>5 (25)</td>
</tr>
<tr>
<td>tory muscles*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-detection with other</td>
<td>1 (20)</td>
<td>3 (50)</td>
<td>9 (45)</td>
</tr>
<tr>
<td>respiratory viruses†</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*When breathing requires extra effort.
†Respiratory syncytial virus, rhinovirus and coronavirus.

(aa 19–109). The C-terminal sequence has a transmembrane domain (aa 110–126) and a putative sorting motif in the cytoplasmic tail (aa 132–136, YLVIL) (Hawkins et al., 1995; Windheim et al., 2004). The HAdV-7m variant retains the first 252 nt of the 16.1K ORF, corresponding to 57% of the protein (Fig. 2b, c). The N-terminal sequence of the 16.1K protein is preserved, along with the signal sequence and part of the hydrophilic luminal region. The C-terminal 62 aa were lost and replaced by a new sequence of 17 aa. Analyses using the TMpred program (Hofmann & Stoffel, 1993) revealed that this new sequence is hydrophobic and may be a new transmembrane domain with a putative prenylation motif at the terminus (CLLL), as revealed by PSORT II (Horton & Nakai 1997). This novel, modified 16.1K protein is probably anchored in the membrane via its C-terminal region but has lost its sorting function. The C-terminal region of the native type 16.1K protein has a distant similarity to the 6.7K protein of species C HAdVs, which is a transmembrane protein that localizes to the endoplasmic reticulum (Hawkins et al., 1995). It collaborates with the receptor internalization and degradation (RID) complex (E310.4K/14.5K) in the degradation of tumour necrosis factor-related apoptosis-inducing ligand receptor 2 (TRAIL-R2) and the inhibition of apoptosis (Benedict, et al., 2001). Although some similarity between 16.1K and 6.7K proteins is mentioned by Hawkins et al. (1995), the functions of 16.1K in HAdV species B infection remain unknown.

The deletion observed in the HAdV-7m variant resulted in the complete loss of ORFs 19K, 20.1K and 20.5K. The 19K protein is an integral membrane glycoprotein that binds MHC class I antigens in the endoplasmic reticulum, thus blocking CTL recognition (Burgert & Blusch, 2000; Horwitz, 2001, 2004). McSharry et al. (2008) describe how this protein also protects against NK cell attack by down-regulating MHC class I chain-related proteins A and B (MICA and MICB) ligands. With the exception of species A and F, all HAdV species express a 19K-like protein (Burgert & Blusch, 2000). Studies of E3 19K function have been conducted with species C adenoviruses by using animal
models, and showed that infection with HAdV mutants with E3 19K deletions produced increased pathogenesis and an enhanced inflammatory response in cotton rats, while deletion mutants of other genes (ORFs 11.6K and 12.5K) produced pneumonia that was indistinguishable from that caused by wild-type virus (Ginsberg et al., 1989, 1991; Prince et al., 1993).

The 20.1K and 20.5K ORFs are specific to HAdVs of species B. The product of the 20.5K ORF was shown to be expressed in two glycosylated forms, 22K and 36K. The function of these proteins remains unknown, but the fact that they are encoded only by HAdVs of species B, which cause the most severe respiratory infections, suggests that these unique proteins could play a role in the virulence of these viruses (Hawkins & Wold, 1995; Hawkins et al., 1995).

The proteins expressed by the E3 cassette are not necessary for the replication of adenovirus in vitro, but apparently are important during natural infection, as they have several roles in the modulation of the immune response. Naturally occurring E3 deletion mutants are rare, and deletion of these genes could have unpredictable effects on pathogenesis and viral clearance (Kajon et al., 2005). Recently Su et al. (2011) reported the discovery of another HAdV-7 E3 deletion mutant, isolated from a child with acute respiratory infection in China, that had preserved only the first two proteins of the E3 region (12.5K and 16.1K). These proteins are conserved among almost all adenovirus species. Surprisingly, this new virus also lost the highly conserved right segment of the E3 region which encodes the RID complex. Further studies on the function of the E3 proteins of species B HAdVs, in particular the 16.1K protein, will be important for gaining further insights into the pathogenesis of these viruses.

Acknowledgements

We thank Dr Teresa C. T. Peret for helpful discussions. This study was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, process 06/56482-0) and the Respiratory and Enteric Virus Branch, CDC, Atlanta, USA. Juliana Cristina Marinheiro received a FAPESP fellowship (03/02055-5).

References


