Open reading frames 1 and 2 of adenovirus region E4 are conserved between human serotypes 2 and 5

Ian Dix and Keith N. Leppard*

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, U.K.

The E4 region of human adenovirus type 2 is predicted to encode seven proteins as judged from its nucleotide sequence and the pattern of differential splicing of its transcript. Two of the open reading frames (ORFs), ORF1 and ORF2, had been identified as being disrupted in the recently published sequence of the related serotype 5 virus. These ORFs were resequenced and found to be intact in the wt300 strain of adenovirus type 5.

The molecular genetics of human adenovirus type 5 (Ad5) have been studied extensively. This work has been underpinned by sequence data derived from the Ad5 genome supplemented by data from the closely related Ad2, for which a complete sequence has been available for a number of years (Roberts et al., 1984). The completion of a composite Ad5 sequence (Chroboczek et al., 1992; GenBank accession no. M73260) has permitted those studies of Ad5 involving previously unsequenced regions to proceed on a more systematic basis.

The E4 region of Ad5 was one of those for which, until recently, only a partial sequence was available. The Ad2 sequence, together with RNA mapping studies in Ad2 (Virtanen et al., 1984; Tigges & Raskas, 1984), predicted the expression of seven distinct proteins from the Ad2 E4 region and, by analogy, from the Ad5 E4 region also. Three of these proteins, the products of open reading frames (ORFs) 3, 6 and 6/7, have been detected in infected cells (Downey et al., 1983; Sarnow et al., 1984; Cutt et al., 1987). The remainder, the products of ORF1, ORF2, ORF3/4 and ORF4, appear to perform no essential function during infections of cultured HeLa cells because mutants unable to express these proteins grow with virtually wild-type kinetics (Halbert et al., 1985; Huang & Hearing, 1989).

Comparison of the published Ad2 and Ad5 E4 region sequences showed two nucleotide substitutions and four deletions in Ad5 ORF1, and four substitutions and one insertion in Ad5 ORF2; these cause frameshifts in both ORF1 and ORF2, and two missense changes in ORF2. However, the high ratio of frameshift to point mutations in these putative coding regions, together with the high degree of sequence conservation as compared with adjacent non-coding regions, suggested that the frameshifts might be sequencing artefacts. To determine whether Ad5 could encode E4 ORF1 and ORF2 products, the region of Ad5 DNA between nucleotides 34790 and 35615 was resequenced.

Relevant pieces of the Ad5 strain wt300 KpnI G genomic DNA fragment were subcloned into plasmids and then M13 vectors mpl8 and mpl9. Sequencing was carried out by the dideoxynucleotide chain termination method using the Sequenase II system (U.S. Biochemicals). As shown in Fig. 1, in each of the four regions where the published Ad5 sequence contained frameshifts in ORF1 or ORF2, the sequence obtained showed the ORFs to be open as in Ad2, with the exception of a silent base change at one of the insertion positions. All except one of the other expected differences between Ad2 and Ad5, including all those in the non-coding regions sequenced, were confirmed by our data, discounting the possibility that the DNA sequenced was of Ad2 origin. The exception was the ORF2 missense mutation at Ad5 nucleotide 34860; our data indicate that strain wt300 ORF2 encodes the same amino acid as that of Ad2 at this position. These changes to the database sequence of Ad5 are summarized in Table 1.

ORF1 and ORF2 of the Ad5 E4 region were originally sequenced by Steenbergh & Sussenbach (1979) using the chemical degradation technique; this sequence was used by Chroboczek et al. (1992) when compiling the complete Ad5 sequence. Our data suggest a revised length of 35938 bp for the Ad5 strain wt300 genome with potentially functional E4 ORF1 and ORF2 regions encoding products identical to, or differing in only one amino acid from, their Ad2 counterparts.

The nucleotide sequence data reported here will appear in the DDBJ, EMBL and GenBank databases under accession number D12587.
Fig. 1. Sequence data for regions of difference between published Ad5 and Ad2 sequences. The regions around (a) nucleotides 35523/22 and 35511/10, (b) nucleotides 35321/20 and (c) nucleotide 34936 are shown. The sequence of the rightward strand deduced from these data is shown to the left of each panel, with the differences from the published Ad5 sequence indicated by arrows.

Table 1. Summary of changes to the published Ad5 sequence

<table>
<thead>
<tr>
<th>Published Ad5 nucleotide position</th>
<th>Published Ad5 sequence*</th>
<th>Revised Ad5 sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>34860</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>34936</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>35320/21</td>
<td>–</td>
<td>CA</td>
</tr>
<tr>
<td>35510/11</td>
<td>–</td>
<td>C</td>
</tr>
<tr>
<td>35522/23</td>
<td>–</td>
<td>A</td>
</tr>
</tbody>
</table>

* Sequence changes indicated are to the leftward strand.

The conservation of ORF1 and ORF2 in two, albeit closely related, serotypes supports the argument that they encode proteins having some function in Ad infections. Since they are not needed for successful infection of HeLa cells, they may be important only in infection of certain cell types or alternatively may be necessary only for infections of the whole organism. Very few of the potential coding regions of Ad5 lack an assigned function; furthering the molecular understanding of this virus by determining those remaining functions is an interesting and important objective for the future.

Ian Dix is the recipient of an SERC research studentship. This work was supported in part by a grant from the MRC.

References


(Received 24 June 1992; Accepted 27 July 1992)