Completion of the Sequence of Bluetongue Virus Serotype 10 by the Characterization of a Structural Protein, VP6, and a Non-structural Protein, NS2

By A. FUKUSHO, 1 Y. YU, 1 S. YAMAGUCHI 1 AND P. ROY 1, 2*

1 University of Alabama at Birmingham, School of Public Health, Department of Environmental Health Sciences, Birmingham, Alabama 35294, U.S.A. and 2NERC Institute of Virology, Mansfield Road, Oxford OX1 3SR, U.K.

(Accepted 21 March 1989)

SUMMARY

The sequence of cDNA clones representing the entire genome of bluetongue virus serotype 10 (BTV-10) has been completed by the analysis of data obtained for the S8 and S9 segments. Each DNA clone has been sequenced completely and the deduced amino acid sequences have been analysed. The sequences of the S8 and S9 gene products as well as another two previously published small gene products (S7 and S10) have been compared with the corresponding size gene products of reovirus type 1. The data do not indicate a relationship between the small proteins of these two viruses except some distant homologies between the BTV VP7 protein and the σ1 protein of reovirus. The characteristics of all the BTV-10 genome segments, the deduced primary gene products and their possible functions are summarized.

INTRODUCTION

Bluetongue virus (BTV) is an arbovirus that is an economically important pathogen of sheep and cattle in many parts of the world including the United States, Africa and Australasia. The virus induces a wide range of clinical and subclinical symptoms. In sheep, infection can be severe but in cattle it is milder. However, infection in cattle during pregnancy can result in abortion, teratogenic defects, or birth of persistently infected animals. Infected vertebrates serve as reservoirs for the maintenance and spread of BTV by Culicoides vectors.

BTV is a member of the Orbivirus genus of the Reoviridae family. The genome of BTV (19218 bp, $M_r$ 13 x 10^6) consists of 10 dsRNA segments (L1 to 3, M4 to 6, S7 to 10) that range in $M_r$ from 2.7 x 10^6 to 0.5 x 10^6. The viral RNA is enclosed by a double capsid protein shell, the outer shell consists of two polypeptides (VP2, VP5), the inner core contains two major (VP3, VP7) and three minor (VP1, VP4, VP6) polypeptides (Verwoerd et al., 1972). Each BTV segment appears to code for at least a single viral polypeptide (Mertens et al., 1984; Huismans, 1979). In addition to the structural proteins, three non-structural polypeptides (NS1, NS2 and NS3; Mertens et al., 1984; van Dijk & Huismans, 1988) have been identified in virus-infected cells. The functions of the non-structural proteins have yet to be established.

We have been involved in developing rational strategies for the identification, control and prevention of BTV infections. Our approach has been to base these strategies on an understanding of the molecular, biochemical and genetic features of the virus. Towards this end we have sequenced cDNA clones of the genomic RNA species. Eight of the 10 dsRNA segments of BTV-10 have already been cloned and sequenced (Purdy et al., 1985, 1986; Ghiasi et al., 1985; Lee & Roy, 1986, 1987; Yu et al., 1987, 1988; Roy et al., 1988). In this paper we complete the sequence analyses by describing the molecular cloning and nucleotide sequence of the remaining two small RNA segments, S8 and S9. The S9 RNA species codes for the inner core protein VP6; the S8 RNA codes for the non-structural protein NS2. The deduced sequences of the S8 and S9 gene products have been analysed. In order to identify any evolutionary relationship between
BTV and reovirus, the prototype virus of the Reoviridae family, amino acid sequence comparisons of these small viral gene products and that of the S7 and S10 products (VP7 and NS3, respectively) have been made. The results of these analyses are reported together with a summary of the complete sequence data of the RNA and proteins of BTV-10.

METHODS

Viruses and cells. United States prototype BTV-10 (CA-8) was plaque-cloned using monolayers of BHK-21 cells. The viral dsRNA was purified (Ritter & Roy, 1988) and the 10 individual RNA segments were separated and isolated as described previously (Purdy et al., 1984).

DNA cloning of the BTV-10 RNA segments. Polyadenylation of BTV-10 dsRNA and synthesis of cDNA copies using an oligo(dT)$_{12-18}$ primer were undertaken as described elsewhere (Purdy et al., 1984). The RNA templates were removed by treatment with 0.5 m-KOH and dsDNA was generated by self-annealing. The products were repaired using the Klenow large fragment of DNA polymerase, followed by 3' tailing with dC and annealing to PstI-cut, dG-tailed, pBR322 plasmid DNA. After transformation, clones containing the viral sequences were recovered and screened by colony hybridization (Grunstein & Hogness, 1975). HinfI restriction patterns of the recombinant plasmids were compared as described previously (Purdy et al., 1984).

RNA gel electrophoresis, blotting and hybridization. Purified BTV-10 RNA was resolved on an agarose gel, blotted onto a GeneScreen membrane (New England Nuclear) and hybridized to nick-translated cloned DNA (Purdy et al., 1984).

Sequencing of BTV DNA clones. The sequences of both DNA clones were mainly determined by the dideoxynucleotide sequencing method (Sanger et al., 1977). For sequencing, appropriate DNA fragments generated by digestion of the cloned DNA with various restriction enzymes (Messing, 1983) were subcloned into M13 vectors (Yanisch-Perron et al., 1985). In addition, analyses were undertaken on a sequential series of overlapping clones produced using the methods described by Dale et al. (1985). However, the terminal sequences of the clones were determined by the method of Maxam & Gilbert (1980) confirming that the clones were full-length.

RESULTS

Molecular cloning of full-length ds cDNAs of BTV RNAs

The strategy used to prepare cDNA from BTV-10 S8 and S9 RNA was similar to that described previously (Purdy et al., 1984, 1985). For the S8 and S9 RNA species, the estimated sizes of the derived cDNA copies (approx. 1000 to 1200 bp) suggested that they were complete, or nearly complete transcripts of the viral RNA species. The cDNA duplexes were each cloned into the PstI site of pBR322, after tailing with dC. Approximately 30 to 50 clones representing each gene were identified by colony hybridization (Grunstein & Hogness, 1975). The HinfI restriction patterns of the clones were determined and two representative clones (A30 and A88) were selected for further analyses. To determine their identity, each clone was individually subjected to Northern blot hybridization using total BTV-10 RNA. The position of each viral RNA segment in the agarose gel was determined by ethidium bromide staining (Purdy et al., 1984). By this procedure it was shown that clone A30 represented S8 RNA and clone A88 represented S9 RNA.

Nucleotide sequences of two of the S RNA segments of BTV-10

To determine the terminal sequences, each DNA clone was excised from pBR322 by PstI digestion and the end sequences were analysed using the Maxam & Gilbert technique (1980). The sequences were compared with the known terminal sequences of the two viral RNA segments (Rao et al., 1983), thereby confirming that clone A88 represented BTV-10 S9 RNA, and that clone A30 represented BTV-10 S8 RNA. Both clones were found to contain the characteristic terminal sequences of BTV-10 RNA segments (see Table 1; GTTAAA.... at their 5' ends and .... CACTTAC at their 3' termini) indicating that each clone was full-length. The sequences of clones A30 (S8) and A88 (S9) were determined by the dideoxynucleotide method (Fig. 1; Sanger et al., 1977). The complete nucleotide sequences of the cDNAs of S8 and S9 in their coding (mRNA) senses are presented in Fig. 2 with the predicted amino acid sequences of the respective single open reading frames shown above the sequences. The two small RNA segments of BTV-10 are 1124 bp (S8) and 1046 bp (S9) long. From these data the $M_r$
<table>
<thead>
<tr>
<th>RNA</th>
<th>5' End sequence</th>
<th>5' Non-coding length</th>
<th>Coding region length</th>
<th>3' Non-coding length</th>
<th>3' End sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>GUUAAAAUGCAUG</td>
<td>11</td>
<td>3906</td>
<td>37</td>
<td>UGAGACCACGCGGAGCCGACGCGCCGCAUUACACUAC</td>
</tr>
<tr>
<td>L2</td>
<td>GUUAAAAGUGUUCUACAG</td>
<td>19</td>
<td>2868</td>
<td>39</td>
<td>UAGGUCUCUGUGACAUGGACCGUAUGCUUACACUAC</td>
</tr>
<tr>
<td>L3</td>
<td>GUUAAUUUCCUGAGCAG</td>
<td>17</td>
<td>2703</td>
<td>52</td>
<td>UAGAUGUCCUGACCGGAUCACUUGACAGCAGGC-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6AAACACACUACUAC</td>
</tr>
<tr>
<td>M4</td>
<td>GUUAAAAACAG</td>
<td>8</td>
<td>1902</td>
<td>41</td>
<td>UAAUGCGUGACUGCUAGGAGGGGCGUUAACACUAC</td>
</tr>
<tr>
<td>M5</td>
<td>GUUAAAAAGUUCUUCUACUCGAGAAGAUG</td>
<td>29</td>
<td>1578</td>
<td>31</td>
<td>UGAAACGCAGGGGGGAGGACCUUACUAC</td>
</tr>
<tr>
<td>M6</td>
<td>GUUAAAAGUUCUCUAGUGUGGACACGCAAACACUG</td>
<td>34</td>
<td>1656</td>
<td>79</td>
<td>UAGGUACUGAUUUUUAGUUEUUUUCUUCUUAUCAUUACU</td>
</tr>
<tr>
<td>S7</td>
<td>GUUAAAUGUUCUAGACAGAUG</td>
<td>17</td>
<td>1047</td>
<td>92</td>
<td>UAGUCACCUUGCACGGGUUGUGGUACAGCGUGUGUGG-</td>
</tr>
<tr>
<td>S8</td>
<td>GUUAAAAUGUUCUAGACAGAUG</td>
<td>19</td>
<td>1071</td>
<td>34</td>
<td>UGAGCAGCAUGGUGGGGGGUAUUACACUAC</td>
</tr>
<tr>
<td>S9</td>
<td>GUUAAAAUGUUCUAGACAGAUG</td>
<td>15</td>
<td>984</td>
<td>47</td>
<td>UAAAAGGGUGCCCGGUAUCUCUUGACGUAGGCGGU-</td>
</tr>
<tr>
<td>S10</td>
<td>GUUAAAAGUUCUAGACAGAUG</td>
<td>19</td>
<td>687</td>
<td>116</td>
<td>UGAGCAGACGUAGGUGUAGUGGUGGACCA-</td>
</tr>
</tbody>
</table>

* Conserved 3' and 5' end sequences are underlined. The first putative translation initiation codon in the mRNA-sense strand of the viral cDNA and termination codons preceding the 3' non-coding region have double underlines.
Fig. 1. The strategy used to determine the sequence of the cDNA clones of two small RNA segments of BTV-10. Clones A30 and A88 represent the RNA segments S8 and S9, respectively. The restriction enzymes used to generate the restriction fragments are shown on the top of the figure. The distance and directions in which individual strands were sequenced are shown by the solid arrows.

values of the dsRNA were calculated to be $7.6 \times 10^5$ for S8 and $7.1 \times 10^5$ for S9, in good agreement with previous estimates.

Characterization of the predicted proteins encoded by two of the small RNA segments of BTV-10

Analyses of the nucleotide sequence of each DNA segment indicated the existence of only one long open reading frame (Fig. 2) on one strand of the viral cDNA (mRNA). Potential translation initiation codons were identified at residues 20 to 22 for S8 and 16 to 18 for S9. Although no analyses were undertaken to determine whether they represented the actual translation initiation sites, each has favourable flanking sequences to initiate protein synthesis according to Kozak's model (Kozak, 1981, 1983). Translation termination codons were located at residues 1091 to 1093 (UGA, S8) and 1000 to 1002 (UAA, S9). The estimated $M_r$ of the 357 amino acid S8 gene product (NS2) is 40999 and that of the 328 amino acid S9 gene product (VP6) is 35750 (Table 2).

The putative NS2 protein encoded by RNA S8 contains many charged residues that contribute to the overall hydrophilic character of the protein as illustrated in Fig. 3(a). The protein has an overall net negative charge of $-3$ at neutral pH values. The amino terminus, and to a lesser extent the carboxy terminus, contains stretches of hydrophobic residues. The middle region of the NS2 molecule (i.e. amino acid residues 125 to 286) contains many charged residues. The NS2 protein has been shown to be a phosphoprotein capable of binding BTV ssRNA (Huismans et al., 1987a). Analyses of the potential secondary structure of the NS2 protein indicate that it is probably abundant in $\beta$ turns (data not shown).

The inner core VP6 protein, encoded by RNA S9, has a net positive charge of $+11$ at neutral pH values. The protein has at least two long stretches of hydrophilic amino acids (residues 30 to 80 and 175 to 230, see Fig. 3(b)). Northern hybridization data involving BTV-10 cDNAs and viral RNA preparations representing 20 of the 24 BTV serotypes have indicated that RNA S9 is highly conserved (Ritter & Roy, 1988). These results suggest that the VP6 polypeptides of all the BTV serotypes are structurally similar.

Comparisons of S RNA gene products of BTV-10 and reovirus subtype 1: evidence for homology between BTV VP7 protein and reovirus $\sigma 1$ polypeptide

The derived amino acid sequences of the BTV-10 S8- and S9-encoded polypeptides and those of the previously sequenced S7- and S10-encoded polypeptides were compared with the amino
Completion of the BTV-10 sequence

Fig. 2. The nucleotide sequences of the cloned S8 and S9 genes of BTV-10. For S8 (a), the open reading frame begins at residues 20 to 22 and terminates with a UGA at residues 1091 to 1093. For S9 (b), the open reading frame begins at residues 16 to 18 and terminates with a UAA at residues 1000 to 1002. The single-letter amino acid codes are shown above the respective DNA codons.

acid sequences of reovirus proteins available in GenBank using the FASTP computer alignment program of Lipman & Pearson (1985). No sequence homology was detected between the BTV RNA S gene products and those of reovirus except the S7 gene product of BTV and the σ1 protein of reovirus type 1 (Cashdollor et al., 1985) which indicated some homology. Although the amino acid sequence similarity was not extensive, it was considered significant since similar alignment comparisons of the other S RNA-encoded polypeptides with those of reovirus did not identify similar levels of protein homology. As shown in Fig. 4, with a minimum number of gaps, 75 amino acid residues out of 340 possible matches are homologous between the two polypeptides (a score of 22% homology). However, when possible conservative mutational
Table 2. Amino acid compositions of the 10 primary gene products of BTV-10*

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>VP1 core</th>
<th>n†</th>
<th>VP2 shell</th>
<th>n</th>
<th>VP3 core</th>
<th>n</th>
<th>VP4 core</th>
<th>n</th>
<th>VP5 shell</th>
<th>n</th>
<th>NS1</th>
<th>n</th>
<th>VP7 core</th>
<th>n</th>
<th>NS2</th>
<th>n</th>
<th>VP6 core</th>
<th>n</th>
<th>NS3</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine (A)</td>
<td>76 58</td>
<td>53 55</td>
<td>63 70</td>
<td>40 61</td>
<td>49 93</td>
<td>41 74</td>
<td>40 115</td>
<td>21 59</td>
<td>31 95</td>
<td>20 87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arginine (R)</td>
<td>91 70</td>
<td>65 68</td>
<td>69 77</td>
<td>54 83</td>
<td>28 53</td>
<td>43 78</td>
<td>24 69</td>
<td>27 76</td>
<td>32 98</td>
<td>11 48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspartic (D)</td>
<td>77 59</td>
<td>69 72</td>
<td>59 65</td>
<td>48 73</td>
<td>26 49</td>
<td>32 58</td>
<td>13 37</td>
<td>29 81</td>
<td>18 55</td>
<td>11 48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asparagine (N)</td>
<td>53 41</td>
<td>41 43</td>
<td>41 46</td>
<td>24 37</td>
<td>16 30</td>
<td>20 36</td>
<td>18 52</td>
<td>12 34</td>
<td>9 27</td>
<td>8 35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cysteine (C)</td>
<td>12 9</td>
<td>16 17</td>
<td>5 6</td>
<td>6 9</td>
<td>3 6</td>
<td>16 29</td>
<td>3 9</td>
<td>7 20</td>
<td>1 3</td>
<td>2 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutamic (E)</td>
<td>84 65</td>
<td>60 63</td>
<td>50 55</td>
<td>48 73</td>
<td>55 105</td>
<td>39 71</td>
<td>14 40</td>
<td>34 95</td>
<td>36 110</td>
<td>15 66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutamine (Q)</td>
<td>42 32</td>
<td>37 39</td>
<td>42 47</td>
<td>8 12</td>
<td>19 36</td>
<td>26 47</td>
<td>20 57</td>
<td>16 45</td>
<td>9 27</td>
<td>10 44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycine (G)</td>
<td>69 53</td>
<td>49 51</td>
<td>43 48</td>
<td>33 50</td>
<td>33 63</td>
<td>32 58</td>
<td>23 66</td>
<td>20 56</td>
<td>40 122</td>
<td>6 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histidine (H)</td>
<td>21 16</td>
<td>29 30</td>
<td>16 18</td>
<td>27 41</td>
<td>17 32</td>
<td>14 25</td>
<td>6 17</td>
<td>4 11</td>
<td>6 18</td>
<td>3 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isoleucine (I)</td>
<td>95 73</td>
<td>70 73</td>
<td>65 72</td>
<td>38 58</td>
<td>48 91</td>
<td>39 71</td>
<td>24 66</td>
<td>19 53</td>
<td>19 58</td>
<td>11 48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leucine (L)</td>
<td>111 85</td>
<td>90 94</td>
<td>82 92</td>
<td>63 96</td>
<td>43 82</td>
<td>42 76</td>
<td>28 80</td>
<td>21 59</td>
<td>16 49</td>
<td>23 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lysine (K)</td>
<td>87 67</td>
<td>61 64</td>
<td>27 30</td>
<td>36 55</td>
<td>40 76</td>
<td>23 42</td>
<td>1 3</td>
<td>29 81</td>
<td>30 91</td>
<td>19 83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methionine (M)</td>
<td>50 38</td>
<td>20 21</td>
<td>35 39</td>
<td>27 41</td>
<td>17 32</td>
<td>23 42</td>
<td>19 54</td>
<td>13 36</td>
<td>6 18</td>
<td>13 57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylalanine (F)</td>
<td>65 50</td>
<td>43 45</td>
<td>39 43</td>
<td>25 38</td>
<td>20 38</td>
<td>26 47</td>
<td>14 40</td>
<td>11 31</td>
<td>2 6</td>
<td>9 39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proline (P)</td>
<td>51 39</td>
<td>36 38</td>
<td>47 52</td>
<td>32 49</td>
<td>15 29</td>
<td>20 36</td>
<td>21 60</td>
<td>17 48</td>
<td>7 21</td>
<td>11 48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serine (S)</td>
<td>83 64</td>
<td>47 49</td>
<td>42 47</td>
<td>37 57</td>
<td>31 59</td>
<td>25 45</td>
<td>11 32</td>
<td>20 56</td>
<td>20 61</td>
<td>21 92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threonine (T)</td>
<td>78 60</td>
<td>48 50</td>
<td>53 59</td>
<td>24 37</td>
<td>22 42</td>
<td>23 42</td>
<td>30 86</td>
<td>13 36</td>
<td>18 55</td>
<td>15 66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tryptophan (W)</td>
<td>11 8</td>
<td>13 14</td>
<td>10 11</td>
<td>15 23</td>
<td>2 4</td>
<td>13 24</td>
<td>5 14</td>
<td>5 14</td>
<td>3 9</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyrosine (Y)</td>
<td>56 43</td>
<td>47 49</td>
<td>39 43</td>
<td>28 43</td>
<td>14 27</td>
<td>28 51</td>
<td>9 26</td>
<td>10 28</td>
<td>4 12</td>
<td>3 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valine (V)</td>
<td>90 69</td>
<td>62 65</td>
<td>73 81</td>
<td>41 63</td>
<td>28 53</td>
<td>27 49</td>
<td>26 74</td>
<td>29 81</td>
<td>21 64</td>
<td>17 74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1302 956</td>
<td>901 654</td>
<td>526 552</td>
<td>349 357</td>
<td>328 229</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net charge</td>
<td>+27.5  +11.5</td>
<td>−5.0</td>
<td>+7.5</td>
<td>−4.5</td>
<td>+2</td>
<td>+1</td>
<td>−3</td>
<td>+11</td>
<td>+5.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>149588</td>
<td>111023</td>
<td>103326</td>
<td>76433</td>
<td>59163</td>
<td>64445</td>
<td>38548</td>
<td>40999</td>
<td>35750</td>
<td>25602</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Location of proteins in the virion are as follows: shell, outer capsid; core, inner capsid; NS, non-structural proteins.
† n, Number per 1000 residues.
Completion of the BTV-10 sequence

Fig. 3. Hydropathic plot and distribution of cysteines for the predicted two small RNA gene products (a NS2; b VP6) of BTV-10. The regions of the predicted proteins with a net hydrophobicity (areas above the centre line) or hydrophilicity (areas below the centre line) as well as the distribution of cysteine residues (vertical bars) are displayed (Kyte & Doolittle, 1982). The plot involves a span setting of 21 amino acids.

Changes (polar, hydrophobic, charged or hydrophilic residues) were included in the comparison (see asterisks in Fig. 4), the indicated homology was much higher (> 50%). It is possible, therefore, that the BTV VP7 and reovirus σ1 proteins may share some functional similarities. However, this has yet to be proved. The σ1 protein is a surface component of the reovirus virion and plays several roles in reovirus infections. These include involvement in determining reovirus tissue tropism (cell attachment) and virulence in a permissive host, triggering host immune responses (both humoral and cellular), as well as exhibiting a haemagglutinin function (Joklik, 1985; Sharpe & Fields, 1985). Although it has been postulated that the BTV VP7 molecules are located on the surface of the core particles, Hyatt & Eaton (1988) have recently reported data which indicate that VP7 may be accessible at the surface of the virus. In this context, it is noteworthy that among the different BTV serotypes, Northern hybridization analyses indicate that the S7 RNA (encoding VP7 protein) is somewhat less conserved than the RNAs coding for the other core proteins, although not to the extent of the genes encoding the outer capsid proteins VP2 and VP5 (Huismans & Cloete, 1987; Squire et al. 1986; Ritter & Roy, 1988). Further analyses on the location and role of the VP7 protein are required.

DISCUSSION

The sequence of the 19218 bp genome of a U.S. strain of BTV-10 ($M$, $13 \times 10^6$) has been completed by this report of the sequence data for cDNA clones representing the BTV S8 and S9 RNA segments. These two RNA species code for the NS2 and VP6 proteins, respectively.
Presented in Table 1 is a summary of data for the coding arrangements of all 10 dsRNA species of BTV-10. Analyses of the cloned DNAs have confirmed the previous RNA studies that demonstrated the existence of conserved sequences at the termini of all the viral RNA species (Rao et al., 1983). Each RNA has highly conserved terminal sequences (i.e. 5' ends of the mRNA sense strands: GUUAAA...; 3' ends: ...CACUUAC). In addition, other features of the sequences proximal to these highly conserved regions can be recognized in most of the RNAs (Table 1). For example nine of the 10 RNAs have another A following the 5' conserved sequence (i.e. residue 7), six of the RNAs have two A residues (residues 7 and 8). At the 5' ends of the mRNA-sense strands the conserved sequences are preceded by an AC in seven of the 10 RNAs, in the other three RNAs (M4, M5, M6) they are preceded by CA or CC (Table 1). Also proximal to the 3' ends of the coding strands are UU dimers in seven of the 10 RNA species (Table 1; L1, L2, M5, M6, S7, S8 and S9). Another noteworthy feature in the 3' non-coding sequences are purine-rich sequences in seven of the RNAs (L3, M4, M5, S7, S8, S9 and S10). The roles of these sequences in RNA transcription, replication or morphogenetic events are not known.

The sizes of the BTV RNAs range from 3954 bp (L1, M4, M5, M6) to 822 bp (S10, M7, 5 x 10^3). The overall base ratio of the genome is 28.1% A, 28.1% U, 21.9% G and 21.9% C. The base ratios of the individual segments are quite similar (i.e. A and U range: 29-3 to 25.4%; G and C range: 24-6 to 20-7%). The 5' non-coding regions range from eight nucleotides (M4) to 34 nucleotides in length (M6). The 3' non-coding sequences are generally longer, i.e. from 31 nucleotides (M5) to 116 nucleotides (S10).

Apart from the L1 segment, the first AUG codon in each of the mRNA-sense strands initiates
Completion of the BTV-10 sequence

a long open reading frame. In the L1 RNA there is an AUG upstream of the initiation of the open reading frame (residues 7 to 9). Whether it is present fortuitously, or affects the translation efficiency of the viral L1 gene by reducing the rate of synthesis of the L1 gene product is not known. However, similar AUG codons were also reported for the termini of L1 RNA species of BTV-1, -11 and Ibaraki virus (Mertens & Sangar, 1985; Rao et al., 1983). The flanking sequences to the AUG codons that initiate the open reading frames have a G at −3 (counting the AUG as +1, +2 and +3) in five of the RNAs, an A in four of the RNAs and a C in the S9 RNA. At the −2 position there is a C in four RNAs, an A in five RNAs and a U in the S8 RNA. At the −1 position there is a C in six of the RNAs, a G in two RNAs and a U or a C in the other RNAs (S9 and S10, respectively). At the +4 positions there are G residues in seven of the RNA species, a U in one (S9) and a C in two (M4, S10). In general, therefore, the flanking sequences agree with the models proposed by Kozak (1981, 1983). All three of the possible translation termination codons are used. Four of the gene products terminate with a UGA codon, four with a UAG codon and two with a UAA codon.

Table 2 presents a summary of the amino acid compositions, sizes and net charge data for all the primary gene products of BTV-10. To aid in these comparisons the number of amino acids per 1000 residues of each protein type has been calculated. From these data the mean number of amino acids for the BTV gene products can be determined. The mean numbers are as follows: Ala 76-7; Arg 72-0; Asp 59-7; Asn 38-1; Cys 11-7; Glu 74-3; Gln 38-6; Gly 59-3; His 22-1; Ile 66-6; Leu 81-3; Lys 59-2; Met 37-8; Phe 37-7; Pro 42.0; Ser 56-2; Thr 53.3; Trp 12.5; Tyr 33.5; Val 67.3. When the individual proteins are compared to these values, several striking primary composition variations are evident. These are indicated in the synopses of the protein and gene data presented below.

(i) L1 gene, VP1 protein. This gene, the largest of the BTV segments, codes for a minor protein, VP1. The protein is present in the core of virions. Hybridization data have indicated that the gene is highly conserved among all BTV serotypes tested and exhibits some sequence relationship with the epizootic haemorrhagic disease virus (EHDV), a related orbivirus (Ritter & Roy, 1988; Kowalik & Li, 1987; Huismans & Cloete, 1987; Mertens et al., 1987; Squire et al., 1986). Evidence has been obtained that indicates that the L1 gene product is a transcriptase component (T. Urakawa et al., unpublished data). The VP1 protein has the highest net positive charge at neutral pH, although it is not more abundant in charged amino acids per unit protein length than other viral gene products. Thus, per 1000 amino acids, there are 145 positively charged residues (R + K + ½H) and 124 negatively charged amino acids (D + E). Relative to the other BTV proteins, the VP1 protein is rich in the hydrophilic amino acids serine and threonine and the aromatic amino acids phenylalanine and tyrosine.

(ii) L2 gene, VP2 protein. The protein VP2 is a major component of the outer shell of the virus. The protein is the principal serotype-specific antigen of BTV. It elicits neutralizing antibodies (Huismans & Erasmus, 1981; Kahlon et al., 1983) and, together with the viral VP5 protein, exhibits the greatest sequence variation between BTV serotypes (Fukusho et al., 1987; Ghiasi et al., 1985; Purdy et al., 1985; Ritter & Roy, 1988; Yamaguchi et al., 1988a, b; Gould, 1988a). VP2 protein contains a high number of conserved cysteine residues, suggesting that it has a highly ordered structure that involves disulphide bonds. The protein is also rich in aromatic amino acids.

Recently Huismans et al. (1987a) reported that shortly after infection, parental BTV virions are first converted to core particles (470S) and then to subcore particles (390S). Core particles lack only the VP2 and VP5 proteins and exhibit a well-defined structure in the electron microscope. Cores have an active RNA transcriptase and are very stable. The subcore particles lack the outer capsid proteins VP2 and VP5, as well as the inner capsid protein VP7. Subcores are unstable and apparently lack transcriptase activity in vitro.

(iii) L3 gene, VP3 protein. The VP3 protein is a major structural protein of the core of BTV virions. It has a highly conserved amino acid sequence as evidenced from sequence analyses of three BTV L3 RNA species (Purdy et al., 1984; Ghiasi et al., 1985; Gould, 1987) and by hybridization data (Roy et al., 1985; Squire et al., 1986; Kowalik & Li, 1987; Huismans & Cloete, 1987; Mertens et al., 1987; Ritter & Roy, 1988). The protein contains group-specific
antigenic determinants (Huismans & Erasmus, 1981). It has a low content of charged amino acids although an average number of serine and threonine residues.

(iv) M4 gene, VP4 protein. The M4 gene is highly conserved among BTV serotypes (Squire et al., 1986; Kowalik & Li, 1987; Huismans & Cloete, 1987; Mertens et al., 1987; Ritter & Roy, 1988). The VP4 protein is a minor component of the inner core of virions. Its function is not known. The VP4 protein of BTV-10 has a high content of charged amino acids (particularly histidine) and tryptophan residues. The protein has a very low content of glutamine, although whether this is particular to the BTV-10 isolate or representative of other BTV serotypes and isolates is not known.

(v) M5 gene, VP5 protein. The VP5 protein is the second major protein forming the outer capsid of BTV virions. On the basis of hybridization data (Squire et al., 1986; Kowalik & Li, 1987; Huismans & Cloete, 1987; Mertens et al., 1987; Ritter & Roy, 1988) the BTV-10 RNA sequence is not highly conserved among other BTV serotypes. The protein is rich in certain non-polar amino acids (e.g. alanine, isoleucine) although it has average numbers of others (e.g. leucine, methionine). It has a low content of tryptophan residues. Recently the complete sequence of this gene of two isolates of serotype 1 has been reported (Gould & Pritchard, 1988; Wade-Evans et al., 1988); when compared to VP5 of BTV-10, a homology of 68% at the nucleotide level and 76% at the amino acid level was observed.

(vi) M6 gene, NS1 protein. The M6 gene appeared to be conserved (80% to 90%) among BTV serotypes although it does not appear to give strong hybridization signals with EHDV RNAs (Squire et al., 1986; Kowalik & Li, 1987; Huismans & Cloete, 1987; Mertens et al., 1987; Ritter & Roy, 1988; Gould 1988b). However, the gene product which is highly conserved is a non-structural protein, designated NS1, and forms tubules in the cytoplasm of BTV-infected cells and in recombinant baculovirus-infected Spodoptera frugiperda cells (Huismans & Els, 1979; Eaton et al., 1988; Urakawa & Roy, 1988). The composition analyses of the NS1 protein indicate that it is particularly rich in cysteine, tryptophan and tyrosine residues. The protein has the lowest numbers of serine plus threonine residues per unit length of any BTV protein. It also has a low content of charged amino acids. Similar compositions were reported for NS1 of BTV-1 (Gould, 1988a). The contribution of the primary structure and composition to the tertiary and quaternary structure and function of the protein, and the formation of tubules in infected cells, needs to be determined.

(vii) M7 gene, VP7 protein. The VP7 protein is a major structural component of the cores of BTV, forming 36% of the core particle protein (Huismans et al., 1987b). It contains group-specific antigenic determinants (Huismans & Erasmus, 1981). The M7 gene is less conserved among the BTV serotypes but exhibits little hybridization homology to the corresponding segments of EHDV (Squire et al., 1986; Kowalik & Li, 1987; Huismans & Cloete, 1987; Mertens et al., 1987; Ritter & Roy, 1988). The primary composition of the VP7 protein differs considerably from those of other BTV gene products. Thus it has a very low number of charged residues per unit length of 1000 amino acids (K + R + \(\frac{1}{2}\)H = 80, D + E = 77; average for the 10 BTV gene products is 142 and 134 respectively). However, it has the highest number of asparagine and glutamine residues per unit length (N + Q = 109, by comparison to an average number of 77 for all the BTV gene products). The protein contains only one lysine residue, a remarkable feature since lysines are relatively abundant in all the other BTV gene products. The protein is also relatively rich in alanine, methionine and proline residues. Although containing more threonine residues per unit length than other BTV gene products this is counterbalanced, in part, by a lower than average representation of serines. The sequence data indicate that the VP7 protein is extremely hydrophobic. The hydrophobic regions of VP7 resemble those associated with membrane-spanning proteins (Kyte & Doolittle, 1982) or the lipid-free protein coats of fd-type bacteriophages (Smilowitz et al., 1972). The importance of the hydrophobic domains of VP7 and its macromolecular interactions with the other BTV proteins remain to be determined.

(viii) M8 gene, NS2 protein. NS2 is a non-structural protein. It is the only viral phosphoprotein that can be identified in BTV-infected cells (Grubman et al., 1983; Huismans et al., 1987a). The NS2 gene appeared to be less conserved in comparison to the two genes
Completion of the BTV-10 sequence

1687

encoding NS1 and NS3 (Ritter & Roy, 1988; Huismans & Cloete, 1987; Mertens et al., 1987; Kowalik & Li, 1987). The protein has the lowest relative number of threonine residues per unit length of any BTV-10 gene product, but an average number of serine residues (Table 2). The overall high number of cysteine residues located in a cluster at the carboxy-terminal end suggests that the protein may have a highly ordered structure (Fig. 3). Overall, the protein is rich in charged amino acids (aspartate, glutamate and lysine residues) although it has a low number of histidine residues and an average number of arginines. Our data have been confirmed by the sequence data relating to the S8 gene recently obtained from the South African isolate of BTV-10 (Hall et al., 1989).

(ix) S9 gene, VP6 protein. The VP6 protein is a minor component of BTV cores. It is a hydrophilic protein (Fig. 3) with quite a unique composition by comparison with other BTV gene products. The VP6 protein has only one cysteine residue and only a few aromatic amino acids per unit length of 1000 amino acids (Table 2; VP6 F + Y = 18, by comparison to the average composition of 70 aromatic residues for all BTV-10 gene products). The protein has a low content of asparagine and glutamine residues. Another striking feature concerning the composition of the VP6 protein is that it is unusually high in glycines (40 out of 328 amino acids, 12%; see Table 2) with 12 residues closely located in the amino-terminal half of the molecule (residues 88 to 116), including five consecutive glycines (see Fig. 3). The protein is the richest of all the BTV gene products in its content of charged amino acids per unit length of 1000 amino acids (R + K + H = 198, D + E = 165). Although the function of the VP6 protein is not known, it is believed to be associated with the viral RNA in the cores of virions (our unpublished observations).

(x) S10 gene, NS3 protein. The smallest gene of BTV-10 contains two possible initiation sites in the same reading frame and estimated Mr values for the predicted proteins (NS3 and NS3A) are 25602 and 24020, respectively (Lee & Roy, 1986; Gould, 1988b). Two proteins with similar sizes were also identified by in vitro translation of the S10 gene (Sangar & Mertens, 1983; Gould, 1988b; van Dijk & Huismans, 1988). Recently we demonstrated that these two non-structural proteins are indeed related to each other and that both are encoded by the S10 gene (French et al., 1989). These proteins are rich in the hydrophilic serine and threonine residues but deficient in arginine, tryptophan, glycine and isoleucine by comparison to other BTV proteins. The function of NS3/NS3A in BTV-infected cells is not known.

In summary, the sequence of the genome of BTV-10 (U.S. strain CA-8) has been completed by the sequence data reported in this paper for some of the small viral RNA species. The data represent the first completed sequence of a member of the Reoviridae and as such are the largest sequence of an RNA virus reported so far. Interestingly some gene product homology has been detected between the reovirus σ1 protein and the structural VP7 protein of BTV-10. The meaning of this homology from an evolutionary or functional standpoint is not known. Whether homologies exist with other reovirus gene products will have to await further sequence data for that virus.

We thank the staff of the NERC Institute of Virology, especially C.D. Hatton for photography, Mrs J. Oswell for typing and J.J.A. Marshall for computer analysis. This work was supported by EEC Contract BAP-0120 U.K. and by Public Health Service grant AI26879 from the National Institutes of Health, U.S.A. and Alabama State grant ARI no. 86-409.

REFERENCES


Completion of the BTV-10 sequence


(Received 16 January 1989)