Key words: dengue virus, flavivirus cDNA, structural proteins

Partial Nucleotide Sequence and Deduced Amino Acid Sequence of the Structural Proteins of Dengue Virus Type 2, New Guinea C and PUO-218 Strains

By A. GRUENBERG, W. S. WOO, A. BIEDRZYCKA AND PETER J. WRIGHT*

Department of Microbiology, Monash University, Wellington Road, Clayton, Victoria 3168, Australia

(Accepted 7 March 1988)

SUMMARY

The nucleotide sequence and the deduced amino acid sequence for the genes encoding the structural proteins of two strains of dengue virus type 2 (DEN-2) were determined from cDNA clones. The genes for C, prM(M) and E proteins were sequenced for the prototype DEN-2 virus, the New Guinea C strain. Also sequenced were the prM(M) and E genes of PUO-218. This strain of DEN-2 was isolated during 1980 in Bangkok and had received a limited number of laboratory passages. Comparisons of the newly determined sequences with those published for the Jamaica 1409 and Puerto Rico PR-159 (S1 vaccine candidate) strains revealed a close relationship between New Guinea C virus and both the Jamaica and PUO-218 viruses (greater than 96% similarity in nucleotides of the E gene), whereas S1 virus was the most divergent.

The dengue viruses belong to the family Flaviviridae. Four dengue serotypes are currently recognized (Porterfield, 1980). The nucleotide sequences of the genomes of strains of dengue virus type 2 (DEN-2) and dengue virus type 4 (DEN-4) have recently been determined (Deubel et al., 1986; Zhao et al., 1986; Mackow et al., 1987; Hahn et al., 1988). For these viruses and other sequenced members of the Flaviviridae, the gene order is 5'-C-prM(M)-E-NS1-NS2a-NS2b-NS3-ns4a-NS4b-NS5-3', with the proteins apparently encoded as a single polyprotein (Rice et al., 1985; Castle et al., 1986; Coia et al., 1988). Cleavage of the polyprotein to form mature viral polypeptides is probably achieved by at least one host protease and one virus-coded protease (Rice et al., 1985). The four antigenic types of dengue virus are endemic in many tropical and subtropical areas, and have been associated with epidemics of disease in Asia, the Pacific Islands, the Caribbean, East Africa and Brazil. Though most of the disease observed is the relatively mild dengue fever, a significant proportion of infections result in severe haemorrhagic fever (DHF). No established vaccine against dengue is available, and it appears that a second infection with a heterologous serotype increases the risk of contracting DHF. It has been suggested that antibody-dependent enhancement of viral infectivity and variation in virulence of naturally occurring strains contribute to the severity of DHF (for review, see Monath, 1986). Thus it is important to compare the biological and molecular properties of a range of viral isolates. Of particular interest are the structural polypeptides (C, prM, M, E), especially the envelope protein E, which mediates cell attachment and contains enhancing and neutralizing antigenic sites. In this paper we provide new sequence information for two DEN-2 strains, namely New Guinea C (NGC) and PUO-218, a Thai isolate. Using the published sequences for the Puerto Rico PR-159(S1) strain (S1) (Hahn et al., 1988) and the Jamaica 1409 strain (JAM) (Deubel et al., 1986), it is now possible to compare structural genes and proteins for four strains of DEN-2 virus differing in origin and passage history.
The NGC strain of virus, isolated in 1944 (Sabin & Schlesinger, 1945), was chosen for sequencing as it is the prototype DEN-2 strain and has better characterized polypeptides than any other DEN-2 strain (Smith & Wright, 1985; Biedrzycka et al., 1987). For the experiments reported here, NGC virus was obtained from Dr B. Gorman at the Queensland Institute of Medical Research (QIMR; Bramston Terrace, Herston, Brisbane, Australia). The virus used to infect Vero cells had been passaged 24 times through suckling mice, plaque-purified twice in equine kidney (EK) cells, passaged a further three times in EK cells and twice in Aedes albopictus C6/36 cells. The PUO-218 strain of virus was originally isolated from a 21 month old child with a symptomatic primary DEN-2 infection by Dr D. S. Burke during the 1980 dengue epidemic in Bangkok. The virus was isolated from plasma by inoculation into Toxorhynchites splendens mosquitoes. Mosquito thorax–abdomen suspensions were passed once in LLC-MK2 cells (D. S. Burke, personal communication; Morens et al., 1987). We obtained this virus from QIMR also, and grew the virus in C6/36 cells without further plaque purification.

Virus was purified by precipitation with polyethylene glycol and sedimentation through sucrose gradients (Smith & Wright, 1985). The synthesis of cDNA for NGC starting with virion RNA and random primers, the cloning of viral inserts into plasmid pUC8, and the sequencing of the inserts by the chain termination method were as described by Biedrzycka et al. (1987). Similar procedures were used for PUO-218, except that a deoxyoligonucleotide primer complementary to 15 bases at the start of the NS1 gene was used to prime first strand cDNA synthesis.

The nucleotide sequence and the deduced amino acid sequence for the region of the NGC genome encoding the C, prM(M), and E proteins are shown in Fig. 1. The nucleotide sequence was determined from a series of overlapping clones ranging in size from 0.2 to 1.1 kb. Thirty-eight percent of the nucleotide sequence was determined from at least two independently isolated clones, and nucleotides 1637 to 1667 were determined by direct dideoxy nucleotide chain termination sequencing of virion RNA. Nucleotide 1 in the NGC sequence shown corresponds to nucleotide 77 counting from the 5' end of the DEN-2 (JAM) sequence (Deubel et al., 1986). The nucleotide sequence for PUO-218 has been determined for the region encoding the carboxy terminus of C and all of prM and E; alignment of this sequence with the NGC sequence is also shown in Fig. 1. The sequence of PUO-218 was obtained predominantly from a single cDNA clone of 2.1 kb. Five percent of the sequence was determined from a second clone (0.6 kb) located at the 3' end of the E gene. Forty-five percent of the sequence was determined from both strands of the cDNA. Nucleotides identical in NGC and PUO-218 are indicated by dots. The start points of E and NS1 for DEN-2 have been determined by amino acid sequencing of the viral proteins (Bell et al., 1985; Biedrzycka et al., 1987), whereas the amino termini of C, prM and M have been assigned by alignment with other flavivirus proteins (Castle et al., 1985; Deubel et al., 1986).

A comparison of the deduced amino acid sequences of the structural proteins for the strains NGC, PUO-218, JAM (Deubel et al., 1986) and S1 (Hahn et al., 1988) is provided in Fig. 2. Some calculations for similarities at the amino acid and nucleotide levels generated by pairwise comparisons between the strains are summarized in Tables 1 and 2. A comparison between the NGC, S1 and JAM strains for the entire structural region (Table 1) indicated that NGC and JAM are more closely related to each other than S1 is to either of them. Analysis of the E region alone, including strain PUO-218 (Table 2), also demonstrated the dissimilarity of S1 to the other strains because the percentage similarities in pairwise comparisons were lowest in those involving the S1 strain i.e. 91.9%, 90.8%, 91.1% (nucleotides) and 97.6%, 97.6%, 97.2% (amino acids). The most similar pair of strains was NGC and PUO-218 (96.2% nucleotides, 98.8% amino acids) followed closely by NGC and JAM (96.1% nucleotides, 98.4% amino acids).

Further analysis of the data presented in Fig. 1 revealed that the longest stretch of conserved nucleotides in NGC and PUO-218 runs from nucleotide 819 to nucleotide 1015, a region that encodes the amino-terminal portion of E. It is intriguing to note that NGC nucleotides 820 to 946 inclusive are also conserved in the JAM strain, and that this is also the longest conserved stretch observed when the nucleotide sequences for the structural genes of the NGC and JAM strains are compared (detailed analysis not shown). However, a comparison of NGC and S1
Table 1. *Similarities (%) between DEN-2 strains for the C, prM(M) and E region*

<table>
<thead>
<tr>
<th>Nucleotides*</th>
<th>NGC</th>
<th>S1</th>
<th>JAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acids</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NGC</td>
<td>–</td>
<td>92.5</td>
<td>96.0</td>
</tr>
<tr>
<td>S1</td>
<td>97.7</td>
<td>–</td>
<td>90.8</td>
</tr>
<tr>
<td>JAM</td>
<td>97.9</td>
<td>96.4</td>
<td>–</td>
</tr>
</tbody>
</table>

* The nucleotide sequences for the S1 and the JAM strains are from Hahn et al. (1988) and Deubel et al. (1986), respectively.

Table 2. *Similarities (%) between DEN-2 strains for the E region only*

<table>
<thead>
<tr>
<th>Nucleotides</th>
<th>NGC</th>
<th>218</th>
<th>S1</th>
<th>JAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acids</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NGC</td>
<td>–</td>
<td>96.2</td>
<td>91.9</td>
<td>96.1</td>
</tr>
<tr>
<td>218</td>
<td>98.8</td>
<td>–</td>
<td>90.8</td>
<td>94.7</td>
</tr>
<tr>
<td>S1</td>
<td>97.6</td>
<td>97.6</td>
<td>–</td>
<td>91.1</td>
</tr>
<tr>
<td>JAM</td>
<td>98.4</td>
<td>98.4</td>
<td>97.2</td>
<td>–</td>
</tr>
</tbody>
</table>

genomes over this region (820 to 946) reveals 12 differences, so that the significance, if any, of the conservation in the NGC, PUO-218 and JAM strains is not yet apparent.

Within the primary structure of prM and E for the four DEN-2 strains, potential glycosylation sites at prM-69, E-67 and E-153 are conserved (Fig. 2). The presence of one glycosylation site in prM and two in E is consistent with the estimates obtained for the number of N-linked glycans contained in each of these glycoproteins (Smith & Wright, 1985). Eighteen cysteine residues are conserved. Also conserved in E are three peptides underlined in Fig. 2; they are peptides E-98 to E-111, E-371 to E-378 and E-416 to E-422. These peptides are conserved in St Louis encephalitis, West Nile, Murray Valley encephalitis, yellow fever (17D), Kunjin and DEN-2 viruses (Trent et al., 1987; Coia et al., 1988). It has been suggested that these peptides may be involved in important biological functions such as binding to cell receptors or membrane fusion (Deubel et al., 1986; Hahn et al., 1987).

In the prM and E proteins of the NGC strain there are two amino acid substitutions which are not found in any of the other three strains. Isoleucine rather than phenylalanine, another non-polar amino acid, is located at E-402; the basic amino acid lysine rather than glutamic acid is at E-126. The nucleotide sequence for the first conservative substitution was obtained from a single clone. However, the sequence for the second substitution was obtained from two independently isolated clones, indicating that it is unlikely to be the result of a cloning artefact. For PUO-218, unique amino acids are located at prM-125, E-141 and E-164; they involve changes of isoleucine for threonine, or of valine for isoleucine.

Similar conservative interchanges are observed elsewhere in Fig. 2. At present it is not possible to assess the effect of any of these amino acid changes on the structure and function of prM and E, due to lack of information on the three-dimensional structures of the proteins and on the locations of important functional domains.

Trent and his colleagues have assigned DEN-2 isolates from different geographic regions to at least six groups of genetic variants or topotypes by mapping of oligonucleotides following digestion of genomic RNA with RNase T1 (Trent et al., 1983; Monath et al., 1986; Kerschner et al., 1986). A major aim of their studies has been to facilitate the identification of virus strains involved in outbreaks of disease. In general, the topotype groupings of the viruses were confirmed by antigen signature analysis (Monath et al., 1986). Our analyses of sequence data provide yet another way of comparing isolates. The results for the E region, which represents only 14% of the whole genome, indicate a close relationship between PUO-218 and NGC, and
Fig. 1. Nucleotide sequences of the NGC and PUO-218 strains of DEN-2 virus in the region of the genome encoding the structural proteins. The deduced amino acid sequence for NGC virus is shown. Potential glycosylation sites are indicated by asterisks.
Fig. 2. A comparison of the deduced amino acid sequences for structural proteins specified by four strains of DEN-2 virus. Potential glycosylation sites are indicated by asterisks. Underlined peptides are highly conserved in the flaviviruses sequenced so far. The amino acid sequences for the SI and JAM strains are from Hahn et al. (1988) and Deubel et al. (1986), respectively.

Between JAM and NGC. The four strains whose structural regions we have compared differ in geographic origin, date of isolation and passage history. The similarities between the four strains based on this part of the genome (Tables 1 and 2) do not closely follow the geographic region of isolation, e.g. Caribbean (JAM, S1) and S.E. Asia (NGC, PUO-218), or time of isolation, e.g. early [NGC (1944), S1 (1969)] and late [PUO-218 (1980), JAM (1983)] or the number of
passages, e.g. high (NGC, S1) and low (JAM, PUO-218). However, the most divergent strain is apparently S1, which has been extensively passaged in cell culture, is a small plaque and temperature-sensitive virus, and is attenuated in rhesus monkeys (Eckels et al., 1976; Harrison et al., 1977). In order to relate changes at the molecular level to differences in virulence of virus strains, it is now important to obtain further comparisons of attenuated strains such as S1 with low passage field isolates associated with defined disease outbreaks.

We are grateful to Mr M. R. Cauchi for technical assistance and to Dr C. R. Parrish for helpful discussions. We also thank Dr J. H. Strauss for providing results before publication. This investigation received financial support from the WHO Programme for Vaccine Development and from the National Health and Medical Research Council of Australia.

REFERENCES


Short communication


(Received 4 January 1988)