Arch Virol (1989) 105: 39-53



# Variation of the nucleotide and encoded amino acid sequences of the envelope gene from eight dengue-2 viruses

J. Blok<sup>1</sup>, S. Samuel<sup>1, \*</sup>, A. J. Gibbs<sup>2</sup>, and U. T. Vitarana<sup>3</sup>

<sup>1</sup> Sir Albert Sakzewski Virus Research Laboratory, Royal Children's Hospital, Brisbane, and
<sup>2</sup> Research School of Biological Sciences, Australian National University, Canberra, Australia
<sup>3</sup> Medical Research Institute, Colombo, Sri Lanka

Accepted January 3, 1989

**Summary.** The nucleotide sequences of the envelope genes from five Thai and three Sri Lankan dengue-2 viruses were determined by sequencing the viral RNA using synthetic oligonucleotide primers. The results were compared with the four published dengue-2 envelope sequences to obtain a classification of these viruses, which showed that the Thai isolates could be divided into two separate groups while the Sri Lankan isolates were distinct. There was no correlation between disease severity and envelope protein sequence, or between year of isolation and sequence. No particular amino acid changes were associated with virulence or a change in hydrophilic region which could perhaps act as an epitope.

### Introduction

Dengue viruses belong to the *Flaviviridae*, a family of closely related viruses transmitted by mosquitoes and ticks, which also include the causal agents of yellow fever and Japanese encephalitis [33]. There are four distinct serotypes of dengue virus and all of these have been isolated from people with classical uncomplicated dengue fever or dengue haemorrhagic fever/dengue shock syndrome (DHF/DSS) [29]. Most of the millions of cases of dengue occurring annually are uncomplicated classical dengue fever, which is a mild febrile illness with headache, associated joint pains and rash. A small percentage of cases develop DHF/DSS, which is usually characterized by fever, thrombocytopenia, increased vascular permeability and can sometimes be fatal [reviewed in 19]. The causes of the more severe symptoms are not fully understood, but it has

<sup>\*</sup> Present address: Arbovirus Research Unit, Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia.

been hypothesized that DHF/DSS results from immune enhancement due to a second dengue virus infection [14]. There are, however, reports of primary cases of DHF/DSS and it has therefore been suggested that dengue viruses may vary in virulence [24].

The more serious form of dengue is a major cause of paediatric morbidity and mortality in several south-east Asian countries [13]. In Thailand, over 20,000 patients with DHF are hospitalized annually [27] although most dengue infections produce asymptomatic or mild disease and only a minority of cases develop the DHF/DSS syndrome [10, 11, 25]. Prospective studies of patients in Bangkok have shown that many cases of DHF/DSS resulted from sequential infections of different dengue virus serotypes [4, 11] especially when the second infection was that of dengue type 2 [28]. Another study in Bangkok showed that remnants of maternal dengue antibodies increased the likelihood of DHF in infants [17].

However, although all four dengue serotypes are endemic in Sri Lanka and there are many cases of sequential infections, some with the secondary serotype being type 2, only three to eleven cases of DHF (average of 6) have been recorded annually since 1980. The incidence of DHF was even lower before this except for the period from 1966–69 when the first major outbreak of dengue, which was laboratory confirmed, occurred. During this epidemic there were between four and 27 cases reported per year, with four deaths in 1966 and only four cases of DHF in 1969. The very low frequency of DHF cases in Sri Lanka is despite the fact that 70% of the dengue fever cases were secondary infections. The reason for the difference in incidence of DHF in Thailand and Sri Lanka is not clearly understood but it may be due to virus virulence.

The technique of antigen signature analysis, in which monoclonal antibodies are used to recognize distinct epitopes on the antigen, has shown variation of dengue-2 viruses [20]. These studies showed few differences in the envelope proteins of dengue virus isolates from the same geographic region, but showed striking antigenic differences between geographic groupings. The results obtained from these antigen signature analyses were compared with the genetic groupings of these dengue-2 viruses based on RNase  $T_1$  oligonucleotide fingerprinting, and it was shown that both techniques could distinguish variants which correlated with geography and epidemiology [20]. Genetic variation within the dengue-2 serotype can therefore be assessed by oligonucleotide fingerprinting and restriction enzyme mapping of the viral RNA [16, 31, 32], but these studies do not examine one particular gene.

In order to understand the epidemiology, strain variation and perhaps pathogenic potential of dengue viruses, the derivaton of gene sequences coding for the major antigenic protein, the envelope (E) protein, is important. The dengue-2 serotype has been implicated as an increased risk factor for DHF/DSS, and there appears to be a difference in this risk factor depending on which country one studies. Therefore, the sequences of the E gene from five Thai and three Sri Lankan dengue-2 virus isolates were determined and are reported in this paper. These are compared with other published dengue-2 envelope protein gene sequences from viruses isolated from the Caribbean [5, 9] as well as the prototype New Guinea C [8] virus.

### Materials and methods

### Virus isolates

The full names, codes, and clinical symptoms caused by the dengue-2 viruses used in this study are listed in Table 1. Strain 16681 was isolated from a child with shock syndrome in Bangkok in 1964 [12]. The other Thai isolates were from blood specimens from patients at Bangkok Children's Hospital in 1980 [32] and these were kindly provided by Dr. Don Burke. Clinical diagnosis and grades of disease severity were assigned by Dr. Suchitra Nimmannitya according to WHO criteria [34]. Two of the strains from Sri Lanka (b, c) were isolated from patients' serum by passage in suckling mice (two to four passages) followed by two passages in mosquito cells; while the other strain (a) was obtained by mosquito (*Aedes aegypti*) inoculation followed by two passages in mosquito cells.

None of the virus stocks were plaque purified before growth in a mosquito cell line, and the number of passages in this cell line was kept to a minimum (less than four passages). In this way we attempted to ensure that the virus population in the original patient dominated the viral genomes we sequenced.

Virus isolate	us Code Year of Country ate isolation of origin		Country of origin	Clinical disease	Patient age	Sero response	Envelope sequence reference
New Guin- ea C	NGC	1944	New Guin- ea	DF	Adult		[8]
16681 D80-038 PUO-280 D80-100 D80-141 PUO-218	16681 D F G H AA	1964 1980 1980 1980 1980 1980	Thailand Thailand Thailand Thailand Thailand Thailand	DSS DHF DF DHF DSS DF	8 3 7 4/12 7 21/12	2° 2° 1° 2° 1°	* * * * [8]
PR159 (S1) D2-1409	S1 JAM	1969 1983	Puerto Rico Jamaica	Vaccine			[9] [5]
SL767 SL77/69 SL1050	a b c	1982 1968 1969	Sri Lanka Sri Lanka Sri Lanka	DF DF DF	11 21 50	2° 2° 1°	* * *

Table 1. Characteristics and codes of the dengue-2 viruses used in this study

*DF* dengue fever, *DHF* dengue haemorrhagic fever, *DSS* dengue shock syndrome \* This paper

#### Growth and purification of viruses and genomic RNA

Viruses were propagated in a continuous line of *Aedes albopictus* cells (clone C6/36) [15]. The cells were grown at 32 °C in RPMI 1640 medium containing 10% foetal calf serum, 100 units/ml penicillin and 100  $\mu$ g/ml streptomycin. Virus inoculation and daily haemag-

glutination monitoring were as described [32], and once viral haemagglutinin was detected, the tissue culture medium containing the mature virions was collected for three consecutive days and centrifuged to remove cells and debris. Virus was harvested from this medium by PEG precipitation (6.6% w/v) and then further purified in 5–50% sucrose gradients. RNA was extracted from the viral pellets as previously described [2], and was ethanol precipitated twice before use in sequence experiments.

#### Nucleotide sequencing

The dengue-2 genomic RNA was sequenced using the dideoxynucleotide chain termination method [26]. The oligonucleotide primers (13 to 15mers) were synthesized in an Applied Biosystems oligonucleotide synthesizer, desalted in a PD-10 Sephadex G-25 M column (Pharmacia) and the first three fractions were dried and then dissolved in 100  $\mu$ l. One microlitre of a 1:100 dilution of the primer was usually used with 0.5  $\mu$ g RNA although the dilution was determined empirically for each primer. The viral RNA and a synthetic oligonucleotide primer were heated in a 90° water bath for one min, quickly chilled in an ice/water bath and sequenced with 1.0 unit of AMV reverse transcriptase (Life Sciences), as described [1].

### Data analysis

Various programmes in the SEQ and STATS libraries of the VAX 11-750 of the Research School of Biological Sciences, A.N.U., Canberra were used to compare the nucleotide sequences and the amino acids they encoded.

#### Results

### Oligonucleotide sequence variation and its effect on priming ability

The E gene sequence of the dengue-2 viruses was obtained by sequencing the viral RNA genome with specific oligonucleotide primers complementary to chosen parts of the sequence of a Jamaican isolate [5]. These selected sequences were spaced about 150 nucleotides apart and eleven primers, each of 15 nucleotides, were synthesized initially. Some of these primers were ineffective in certain regions of the E gene, and so a further seven primers, each of 13 nucleotides were synthesized. The decrease in length of the oligonucleotides was to save costs, because it was found that the shorter primers were as efficient and specific as the larger ones; the principal variable was the chosen genomic sequence.

Table 2 shows the variation in the primer sequences and their influence on the ability to obtain sequence information. It is clear that a perfect complementary sequence is not absolutely required for binding and subsequent priming. However, more than one mismatch in any one primer binding site generally limits the chances of obtaining data. For example, mismatches in primers 3, 5, 7, and 8 (in Table 2) resulted in partial priming which gives some sequence data but with background sequences present, nonspecific priming which yields no sequence data, or no priming at all. There was one case (primer 12 in Table 2) where two pyrimidine transitions in the viral RNA did not disrupt the hydrogen bonding between the primer and viral RNA, thereby allowing the priming and

Oligonucleotide primer sequence 5' to 3'	Nucleotide changes in the complement of the dengue-2 viral RNA							
1. CCCATCCTCTGTC	CCCATCCCCTGTC-PP (2)*	CCCATCCTCGGTC-PP (1)						
2. CTGGTTGCACGAC	CTGGTTGCACGAT-PP (3), P (1)	_						
3. ACTCCATCGTGAC	ACTGCAGCGTGAC-PP (1)	ACTCCATTGTGAC-P (1)						
	ACTCCATCGTGAT-N (1)							
4. ATTGCCTATGCAC	ATTGACTATGCAC-N (1)	ATTGCCTGTGCAC-N (4)						
5. AAAAGCAGCCCCA	GAAAGCAGCCCCA-P (1)	GAAGGCAGCTCCA-N (2)						
	– AAAGGTAGCCCCA-N (1)	AAAGGCAGCTCCA-N (1)						
6. AGGTGCTACGTGA	AGGTGCTGCGTGA-P (5)							
7. CGCTCCTCTCATT	TGCTCCTCTCATT-NSP (1)	GGCTCCTCTCATT-PP (1)						
<u> </u>	CGCCCCCTCATT-NSP (2)	CGCTCCCCTCATT-PP (1)						
8. TTGTGTGTCCGCTCC	TTGTGTATCCGCTCC-PP (1)	TTGTGTATCTGCTCC-PP (1)						
	TTGTGTGTCAGCTCC-PP (1)	TTGTGTGTGTCC-PP (1)						
9. TTTGATTTCCTTGCC	TTTTATTTCCTTGCC-PP (2), P (3)							
	TTTGATTTCTTTGCC-P (1)							
10. GGCCCCTTCTTGAGA	AGCCCCTTCTTGGGA-P (6)							
11. TGACATTCCTTTGAG	TGACATCCCTTTGAG-P (1)							
12. GCGACCTAAGACATG	GCGGCCTAAGACATG-P (2)	GCGGCCTAAAACATG-P (2)						
	GCGACCTAACACATG-P (1)							
13. TAT/CGTTGACTGGGCT	TAT/CTTTGACTGGGCT-P (1)							
14. GTGGAGAGCCTTTCC	GTGGAGAGCTTTTCC-N (2), PP (1)	)						
	GTGGAGAGCCTTCCC-P (1)							

г ц.	•	<b>C</b>		·		1	- 4	1		- CC +		
lahle	Ζ.	Sequence	variation	1n	primer	binding	sites	and	IIS	effect	on	sequencing
1 4010	_	009 401100	,	***		VIII GIII B	01000	****	100	011000	••••	oud annung

P Primes to give sequence information

N No priming

PP Partial priming to give some sequence information but background banding present NSP Non-specific priming with multiple bands on sequence gels and no sequence information obtained

\* In parenthesis, number of viral RNAs which gave the priming results

subsequent sequencing reaction to proceed. Pyrimidine transitions (Py) in the viral RNA were generally very well tolerated, while purine (Pu) transitions and Py/Pu or Pu/Py transversions gave variable results. The variation in priming results was also due to the purity of the viral RNA since three primers with single mismatches (primers 2, 9, and 14 in Table 2), gave varying amounts of sequence data depending on the quality of the viral RNA. The primer binding ability could not be predicted before attempted sequence analysis as there did not appear to be any rules which correlated mismatching and priming ability.

10001			u a la dana					
D F G H b C	U	A A A A A A A A A	U A C C	A A A A A A A	UUULAKKAKK	C C C C	C U	U
16681 D F G	GUGACGACGAUG	GCAAAAAACAAA	CCAACAUUGGA	AUUUUGAAC	UGAUAAAAAAC	AGAAGOCAAAC	AGCCUGCCACCCUA. U	AGGAAGUACUGU <sup>180</sup>
a b c	A				c	J G	A UA A U	U
16681 D F G H a	AUAGAGGGAAAAG A U G A A A C	CUAACCAACACA G U G	IACAACA <sup>I</sup> GAAUC G G C	UCCCUCCC U U U U U U	CAACACAAGG U C U	GAACCCAGCC G U	uaaaugaagaagaag c	GACAAAAAGGUUC <sup>27</sup> UGG UG UGG U
b c		G G	G U	A	C A G	AGA U		G
16681 D F G H a b	GUCUGCAAACACI U	G UCCAUGGUAGAC	G G G G	JAAAUGGAU	GUGGACUAUU U C U U C U U	A A A	C	UC C C UC UC UC UC UC UC
16681 D F G H a b c	UGCAAAAAGAACA	AUGGAAGG <mark>A</mark> AAA A A	GUUGUGCAACC A C A C A C A C C C C C	U U U U	uggaaUa <sup>l</sup> cacx U	cauuqudauaaa C G C C	CACCUC <sup>I</sup> ACUCAGGG	GAAGAGCAUGCA*50 G G G C U AG
16681 D F G H a b C	GUCGGAAAUGACA	ACAGGAAA <sup>I</sup> ACAU	GGCAAG <sup>I</sup> GAAAU A	ICAAAAUAA4 A G A G A G A G	u u A	UCCACCACAG U U U U U U U U U U U U U U U U U U U	AAGCAGAAUUGACAG C G A	Edulaucgcacus 4 e C C C C C C C
16681 D F G H a b c	GUCACAAUGGAGR G G G G A Q U	UGCUCUCCÁAGA G G G G G G G G	ACGGGCCUCGA A G A U G A A A A A A G	ACUUCAAUG	AGAUGGUGUC C A	ACUGCAG <sup>A</sup> UGG A A A A	AAAAUAAAGCUUGG G C G C G C C G	CUGGUGCACAGG® 3 0 A U A U A CA U A J
16681 D F G H a b c	CAAUGGUUCCUAC	GACCUGCCGUUA A A G	CCAUGGUUGCC C A C C A C A C U C U C	XCGGAGCGG A A A	ACACACA <sup>1</sup> AGGC A U A A A A A U A A A A A A A A A A A	RICAAAUUGGAI A A A	JACAGA <sup>I</sup> AAGAGACAT A G	C
16681 D F G H a b C	AAAAAUCCCCAUC C	gogaagaa'acag A A	GAUGUUGU C G C A	UUUACGAUX C G G	CCCAAGA <sup>I</sup> AGGC	GCCAUGCACA U U U	CAGCACUUACAGGGC A A C A C A G C C A U	G

16681 D F	CAAAUGUO	CAUCAGGAA	ACUUACUCI	JUCACAG	GACAUCU	CAAGU	GCAGGC	UGAC	GAAUG	GACA	AGCUAC A	AGCU	CAAAGG	iaaugu	CAUA	CUCU» o o
G H a b c	G G G	U	A C G G A G		С				G		A A A A			G		
16681 D F	AUGUOCA( U	CAGGAAAGU	UUAA <sup>l</sup> AGUU A	JUGAAGG	AAAUAGC	CAGAAA	CACAAC	aluco	JAACA G	AUAG	JUAUCA	GAGU	CAAUA	UGAAG	GGGA	CGGCa a a
G H a b c	บ บ บ	с	A G G						G				A A A		A	U U U U
16681	UCUCCAU	GCAAGAUCC	CUUUUGAG	UAAUGG	AUUUGGA	AAAAA	GACAUG	ม่อม	JAGGU	cada	JGAUUA	cadu	CAACOC	AAUUG	UGAC	AGAA1 0 8 0
D F G H	~	A U A U A U		C C				U	c c		c c	1	U A. U	G C	C A	
a b c	c	U A U		С	С			ប G	C C					с с	U A A	
16681 D F	AAAGAUA	GCCCAGUCA	ACAUAGAA	CAGAAC	CUCCAUU	JCCGGAG	ACAGOU	JACAU	JCAUC.	AUAGO	GAGUAG	AGCCX	GGACA	ACUGA U	AGCU A	CAAC <sup>1170</sup>
G H a b c	С	υ	A G			U		υ				A		A UA U GU	A A	
16681 D F	UGGUUUA/	AGAAAGGAA	GUUCUAUCX C	GCCAAA	UGUUUGA	IGACAA	CAAUGA	A A	XGCCG A	AAGA	JAAUQG	ocalu	JUUAGO G	RUGACA	CAGC	CUGG1 2 6 0
G H a b c	С	A G	C C A	G	A	A		A A A	A A C A A A	Α			G C	A		U U
16681 D F	GAUUUUG	GAUCCUUGG C C	GAGGAGUGI	JUACAU	CUAUAGO	AAAGG A A	cucucc	ACC	AGUC	uuula	GAGCAA	uculai	JGGAGC	neccn	UCAG U	FUGGG1 3 5 0
H a b c		C C C C C C		С	с	G			ប ប ប ប	c c	U		G G G	A U	U	с
16681	CT IL LICAL K		AAAUCCUCA	UAGGAG	UCAUUAU	ICACAU	GGAUAG	GAAI	JGAAU	ucalco	CAGCA	ccuc	ACUGUC	UGUGA	CACU	AGUA1 4 4 0
D F	dooodato			С							U			U		
D F G H a b c	c c c c c c	C	A	c c	UG C C	с			A A A A		U U			บ 2 บ บ บ บ บ	υ	
D F G H a b c 16681 D F	C C C C C UUGGUGG	C GAAUUGUGA G C A	A CACURUAU C	c c Jugggałg	UG C C UCAUGG CU	C JGCAGG U	U U		A A A		U U			ບ ເບ ບ ບ ບ ບ ບ	υ	



#### J. Blok et al.: Dengue-2 envelope gene variation

## Nucleotide sequences of the E gene of dengue-2 viruses from Thailand and Sri Lanka

Figure 1 shows the nucleotide sequence of the E gene of the dengue-2 16681 virus isolate, as well as the nucleotides that differ from this sequence for the four Thai (D, F, G, H) and the three Sri Lankan (a, b, c) virus isolates. This figure shows that the nucleotide differences are scattered throughout the length of the gene, and there appears to be no particular region of hypervariability. The numbers of nucleotide differences between each of these viruses and the published dengue-2 viruses JAM [5], S1 [9], AA [8], and NGC [8] are summarized in Table 3. These results indicate that the S1 virus [9], which is a live attenuated vaccine strain from the Puerto Rican isolate PR159, is less closely related to the other dengue-2 sequences showing an average of 9.5% nucleotide sequence difference, compared with the range from 1.5% to 8.7% (average 5.3%) difference among the other dengue-2 virus isolates. There are three pairs of viruses which show very few nucleotide differences; (i) F and G, which are both isolates from the 1980 Bangkok epidemic, show 98.5% sequence similarity, (ii) H and JAM, which represent different geographic regions, show a 98.5% similarity, and (iii) b and NGC, which represent different geographic regions as well as being isolated 24 years apart, show a 98.1% similarity. The E gene sequence of the 1980 Thai isolate AA [8] appears to be similar to three other

Virus isolates <sup>1</sup>													
	16681	D	F	G	Н	AA <sup>2</sup>	а	b	c	<b>S</b> 1 <sup>3</sup>	JAM <sup>4</sup>	NGC <sup>2</sup>	
16681		62	45	49	76	34	93	74	90	131	75	48	
D	12		61	60	48	61	106	85	100	146	53	70	
F	10	11	-	22	94	32	98	85	112	147	90	69	
G	11	9	8	_	82	33	106	83	109	140	90	69	
Η	11	5	12	10	_	82	105	76	95	139	22	61	
AA <sup>2</sup>	6	8	5	5	9		99	81	102	137	79	57	
а	16	16	17	17	17	12	_	101	123	145	109	93	
b	19	18	22	18	17	17	19		84	140	79	28	
с	14	15	18	16	14	13	17	15	_	153	95	68	
<b>S</b> 1 <sup>3</sup>	14	14	15	15	13	12	21	21	21	_	133	121	
JAM <sup>4</sup>	10	6	13	12	5	8	16	18	15	14	_	59	
NGC <sup>2</sup>	8	10	11	11	9	6	13	12	11	12	8	_	

 Table 3. Nucleotide and predicted amino acid differences among twelve dengue-2 virus isolates in the gene coding for the envelope glycoprotein

Numbers above the diagonal represent nucleotide differences among the dengue-2 viruses, while those below the diagonal represent amino acid differences

<sup>1</sup> The full names of the viruses which are represented by codes are listed in Table 1

<sup>2</sup> From Gruenberg et al. [8]

<sup>3</sup> From Hahn et al. [9]

<sup>4</sup> From Deubel et al. [5]

16681 D	MRCIGMSNRDFVEGV	SGGSWVDIVL	EHGSCVTT	MAKNKP	LDFELIKT	EAKQPAT	LRKYCII	SAKLTI	<u>177</u> 7E7 1
F	Ī						M		
G н	I								
a	I L	S							A
b	IA	S					1		D
c	IA	S			N				
16681 D	SRCPTQGEP <sup>I</sup> SLNEEQ	DKRF <sup>I</sup> VCKHSM L	VDRGWGN	GCGLFGK	GGIVTCAMF	RCKKNME T	GKVVQP I	ENLEY	rivit <sup>1 4 2</sup>
F		_		F	:	T	I		V
G u		L				T T	IT		V
a	L					T	1		
b	-	L				т к			
c						T K			
16681	PHSCREHAVCNDTCK	HCKETKTTPC	SSTTEAE	TOVOTV	TMRCSPRTG	DENEMV	LLOMEN	KAWI.VI	HROWF213
D	I HOOHEMATOMDI OK	V	I				D		
F		Y	IA						
G		V	I		v		n		
n a	S	v	T				U	P	
b	-	•	ī	I			D	•	
c			M						
16681 D F G H	LDLPLPWLPGADTQG	SNWIQKETLV	I TFKNPHA	ĸĸġdvvv	LGSQEGAMH	TALTGAT	EIQMSS	GNLLF'	I TGHLK <sup>2 B 4</sup>
a	-								
D C	Т							м	
·									
16681	CRLRMDKLQLKGMSY	SMCTGKFKV	KEIAETQ	HGTIVIR	VQYEGDGSP	CKIPFEI	MDLEKR	HVLGR	LITVN <sup>355</sup>
D F		1							
Ğ									
H		I							
a L					P				
с С					r				
16681 D	IVTEKDSPVNIEAE	L PPFGDSYIII	GVEPGQL	KLNWFKK	GSSIGQMFE	TTMRGAK	RMAILG	DTAVD	FGSLG <sup>426</sup>
G	K								
H									
a h					т			N	
c c					T				
16681	GVFTSIGKALHQVFG	AIYGAAFSC	SWTMKIL	IGVIITW	IGMNSRSTS	LSVTLVL	VGIVTL V	YLGVM	VQA <sup>4 9 5</sup> I.
D						-			-
D F						P			
D F G		_			I	P S		_	
D F G H		Т		v	I I	P S S	V V	AF	
D F G H a b		Т		V	I I I	Р 5 5 5 5	V V L	A F F	

Fig. 2. Envelope protein sequences derived by the translation of the nucleotide sequences shown in Fig. 1 of eight dengue type 2 viruses. Only the amino acids that differ from the 16681 sequence are given for the other viruses, whose full names are described in Table 1. The two potential glycosylation sites (NXT/S) at positions 67 and 153 are underlined; ▼ positions (308 and 491) which gave the highest correlations in the amino acid/symptoms comparisons

isolates from Thailand, the 16681 isolated in 1964 (97.7% similarity) and the 1980 isolates F and G (both show a 97.8% nucleotide sequence similarity).

### Amino acid differences in the E protein of the dengue-2 serotype

The encoded amino acid sequence of the envelope protein from the 16681 virus isolate is shown, in the standard one letter code, in Fig. 2 and only the differences are noted for the other isolates (D, F, G, H, a, b, c—see Table 1 for full names of viruses). The two potential glycosylation sites (NTT at positions 67–69 and NDT at position 153–155) are conserved among all of the dengue-2 viruses sequenced so far. Various amino acid differences are scattered throughout the protein and there is a cluster of conservative changes in the C-terminal portion, which is the hydrophobic membrane anchor of the E protein. Table 3 summarizes the amino acid differences between the Thai and Sri Lankan isolates as well as the published JAM [5], S1 [9], AA [8], and NGC [8] sequences. These results indicate that variability in protein sequence exists (1.0% to 4.0% amino acid sequence difference) although it is far less than the nucleotide divergence which ranges from 1.5% to 10.2% difference. The three pairs which were very closely related at the nucleotide level are not as closely related at the amino acid level indicating that several of the changes are not silent mutations.

#### Relationships

Dendrograms were calculated from the matrices of nucleotide and encoded amino acid differences (Table 3) by the unweighted pair-group method using arithmetic averages (UPMGA) [23, 30]. This UPMGA strategy has been found to be a good and simple method for calculating phylogenetic trees when there is a constant rate of divergence in all lineages [30]. The resulting dendrograms are closely similar in structure (Fig. 3a and b), even though the ranges of differences are not the same, and they show that some of the isolates fall into two clusters. One contains the New Guinea isolate (NGC) and four of the Thai isolates (16681, AA, F, G), and the other contains the Jamaican isolate (JAM) and the remaining two Thai isolates (D and H). The vaccine strain (S1) derived from the Puerto Rican isolate PR159 and two of the Sri Lankan isolates (a and c) are distinct from the cluster and from one another. The third Sri Lankan isolate (b) is anomalous, grouping with the NGC virus in the nucleotide classification and with the Sri Lankan isolate c in the amino acid classification. In another classification of the nucleotide sequences from which the third nucleotide of each codon had been removed, isolate b classified as in the amino acid classification. This indicated that its anomalous behaviour reflected third codon position similarities with NGC that did not correlate with first and second codon position similarities with isolate c.

A programme, SEQCORR, was used to scan the aligned nucleotide and amino acid sequences and seek correlations between the differences at each position and particular characteristics of each dengue virus isolate. Firstly, Dengue-2 envelope gene variation



Fig. 3. Dendrograms illustrating the relationship of **a** the envelope gene sequences and **b** the encoded envelope glycoprotein sequences of 12 dengue type 2 viruses. The full names of the viruses are shown in Table 1

correlations were sought with the severity of the symptoms shown by the patient from whom the isolate was obtained (Table 1). For this purpose, DHF and DSS isolates were pooled and compared with DF isolates. Secondly, correlations were sought with the presence of a primary or secondary serological response in the patient (Table 1). No significant correlations were found; however, the highest correlation in the symptom/amino acid comparison was at position 308 and 491, where two of the DHF/DSS viruses had isoleucine and alanine residues respectively, and the remainder (two DHF/DSS isolates and six DF isolates) had valine residues (see Fig. 2). A further analysis of the amino acid sequence and severity of symptoms was carried out using standard groupings of amino acid (e.g., hydrophobic, polar, aromatic or charged). No further correlations other than the two positions (308 and 491) were found. A graph plotting pairwise nucleotide and amino acid differences (Fig. 4) shows no comparable distinction between DHF/DSS isolates and DF isolates. This is different from results obtained for influenza virus haemagglutinins, where the human or avian origin of these viruses can be distinguished by the relative number of nucleotide and amino acid differences of the haemagglutinin sequences [7]. There seemed to be stronger selection against amino acid changes in bird influenza haemagglutinins than in those from human beings.

In addition, comparisons using the SEQCORR programme found no correlation between the position where there was amino acid variability and the



Fig. 4. A graph plotting pairwise nucleotide and amino acid difference of the DHF/DSS and DF isolates. The circled dots represent DHF/DSS<sup>2</sup> comparisons and this graph shows that these pairs do not form an unusual subset

hydrophilicity of those positions; the variability was calculated by the traditional heterozygosity index [6] and the hydrophilicity by the measures of Kyte and Doolittle [18]. An analysis of the total nucleotide difference between each pair of isolates and the number of years between the times of isolation showed no correlation, the contrary of that found for influenza A orthomyxovirus genes [3].

#### Discussion

The results presented in this paper show that the envelope gene of dengue serotype 2 viruses isolated from patients with clinical symptoms of dengue fever, DHF or DSS from several countries and various years of isolation varies by 1.5% to 10.3% in its nucleotide sequence. Previous studies of dengue viruses have established that clusters of genetic variants occur in this serotype [16, 20,31, 32], and that viruses representing different clusters can occur within one geographic region [31, 32]. The classification of the dengue-2 viruses according to their envelope gene sequences (as shown in the dendrograms in Fig. 3) are complementary to results obtained by RNase T<sub>1</sub> oligonucleotide and restriction enzyme (RE) mapping of the viral RNA genomes. Walker et al. [32] showed that the Thai isolates G and F were closely related (86.5% homology by oligonucleotide fingerprinting and 96.4% by RE mapping) while the Thai isolate H was less closely related to G (58.0% homology by oligonucleotide fingerprinting and 60.7% by RE mapping). These studies also showed that the 16681 virus was more closely related to G (68.8% and 83.6% homology by oligonucleotide and RE mapping, respectively). Based on the genetic mapping and monoclonal antibody studies (solid phase radioimmune and indirect immunofluorescence assays), these authors suggested that the H isolate represented a minor cluster of variants which was different from the other 1980 Bangkok dengue viruses examined.

The classification of the Thai isolates into two groups, where one group contains the 1983 Jamaican isolate (JAM) is similar to results obtained by

Monath et al. [20] using antigen signature analysis. These authors found that the 1981 Jamaican viruses they examined had a relatively close antigenic relationship with several Thai viruses examined, and this was also established by oligonucleotide fingerprinting [20].

Analyses of the sequences for correlations with disease severity, which might confer virulence to the envelope glycoprotein linear sequence, showed that no one amino acid could be associated with the DHF/DSS syndrome. The sample of viruses reported is small but it is representative of the different dengue disease states, including primary and secondary seroresponses. This lack of correlation between sequence and disease symptoms is similar to that found by RNA genome mapping of some of these viruses [32], although Morens and Halstead [21] have reported a good correlation between disease severity and antigenic structure with a similar set of viruses. These latter studies used monoclonal antibodies to a dengue-4 virus to determine the antigenic variation in several dengue-2 viruses from the 1980 Bangkok epidemic, by antibody-dependent infection enhancement assays. These monoclonal antibodies preferentially enhanced the infectivity in vitro of viruses isolated from patients with DHF or DSS, and the authors reported very little difference in reactivity of the D, G, and H isolates but found a marked difference in the dengue fever isolate F. However, the actual protein the dengue-4 monoclonal antibodies are specific for was not determined and similar studies with dengue-2 monoclonal antibodies were unsuccessful in differentiating viruses with respect to disease severity [22]. This suggests that factors other than the linear amino acid sequence of the envelope glycoprotein, such as tertiary or specific steric arrangements of epitopes on the virion or infected cell surface other than those predicted from the E gene sequence, may define virulence determinants.

### Acknowledgements

We would like to thank Robin Hay and Sean McWilliam for expert technical assistance, Drs. A. Nisalak and E. Henchal for recovering some of the Sri Lankan isolates during transport, and Dr. Barry Gorman for providing the dengue virus stocks for some of the isolates. This research was supported by the National Health and Medical Research Council of Australia and the World Health Organization as part of its Programme of Vaccine Development.

#### References

- 1. Blok J, Air GM (1980) Comparative nucleotide sequences at the 3' end of the neuraminidase gene from eleven influenza type A viruses. Virology 107: 50-60
- Blok J, Henchal EA, Gorman BM (1984) Comparison of dengue viruses and some other flaviviruses by cDNA-RNA hybridization analysis and detection of a close relationship between dengue virus serotype 2 and Edge Hill virus. J Gen Virol 65: 2173-2181
- 3. Buonagurio DA, Nakada S, Parvin JD, Krystal M, Palese P, Fitch WM (1986) Evolution of influenza A viruses of 50 years: rapid uniform rate of change in NS genes. Science 232: 980–982

#### J. Blok et al.

- 4. Burke DS, Nisalak A, Johnson DE, McNScott R (1988) A prospective study of dengue infections in Bangkok. Am J Trop Med Hyg 38: 172–180
- Deubel V, Kinney RM, Trent DW (1986) Nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue type 2 virus, Jamaica genotype. Virology 155: 365-377
- 6. Falconer DS (1964) Introduction to quantitative genetics. Oliver and Boyd, London, p 85
- Gibbs A (1987) Molecular evolution of viruses; 'trees', 'clocks' and 'modules'. J Cell Sci [Suppl] 7: 319–337
- Gruenberg A, Woo WS, Biedrzycka A, Wright PJ (1988) Partial nucleotide sequence and deduced amino acid sequence of the structural proteins of dengue virus type 2, New Guinea C and PUO-218 strains. J Gen Virol 69: 1391–1398
- 9. Hahn YS, Galler R, Hunkapiller T, Dalrymple JM, Strauss JH, Strauss EG (1988) Nucleotide sequence of dengue 2 RNA and comparison of the encoded proteins with those of other flaviviruses. Virology 162: 167–180
- Halstead SB, Scanlon JE, Umpaivit P, Udomsakdi S (1969) Dengue and Chikungunya virus infection in man in Thailand, 1962–1964. IV. Epidemiologic studies in Bangkok metropolitan area. Am J Trop Med Hyg 18: 997–1021
- 11. Halstead SB, Nimmannitya S, Cohen SN (1970) Observations related to pathogenesis of dengue hemorrhagic fever. IV. Relation of the disease severity to antibody response and virus recovered. Yale J Biol Med 42: 311–328
- 12. Halstead SB, Simasthien P (1970) Observations related to the pathogenesis of dengue hemorrhagic fever. II. Antigenic and biologic properties of dengue viruses and their association with disease response in the host. Yale J Biol Med 42: 276–292
- 13. Halstead SB (1980) Dengue hemorrhagic fever. A public health problem and a field for research. Bull WHO 58: 1-21
- 14. Halstead SB (1981) The pathogenesis of dengue: molecular epidemiology of infectious disease. Am J Trop Med Hyg 114: 632-648
- 15. Igarashi A (1978) Isolation of a Singh's *Aedes albopictus* cell clone sensitive to dengue and Chikungunya viruses. J Gen Virol 40: 531-544
- Kerschner JAH, Vorndam AV, Monath TP, Trent DW (1986) Genetic and epidemiological studies of dengue type 2 viruses by hybridization using synthetic deoxyoligonucleotides as probes. J Gen Virol 67: 2645–2661
- Kliks SC, Nimmannitya S, Nisalak A, Burke DS (1988) Evidence that maternal dengue antibodies are important in the development of dengue hemorrhagic fever in infants. Am J Trop Med Hyg 38: 411–419
- 18. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132
- 19. Monath TP (1986) Pathology of flaviviruses. In: Schlesinger S, Schlesinger M (Eds) The Togaviridae and Flaviviridae. Plenum, New York, pp 375–440
- Monath TP, Wands JR, Hill LJ, Brown NV, Marciniak RA, Wong MA, Gentry MK, Burke DS, Grant JA, Trent DW (1986) Geographic classification of dengue-2 virus strains by antigen signature analysis. Virology 154: 313–324
- 21. Morens DM, Halstead SB (1987) Disease severity-related antigenic differences in dengue 2 strains detected by dengue 4 monoclonal antibodies. J Med Virol 22: 169–174
- 22. Morens DM, Larsen LK, Halstead SB (1987) Study of the distribution of antibodydependent enhancement determinants on dengue 2 isolates using dengue 2-derived monoclonal antibodies. J Med Virol 22: 163–167
- Rohlf FJ, Wooten MC (1988) Evaluation of the restricted maximum-likelihood method for estimating phylogenetic trees using simulated allele-frequency data. Evolution 42: 581–595

- 24. Rosen L (1977) *The Emperor's New Clothes* revisited, or reflections on the pathogenesis of dengue hemorrhagic fever. Am J Trop Hyg Med 26: 337–343
- Russell PK, Yuill TM, Nisalak A, Udomsakdi S, Gould DJ, Winter PE (1968) An insular outbreak of dengue hemorrhagic fever. II. Virologic and serologic studies. Am J Trop Med Hyg 17: 600–608
- 26. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5467
- Sangkhawibha N (1982) Viral diseases in Thailand—a national overview. In: Mackenzie JS (Ed) Viral diseases in Southeast Asia and the Western Pacific. Academic Press, New York, pp 217–223
- Sangkhawibha N, Rojanasuphot S, Ahandrik S, Viriyapongse S, Jatanasen S, Salitul V, Phanthumachinda B, Halstead SB (1984) Risk factors in dengue shock syndrome: a prospective epidemiologic study in Rayong Thailand. I. The 1980 outbreak. Am J Epidemiol 120: 653–669
- 29. Schlesinger RW (1977) Dengue viruses. Springer, Wien New York [Gard S, Hallauer C (eds) Virology monographs, vol 16]
- 30. Sneath PHA, Cokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. Freeman, San Francisco, p 230
- 31. Trent DW, Grant JA, Rosen L, Monath TP (1983) Genetic variation among dengue 2 viruses of different geographic origin. Virology 128: 271–284
- 32. Walker PJ, Henchal EA, Blok J, Repik PM, Henchal LS, Burke DS, Robbins SJ, Gorman BM (1988) Variation in dengue type 2 viruses isolated in Bangkok during 1980. J Gen Virol 69: 591-602
- Westaway EG, Brinton MA, Gaidamovich SYa, Horzinek MC, Igarashi A, Kääriäinen L, Lvov DK, Porterfield JS, Russell PK, Trent DW (1985) Flaviviridae. Intervirology 24: 183–192
- 34. World Health Organization (1975) Technical guides for diagnosis, treatment, surveillance, prevention and control of dengue hemorrhagic fever. WHO, Geneva

Authors' address: Dr. J. Blok, Sir Albert Sakzewski Virus Research Laboratory, Royal Children's Hospital, Herston Road, Brisbane, Queensland 4029, Australia.

Received November 8, 1988