

NUCLEOTIDE SEQUENCE AND DEDUCED AMINO ACID SEQUENCE OF THE NONSTRUCTURAL PROTEINS OF DENGUE TYPE 3 VIRUS, BANGKOK GENOTYPE

WH Attatippaholkun¹, MK Attatippaholkun², A Nisalak³, DW Vaughn³ and BL Innis⁴

¹Department of Clinical Chemistry, Faculty of Medical Technology, Mahidol University; ²Thailand Institute of Scientific and Technological Research, Bangkok, Thailand; ³Department of Virology, US Army Medical Component, Armed Forces Research Institute of Medical Science, Bangkok, Thailand;

⁴Department of Virus Diseases, Walter Reed Army Institute of Research, Washington DC, USA

Abstract. The nucleotide sequence of the nonstructural protein gene (1,610 bases) of dengue 3 virus (Bangkok genotype; CH53489 isolated in 1973) has been determined in both forward and reverse directions. The PCR based cycle sequencing technique by the enzymatic method of Sanger *et al* using a sequencing primer 5'-end labeled with γ -³²P-ATP was the method of our choice for sequence analysis. Two cDNA templates were prepared by RT-PCR technique starting from the nucleotides 6,306-6,969 and 6,925-7,915 of the dengue 3 genome with the lengths of 663 and 990 base pairs respectively. In our cycle sequencing experiments, it has been observed that the substitution of 7-deaza-dG for dG in DNA eliminated most of the secondary structures that produce gel artifacts. The final sequence results of these two cDNA templates were established from their sequence data determined on both strands in opposite directions. Alignment between the newly established nucleotide sequences as well as their deduced amino acid sequences of the Bangkok dengue 3 (CH53489) virus and the published sequence data of the dengue 3 prototype (H87) was manipulated by the PC-DOS-GIBIO DNASIS TM 06-00 software. The homology of the nucleotide sequences between the two dengue 3 viruses was 96.65%. The deduced amino acid sequence from nucleotides 6,306-7,915 of the two viruses showed conserved amino acids of the nonstructural protein NS4a and 6 amino acid changes in NS4b and NS5.

INTRODUCTION

Dengue viruses is a growing health problem in much of the tropical world. In Southeast Asia, the frequency of epidemics is increasing; the trend is toward larger epidemics and more severe disease. Because control of the dengue vector *Aedes aegypti* has almost completely lapsed on a world-wide scale, dengue is now the most prevalent serious arthropod-borne viral disease of humans, a distinction not likely to change.

Dengue viruses are members of the Flaviviridae and can be divided into four antigenic serotypes (types 1 to 4) which are distinguished by serological cross-reactivity. The four dengue virus serotypes have different pathogenicities for humans, ranging from nondescript febrile illness (DF) to the dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS). Temporal and geographic distribution of specific dengue virus genotypes have been determined with serologic and molecular markers (Monath *et al*, 1986), (Trent *et al*, 1983). In addition, it is important to obtain genetic information in all the dengue virus genotypes in order to understand the causes of the severe diseases and to develop an effective dengue vaccine.

The dengue virus genome consists of a single positive-strand RNA molecule that is about 11 kilobases (Kb) in length and lacks a 3'-poly (A) tract. It encodes a single polypeptide of about 3,400 amino acid residues in the gene order 5'-C'-prM-M-E-NS1-NS2a-NS2b-NS3-NS4a-NS4b-NS5-3' (Rice *et al*, 1985). Individual proteins appear to be processed by post-translational cleavage of the polypeptide at specified sites recognized by cellular or virus-encoded proteases (Rice *et al*, 1985; Speight *et al*, 1988). The size and functions of the virion structural proteins have been characterized (Rice *et al*, 1985); however, the functions of the nonstructural proteins have not been assigned. Proteins NS3 and NS5 may have enzymatic roles in RNA replication (Mackow *et al*, 1987). The 5'- and 3'-extremities of the flavivirus genome are not translated and contain stem-and loop-structures that may be involved in regulation of transcription or replication (Rice *et al*, 1985; Wengler and Castle, 1986; Sumiyoshi *et al*, 1987).

Dengue 3 viruses have been isolated from patients in Bangkok every year since 1973 when a longitudinal study of dengue in hospitalized patients was begun by the Department of Virology, Armed Forces Research Institute of Medical Science (AFRIMS). Beginning in 1984, the proportion

of dengue-3 isolates responsible for severe disease rose dramatically, culminating in the massive epidemic of 1987-1988. An analysis of frequency of death among patients infected with each of the 4 dengue serotypes over an 18 year period from 1973-1990 identified a greater than expected frequency of death among patients infected with dengue 3, especially during the 1987-1988 epidemic.

The complete nucleotide sequence of the dengue-3 prototype, strain H87, has already been published since 1990 (Osatomi *et al*, 1990). We now report the nucleotide sequence and deduced amino acid sequence of the genome encoding the nonstructural proteins, NS4a, NS4b and NS5 of the Bangkok dengue 3 strain CH53489 isolated in 1973 and compare it with the dengue 3 prototype, H87 (Osatomi *et al*, 1990). These have provided new information on flavivirus evolution and replication.

MATERIALS AND METHODS

Virus culture and RNA extraction

The dengue 3 virus, strain CH53489 (provided by Dr BL Innis, AFRIMS Bangkok) was cultured in *Aedes albopictus* C6/36 cells in the laboratory of Dr DW Vaughn (Virology Department, AFRIMS Bangkok). Total genomic RNA was extracted from the frozen viruses using the acid guanidinium thiocyanate-phenol-chloroform method (Chomczynski and Sacchi, 1987).

Viral cDNA synthesis

Two cDNA templates were synthesized from viral genomic RNA by RT-PCR (Reverse transcription and Polymerase Chain Reaction) technic using AMV reverse transcriptase for cDNA synthesis and Taq DNA polymerase for PCR. Two pairs of primers were used in the RT-PCR to synthesize and amplify their designed PCR products at 663 (the nucleotides 6,306-6,969) and 990 (the nucleotides 6,925-7,915) base pairs respectively.

Sequence analysis of the Bangkok dengue 3 (CH53489)

The cycle sequencing by the modified enzymatic method of Sanger *et al* (Sanger *et al*, 1977; Innis *et al*, 1988; Murray, 1989) using a sequencing primer 5'-end labelled with γ - 32 P-ATP was chosen for sequence analysis. The nucleotide sequences have been determined in both forward and reverse directions. The strategy of primer design for nucleotide sequence walking of the Bangkok Dengue 3 (CH53489) RNA genome in both directions was illustrated in Fig 1. In all our sequencing experiments, the cycle sequencing method utilized the Amplicycle Sequencing kit (Perkin Elmer, USA) which used AmpliTaq DNA polymerase CS for generating high-quality sequencing ladders. Three sequencing gels of each sample were run, using Sequi-GenII Sequencing Cell and a Power Pac 3000

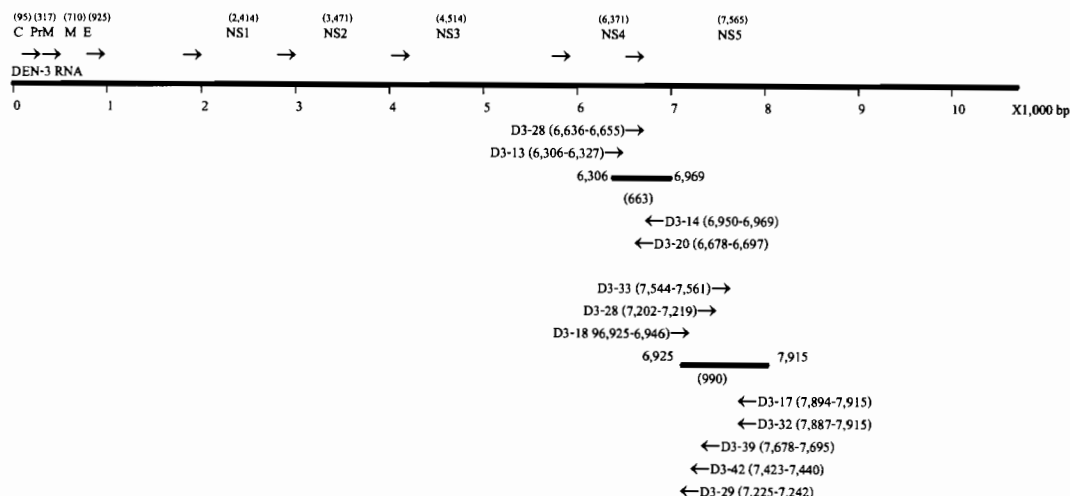


Fig 1-The strategy of the cDNA template preparation as well as the primer design for nucleotide sequence walking on the RNA genome of the Bangkok dengue 3 virus (CH53489).

1	AT	ACC	CCC	ACT	TAT	TCA	GAT	CCC	TTG	GCA	CTC	AAG	GAA	TTC	AAG	GAT	47
1		A	R	T	T	S	D	P	L	A	L	E	E	F	E	D	15
48	TTT	GCA	GCT	GGC	AGA	AGA	TCA	ATC	CCC	CTT	GAT	CTT	GTG	ACA	GAA	ATA	95
48	P	A	A	G	R	K	S	I	A	L	D	L	V	T	E	I	31
96	GGA	AGA	GTG	CTT	TCA	TAT	TTA	CCC	CAC	AGA	ACG	AGA	AAC	CCC	CTG	GAC	143
96	G	R	V	P	S	H	L	A	H	R	T	R	N	A	L	D	47
144	AAT	TTG	GTG	ATG	CTG	CAC	ACG	TCA	GAA	CAT	GCC	GGT	AGG	AGC	TAC	AGG	191
144	H	L	V	T	H	L	H	T	S	E	H	G	G	T	A	R	63
192	CAT	GCA	GTG	GAG	GAA	CTA	CCA	GAA	ACA	ATG	GAA	ACA	CTC	CTA	CTC	TTG	239
192	H	A	V	E	E	L	P	E	T	H	E	T	L	L	L	L	79
240	GGA	CTC	ATG	ATC	TTG	TTA	ACA	GGT	GGA	ATG	CTG	TTC	TTG	ATA	TCA	287	
240	G	L	H	I	L	L	T	G	G	A	M	L	F	L	I	S	95
288	GCT	AAA	GGG	ATT	GGA	AAG	ACT	TCA	ATA	GGA	CTC	ATC	TGT	TTA	ATT	GCT	335
288	G	E	G	I	G	E	T	S	I	G	L	I	C	V	I	A	111
336	TCC	AGT	GGC	ATG	TTA	TGG	ATG	GCC	GAT	CTC	CCA	CTT	GAA	TGG	ATG	GCG	383
336	S	S	G	H	L	M	H	A	D	V	P	L	Q	W	I	A	127
384	TGG	GCT	ATA	ATC	CTG	GAA	TTT	TTT	ATG	ATG	TTG	CTT	ATA	CCA	GAA	431	
384	S	A	I	V	L	E	F	F	M	H	V	L	I	P	E		143
432	CCA	GAA	EAG	CAG	AGA	ACT	CCC	GAA	CAC	AAC	CAA	CTC	GCA	TAT	CTC	GTG	479
432	P	E	E	K	Q	R	T	P	Q	D	N	Q	L	A	Y	V	189
480	ATA	GCG	ATA	CTT	ACA	CTG	CTC	GCA	ATA	ACA	CCC	AAT	GAA	ATG	GGA	527	
480	I	G	I	L	T	L	A	A	I	V	A	A	N	E	N	G	175
528	CTG	TTG	GAA	ACT	TCA	AAG	AGA	GAT	TTA	GGA	ATG	TCT	AAA	GAA	GCG	575	
528	L	L	E	T	T	E	R	D	L	G	M	S	K	E	P	G	191
576	GTT	GTT	TCT	CCA	ACC	AGC	TAT	TTA	GAT	ATA	GCA	CTC	ACA	GCA	TCA	623	
576	V	V	S	P	T	S	Y	L	D	V	D	L	H	P	A	S	207
624	GCC	TGG	ACA	TTG	TAC	GCT	GTG	GCC	ACA	ACA	CTA	ATA	AGA	CCA	ATG	TTG	671
624	A	M	T	L	Y	A	V	A	T	Z	V	I	T	P	H	L	223
672	AGA	CAT	ACC	ATA	GAG	AAT	TCC	ACA	AAT	GTG	TCC	CTG	GCA	GCT	ATA	719	
672	R	H	T	I	E	N	S	T	A	N	V	S	T	P	H	A	239
720	GCT	AAC	CAG	GCA	GTG	CTG	CTG	ATG	GCC	TTA	GAC	AAA	GGA	TGG	CCA	ATA	767
720	A	N	G	A	V	L	H	G	L	D	E	G	W	P	I		255
768	TGG	AAA	ATG	GAC	TTA	GCC	ATC	CCA	CTA	CTG	GCA	CTG	GCG	TGC	TAT	TCA	816
768	S	E	H	D	L	G	V	P	L	G	C	G	T	C	Y	S	271
816	GAA	GTG	AAC	CCA	CTG	ATC	CTC	ACA	GCG	GCA	ATA	CTT	TTG	CTA	GCT	ACA	863
816	Q	V	N	P	L	T	L	T	A	A	V	L	L	L	A	T	267
864	CAT	TAT	GCT	ATT	ATA	GCT	CCA	GGA	TTG	CAG	GCA	AAA	GCC	ACT	GTT	GAA	911
864	H	Y	A	I	I	G	P	G	L	Q	A	E	A	T	R	E	303
912	GCT	CAA	AAA	AGG	ACA	GCT	GCT	GGA	ATA	ACG	AAG	AAC	CCA	ACG	GTG	GAT	959
912	A	Q	E	R	T	A	A	G	I	T	E	N	P	T	V	D	319
960	GCG	ATA	ATG	ACA	ATA	GAC	CTA	GAT	CTT	GTA	ATA	TAC	GAT	TCA	AAA	TTC	1007
960	G	I	H	T	I	D	L	D	P	V	I	T	C	A	K	F	335
1008	GAA	AAG	CAA	CTA	GGA	CAG	GTT	ATG	CTC	CTA	CTT	TGT	GCA	GTT	CAA	1085	
1008	E	E	K	L	G	Q	V	H	L	L	V	L	C	A	V	Q	351
1056	CTT	TTG	TTA	ATG	AGA	ACA	TCA	TGG	GCC	TTG	TCT	GAA	GCT	CTA	ACC	CTA	1103
1056	L	L	L	L	H	R	T	S	M	A	L	T	S	E	A	L	367
1104	GCC	ACA	GGA	CCA	ATA	ACA	CTC	TGG	GAA	GGA	TCA	CGT	GAA	AAG	TTT	1151	
1104	A	T	G	P	T	I	A	V	S	H	A	N	I	F	R	G	383
1152	TGG	AAC	ACC	ACG	ATA	GCT	GTT	TCC	ATG	GCT	AAC	ATC	TTT	AGA	GGG	AGC	1199
1152	M	N	T	T	I	A	V	S	H	A	N	I	F	R	G	S	399
1200	TAT	TTA	GCA	GGA	ACT	GGG	CTT	GCT	TTT	TCT	ATC	ATG	AAA	TCA	GTT	GGA	1247
1200	Y	L	A	G	A	G	L	A	F	S	I	H	K	S	V	G	415
1248	ACA	GGA	AAG	AGA	GGA	ACG	GGG	TGA	CAA	GCC	GAA	ACC	TTA	GGA	GAA	AAG	1295
1248	T	G	E	R	G	T	G	S	Q	G	S	T	L	G	E	E	431
1296	TGG	AAA	AAG	AAA	TTA	AAT	CAG	TTA	TCC	CGG	AAA	GAG	TTT	GAC	CTT	TAC	1343
1296	M	K	K	E	L	N	Q	G	L	S	R	K	E	F	D	L	447
1344	AAG	AAA	TCT	GGA	ATC	ACT	GAA	GTG	GAT	AGA	GAA	CCC	AAA	GAA	GGG	1391	
1344	K	K	S	Q	I	T	E	V	D	R	T	E	A	K	E	G	463
1392	TTG	AAA	AGA	GGA	ATA	ACA	GCT	GAT	GCC	GTG	TCC	AGA	GCC	AGC	GCA	1439	
1392	L	E	R	G	E	I	T	R	H	A	V	S	R	G	S	A	479
1440	AAA	CTT	CAA	TGG	TTG	CTG	GAG	AGG	AAC	ATG	CTC	ATT	CCC	GAA	GGA	AGA	1487
1440	E	L	Q	W	F	V	E	R	N	H	V	I	P	E	Q	R	495
1488	GTG	ATA	GAC	TTA	GCC	TGT	GGA	AGA	GGA	GCC	TGG	TCA	TAT	TAC	TGC	ACA	1535
1488	V	I	D	L	Q	C	G	R	G	G	M	S	Y	Y	C	A	511
1536	GGA	TTG	AAA	AAA	ATT	ACA	GAA	GTG	GGA	TAC	ACA	AAA	GGC	GGC	GGC	ACA	1583
1536	G	L	E	E	K	V	T	E	V	R	G	Y	T	E	K	G	527
1584	GGA	CAC	GAA	GAA	CTA	CCT	ATG	TCT									1610
1584	G	H	E	E	P	V	P	H	S								536

Fig 2-The nucleotide and deduced amino acid sequences of the nonstructural proteins NS4a, NS4b and NS5 of the Bangkok dengue 3 virus (CH53489).

power supply (BIORAD, USA). For autoradiography, the ³²P-sequencing gel was exposed to Dupont REFLECTION Autoradiography film with the REFLECTION intensifying screen (Dupont NEN, USA).

RESULTS

The nucleotide sequence of the nonstructural protein gene (1,610 bases) of the Bangkok dengue

3 virus (CH53489) was determined in forward and reverse directions following the strategy of the cDNA template preparation as well as the primer design for nucleotide sequence walking on the RNA genome of the virus shown in Fig 1. The nucleotide sequence from 6,306 to 7,915 and the deduced amino acids of the nonstructural proteins NS4a, NS4b and NS5 of the Bangkok dengue 3 virus were established (Fig 2). Alignment between the newly established nucleotide sequences of the Bangkok dengue 3 virus and the published nucleotide sequence of the dengue 3 prototype virus (H87) (Osatomi *et al*, 1990) was manipulated by the PC-DOS-GIBIO DNASIS TM 06-00 software. The homology of the nucleotide sequences between the two dengue 3 viruses revealed 96.65 % (Fig 3). The deduced amino acid sequence from the nucleotide 6,306-7,915 of the two viruses showed conserved amino acids of the nonstructural protein NS4a and 6 amino acid changes in NS4b and NS5 (Fig 4).

DISCUSSION

The primary structure of the nucleotide sequence from 6,306-7,915 of the Bangkok dengue 3 (CH53489) genome coding for its nonstructural proteins NS4a, NS4b and NS5 was reported (Fig 2). Highly homologous nucleotide and amino acid sequences were observed among the Bangkok dengue 3 virus (CH53489) and the prototype dengue 3 virus (H87), (Osatomi and Sumiyoshi, 1990). The homology of the nucleotide sequences between the two dengue 3 viruses revealed 96.65% (Fig 3). The results showed that the degree of similarity varied substantially for specific coding regions. Alignments of the deduced amino acid sequences coding from the nucleotides 6,306-7,915 of these viruses indicated that the amino acid sequence coding from the nucleotide 6,306 to 7,141 was completely conserved whereas 6 amino acids were changed at the nucleotide 7,142; 7,163; 7,244; 7,397; 7,529 and 7,718 (Fig 4). Our data suggested that the amino acid sequence of NS4a protein was more conserved than those of NS4b and NS5 proteins.

Variations in the dengue 3 genome have previously been analyzed (Chungue *et al*, 1993; Lanciotti *et al*, 1994). Genetic evolution among the prM/M and E structural protein genes of dengue 3 viruses has occurred independently within geographical regions where the viruses are endemic (Lanciotti *et al*, 1994). In spite of this evolution, the prM/M

H87	6306	6315	6325	6336	6345	6355	6366
	ATGCCGCACTTATTGAGATCTTTAGCACTCAAGAAATCAAGGATTTGCAAGTGGCA						
CH53489	6306	6315	6325	6336	6345	6355	6366
	ATGCCGCACTTATTGAGATCTTTAGCACTCAAGAAATCAAGGATTTGCAAGTGGCA						
	6368	6375	6385	6395	6405	6415	6425
	GAAAGTCAATGGCCCTTGATCTTTGAGCAAAATAGGAGAGTGCCTTCACACTTAAGCCC						
	6368	6375	6385	6395	6405	6415	6425
	GAAAGTCAATGGCCCTTGATCTTTGAGCAAAATAGGAGAGTGCCTTCACACTTAAGCCC						
	6426	6435	6445	6455	6465	6475	6485
	ACAGAAAGGAAAGCGCCCTGGCAATTTGGTGATGCTGCACAGTCAGAACATGGCGGTA						
	6426	6435	6445	6455	6465	6475	6485
	ACAGAAAGGAAAGCGCCCTGGCAATTTGGTGATGCTGCACAGTCAGAACATGGCGGTA						
	6486	6495	6505	6515	6525	6535	6545
	GGGCGTACAGGATGAGTGGAGGAACTTACAGAAAGGATGGAAACACTCTTACTCGCTTG						
	6486	6495	6505	6515	6525	6535	6545
	GGGCGTACAGGATGAGTGGAGGAACTTACAGAAAGGATGGAAACACTCTTACTCGCTTG						
	6546	6555	6565	6575	6585	6595	6605
	GACTGATGATCTTTTAAAGGTTGGAGCAATGCTCTTCTGATATCAGGTAAGGGGATG						
	6546	6555	6565	6575	6585	6595	6605
	GACTGATGATCTTTTAAAGGTTGGAGCAATGCTCTTCTGATATCAGGTAAGGGGATG						
	6606	6615	6625	6635	6645	6655	6665
	GAAAGACTTCAATAGGACTCATCTGTGTAATTTGCTTCCAGTGGCAATTTGATGGATGGCTG						
	6606	6615	6625	6635	6645	6655	6665
	GAAAGACTTCAATAGGACTCATCTGTGTAATTTGCTTCCAGTGGCAATTTGATGGATGGCTG						
	6666	6675	6685	6695	6705	6715	6725
	ATGTCCCACTTCCAAATGAGTCCGCTATAGTCCGCTATAGTCCGCTATAGTCCGCTATAGTCC						
	6666	6675	6685	6695	6705	6715	6725
	ATGTCCCACTTCCAAATGAGTCCGCTATAGTCCGCTATAGTCCGCTATAGTCCGCTATAGTCC						
	6726	6735	6745	6755	6765	6775	6785
	TCATACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG						
	6726	6735	6745	6755	6765	6775	6785
	TCATACCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG						
	6786	6795	6805	6815	6825	6835	6845
	TAGGCATACCTTACATTTGGCTGCAATAGTGGGCGCAATGAAATGGAGCTGTTGGAACTTA						
	6786	6795	6805	6815	6825	6835	6845
	TAGGCATACCTTACATTTGGCTGCAATAGTGGGCGCAATGAAATGGAGCTGTTGGAACTTA						
	6846	6855	6865	6875	6885	6895	6905
	CAAGAGAGATTTAGGAATGTTCTAAAGAACCAAGTGTGTTTCTCCAAACAGCTATTGG						
	6846	6855	6865	6875	6885	6895	6905
	CAAGAGAGATTTAGGAATGTTCTAAAGAACCAAGTGTGTTTCTCCAAACAGCTATTGG						
	6906	6915	6925	6935	6945	6955	6965
	ATGTGGACTTGCACCCAGCATGAGCTGGCAATTTGTAAGCGCTGGCCACCAACAGTAATAA						
	6906	6915	6925	6935	6945	6955	6965
	ATGTGGACTTGCACCCAGCATGAGCTGGCAATTTGTAAGCGCTGGCCACCAACAGTAATAA						
	6966	6975	6985	6995	7005	7015	7025
	CACCAATGTTGAGACACCAATAGAGAAATTCACAGCAAAATGTTGCTCCGAGGACATAG						
	6966	6975	6985	6995	7005	7015	7025
	CACCAATGTTGAGACACCAATAGAGAAATTCACAGCAAAATGTTGCTCCGAGGACATAG						
	7026	7035	7045	7055	7065	7075	7085
	CTAACAGGCAAGTGTCTGATGGGTTAGAGCAAAAGATGGCGATATGAAATGGAGT						
	7026	7035	7045	7055	7065	7075	7085
	CTAACAGGCAAGTGTCTGATGGGTTAGAGCAAAAGATGGCGATATGAAATGGAGT						
	7086	7095	7105	7115	7125	7135	7145
	TGGGCGTACCACTATTGGCACTGGGTTGCTATTCAAGTGAACCACTAACTCTTATAG						
	7086	7095	7105	7115	7125	7135	7145
	TGGGCGTACCACTATTGGCACTGGGTTGCTATTCAAGTGAACCACTAACTCTTATAG						
	7146	7155	7165	7175	7185	7195	7205
	GGGCACTACTTTTCTGATGCTACACATTATGCAATTATAGTCCAGGATTTGCAAGGCAAAA						
	7146	7155	7165	7175	7185	7195	7205
	GGGCACTACTTTTCTGATGCTACACATTATGCAATTATAGTCCAGGATTTGCAAGGCAAAA						
	7206	7215	7225	7235	7245	7255	7265
	GCCACTGTTGAAGCTCAGAAAAAGGACAGCTGCTGGAAATGAAGAATCCAAAGGATGAT						
	7206	7215	7225	7235	7245	7255	7265
	GCCACTGTTGAAGCTCAGAAAAAGGACAGCTGCTGGAAATGAAGAATCCAAAGGATGAT						
	7266	7275	7285	7295	7305	7315	7325
	GGAATATGCAATAGACCTAGATCTGTATATATGATTCACAAATTTGAAAGCAACTTA						
	7266	7275	7285	7295	7305	7315	7325
	GGAATATGCAATAGACCTAGATCTGTATATATGATTCACAAATTTGAAAGCAACTTA						
	7326	7335	7345	7355	7365	7375	7385
	GGACAGGTTATGCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG						
	7326	7335	7345	7355	7365	7375	7385
	GGACAGGTTATGCTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG						
	7386	7395	7405	7415	7425	7435	7445
	GCTTGTGTGAAGCTCTAACCTAGCCACAGGACCAATTAACCAACTCTGGGAAGGATCA						
	7386	7395	7405	7415	7425	7435	7445
	GCTTGTGTGAAGCTCTAACCTAGCCACAGGACCAATTAACCAACTCTGGGAAGGATCA						
	7446	7455	7465	7475	7485	7495	7505
	CTGGGAAGTCTTGGAAACCAAGATGCTGCTTCCATGCGCAACTCTTTAGAGGAGGAC						
	7446	7455	7465	7475	7485	7495	7505
	CTGGGAAGTCTTGGAAACCAAGATGCTGCTTCCATGCGCAACTCTTTAGAGGAGGAC						
	7506	7515	7525	7535	7545	7555	7565
	TATTTAGCAGGAGCTGGGCTGCTCTTCTATCATGAATCAGTTGGAAACAGGAAAGAGA						
	7506	7515	7525	7535	7545	7555	7565
	TATTTAGCAGGAGCTGGGCTGCTCTTCTATCATGAATCAGTTGGAAACAGGAAAGAGA						
	7566	7575	7585	7595	7605	7615	7625
	TATTTAGCAGGAGCTGGGCTGCTCTTCTATCATGAATCAGTTGGAAACAGGAAAGAGA						
	7566	7575	7585	7595	7605	7615	7625
	TATTTAGCAGGAGCTGGGCTGCTCTTCTATCATGAATCAGTTGGAAACAGGAAAGAGA						
	7626	7635	7645	7655	7665	7675	7685
	TATTTAGCAGGAGCTGGGCTGCTCTTCTATCATGAATCAGTTGGAAACAGGAAAGAGA						
	7626	7635	7645	7655	7665	7675	7685
	TATTTAGCAGGAGCTGGGCTGCTCTTCTATCATGAATCAGTTGGAAACAGGAAAGAGA						
	7686	7695	7705	7715	7725	7735	7745
	GCCAAAGAGGTTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG						
	7686	7695	7705	7715	7725	7735	7745
	GCCAAAGAGGTTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG						
	7745	7755	7765	7775	7785	7795	7805
	AAACTTCAATGTTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG						
	7745	7755	7765	7775	7785	7795	7805
	AAACTTCAATGTTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG						
	7805	7815	7825	7835	7845	7855	7865
	GGCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG						
	7805	7815	7825	7835	7845	7855	7865
	GGCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG						
	7865	7875	7885	7895	7905	7915	7925
	CGAGGATACACAAAAGGGGGCCAGGACACGAAAGAACAGTACCTATGCT						
	7865	7875	7885	7895	7905	7915	7925
	CGAGGATACACAAAAGGGGGCCAGGACACGAAAGAACAGTACCTATGCT						

Fig 3-Alignment of the nucleotide sequences from the nucleotides 6,306-7,915 on the RNA genome between the Bangkok dengue 3 virus (CH53489) and the published nucleotide sequence data of the prototype dengue 3 virus (H87).

GENE SEQUENCE OF DENGUE 3 VIRUS

```

H87      2072'  ARTYSDPLALKEFKDFAAGRKSIALDLVT
*****
CH53489  2072"  ARTYSDPLALKEFKDFAAGRKSIALDLVT

2101'  EIGRVPSHLAHRTRNALDNLVMLHTSEHGGRAYRHAEEELPETMETLLLLGLMILLTGA
*****
2101"  EIGRVPSHLAHRTRNALDNLVMLHTSEHGGRAYRHAEEELPETMETLLLLGLMILLTGA

2161'  MLFLISGKGIGKTSIGLICVIASSGMLWMADVPLQWIASAIVLEFFMMVLLIPEPEKQRT
*****
2161"  MLFLISGKGIGKTSIGLICVIASSGMLWMADVPLQWIASAIVLEFFMMVLLIPEPEKQRT

2221'  PQDNQLAYVVGIGILTAAIVAANENGLETTKRDGLMSKEPGVVSPTSYSYLDVDLHPASAW
*****
2221"  PQDNQLAYVVGIGILTAAIVAANENGLETTKRDGLMSKEPGVVSPTSYSYLDVDLHPASAW

2281'  TLYAVATTVITPMLRHTIENSTANVSLAAIANQAVVLMGLDKGWFISKMDLGVPLALGC
*****
2281"  TLYAVATTVITPMLRHTIENSTANVSLAAIANQAVVLMGLDKGWFISKMDLGVPLALGC

2341'  YSQVNPLTLIAAVLLLVTHYAIIGPGLQAKATREAQKRTAAGIMKNPTVDGIMTIDLDPV
*****
2341"  YSQVNPLTLIAAVLLLVTHYAIIGPGLQAKATREAQKRTAAGITKNPTVDGIMTIDLDPV

2401'  IYDSKFEKQLGQVMLLVLCVQLLLMRTSWALCEVLTATGPITTLWEGSPGKFWNTTIA
*****
2401"  IYDSKFEKQLGQVMLLVLCVQLLLMRTSWALCEALTATGPITTLWEGSPGKFWNTTIA

2461'  VSMANIFRGSYLAGAGLALSINKSVGTGKRGTSQGETLGEKWKKKLNQLSRKEFDLYK
*****
2461"  VSMANIFRGSYLAGAGLAFSINKSVGTGKRGTSQGETLGEKWKKKLNQLSRKEFDLYK

2521'  SGITEVDRTAEKGLKRGIEITHAVSRGSAKLQMFVERNMVIPEGRVIDLCCGROGWSYY
*****
2521"  SGITEVDRTAEKGLKRGIEITHAVSRGSAKLQMFVERNMVIPEGRVIDLCCGROGWSYY

2581'  CAGLKKVTEVRGYTKGGPGHEEPVMS 2607'
*****
2581"  CAGLKKVTEVRGYTKGGPGHEEPVMS 2607"

```

A

```

                2100
H87      ARTYSDPLA LKEFKDFAAG RKSIALDLVT EIGRVPSHLA HRTNALDNLV VMLHTSEHGGRAYRHAEEEL PETMETLLLL GLMILLTGA
CH53489  .....>

                2200
H87      MLFLISGKGI GKTSGLICV IASSGMLWMA DVPLQWIASA IVLEFFMMVL LIPEPEKQRT PQDNQLAYVV IGILTAAIV AANENGLETT
CH53489  .....>

                2300
H87      TKRDGLMSKE PGVVSPTSYSY LDVDLHPASAW TLYAVATTVI TPMLRHTIEN STANVSLAAI ANQAVVLMGL DKGWFISKMD LGVPLALGC
CH53489  .....>

                2400
H87      YSQVNPLTLI AAVLLLVTHY AIIGPGLQAK ATREAQKRTA AGIMKNPTVD GIMTIDLDPV IYDSKFEKQL GQVMLLVLCV VQLLLMRTSW
CH53489  .....T.....A.....T.....>

                2500
H87      ALCEVLTAT GPITTLWEGS PGKFWNTTIA VSMANIFRGS YLAGAGLALS IMKSVGTGKR GTGSQGETLG EKWKKKLNQL SRKEFDLYK
CH53489  .....A.....F.....>

                2600
H87      SGITEVDRTAE KGLKRGIEI THAVSRGSA KLQMFVERNM VIPEGRVIDL CCGRGOWSYY CAGLKKVTEV RGYTKGGPGH KEPVMS
CH53489  .....R.....>

```

B

Fig 4 a,b—Alignment of amino acid of the nonstructural proteins NS4a, NS4b and NS5 between the Bangkok dengue 3 virus (CH53489) and the published amino acid sequence data of the prototype dengue 3 virus (H87).

and E proteins have retained an amino acid sequence similarity greater than 95% over the 36 year period studied (Lanciotti *et al*, 1994). These data

suggest that domains responsible for predicted flavivirus protein architecture and/or biological function are strictly conserved. It is likely that the

random mutation rate for dengue virus RNA is similar to that of other RNA viruses (Holland *et al*, 1982), however, the mutations in the dengue 3 virus genome appear to be buffered by selective pressure. Dengue viruses, and other arthropod-borne viruses, seem to have a genetic stringency imposed on them because they replicate in both arthropod and vertebrate hosts (Weaver *et al*, 1991). The sequence information of the Bangkok dengue 3 virus (CH53489) presented herein permits selection of conserved and variable nucleotide regions in dengue 3 viral RNA for synthesis of complementary oligonucleotides that could be useful in strain typing and evolution studies.

ACKNOWLEDGEMENTS

This research was supported under Grant No. 493-5600-G-00-3461. Program in Science and Technology Cooperation (PSTC), Human Capacity Development, Bureau for Global Programs, Field Support and Research, US Agency for International Development (USAID). The authors are grateful to Ms Ganlayarat Sirisripetch and Ms Yupawadee Teamkaew for their skillful research assistance and also thank Ms Ganlayarat Sirisripetch and Mr Benjapon Samranthin for preparing the manuscript.

REFERENCES

- Chomczynski P, Sacchi N. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987 ; 162 : 156-9.
- Chungue E, Deubel V, Cassar O, Laille M, Martin PMV. Molecular epidemiology of dengue 3 viruses and genetic relatedness among dengue 3 strains isolated from patients with mild or severe form of dengue fever in French Polynesia. *J Gen Virol* 1993 ; 74 : 2765-70.
- Holland SK, Horodyski F, Grabau E, Nichol S, Vandepol F. Rapid evolution of RNA genomes. *Science* 1982; 215 : 1577-85.
- Innis MA, Myambo KB, Gelfand DH, Brow MAD. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction- amplified DNA. *Proc Natl Acad Sci USA* 1988 ; 85 : 9436-40.
- Lanciotti RS, Lewis JG, Gubler DJ, Trent DW. Molecular evolution and epidemiology of dengue 3 viruses. *J Gen Virol* 1994 ; 75 : 65-75.
- Mackow E, Makino Y, Zhao B, *et al*. The nucleotide sequence of dengue 4 virus. Analysis of genes coding for nonstructural proteins. *Virology* 1987 ; 159 : 217-28.
- Monath TP, Wanos JR, Hill LJ, *et al*. Geographic classification of dengue-2 virus strains by antigen signature analysis. *Virology* 1986 ; 154 : 313-24.
- Murray V. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res* 1989 ; 7 : 438-42.
- Osatomi K, Sumiyoshi H. Complete nucleotide sequence of dengue type 3 virus genome RNA. *Virology* 1990; 176 : 643-7.
- Rice CM, Lenches EM, Eddy SR, *et al*. Nucleotide sequence of yellow fever virus. Implications for flavivirus gene expression and evolution. *Science* 1985 ; 229 : 726-33.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 1977 ; 74 : 5463-7.
- Speight G, Coia G, Parker MD, Westaway EG. Gene mapping and positive identification of the nonstructural proteins NS2a, NS2b, NS3, NS4b and NS5 of the flavivirus Kunjin and their cleavage sites. *J Gen Virol* 1988 ; 69 : 23-34.
- Sumiyoshi H, Mori C, Fuke I, *et al*. Complete nucleotide sequence of Japanese encephalitis virus genome RNA. *Virology* 1987 ; 161 : 497-510.
- Trent DW, Kinney RM, Johnson BJ, *et al*. Genetic variation among dengue 2 viruses of different geographic origin. *Virology* 1983 ; 128 : 271-84.
- Weaver SC, Scott TW, Rico-Hesse R. Molecular evolution of eastern equine encephalomyelitis virus in North America. *Virology* 1991 ; 182 : 774-84.
- Wengler G, Castle E. Analysis of structural properties which possibly are characteristic for the 3'-terminal sequence of the genome RNA of flaviviruses. *J Gen Virol* 1986 ; 67 : 1183-8.