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

WH Attatippaholkun[1], MK Attatippaholkun[2], A Nisalak[3], DW Vaughn[3] and BL Innis[4]

[1]Department of Clinical Chemistry, Faculty of Medical Technology, Mahidol University; [2]Thailand Institute of Scientific and Technological Research, Bangkok, Thailand; [3]Department of Virology, US Army Medical Component, Armed Forces Research Institute of Medical Science, Bangkok, Thailand; [4]Department 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 technic by the enzymatic method of Sanger *et al* using a sequencing primer 5'-end labeled with $\gamma$-$^{32}$P-ATP was the method of our choice for sequence analysis. Two cDNA templates were prepared by RT-PCR technic 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 polyprotein 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 polyprotein 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 (Chomczymski 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 $\gamma$-$^{32}$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

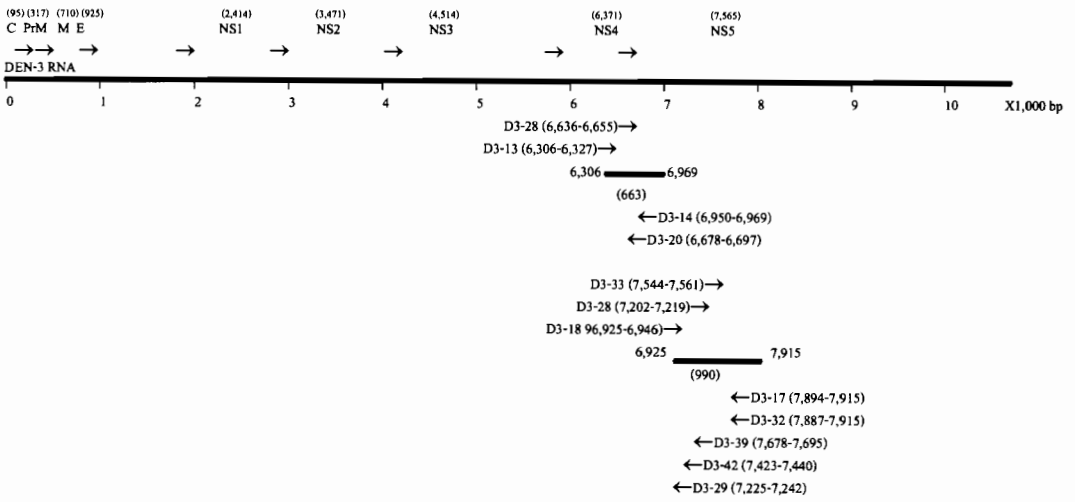


Fig 1-The strategy of the cDNA template preparation as well as the primer design for nucleotide sequence walking on the RNA genome of the Bangkok dengue 3 virus (CH53489).

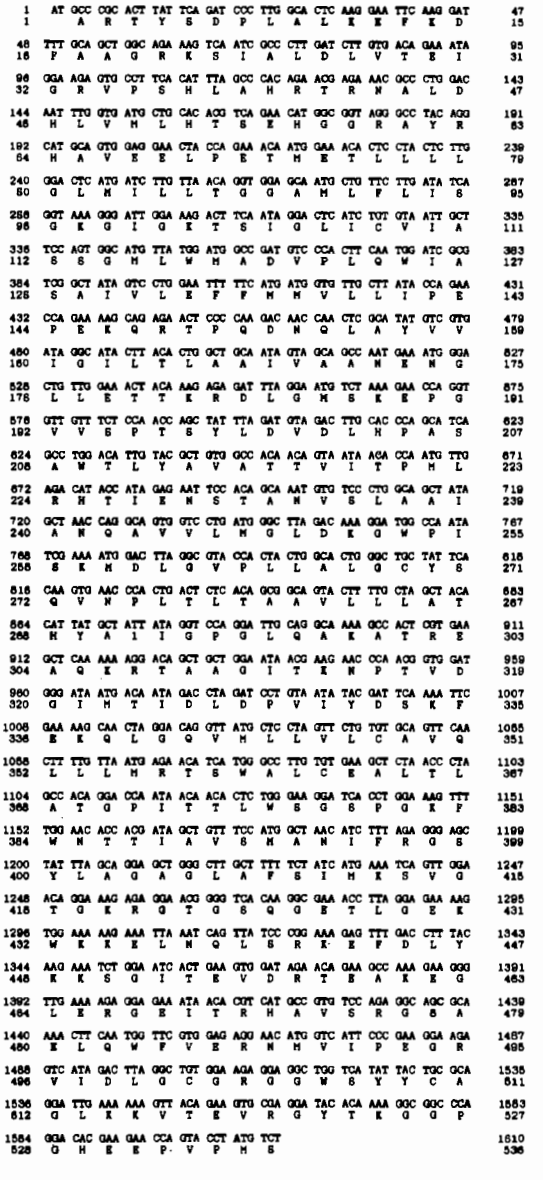

Fig 2-The nucleotide and deduced amino acid sequences of the nonstructural proteins NS4a, NS4b and NS5 of the Bangkok dengue 3 virus (CH53489).

power supply (BIORAD, USA). For autoradiography, the $^{32}$P-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). Inspite of this evolution, the prM/M

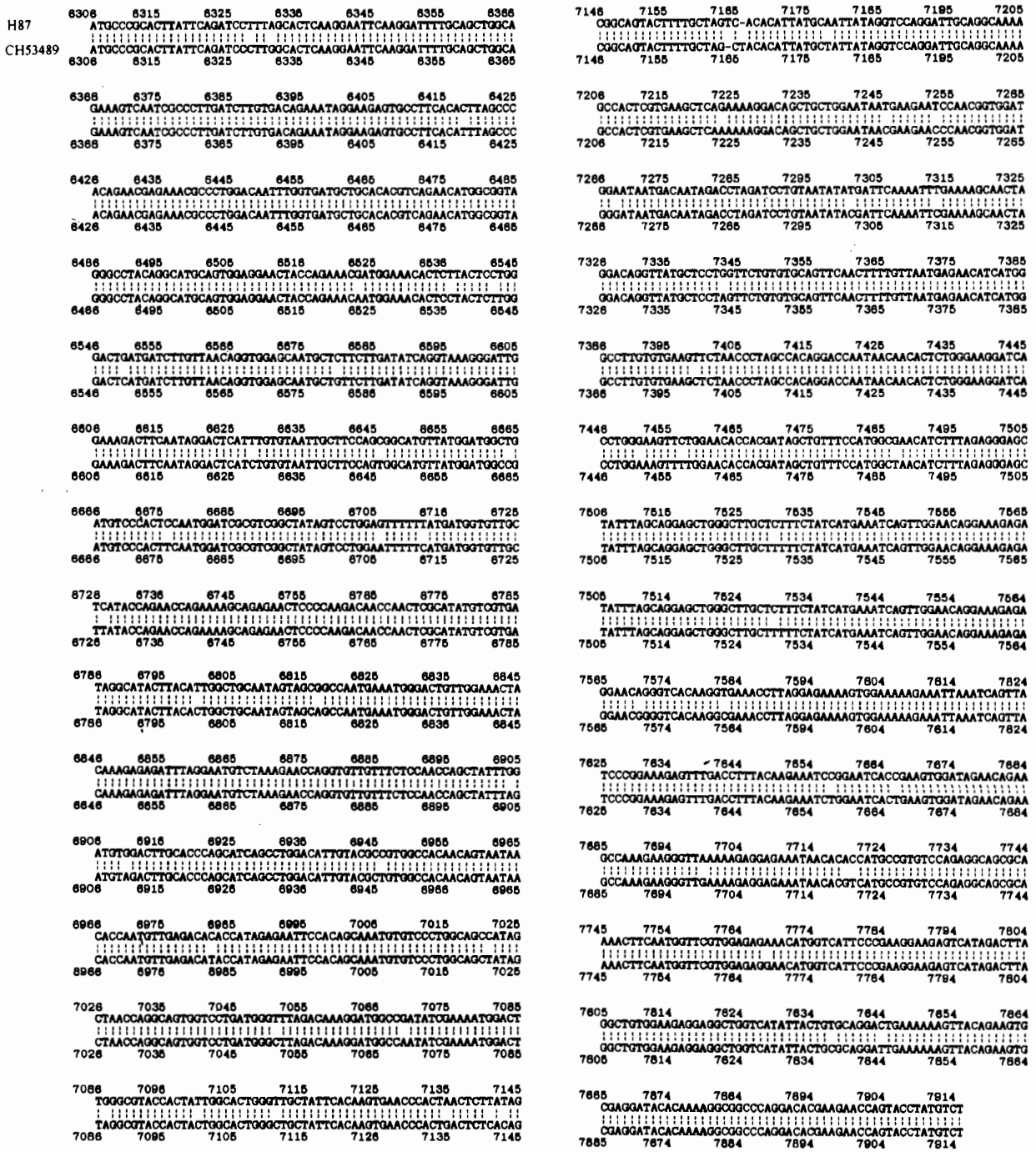


Fig 3-Alignment of the nucleotide sequences from the nucleotides 6,306-7,915 on the RNA genome between the Bangkok dengue 3 virus (CH53489) and the published nucleotide sequence data of the prototype dengue 3 virus (H87).

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

2101' EIGRVPSHLAHRTRNALDNLVMLHTSEHGGRAYRHAVEELPETMETLLLLGLMILLTGGA
      ************************************************************
2101" EIGRVPSHLAHRTRNALDNLVMLHTSEHGGRAYRHAVEELPETMETLLLLGLMILLTGGA

2161' MLFLISGKGIGKTSIGLICVIASSGMLWMADVPLQWIASAIVLEFFMMVLLIPEPEKQRT
      ************************************************************
2161" MLFLISGKGIGKTSIGLICVIASSGMLWMADVPLQWIASAIVLEFFMMVLLIPEPEKQRT

2221' PQDNQLAYVVIGILTLAAIVAANEMGLLETTKRDLGMSKEPGVVSPTSYLDVDLHPASAW
      ************************************************************
2221" PQDNQLAYVVIGILTLAAIVAANEMGLLETTKRDLGMSKEPGVVSPTSYLDVDLHPASAW

2281' TLYAVATTVITPMLRHTIENSTANVSLAAIANQAVVLMGLDKGWPISKMDLGVPLLALGC
      ************************************************************
2281" TLYAVATTVITPMLRHTIENSTANVSLAAIANQAVVLMGLDKGWPISKMDLGVPLLALGC

2341' YSQVNPLTLIAAVLLLVTHYAIIGPGLQAKATREAQKRTAAGIMKNPTVDGIMTIDLDPV
      *********.******.********************** ****************
2341" YSQVNPLTLTAAVLLLATHYAIIGPGLQAKATREAQKRTAAGITKNPTVDGIMTIDLDPV

2401' IYDSKFEKQLGQVMLLVLCAVQLLLMRTSWALCEVLTLATGPITTLWEGSPGKFWNTTIA
      ************************************.************************
2401" IYDSKFEKQLGQVMLLVLCAVQLLLMRTSWALCEALTLATGPITTLWEGSPGKFWNTTIA

2461' VSMANIFRGSYLAGAGLALSIMKSVGTGKRGTGSQGETLGEKWKKKLNQLSRKEFDLYKK
      ******************.*****************************************
2461" VSMANIFRGSYLAGAGLAFSIMKSVGTGKRGTGSQGETLGEKWKKKLNQLSRKEFDLYKK

2521' SGITEVDRTEAKEGLKRGEITHHAVSRGSAKLQWFVERNMVIPEGRVIDLGCGRGGWSYY
      **********************.*****************************************
2521" SGITEVDRTEAKEGLKRGEITRHAVSRGSAKLQWFVERNMVIPEGRVIDLGCGRGGWSYY

2581' CAGLKKVTEVRGYTKGGPGHEEPVPMS 2607'
      ****************************
2581" CAGLKKVTEVRGYTKGGPGHEEPVPMS 2607"
```

A

```
                                2100
                                 *
H87       ARTYSDPLA LKEFKDFAAG RKSIALDLVT EIGRVPSHLA HRTRNALDNL VMLHTSEHGG RAYRHAVEEL PETMETLLLL GLMILLTGGA
CH53489   ......... ......... ......... ......... ......... ......... ......... ......... .........>

                                          2200
                                           *
H87       MLFLISGKGI GKTSIGLICV IASSGMLWMA DVPLQWIASA IVLEFFMMVL LIPEPEKQRT PQDNQLAYVV IGILTLAAIV AANEMGLLET
CH53489   ......... ......... ......... ......... ......... ......... ......... ......... .........>

                                                    2300
                                                     *
H87       TKRDLGMSKE PGVVSPTSYL DVDLHPASAW TLYAVATTVI TPMLRHTIEN STANVSLAAI ANQAVVLMGL DKGWPISKMD LGVPLLALGC
CH53489   ......... ......... ......... ......... ......... ......... ......... ......... .........>

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

                                                                        2500
                                                                         *
H87       ALCEVLTLAT GPITTLWEGS PGKFWNTTIA VSMANIFRGS YLAGAGLALS IMKSVGTGKR GTGSQGETLG EKWKKKLNQL SRKEFDLYKK
CH53489   ....A..... ......... ......... ......... ........F. ......... ......... ......... .........>

                                                                                  2600
                                                                                   *
H87       SGITEVDRTE AKEGLKRGEI THHAVSRGSA KLQWFVERNM VIPEGRVIDL GCGRGGWSYY CAGLKKVTEV RGYTKGGPGH EEPVPMS
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 MsGanlayarat 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

Chomczymski 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.