NUCLEOTIDE SEQUENCES OF THE GENES CODING FOR THE MIDDLE SURFACE PROTEIN FROM HEPATITIS B VIRUS GENOMES SUBTYPES *adr* AND *adw* ISOLATED IN THAILAND

DUANTHANORM THAWARANANTHA[1], KRUAVON BALACHANDRA[1], CHUENCHIT BOONCHIRD[2], CHEDSADAPORN PITUKSUTEEPONG[1], JITTAPORN WATANASEREE[3] AND SOMSAK PANTUWATANA[4]

[1] *National Institute of Health, Department of Medical Sciences, Nonthaburi, 11000.*

[2] *Department of Biotechnology, Faculty of Science, Mahidol University.*

[3] *Government Pharmaceuticals Organization.*

[4] *Department of Microbiology, Faculty of Science, Mahidol University, Bangkok 10400, Thailand.*

(Received April 14, 1997)

ABSTRACT

Viral genomes of 8 isolates of hepatitis B virus (HBV) derived from blood specimens were used to sequence of pre-S2 and S regions. Four samples of each subtype adr and adw were used. It was found that there was quite high homogeneity among four samples of the same subtype, that is, 99.6% among samples of adr *subtype and 98.7% among samples of* adw *subtype. Although, the two subtypes showed many nucleotide variations, all of them presented the pattern of wild type DNA. Results showed that the isolated DNA of subtype adw could be classified as genotype B whereas those of subtype adr were classified as genotype C. Concerning the most common antigenic determinant* a, *no variation as vaccine-induced escape mutant was observed in all eight samples, for example, at amino acids 126 and 145 of HBsAg.*

INTRODUCTION

Hepatitis B is a widespread and serious human disease that requires an extensive program of vaccination. The progress in recombinant DNA methods have provided an alternative approach for the development of vaccine to support of global immunization activities. It has been demonstrated that bacteria, yeast and mammalian cells transformed with appropriate expression vectors are able to synthesize hepatitis B surface antigen (HBsAg).

Hepatitis B viral (HBV) DNA is a partially double stranded circular molecule. The long strand is approximately 3,200 nucleotides while the short strand varies from 1,700 to 2,800 nucleotides in different molecules[1]. HBsAg proteins are encoded by two adjacent regions of the HBV envelope gene, the pre-S which comprises of the pre-S1 and the pre-S2, and the S region. A protein of 226 amino acids is the major protein constituent of the HBV envelope that is encoded by the S gene. A "middle" protein carrying an additional 55 amino acids is encoded by the pre-S2 portion while a "large" protein carrying an extra 108 amino acids is encoded by the pre-S1 portion. The pre-S2 region is the binding site for polymerized human serum albumin that is believed to be involved in the attachment of HBV to human hepatocyte[2,3]. However, Neurath *et al.*,[4] reported that immunodominant epitopes in the pre-S2 region were recognized by human antibodies to HBV, and these epitopes enhance the immunogenicity of HBsAg in mice[5]. Additionally, antibodies to synthetic peptides from the pre-S2 region are virus-neutralizing and protective antibodies in chimpanzees[6]. Moreover, anti-pre-S2 antibodies seem to play a role in immunological clearance of infectious HBV[7].

Although the subtype of hepatitis B surface antigen (*adr, adw, ayr, ayw*) may vary among different geographical isolates of the virus, all share a common antigenic determinant ***a***. Production of antibodies to this epitope appears to mediate cross-protection against all subtypes. The epitope ***a*** is located in the amino acid region of residues 124-147 of the surface protein[8]. There have been several reports[8,9,10,11] which inferred that mutations occurred in the gene coding for the ***a*** epitope resulted in significant reduction of antigenicity and immunoreactivity of the surface protein.

These studies were focused on nucleotide sequences. Since the HBV DNA clones could be used in a certain purpose as recombinant vaccine production, thus, the pre-S2 and S regions that would be an inserted portion in gene expression were studied. Additionally, it is of interest to compare these regions of subtype *adr* and *adw*, since it is known that there are only these two subtypes that are endemic in Thailand.

MATERIALS AND METHODS

Preparation of HBV-DNA

The HBV DNA was isolated from the Dane particles of HBV isolated from blood samples. The isolated viral DNA was subsequently used to construct recombinant plasmids resulting in several plasmids that carrying the entire HBV genome of subtype *adr*. The constructed recombinant plasmids, pDKC1, pDKC2, pDKC3 and pPM5 were selected. The recombinant plasmids pDKC1, pDKC2 and pDKC3 had been prepared and described by Balachandra *et al.*[12], and the pPM5 had been prepared and described by Boonchird *et al.*[13] The four viral genomes of subtype *adw* were prepared from those isolates obtained from virus-positive sera as described by Thawaranantha *et al.*[14] These specimens were designated as No. 240, 508, 540, and 562. The method for subtyping of all HBV genomes was done by the method of PCR as described by Thawaranantha *et al.*[14]

Amplification of the pre-S2 and the S regions

The portions of pre-S2 and S were amplified from HBV DNA clones by the method of polymerase chain reaction as described by Thawaranantha *et al.*,[14] except that the primers used were PS-X and PS-8. The nucleotide sequences of the primers used are shown in Table 1. After amplification, the amplified fragments were purified by the methods of the low-melting point temperature agarose gel electrophoresis as described by Maniatis *et al.*[15]

Cycle sequencing

The cycle sequencing reaction of each DNA sample was performed following the protocol of ABI PRISM Dye Terminator, Perkin-Elmer Corporation, USA. In brief, 0.5 μg DNA and 3.2 pmole of a sequencing primer were added to 8.0 μl of the Terminator Ready Reaction Mix to make a total volume of 20 μl. The reaction mixture was allowed to proceed in a sequence of thermal cycling at 96°C for 10 sec, 50°C for 15 sec, and 60°C for 4 min. The thermal cycling reaction was allowed to complete 25 cycles. The sequencing primers used were PS-X and PS-5 while the sequencing primers used for the complementary strand were PS-8 and PS-4. The nucleotide sequences of the primers used are shown in Table 1. The extension fragments were purified from unincorporated fluorescent-labeled reagent in the Terminator Ready Reaction Mix by extraction with the mixture of 100 μl chloroform and 100 μl of phenol-water-chloroform solution. The phenol-water-chloroform solution was consisting of phenol, water and chloroform in a ratio of 68:18:14. The sample was centrifuged at 15,000 rpm for 3 min and the upper phase

Table 1. Nucleotide sequences of primers for amplification and cycle sequencing of pre-S2 and S genes.

Primer	Nucleotide sequence(5' - 3')	Nucleotide position[a]
PS-X	TCCACCAATCGGCAGTCAGG	3133-3152
PS-5	GCCTCATCTTCTTGTTGG	423-440
PS-4	ATGAGGCATAGCAGCAGGATG	408-429
PS-8	GTACCCCAACTTCCAATTAC	886-905

a : Nucleotide positions were numbered from a hypothetical EcoRI site.

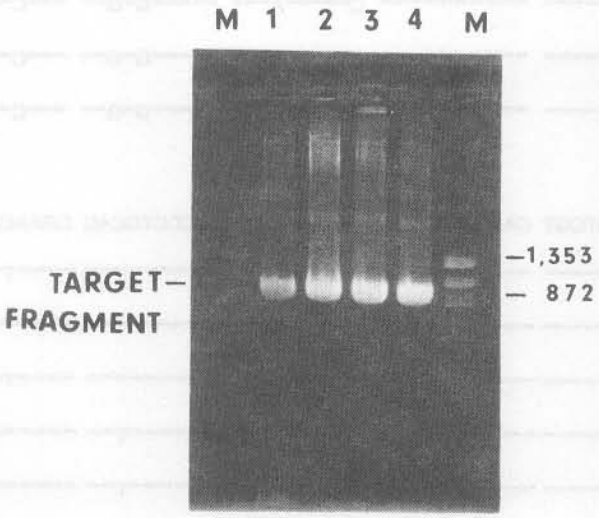

Fig. 1. Gel electrophoresis of the amplified fragments for sequencing of cloned and isolated HBV DNA shown in lane 1, 2, 3, 4.
M is the molecular weight marker (bp).

```
pDKC1  :   1  ATGCAGTGGA ATTCCAGCAC ATTCCACCAA GCTCTGCTAG ATCCCAGAGT GAGGGGCCTA  60
pDKC2  :      ---------- -C-------- ---------- -T-------- ---------- ----------
pDKC3  :      ---------- ---------- ---------- ---------- ---------- ----------
pPM5   :      ---------- -C-------- ---------- ---------- ---------- ----------
No. 240 :     ---------- -C----C--- T-G------- A----T-A-- ---------- C----C---G
No. 508 :     ---------- -C----C--- T--------- A----T-A-- ---------- C----C---G
No. 540 :     ---------- -C----C--- T--------- A----T-A-- ---------- C----C---G
No. 562 :     ---------- -C----C--- T--------- A----T-A-- ---------- C----C---G

pDKC1  :  61  TACTTTCCTG CTGGTGGCTC AAGTTCCGGA ACAGTAAACC CTGTTCCGAC TACTGCCTCT 120
pDKC2  :      ---------- ---------- ---------- -T-------- ---------- ----------
pDKC3  :      ---------- ---------- ---------- ---------- ---------- ----------
pPM5   :      ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :     ---------- ---------- C-----A--- -----G-G-- ---C--A--A C----T----
No. 508 :     ---------- ---------- C-----A--- -----G-G-- ---C--A--A -----T----
No. 540 :     ---------- ---------- C-----A--- -----G-G-- ---C--A--A -----T----
No. 562 :     ---------- ---------- C-----A--- -----G-G-- ---C--A--A -----T----

                                                                    →S
pDKC1  :  121 CCCATATCGT CAATCTTCTC GAGGACTGGG GACCCTGCAC CGAACATGGA GAGCACCACA 180
pDKC2  :      ---------- ---------- ---------- ---------- ----T----- ----------
pDKC3  :      ---------- ---------- ---------- ---------- ---------- ----------
pPM5   :      ---------- ---------- ---------- ---------- -----T---- ----------
No. 240 :     G--------- -------A-- ---------- -------T-- ---------- --A--T-G--
No. 508 :     G--------- -------A-- --A------- -------T-- ---------- --A--T-G--
No. 540 :     G--------- -------A-- --A------- -------T-- ---------- --A--T-G--
No. 562 :     G--------- -------A-- --A------- -------T-- ---------- --A--T-G--
```

Fig.2. Nucleotide sequences of the pre-S2 and S regions of HBV DNA clones:pDKC1, pDKC2, pDKC3 and pPM5, and isolated genomes: No. 240, 508, 540, and 562. The sequences were aligned with the pDKC1 sequence using the first A residue of pre-S2 as nucleotide 1.

```
pDKC1  : 181 TCAGGATTCC TAGGACCCCT GCTCGTGTTA CAGGCGGGGT TTTTCTTGTT GACAAGAATC 240
pDKC2  :     ---------- ---------- ---------- ---------- ---------- ----------
pDKC3  :     ---------- ---------- ---------- ---------- ---------- ----------
pPM5   :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :    ------C--- ---------- ---------- -----G----- ---------- -----A----
No. 508 :    ------C--- ---------- ---------- ---------- ---------- -----A----
No. 540 :    ------C--- ---------- ---------- ---------- ---------- -----A----
No. 562 :    ------C--- ---------- ---------- ---------- ---------- -----A----

pDKC1  : 241 CTCACAATAC CACAGAGTCT AGACTCGTGG TGGACTTCTC TCAATTTTCT AGGGGGAGCA 300
pDKC2  :     ---------- ---------- ---------- ---------- ---------- ----------
pDKC3  :     ---------- ---------- ---------- ---------- ---------- ----------
pPM5   :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :    ---------- ---------- ---------- ---------- ---------- -----A----
No. 508 :    ---------- ---------- ---------- ---------- ---------- -----A----
No. 540 :    ---------- ---------- ---------- ---------- ---------- -----A----
No. 562 :    ---------- ---------- ---------- ---------- ---------- ----------

pDKC1  : 301 CCCACGTGTC CTGGCCAAAA TTTGCAGTCC CCAACCTCCA ATCACTCACC AACCTCTTGT 360
pDKC2  :     ---------- ---------- -----G----- ---------- ---------- ----------
pDKC3  :     ---------- ---------- -----G----- ---------- ---------- ----------
pPM5   :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :    ---GT----- T--------- --C------- -----T----- G--------- ----------
No. 508 :    ---GT----- T--------- --C------- -A---T----- G--------- -----G----
No. 540 :    ---GT----- T--------- --C------- -A---T----- G--------- -----G----
No. 562 :    ---GT----- T--------- --C------- -A---T----- G--------- -----G----
```

Fig.2. (continued)

```
pDKC1  : 361 CCTCCAATTT GTCCTGGTTA TCGCTGGATG TGTCTGCGGC GTTTTATCAT CTTCCTCTTC 420
pDKC2  :     ---------- ---------- ---------- ---------- ---------- ----------
pDKC3  :     ---------- ---------- ---------- ---------- ---------- -----C----
pPM5   :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :    ---------- ---------- ---------- ---------- ---------- --------G-
No. 508 :    ---------- ---------- ---------- ---------- ---------- --------G-
No. 540 :    ---------- ---------- ---------- ---------G ---------- --------G-
No. 562 :    ---------- ---------- ---------- ---------- ---------- --------G-

pDKC1  : 421 ATCCTGCTGC TATGCCTCAT CTTCTTGTTG GTTCTTCTGG ACTACCAAGG TATGTTGCCC 480
pDKC2  :     ---------- ---------- ---------- ---------- ---------- ----------
pDKC3  :     ------G--- ---------- ---------- ---------- ---------- ----------
pPM5   :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :    ---------- ---------C ---------- ---------- ----T----- ----------
No. 508 :    ---------- ---------- ---------- ---------- ----T----- ----------
No. 540 :    ---------- ---------- ---------- ---------- ----T----- ----------
No. 562 :    ------G--G ---------- ---------- ---------- ---------- ---T------

pDKC1  : 481 GTTTGTCCTC TACTTCCAGG AACATCAACT ACCAGCACGG GACCATGCAA GACCTGCACG 540
pDKC2  :     ---------- ---------- ---------- ---------- ---------- ----------
pDKC3  :     ---------- ---------- ---------- ---------- ---------- ----------
pPM5   :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :    ---------- --A------- -T-------A --------C- ---------- A--------A
No. 508 :    ---------- ---------- -T-------A --------C- ---------- A--------A
No. 540 :    ---------- --A------- -T-------A --------C- ---------- A--------A
No. 562 :    -----G--C- ---------- -T-------A --------C- ---------- A--------A
```

Fig.2. (continued)

```
pDKC1  : 541 ATTCCTGCTC AAGGAACCTC TATGTTTCCC TCTTGTTGCT GTACAAAACC TTCGGACGGA 600
pDKC2  :     ---------- ---------- ---------- ---------- ---------- ----------
pDKC3  :     ---------- ---------- ---------- ---------- ---------- ----------
pPM5   :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :    -C-------- ---------- ---------- --A------- ---------- -A--------
No. 508 :    -C-------- ---------- ---------- --A------- ---------- -A--------
No. 540 :    -C-------- ---------- ---------- --A------- ---------- -A--------
No. 562 :    -C-------- ---------- ---------- --A------- ---------- -A--------

pDKC1  : 601 AATTGCACTT GTATTCCCAT CCCATCATCT TGGGCTTTCG CAAGATTCCT ATGGGAGTGG 660
pDKC2  :     ---------- ---------- ---------- ---------- ---------- ----------
pDKC3  :     ---------- ---------- ---------- ---------- ---------- ----------
pPM5   :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :    --C-----C- ---------- ---------- ---------- ---A--A--- ----------
No. 508 :    --C-----C- ---------- ---------- ---------- ---A--A--- ----------
No. 540 :    --C-----C- ---------- ---------- ---------- ---A--A--- ----------
No. 562 :    --C-----C- ---------- ---------- ---------- ---A--A--- ----------

pDKC1  : 661 GCCTCAGTCC GTTTCTCCTG GCTCAGTTTA CTAGTGCCAT TTGTTCAGTG GTTCGTAGGG 720
pDKC2  :     ---------- ---------- ---------- ---------- ---------- ----------
pDKC3  :     ---------- ---------- ---------- ---------- ---------- ----------
pPM5   :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :    ---------- -------T-- ---------- ---------- ---------- ----------
No. 508 :    ---------- -------T-- ---------- ---------- ---------- ----------
No. 540 :    ---------- -------T-- ---------- ---------- ---------- ----------
No. 562 :    ---------- -------T-- ---------- ---------- ---------- ----------
```

Fig.2. (continued)

```
pDKC1 : 721 CTTTCCCCCA CTGTTTGGCT TTCAGTTATA TGGATGATGT GGTATTGGGG GCCAAGTCTG 780
pDKC2 :     ---------- ---------- ---------- ---------- ---------- ----------
pDKC3 :     ---------- ---------- ---------- ---------- ---------- ----------
pPM5  :     ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :   ----------    ----C----- ---------- ---------- ---------- ----------
No. 508 :   ----------    ----C----- ---------- ---------- ---T------ ----------
No. 540 :   ----------    ----C----- ---------- ---------- ---T------ ----------
No. 562 :   ----------    ----C----- ---------- ---------- ---T------ ----------

                                                                          .840
pDKC1 : 781 TACAACATCT TGAATCCCTT TTTACCTCTA TTACCAATTT TCTTTTGTCT TTGGGTATAC ATTTAA
pDKC2 :     ---------- ---------- ---------- ---------- ---------- ---------- ------
pDKC3 :     ---------- ---------- ---------- ---------- ---------- ---------- ------
pPM5  :     ---------- ---G------ ---------- ---------- ---------- ---------- ------
No. 240 :   ---------- ---G------ -A-G--G--G ---------- ---------- ---------- ------
No. 508 :   ---------- ---G------ -A-G--G--G ---------- ---------- ---------- ------
No. 540 :   ---------- ---G------ -A-G--G--G ---------- ---------- ---------- ------
No. 562 :   ---------- ---G------ -A-G--G--G ---------- ---------- ---------- ------
```

Fig.2. (continued)

```
                →PRE-S2                          30                          50          HBsAg
                                                                                          1
pDKC1 :  1 MQWNSSTFHQ ALLDPRVRGL YFPAGGSSSG TVNPVPTTAS PISSIFSRTG DPAPNMESTT  5
pDKC2 :    ---------- V--------- ---------- ---------- ---------- ----------
pDKC3 :    ---------- ---------- ---------- ---------- ---------- ----------
pPM5  :    ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :  -----T-C-- T-Q-----A- ---------- --S-AQN-V- A----L---- --V----NI A
No. 508 :  -----T----- T-Q-----A- ---------- --S-AQN-V- A------K-- --V----NI A
No. 540 :  -----T----- T-Q-----A- ---------- --S-AQN-V- A------K-- --V----NI A
No. 562 :  -----T----- T-Q-----A- ---------- --S-AQN-V- A------K-- --V----NI A

pDKC1 :  6 SGFLGPLLVL QAGFFLLTRI LTI PQSLDSW WTSLNFLGGA PTCPGQNLQS PTSNHSPTSC 65
pDKC2 :    ---------- ---------- ---------- ---------- ---------- ----------
pDKC3 :    ---------- ---------- ---------- ---------- --------E- ----------
pPM5  :    ---------- ---------- ---------- ---------- --------E- ----------
No. 240 :  --L------- -G------K- ---------- ---------T -V-L---S-- -I-S------
No. 508 :  --L------- --------K- ---------- ---------T -V-L---S-- QI-S----C-
No. 540 :  --L------- --------K- ---------- ---------T -V-----S-- QI-S----C-
No. 562 :  --L------- --------K- ---------- ---------- -V-----S-- QI-S----C-

pDKC1 :  66 PPICPGYRWM CLRRFIIFLF I LLLCL I FLL VLLDYQGMLP VCPLLPGTST TSTGPCKTCT 125
pDKC2 :    ---------- ---------- ---------- ---------- ---------- ----------
pDKC3 :    ---------- --------P- --V------- ---------- ---------- ----------
pPM5  :    ---------- ---------- ---------- ---------- ---------- ----------
No. 240 :  ---------- ---------C ------T--- ---------- ----I --S-- ----------
No. 508 :  ---------- ---------C ---------- ---------- -------S-- ----------
No. 540 :  ---------- ---G-----C ---------- ---------- ----I --S-- ----------
No. 562 :  ---------- ---------C --VV------ -------I -- -W-----S-- ----------
```

Fig. 3. Deduced amino acids sequences of the pre-S2 and S regions from nucleotide sequences of pDKC1, pDKC2, pDKC3, pPM5, No. 240, 508, 540, and 562.

Underlines indicated the antigenic determinant ***a*** region. The sequences were alligned with the pDKC1 sequence using the first M residues of pre-S2 and HBsAg as amino acid 1 of pre-S2 and HBsAg, respectively.

```
pDKC1  : 126 I PAQGTSMFP SCCCTKPSDG NCTCI PI PSS  WAFARFLWEW  ASVRFSW LS L  LVPFVQWFVG 185
pDKC2  :     ---------- ---------- ---------- ----------  ----------  ----------
pDKC3  :     ---------- ---------- ---------- ----------  ----------  ----------
pPM5   :     ---------- ---------- ---------- ----------  ----------  ----------
No. 240 :    T--------- -------T-- ---------- ----KY----  ----------  ----------
No. 508 :    T--------- -------T-- ---------- ----KY----  ----------  ----------
No. 540 :    T--------- -------T-- ---------- ----KY----  ----------  ----------
No. 562 :    T--------- -------T-- ---------- ----KY----  ----------  ----------

                                                                    226
pDKC1  : 186 LSPTVWLSVI WMMWYWGPSL YNI LNPFLPL LPI FFCLWVY  I *
pDKC2  :     ---------- ---------- ---------- ----------  --
pDKC3  :     ---------- ---------- ---------- ----------  --
pPM5   :     ---------- ---------- ----S----- ----------  --
No. 240 :    ---------- ---------- ----S--M-- ----------  --
No. 508 :    ---------- ----F----- ----S--M-- ----------  --
No. 540 :    ---------- ----F----- ----S--M-- ----------  --
No. 562 :    ---------- ----F----- ----S--M-- ----------  --
```

Fig.3. (continued)

material was collected. The sample of extension fragments was precipitated by a mixture of 15 μl of 2 M sodium-acetate pH 4.5 and 300 μl of 100% ethanol. The precipitated sample was centrifuged at 15,000 rpm for 15 min and the pellet was collected.

The sample of extension fragments was subsequently redissolved in a mixture of deionized formamide and EDTA. The mixture was consisting of deionized formamide and 50 mM EDTA pH 8.0 in a ratio of 5:1. The sample was electrophoreses and analyzed through the model 373 DNA Sequencer, Perkin-Elmer Corporation, USA.

RESULTS

Four clones of HBV DNA subtype *adr* in *E. coli*, and four isolated viral genomes of HBV subtype *adw* were analyzed for the nucleotide sequences of the regions of pre-S2 and S genes. To accomplish this, the combining portions of pre-S2 and S genes were prepared from each whole genome sample by amplification with the primers PS-X and PS-8. It was found that the amplified product was approximately 870 bp when it was comparatively estimated from the molecular weight markers. Results are summarized in Figure 1. The fragments were subsequently purified before gene sequencing by the technique of low-melting temperature agarose gel electrophoresis.

In this study, the strategy for DNA sequence analysis was cycle sequencing by using of specific sequencing primers, PS-X and PS-5. The PS-X primer started to sequence before the starting codon of pre-S2 whereas the PS-5 primer started to sequence from the overlapping region done by PS-X and stopped beyond the end of S-region. In contrary, the primers PS-8 and PS-4 were used for the complementary strand. Before sequencing, the fragments were prepared as 3 or 4 individual amplified products from each sample DNA. It was found that there were 846 nucleotides in the two adjacent regions, e.g., 165 nucleotides in pre-S2 gene and 681 nucleotides in S gene, and the numbers were not different among the eight genomes. According to the nucleotide sequence, the amino acid contents could be deduced as 55 residues from pre-S2 and 226 residues from S region. Results of the nucleotide sequences are summarized in Figure 2. The nucleotide sequences of all 4 clones of subtype *adr*, that is, pDKC1, pDKC2, pDKC3, and pPM5, showed almost the same which very few nucleotides were different, and these revealed as high homogeneity as 99.6%. Isolated genomes of subtype *adw* No. 240, 508, 540, and 562 were also shown high homogeneity among themselves as 98.7%. The deduced amino acid sequences of the eight genomes are summarized in Figure 3. Among 55 amino acids of pre-S2, there were as many as 11 residues that all four genomes of subtype *adw* were the same and completely different from those in *adr*. The most conserved region in pre-S2 of the two subtypes was ranged from residues 21 to 32. The key amino acid residues that specify *r* or *w* determinants were observed as residues 4, 47, 113, 126, 160 and 213 as shown in Figure 3.

DISCUSSION

In this study, the nucleotide and deduced amino acid sequences of two subtypes of HBV, that is *adr* and *adw*, was analyzed and compared. The pre-S2 and S genes of HBV isolated in Thailand were amplified and sequenced. The samples were prepared from four recombinant clones of HBV subtype *adr* and four isolated DNA of subtype *adw*. It has been shown that the error rate due to dNTP misincorporation done by thermostable DNA polymerases is as high as around 300×10^{-6} base insertions. Thus, the fidelity of the amplified products in the case of the conservation to its template should be reliable. According to the portion studied which was less than 1 kilobase, the error rate should be very rare. Before sequencing, the fragments were

prepared as 3 or 4 individual amplified products from each sample DNA. From the study, there were 846 nucleotides in the two adjacent regions, e.g., 165 nucleotides in pre-S2 gene and 681 nucleotides in S gene, and the numbers were not different among the eight genomes. Due to the nucleotide sequence, the amino acid contents could be deduced as 55 residues from pre-S2 and 226 residues from S.

It has been shown that pre-S2 antigen is highly immunogenic in mice and its presence together with S protein may potentiate the immune response to S candidate for a synthetic peptide vaccine[17,18]. The pre-S2 gene which next to S was included in gene cloning for further investigations and applications[12,13]. Thus, not only the S gene but also pre-S2 was intended to analyze the sequences.

It was shown that all individual clones of each subtype had high homogeneity among themselves with the degree of homogeneity was above 98 %. It was obvious that there were many point nucleotide differences between the groups of two subtypes. At each point mutation, all four genomes of one subtype showed the same among themselves with the degree of homogeneity about 93 % and differed from the other subtype as shown in Figure 2.

The deduced amino acid sequences from the pre-S2 and S genes were analyzed. Among 55 amino acids of pre-S2 region, there were as many as 11 residues of all four genomes of subtype *adw* were the same and completely different from those in *adr*. The most conserved region in pre-S2 of the two subtypes was ranged from residues 21 to 32. There were many point mutations observed in the S region that differentiated those in *adw* from *adr* (as shown in Figure 2) resulted in amino acid variations (see Figure 3). The key amino acid residues that specify *r* or *w* determinants were reported as residues 4, 47, 110, 113, 126, 160, 207, and 213[19]. In this study, such distinctive amino acids were observed as residues 4, 47, 113, 126, 160 and 213 as shown in Figure 3.

Along with serological or molecular subtyping, HBV genomes can be classified into 6 genomic groups, that are A, B, C, D, E and F[20,21]. It was demonstrated that all four genomes of HBV DNA subtype *adr* could be identified as genotype C when they were compared with the sequence of each genotype described by Norder *et al.*[21]. All four isolated genomes of HBV subtype *adw* were classified as genotype B. This is based on the fact that almost all key amino acids are exactly the same as those residues, that are 5, 8, 24, 45, 56, 59, 64, 85, described by Norder *et al.*[21]. This genotype B has also been found in Japan, China, Hawaii and Indonesia[21].

The antigenic determinant ***a*** which is represented by amino acids 124-147 is known as a site of neutralization antibody binding, or the protective anti-HBs response in individual is generally detected against this common group-specific determinant. This region of all eight genomes was also observed in this study. All 24 amino acid residues showed the same as those of wild type HBV-DNA[22,23]. Recently, Carman *et al.*,[8] have reported a vaccine-induced escape mutant of HBV that expressed an altered ***a*** determinant of HBsAg in a vaccinated child. This alteration consisted of a single amino acid substitution from glycine to arginine at position 145. Furthermore, the alteration was also reported at residue 126 by substitution of serine or asparagine of the wild type virus to isoleucine or threonine[10,11]. In this study, such alteration in the determinant ***a*** as vaccine-escape mutation was not observed in clones pDKC1, pDKC2, pDKC3, and pPM5 at both positions 126 and 145. This wild type of the determinant ***a*** was also observed in the rest isolated genomes No. 240, 508, 540 and 562.

This study showed nucleotide sequences and deduced amino acid sequences of pre-S2 and S among eight genomes of 2 subtypes of HBV, *adr* and *adw*, isolated in Thailand. The pre-S2 and S regions of these isolated HBV genomes showed the sequences of wild type virus,

especially the S region when compared with those reported by others (22,23). No variation as vaccine-induced escape mutation was found among these eight genomes. These four clones which carrying the pre-S2 and S regions can be used to insert into expression vector to produce surface and middle-surface proteins for recombinant vaccine production and diagnostic purposes. The study confirms that these clones are appropriate for such applications based on sequence analysis. Moreover, this study revealed the comparison of pre-S2 and S between genomes subtype *adr* and *adw* isolated in Thailand.

ACKNOWLEDGMENT

This project was supported by the grant of the Government Pharmaceutical Organization.

REFERENCES

1. Landers, T.A., Greenberg, H.B., and Robinson, W.S. (1977). *J. Virol.* **23**, 368.
2. Pontisso, P., Petit, M.A., Bankowski, M. J., and Peeples, M.E. (1989). *J. Virol.* **36**, 1981.
3. Krone, B., Lenz, A., Heermann, K.H., Seifer, M., Xuangyong. L., and Gerlich, W.H. (1990). *Hepatology* **11**, 1050.
4. Neurath, A. R., Kent, S.B.H and Strick, N. (1984). *Science* **224**, 392.
5. Milich, D. R., thornton, G. B., Neurath, A. R. *et al.*, (1985) *Science* **228**, 1195.
6. Neurath, A. R., Kent, S.B.H., Parker, K., *et al.*, (1986). *Vaccine* **4**, 35.
7. Budkowska, A., Briantais, M. J., Dubreuil, P., Capel, R., Grangeot-keros, L., and Pillot, J. (1985). Ann. Inst. Pasteur Immunol. 136D, 57.
8. Carman, W.F., Zanetti, A. R., Karayianis, *et al.*, (1990). *Lancet* **336**, 325.
9. Fujii, H., Moriyama, K., Sakamoto, *et al.*, (1992). *Biochem. Biophys. Res. Comm.* **184**,1152.
10. Yamamoto, K., Horikita, M., Tsuda, F., *et al.*, (1994). *J. Virol.* **68**, 2671.
11. Okamoto, H., Yano, K., Nazaki, Y., *et al.*, (1992). *Pediatr. Res.* **32**, 264.
12. Balachandra, K., Thawaranantha, D., Boonchird, C., *et al.*, (1996). *J. Sci. Soc. Thailand* **22**, 325.
13. Monkongdee, P., Boonchird, C., Balachandra, K., *et al.*, (1995). Abstract in International Conference on Biotechnology Research and Application for Sustainable Development, Bangkok, Thailand. Aug 7-10, 1995. pp 202.
14. Thawaranantha, D., Balachandra, K., Watanaseree, J., *et al.*, (1994). *J. Sci. Soc. Thailand* **20**, 115.
15. Maniatis, T., Fritsch, E. F., and Sambrook, J. (eds) (1989). Molecular Cloning, 2nd edition. Coldspring Harbor Laboratory Press.
16. Milich, D.R., Mclachalan, A., Chisari, F. V., Kent, S. B. H., and Thornton, G.B. (1986). *J. Immunol* **137**, 315
17. Neurath, A. R., Kent, S.B.H., Strick, N., Taylor, P., and Stevens, C.E. (1985). *Nature (London)* **315**, 154.
18. Imai, M., Yanase, Y.,Nojiri, T., Miyakawa, Y., and Mayumi, M. (1979). *Gastroenterology* **76**, 242.
19. Okamoto, H., Imai M., Shimozaki, M., *et al.*, (1986). *J. Gen. Virol.* **67**, 2305.
20. Norder, H., Hammas, B., Lofdahl, S., Courouce, A. M., and Magnius, L.O. (1992). *J. Gen. Virol.* **73**, 1201.
21. Norder, N., Hammas, B., Lee, S. D., *et al.*, (1993). *J. Gen. Virol.* **74**, 1341.
22. Ono, Y., Onda, H., Sasada, R., Igarashi, K., Sugion, Y., and Nishioka, K. (1983). *Nucleic Acids. Res.* **11**, 1747
23. Kobayashi, M., and Koike, K. (1984). *Gene.* **30**, 227.

บทคัดย่อ

ได้นำยีโนมของไวรัสตับอักเสบบี 8 ตัวอย่างมาศึกษาลำดับนิวคลีโอไทด์ ช่วง Pre-S2 และ S โดยเป็นตัวอย่างที่เลือกจาก subtype *adr* และ *adw* ที่ตรวจพบในประเทศไทยอย่างละ 4 ตัวอย่าง พบว่า ลำดับนิวคลีโอไทด์ในแต่ละกลุ่มของ subtype มีความคล้ายคลึงกัน (homogeneity) มาก คือ ร้อยละ 99.6 ในกลุ่มของ subtype *adr* และ ร้อยละ 98.7 ในกลุ่ม subtype *adw* ในส่วนของยีนช่วง Pre-S2 และ S ที่ศึกษา พบว่า ทุกตัวอย่างแสดงลักษณะของสายพันธุ์ปกติ ซึ่งระหว่าง subtype *adr* และ *adw* มีนิวคลีโอไทด์ที่แตกต่างกันหลายตำแหน่ง เมื่อแบ่งแยกตัวอย่างเหล่านี้เป็นจีโนไทป์ (genotype) พบว่า subtype adw จัดเป็นจีโนไทป์ B ส่วนของ subtype *adr* เป็นจีโนไทป์ C เมื่อพิจารณาช่วง antigenic determinant ***a*** จากการแปลผลเป็นลำดับกรดอมิโน ของ HBsAg พบว่ายีโนมทั้ง 8 ตัวอย่างนี้ไม่มีการกลายพันธุ์ในลักษณะของ vaccine-induced escape mutant ซึ่งจะพบการกลายพันธุ์ที่กรดอมิโนตำแหน่ง 126 หรือ 145