CLONING AND SEQUENCE ANALYSIS OF HEPATITIS B VIRUS GENOME OF *ADR* SUBTYPE ISOLATED IN THAILAND

PATAMA MONKONGDEE[a], CHUENCHIT BOONCHIRD[b], KRUAVON BALACHANDRA[c], DUANTHANOM THAWARANANTHA[c], JITTAPORN WATANASEREE[d] AND SOMSAK PANTUWATANA[a]

[a] *Department of Microbiology, Faculty of Science, Mahidol University, Bangkok 10400*

[b] *Department of Biotechnology, Faculty of Science, Mahidol University, Bangkok 10400*

[c] *National Institue of Health, Department of Medical Science, Nonthaburi 11000*

[d] *Government Pharmaceutical Organization, Bangkok 10400, Thailand.*

(Received July 22, 1998)

ABSTRACT

Hepatitis B virus (HBV) DNA derived from Dane particle of subtype **adr** *was successfully cloned into the plasmid pBluescript. The virus was recovered from blood specimens donated to the Thai Red Cross. Its complete nucleotide sequence was determined. The 3,215 bp sequence showed the presence of open reading frames (ORF) including the genes coding for hepatitis B surface antigen (HBsAg) and pre-hepatitis B surface antigen (pre-S), P protein, hepatitis B core antigen (HBcAg) and X proteins. The HBsAg had 226 amino acids while the pre-S comprised of 174 amino acids. The P polypeptide comprised of 843 amino acids whereas HBcAg was consisting of 212 amino acids. The X protein comprised of 154 amino acids. It was found that all of the direct repeat of the undecanucleotide sequence (DR1) and the two small direct repeat sequences (DR2 and DR3) were characteristic structures of HBV DNA.*

Analysis of deduced amino acid sequences in the HBsAg revealed that the 9 amino acids of antigenic determinant **a**, *139-CTKPSDGNC-147, of this clone were conserved. The Cys-124 and Cys-147 together with Pro-142 required for full antigenicity were also observed. According to comparison of the amino acid sequences in the HBsAg sequence, the genotype of cloned HBV was classified as group* **C**. *In addition, the presence of Val-177 and Pro-178 in the HBsAg indicated that the cloned HBV had* **q** *subtype determinant. The amino acid sequences proposed to be immunodominant in the pre-S2 region, 131-LLDPRVRGLYFPAG-144, were found.*

INTRODUCTION

An immunization against hepatitis B virus (HBV) infection has been used as an effective measure in preventing the establishment of chronic HBV infection[1]. The vaccine contains the 22 nm subviral hepatitis B surface antigen (HBsAg) particle that consists mainly of the 226 amino acids of S protein (short HBsAg or SHBsAg). The S protein is a major polypeptide of the viral surface protein. The protective antibodies induced by vaccine are found directed against the common group determinant, ***a***, shared by all known subtypes of HBV. This group determinant, ***a***, is located in a cysteine cluster between amino acid 139 and 147 of the S protein which are conserved in all known serotypes of HBV. At present, the major source of commercially available vaccine has been produced by recombinant DNA technology using yeast expression system. The immunogen from cloned SHBsAg gene, second generation vaccine, has been demonstrated to give protective immunity[2]. However much attention has been

focused on the pre-S2 region which is in the middle HBsAg (MHBsAg). The pre-S2 region consisting of 55 amino acids was speculated to play a role in virus attachment to host[3-5] because it binds to polymerized human serum albumin (pHSA) *in vitro*[6,7] and may mediate viral uptake by albumin specific receptor[6,8]. The pre-S2 region is significantly more immunogenic than the SHBsAg in terms of primary antibody production *in vivo* even in cases of non-responsiveness to the SHBsAg[9-10]. Antibodies to pre-S2 epitope[11-13] and antibodies blocking the pHSA binding of HBV has been described in the early phase of recovery from acute infection[14]. The immunodominance of the pre-S2 region was recently mapped to be between amino acids 131-144[15]. Therefore the immunodominant nature of the pre-S2 may have promise as a immunogen for the third generation vaccine or for diagnostic reagents.

In an attempt to produce the vaccine for prophylaxis of HBV infection among Thai population, it is worthwhile to use local strains of virus for the production of vaccine. In Thailand, we found ***adr*** and ***adw*** subtypes in the proportion of 81.0% and 19.0%, repectively[16]. There is a need to clone the gene that encodes surface antigen and determine its nucleotide sequence to ensure that it codes the conserved determinant ***a*** and is immunodominant in the pre-S2 region. The nucleotide sequences of HBV DNA obtained from the local strain in Thailand has never been reported. Therefore it is worthwhile to clone and study the nucleotide sequence of the complete genome including the gene that encodes surface antigen. In this study the cloning and sequencing of HBV DNA subtype ***adr*** have been accomplished. The comparison of the nucleotide sequence of the Thai isolate with those previously reported was also studied.

MATERIALS AND METHODS

Strains and plasmid

The local isolates of HBV were recovered from the blood specimens of HBV carriers donated to the Thai Red Cross. One isolate which was previously identified as ***adr*** subtype by polymerase chain reaction[16] was used as a source for preparation of HBV DNA. *Escherichia coli* XL-1 Blue [*recA, recA1, lac, endA1, gyrA96, thi, hsdR17, supE44, relA, (F', proAB, lacI^q, lacZDM15, Tn10)*] obtained from Stratagene^R was used for plasmid amplification and preparation of double-stranded template for DNA sequencing. A plasmid pBluescript (pBS) provided by Stratagene^R was used for cloning of the whole genome of HBV-DNA and preparation of nested deletion of the HBV DNA.

DNA preparation and *E. coli* transformation

The complete double-stranded HBV DNA was used for gene cloning. The incomplete double stranded HBV DNA in the Dane particles was subjected to endogenous DNA polymerase reaction and subsequently purified by the method as described by Balachandra *et al.*[17] *E. coli* plasmid DNA was prepared by the alkaline method [18] or the rapid boiling method [19]. *E. coli* was transformed according to the method as described by Sambrook *et al.*[20]

Cloning of the HBV DNA in *E. coli*

The complete double-stranded DNA was digested with *Bam*HI and ligated to the plasmid vector pBSII(+SK) that had been digested with *Bam*HI and dephosphorylated by alkaline phosphatase. The constructed plasmid was used to transform *E. coli* XL-1 Blue to ampicillin resistance. The transformants harboring HBV DNA were selected by blue-white selection on X-gal medium plate and screened as positive clones by colony hybridization. The positive

clones, which were thought to contain the full length HBV DNA of about 3.2 kb, were analysed by restriction endonuclease analysis with *Bam*HI and confirmed by Southern blot analysis. One clone, named pPM13, was chosen for DNA sequencing.

DNA sequencing

In order to sequence both strands, the 3.2 kb *Bam*HI fragment of HBV DNA was subcloned into the plasmid vector pBS(-KS) at the *Bam*HI site. The new clone was named pPM31. A nested series of deletion derivatives (28 in total), each shorter in size than the previous one by approximately 300 bp were generated from pPM13 and pPM31 with exonuclease III. The ends were made blunt by mung bean nuclease and then self ligated. After transformation into *E. coli*, the deleted clones were screened by minipreparation. Both DNA strands were sequenced by the dideoxy chain termination method[21] using reverse primer or M13 primer for templates obtained from pPM13 [pBSII(+SK)] or pPM31 [pBS(-KS)], respectively. DNA sequence data were analyzed with the DNAsis™ program (Hitachi Software Engineering Co. Ltd.).

Colony hybridization and Southern blot analysis

The colonies on X-gal medium plates were transferred to a nylon membrane by the method as described by Sambrook *et al.*[20] The plasmids which contained approximately 3.2 kb insert were digested with various restriction enzymes and subjected to agarose gel electrophoresis. The gel was blot-transferred onto a nitrocellulose membrane according to the method described by Southern[22]. The colonies on membrane or the blot were hybridized with HBV DNA probe (obtained from other subtype clone), labeled with ^{32}P by Prime-a-Gene labeling system (Promega®), and autoradiographed at -70°C with intensified screen.

RESULTS AND DISCUSSION

Cloning of the HBV DNA

In order to clone the full length of HBV DNA, the restriction enzymes that digested only one position in the viral genome were searched from previous reports. It was shown that the HBV DNA of subtype ***adr*** had a unique *Bam*HI site[23,24]. In addition, our ***adr*** HBV DNA was digested with *Bam*HI at one position (data not shown). The completed double-stranded DNA was therefore cloned at the *Bam*HI site of the plasmid pBS. The recombinant plasmids which contained about 3.2 kb inserted were chosen by the size selection of inserts. Southern blot analysis confirmed that the selected clone, pPM13, contained full length HBV DNA (Fig. 1). Restriction endonuclease analysis of this clone revealed that the HBV DNA of ***adr*** subtype isolated in Thailand contained different restriction sites (Fig. 2) from those previous reported[23,24]. The difference in restriction sites was probably due to the geographic distributions among the same HBV subtypes[25].

DNA sequence of the cloned HBV DNA

A nested series of deletions of 3.2 kb *Bam*HI DNA fragment in the plasmid pBS, differing in size by approximately 300 kb, were created by exonuclease III and mung bean nuclease treatment, as shown in Fig. 2. Both strands of the insert were sequenced by the dideoxy chain determination method[21] using double-stranded templates derived from deleted clones of pPM13 and pPM31. As shown in Fig. 3, the sequence of the cloned ***adr*** HBV DNA was 3,215 nucleotides long. The nucleotide 1-6, GAACTC, in Fig. 3 corresponded to the unique *Eco*RI site in other subtypes. For ease of comparison, G (nucleotide 1) in the *Eco*RI-like sequence was

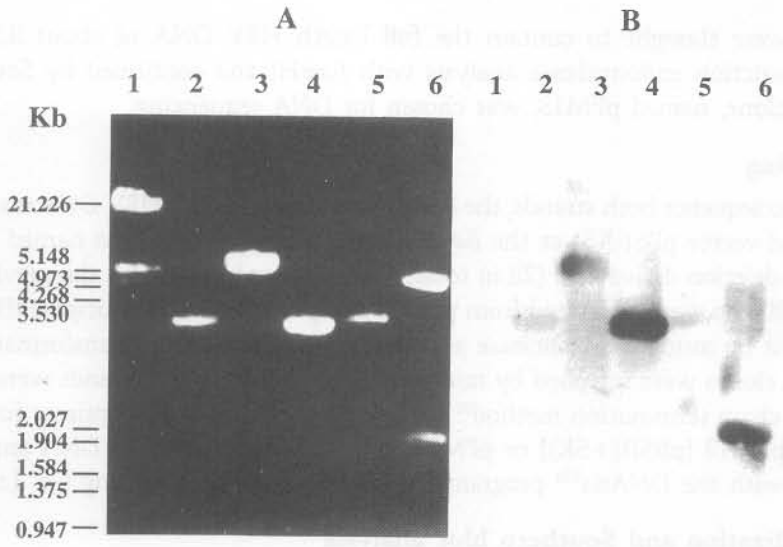

Fig.1 Agarose gel electrophoresis and Southern blot hybridization of the cloned HBV DNA (pPM13).

A : Agarose gel electrophoresis : lane 1, l *Hin*dIII-*Eco*RI marker; lanes 2 and 5, pBSII(+SK) cut with *Sac*I and *Xho*I, respectively; lanes 3,4 and 6, pPM13 cut with *Sac*I, *Bam*HI and *Xho*I, respectively. **B** lanes 1-6 : Southern blot of **A**. The probe is ^{32}P-labeled HBV DNA fragment.

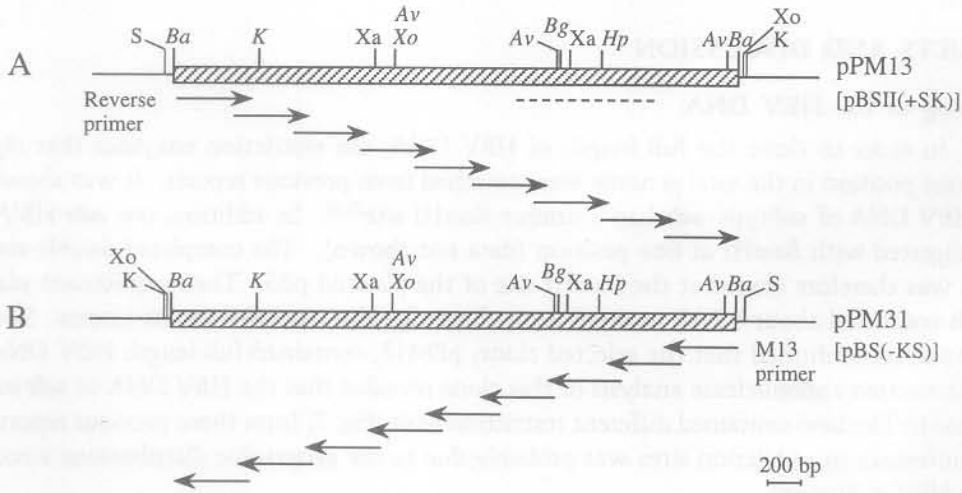

Fig.2 Restriction map of the 3.2 kb *Bam*HI fragment and sequencing strategy.

Bars represent cloned HBV DNA. Solid lines represent plasmid pBS. Broken line indicates the region where probe hybridizes to the cloned HBV DNA. Arrows indicate the DNA regions sequenced from deletion clones. The deleted portions of the fragment corresponds to the distance between the 5' or 3' end of the gene and the beginning of the arrow. Reverse or M13 primers were hybridized with the templates prepared from pPM13 **(A)** or pPM31 **(B)**, respectively. The different restriction sites from those previous reported are shown in italic letters. Restriction site abbreviation : Av, *Ava*I; Ba, *Bam*HI; Bg, *Bgl*II; Hp, *Hpa*I; K, *Kpn*I; *Sac*I, S; Xa, *Xba*I; Xo; *Xho*I.

used as the starting point in our sequence[26]. The nucleotide sequence of our clone was compared with those deposited in the GenBank using two by two analysis. It was found that our sequence showed high homology of about 95.1-96.4% with 13 sequences of ***adr*** subtype. However some of the reported sequences varied in the number of nucleotides from 3,188 to 3,221. Comparison with 6 reported sequences having the same number of nucleotide showed that our sequence showed 95.6% homology with the reported sequence accession number M12906, 96.5% with D23684, 96.2% with D00630, 96.0% with D23681, 95.9% with D23682 and 95.7% with D23683 (data not shown). These sequence variations of 3.5-4.3% were less than 5.6% which was previously reported to be the variation within the same subtypes[25]. The result of nucleotide sequence and restriction site comparison between our clone and those in the reported sequences revealed that the cloned HBV DNA of ***adr*** subtype in Thailand was diverse from the same subtype of different geographic regions.

Identification of viral coding regions

A computer program DNAsis™ was used to determine ATG triplets and termination codons on both strands of DNA, sequence homology and repetitive structures. It was shown that the four major open reading frames (ORF) were found at different phase on one strand (Fig. 3). They had coding capacities of 226, 183, 843 and 154 amino acids and were placed as genes for HBsAg, HBcAg, P and X, respectively.

HBsAg gene: The HBsAg gene was coded by 678 bp resulting in 226 amino acids. It was located between bp 158 and 835 and was preceded by a pre-S region located between bp 2,851 to 3215 and bp 1 to 157. The pre-S region was encoded for additional 174 amino acids. It was divided into pre-S1 region with 119 amino acids and pre-S2 region with 55 amino acids.

HBcAg gene: The HBcAg gene was encoded by a sequence of 1,549 bp that located between bp 1,904 and 2,452. This sequence had the coding capacity for 183 amino acids. It was preceded by a pre-C region of 87 bp that located between bp 1,817 to 1,903. This gene was encoded for 29 amino acids.

Gene P: The P gene was coded by 2,529 bp that located between bp 2,310 to 3215 and bp 1 to 1,623. It was encoded for 843 amino acids.

Gene X: The X gene was coded by a sequence located between bp 1,377 to 1,838. It was encoded for 154 amino acids.

Direct repeat sequence (DR): The DR sequence of undecanucleotides, "TTCACCTCTGC" or DR1, was found between bp 1,593 to 1,603 and 1,827 to 1,837. This presumably was the location of the short (S) and long (L) strand on the 5' end, respectively. In addition, other two small direct repeat sequences, "AAAGACT" or DR2, were found between bp 1,713 to 1,719 and 1,728 to 1,734. The DR3 sequence, "GGGAGGAG," was found near the middle of the two DR1 that were located at positions between bp 1,735 to 1,742 and 1,747 to 1,754. The sequence "GATTCGG" was found between bp 1,754 to 1,760 just downstream from the DR3. It was found to be inversely repeated (I) at position 1,837 to 1,843, except that one nucleotide was diverse "CCTAATC", where gene X terminated. The presence of DR sequences revealed characteristic structures of HBV DNA in our clone.

In conclusion the organization of viral coding region in our cloned HBV DNA was found to be conserved[27].

Characterization of the HBsAg gene product

It was shown previously that the antigenic ***a*** determinant of HBV DNA was on the

```
W N S S T F H Q A L L D P R V R G L Y F P A G G S S S
  E L Q H I P P S S A R S Q S E G P I L S C W W L K F
GAACTCCAGCACATTCCACCAAGCTCTGCTAGATCCCAGAGTGAGGGGCCTATACTTTCCTGCTGGTGGCTCAAGTTCCG
        10        20        30        40        50        60        70        80

                                                                    HBsAg
 G T V N P V P T T A S P I S S I F S R T G D P A P N (M)
R N S K P C S D Y C L S H I V N L L E D W G P C T E Y
GAACAGTAAACCCTGTTCCGACTACTGCCTCTCCCATATCGTCAATCTTCTCGAGGACTGGGGACCCTGCACCGAATATG
        90       100       110       120       130       140       150       160

  E S T T S G F L G P L L V L Q A G F F L L T R I L T
  G E H H I R I P R T P A R V T G G V F L V D K N P H N
GAGAGCACCACATCAGGATTCCTAGGACCCCTGCTCGTGTTACAGGCGGGGTTTTTCTTGTTGACAAGAATCCTCACAAT
        170       180       190       200       210       220       230       240

I P Q S L D S W W T S L N F L G G A P T C P G Q N L Q
  T T E S R L V V D F S Q F S R G S T H V S W P K F A
ACCACAGAGTCTAGACTCGTGGTGGACTTCTCTCAATTTTCTAGGGGGAGCACCCACGTGTCCTGGCCAAAATTTGCAGT
        250       260       270       280       290       300       310       320

 S P T S N H S P T S C P P I C P G Y R W M C L R R F I
V P N L Q S L T N L L S S N L S W L S L D V S A A F Y
CCCCAACCTCCAATCACTCACCAACCTCTTGTCCTCCAATTTGTCCTGGTTATCGCTGGATGTGTCTGCGGCGTTTTATC
        330       340       350       360       370       380       390       400

  I F L F I L L L C L I F L L V L L D Y Q G M L P V C
  H L P L H P A A M P H L L V G S S G L P R Y V A R L S
ATCTTCCTCTTCATCCTGCTGCTATGCCTCATCTTCTTGTTGGTTCTTCTGGACTACCAAGGTATGTTGCCCGTTTGTCC
        410       420       430       440       450       460       470       480

P L L P G T S T T S T G P C K T C T I P A Q G T S M F
  S T S R N I N Y Q H G T M Q D L H D S C S R N L Y V
TCTACTTCCAGGAACATCAACTACCAGCACGGGACCATGCAAGACCTGCACGATTCCTGCTCAAGGAACCTCTATGTTTC
        490       500       510       520       530       540       550       560

 P S C C C T K P S D G N C T C I P I P S S W A F A R F
S L L L L Y K T F G R K L H L Y S H P I I L G F R K I
CCTCTTGTTGCTGTACAAAACCTTCGGACGGAAATTGCACTTGTATTCCCATCCCATCATCTTGGGCTTTCGCAAGATTC
        570       580       590       600       610       620       630       640

  L W E W A S V R F S W L S L L V P F V Q W F V G L S
  P M G V G L S P F L L A Q F T S A I C S V V R R A F P
CTATGGGAGTGGGCCTCAGTCCGTTTCTCCTGGCTCAGTTTACTAGTGCCATTTGTTCAGTGGTTCGTAGGGCTTTCCCC
        650       660       670       680       690       700       710       720

P T V W L S V I W M M W Y W G P S L Y N I L S P F L P
  H C L A F S Y M D D V V L G A K S V Q H L E S L F T
CACTGTTTGGCTTTCAGTTATATGGATGATGTGGTATTGGGGGCCAAGTCTGTACAACATCTTGAGTCCCTTTTTACCTC
        730       740       750       760       770       780       790       800

L L I P F F C L W V Y I (*) HBsAg
S I T N F L L S L G I H L N P N K T K R W G Y S L N F
TATTACCAATTTTCTTTTGTCTTTGGGTATACATTTAAACCCTAATAAAACCAAACGTTGGGGCTATTCCCTTAACTTCA
        810       820       830       840       850       860       870       880

 M G Y V I G S W G T L P Q E H I V Q K L K Q C F R K L
TGGGATATGTAATTGGAAGTTGGGGTACCTTACCACAGGAACATATTGTACAAAAATTGAAACAATGTTTTCGGAAACTT
        890       900       910       920       930       940       950       960

  P I N R P I D W K V C Q R I V G L L G F A A P F T Q
CCTATAAATAGACCTATTGATTGGAAAGTATGTCAAAGAATTGTGGGTCTTCTGGGCTTTGCTGCCCCTTTTACACAATG
        970       980       990      1000      1010      1020      1030      1040

C G Y P A L M P L Y A C I Q A K Q A F T F S P T Y K A
TGGTTACCCAGCATTAATGCCTTTATATGCATGTATACAAGCTAAGCAGGCTTTCACTTTCTCGCCAACTTACAAGGCCT
      1050      1060      1070      1080      1090      1100      1110      1120

 F L C K Q Y L N L Y P V A R Q R S G L C Q V F A D A T
TTCTGTGTAAACAATATCTGAACCTTTACCCCGTTGCTCGGCAACGGTCAGGTCTTTGCCAAGTGTTTGCTGACGCAACC
      1130      1140      1150      1160      1170      1180      1190      1200
```

Fig.3(1)

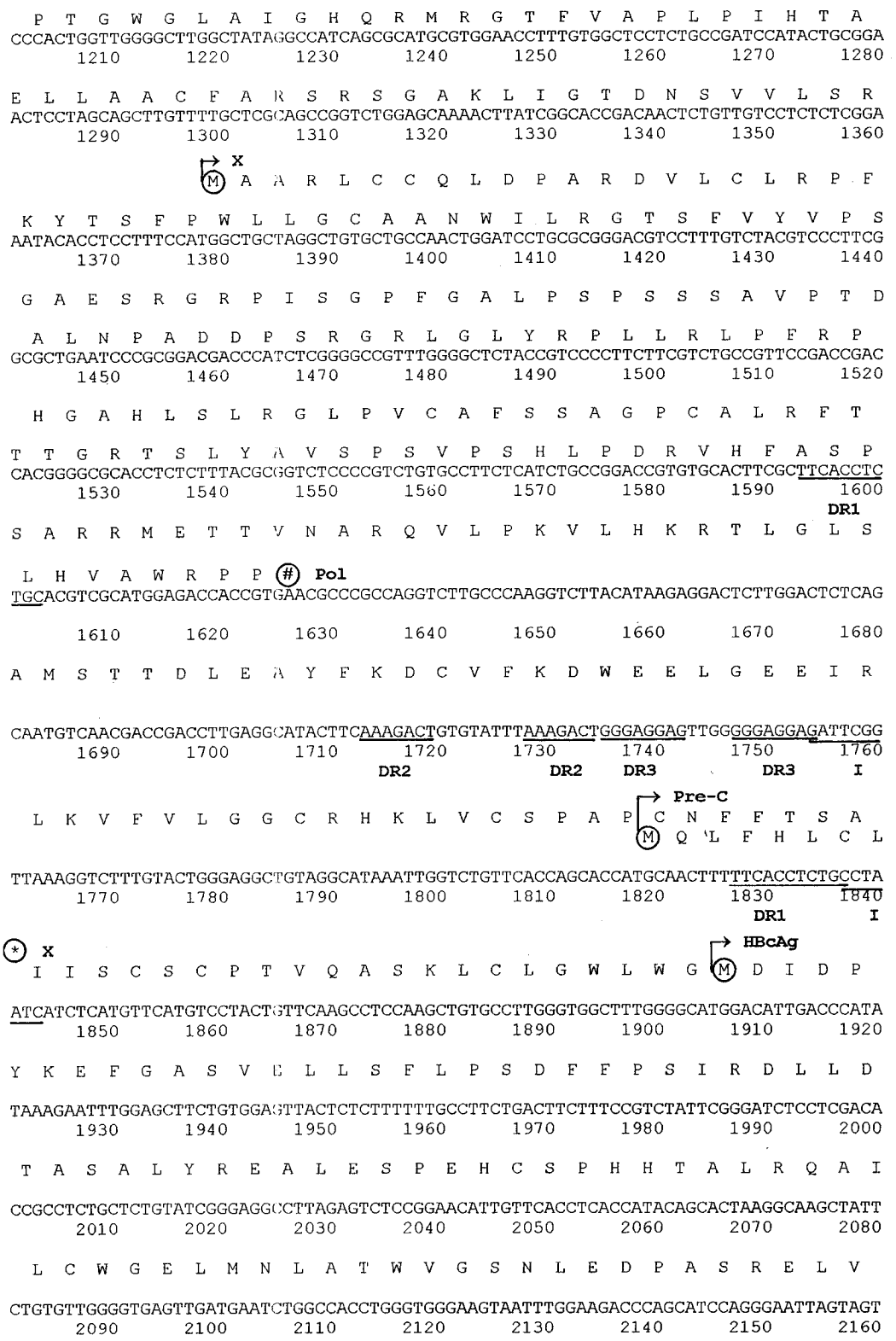
P T G W G L A I G H Q R M R G T F V A P L P I H T A
CCCACTGGTTGGGGCTTGGCTATAGGCCATCAGCGCATGCGTGGAACCTTTGTGGCTCCTCTGCCGATCCATACTGCGGA
1210 1220 1230 1240 1250 1260 1270 1280
E L L A A C F A R S R S G A K L I G T D N S V V L S R
ACTCCTAGCAGCTTGTTTTGCTCGCAGCCGGTCTGGAGCAAAACTTATCGGCACCGACAACTCTGTTGTCCTCTCTCGGA
1290 1300 1310 1320 1330 1340 1350 1360
X
M A A R L C C Q L D P A R D V L C L R P F
K Y T S F P W L L G C A A N W I L R G T S F V Y V P S
AATACACCTCCTTTCCATGGCTGCTAGGCTGTGCTGCCAACTGGATCCTGCGCGGGACGTCCTTTGTCTACGTCCCTTCG
1370 1380 1390 1400 1410 1420 1430 1440
G A E S R G R P I S G P F G A L P S P S S S A V P T D
A L N P A D D P S R G R L G L Y R P L L R L P F R P
GCGCTGAATCCCGCGGACGACCCATCTCGGGGCCGTTTGGGGCTCTACCGTCCCCTTCTTCGTCTGCCGTTCCGACCGAC
1450 1460 1470 1480 1490 1500 1510 1520
H G A H L S L R G L P V C A F S S A G P C A L R F T
T T G R T S L Y A V S P S V P S H L P D R V H F A S P
CACGGGGCGCACCTCTCTTTACGCGGTCTCCCCGTCTGTGCCTTCTCATCTGCCGGACCGTGTGCACTTCGCTTCACCTC
1530 1540 1550 1560 1570 1580 1590 1600
DR1
S A R R M E T T V N A R Q V L P K V L H K R T L G L S
L H V A W R P P # Pol
TGCACGTCGCATGGAGACCACCGTGAACGCCCGCCAGGTCTTGCCCAAGGTCTTACATAAGAGGACTCTTGGACTCTCAG
1610 1620 1630 1640 1650 1660 1670 1680
A M S T T D L E A Y F K D C V F K D W E E L G E E I R
CAATGTCAACGACCGACCTTGAGGCATACTTCAAAGACTGTGTATTTAAAGACTGGGAGGAGTTGGGGGAGGAGATTCGG
1690 1700 1710 1720 1730 1740 1750 1760
DR2 DR2 DR3 DR3 I
Pre-C
L K V F V L G G C R H K L V C S P A P C N F F T S A
M Q L F H L C L
TTAAAGGTCTTTGTACTGGGAGGCTGTAGGCATAAATTGGTCTGTTCACCAGCACCATGCAACTTTTTCACCTCTGCCTA
1770 1780 1790 1800 1810 1820 1830 1840
DR1 I
X
HBcAg
I I S C S C P T V Q A S K L C L G W L W G M D I D P
ATCATCTCATGTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGGCATGGACATTGACCCATA
1850 1860 1870 1880 1890 1900 1910 1920
Y K E F G A S V E L L S F L P S D F F P S I R D L L D
TAAAGAATTTGGAGCTTCTGTGGAGTTACTCTCTTTTTTGCCTTCTGACTTCTTTCCGTCTATTCGGGATCTCCTCGACA
1930 1940 1950 1960 1970 1980 1990 2000
T A S A L Y R E A L E S P E H C S P H H T A L R Q A I
CCGCCTCTGCTCTGTATCGGGAGGCCTTAGAGTCTCCGGAACATTGTTCACCTCACCATACAGCACTAAGGCAAGCTATT
2010 2020 2030 2040 2050 2060 2070 2080
L C W G E L M N L A T W V G S N L E D P A S R E L V
CTGTGTTGGGGTGAGTTGATGAATCTGGCCACCTGGGTGGGAAGTAATTTGGAAGACCCAGCATCCAGGGAATTAGTAGT
2090 2100 2110 2120 2130 2140 2150 2160

Fig.3(2)

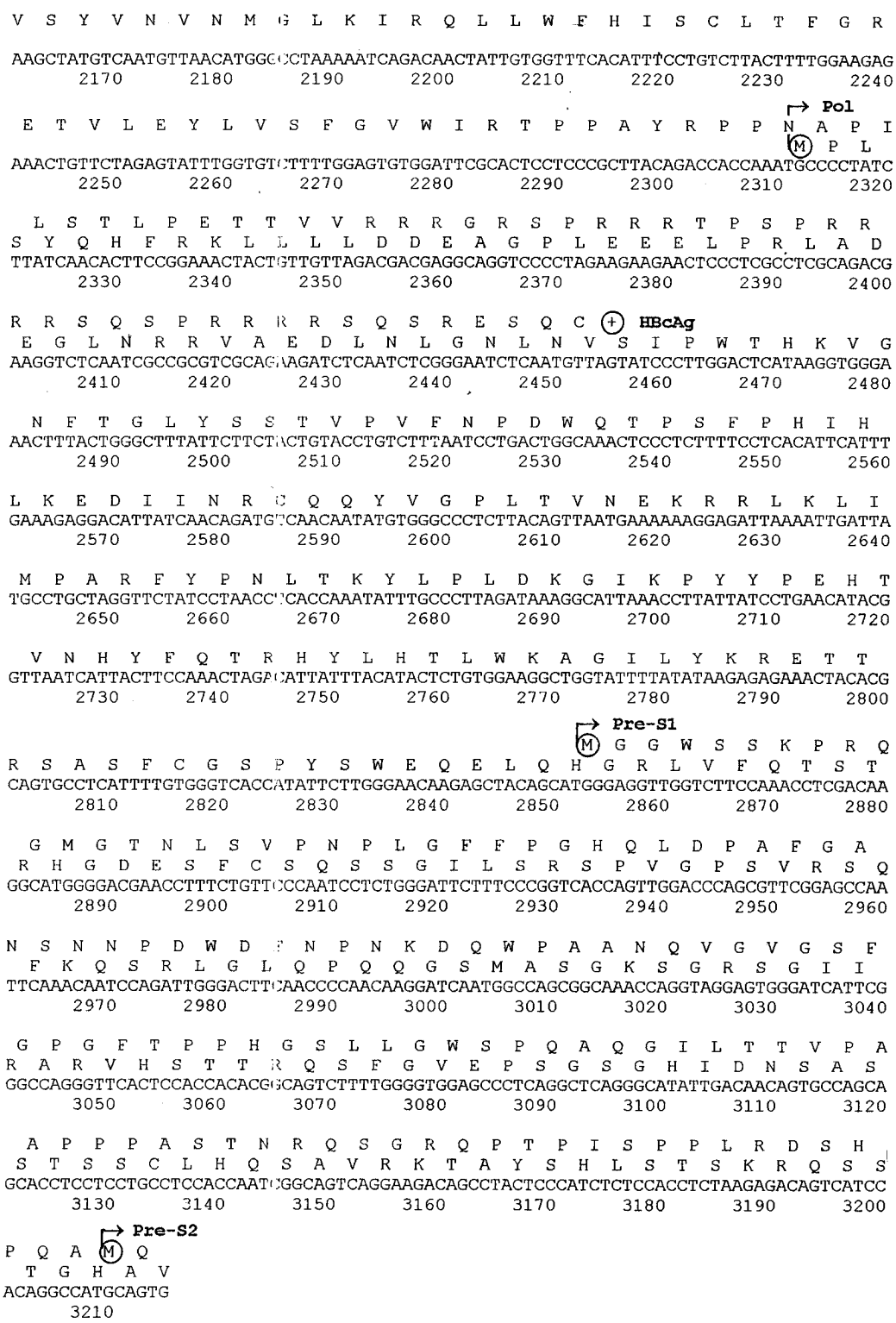

Fig.3(3) Complete nucleotide sequence of the cloned HBV DNA and its protein coding.
There are 3,215 nucleotides in length. This sequence is numbered from the arbitrary *Eco*RI cleavage site. Arrows and open circles indicate the beginning and ending of each coding regions of HBsAg, HBcAg, viral DNA polymerase (Pol) and X-protein (X). Coding regions of pre-S and pre-C are indicated only the beginning. The direct repeat sequences (DR) and inverted sequences (I) are underlined.

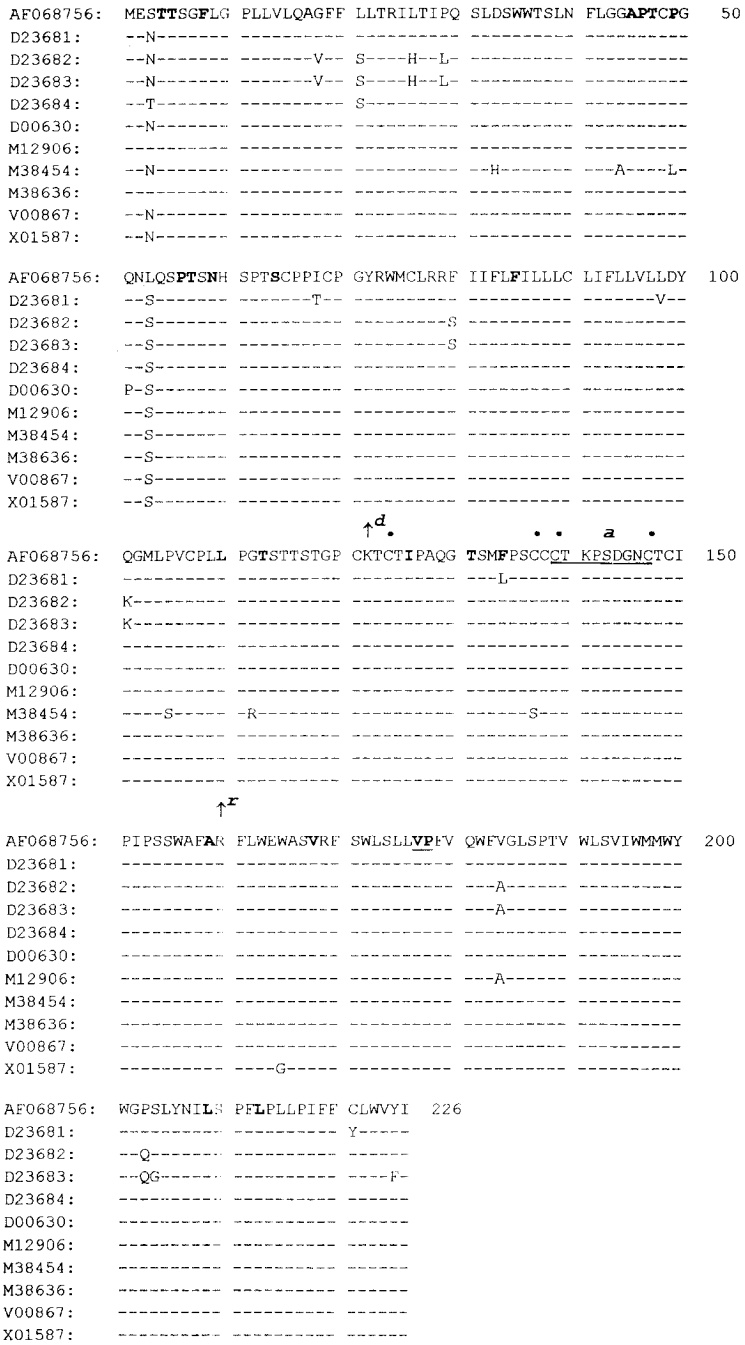

Fig.4 Comparison of deduced amino acid sequence of HBsAg from cloned HBV DNA with published sequence of *adr* subtype.

Sequence in line 1 is derived from this study. The sequence has been submitted to GenBank and its accession number is AF068756. Those in line 2-11, which only different amino acids from line 1 are written, are obtained from GenBank. The accession number of the sequences are shown on the left. The amino acids of proposed ***a*** determinant are underlined. The subtype determinant ***d*** (Lys -122) and ***r*** (Arg -160) are indicated by arrows. The cysteine cluster between amino acid 124-147 are marked with closed circles. Amino acids of HBsAg at 21 group specific positions within group **C** are shown in bold. Val-177 and Pro-178 expressing ***q*** determinant are shown in bold and underline.

```
                ↱ Pre-S1
AF068756:  MGGWSSKPRQ GMGTNLSVPN PLGFFPGHQL DPAFGANSNN PDWDFNPNKD   50
D23681:    ---Y------ H--------- ------D--- ---------- ----------
D23682:    ---------- ---------- ------D--- ---------- ----------
D23683:    ---------- ---------- ------D--- ---------- ----------
D23684:    ---------- ---------- ------D--- ---------- ----------
D00630:    ---------- ---------- ------D--- ---------- ----------
M12906:    ---------- ---------- ------D--- ---------- ----------
M38454:    ---------- ---------- ------D--- ---------- ----------
M38636:    ---------- ---------- ------D--- ---------- ----------
V00867:    ---------- ---------- ------D--- ---------- ----------
X01587:    ---------- ---------- ------D--- --------H- ----------

AF068756:  QWPAANQVGV GSPGPGFTPP HGSLLGWSPQ AQGILTTVPA APPPASTNRQ  100
D23681:    H--E-----A -A-------- --G------- ---------V ----------
D23682:    H--E-----A -A-------- --G------- ---T------ ----------
D23683:    H--E-----A -A-------- --G------- ---T------ ----------
D23684:    H--E-----A -A--------- --G------- ---A-----V ----------
D00630:    H--E-----A -A-------- --G------- ---------- ----------
M12906:    H--E-IK--A -D-------- --G------- ---------- ----V-----
M38454:    H--E-----A -A-------- --G------- ---------V ----------
M38636:    R--E-----A -A----YP-- --G------- ---------- ----------
V00867:    ---E-----A -A-------- --G------- ---------- ----------
X01587:    H--E-----A -A-------- --G------- ---V-----V ----------
                                ↱ Pre-S2
AF068756:  SGRQPTPISP PLRDSHPQAM QWNSSTFHQA LLDPRVRGLY FPAGGSSSGT  150
D23681:    ---------- ---------T ----T----- ---------- ----------
D23682:    ---------- ---------- ----T----- ---------- ----------
D23683:    ---------- ---------T ----T----- ---------- ----------
D23684:    ---------- ---------I ----T----- ---------- ----------
D00630:    ---------- ---------- ----T----- V--------- ----------
M12906:    ---------- ---------- ----T----- ---------- ----------
M38454:    ---------- ---------- ----T----- ---------- ----------
M38636:    ---------- ---------- ----T----V ---------- --P-------
V00867:    ---------- ---------- ----T----- ---------- ----------
X01587:    ---------- ---------- ----T----- ---------- ----------

AF068756:  VNPVPTTASP ISSIFSRTGD PAPN  174
D23681:    ---------- ---------- ----
D23682:    ------I--- ---------- ----
D23683:    ------I--- ---------- ----
D23684:    ---------- ---------- ----
D00630:    ---------- ---------- ----
M12906:    -------V-- ---------- ----
M38454:    ---------- ---------- ----
M38636:    ----A----- ---------- ----
V00867:    ---------- ---------- ----
X01587:    ---------- ---------- ----
```

Fig.5 Comparison of deduces amino acid sequence of pre-S region from cloned HBV DNA with published sequence of *adr* subtype.

The sources of sequence in line 1-11 are the same as in Fig. 4. The initiation codon of prë-S1 and pre-S2 are indicated by arrows. The proposed immunodominant in the pre-S2 region is underlined.

polypeptide of 226 residues encoded by the HBsAg gene. The variability of HBsAg subtype, ***d*** or ***w*** and ***w*** or ***r***, was attributed to variation in this polypeptide. The comparison of the deduced amino acid sequence of our cloned HBsAg gene with 10 of those reported in the GenBank were carried out by two by two analysis. The results were summarized in Fig. 4. It was found that our cloned HBsAg showed high homology of about 95.6-99.6% with the reported sequence. Furthermore the 9 amino acids from 139 to 147 (CTKPSDGNC) being significant for ***a*** determinant[28] were conserved in our clone. This determinant epitope was located on a predicted double loop structure within domains bordered by amino acids 120-147[29]. The cysteine residues at positions 124, 137, 139 and 147 proposed to play a role in forming loop by disulfide bridge and the presence of Cys-124, Cys-147 and Pro-142 required for full antigenicity[30] were observed. The Gly-145 has been reported to be an important part for expression of group specific determinant. Its alteration resulted in an escape mutation or vaccine escape mutation[31-35]. Yamamoto *et al.*[35] found that the mutation of Ile-126 to Ser-126 was also responsible for the loss of group specific determinant. In this study, both Ile-126 and Gly-145 were conserved. The exclusive subtype determinant, lysine-122 for ***d*** and arginine-160 for ***r***[36-37], clearly indicated that our cloned HBV DNA was ***adr*** subtype.

With the emergence of a number of HBV DNA sequences, 6 genotypes were defined based on the HBsAg gene sequence and were designated as genotype A-F[37]. The availability of the HBsAg gene was found to be related to molecular epidemiology of HBV. Genomic encoding ***adr*** subtype was found in group C which was confined to the original population of Far East. As shown in Fig. 4, the amino acids at 21 specific positions[38], i.e., 4, 5, 8, 45, 46, 47, 49, 56, 57, 59, 64, 85, 110, 113, 124, 131, 134, 159, 168, 209, 213 (shown in bold) were found along the protein sequence[38]. This result indicated that our cloned HBV DNA was classified as group C.

Based on the amino residues at position 177 and 178 in the HBsAg gene[39], those with Val-177 and Pro-178 were classified as ***q*** subtype determinant. These amino acids were found in all HBV subtypes (***adr***, ***adw2***, ***ayr***, ***ayw1***, ***ayw2***, ***ayw3*** and ***ayw4***) except subtype ***adrq-*** and ***adw4q-*** which had Ala-177 and Glu-178, respectively. Based on this criteria, our cloned HBV DNA had ***q*** subtype determinant.

Characterization of the pre-S region gene product

The deduced amino acid sequence in the pre-S region was also compared with 10 reported sequences by the same analysis. As shown in Fig. 5, the homology between amino acids in our cloned pre-S region and those in the reported sequences was about 93.6-96.6%. In addition the variation in the pre-S1 was found to be more often than in the pre-S2 region. It had been reported that sequences of pre-S were much less conserved than sequences of HBsAg[40].

It was previously identified that the immunodominant part of pre-S2 region was between amino acid 123-145[41]. The important amino acid sequence in this region was defined as 135-L(Q/L)DPRVRGLY(F/L)PAG-144 by Meisel *et al.*[15] In our pre-S2 region, the sequence 135-LLDPRVRGLYFPAG-144 was found. The result indicated that our cloned HBV DNA contained conserved immunodominant.

CONCLUSION

In this study the HBV DNA isolated from infectious virions found in HBV carrier was cloned and sequenced. It was 3,215 bp long. The analysis of HBsAg gene at the nucleotide level showed that the virus was ***adr*** subtype and the amino acid sequence in the antigenic ***a***

determinant and immunodominant parts were found to be conserved. We believe that this clone would be suitable for the production of recombinant vaccine.

ACKNOWLEDGEMENTS

This work was supported by grant from the Government Pharmaceutical Organization (GPO) and partly by NSTDA grant to Miss Patama Monkongdee.

REFERENCES

1. Fortuin, M., Chotard, J., Jack, A.D., Maine, N.P., Mendy, M., Hall., A.J., Inskip, H.M., George, M.O. and Whittle, H.C. (1993). *Lancet* **341**, 1129-1131.
2. Ellis, R.W. and Kniskern, P.J. (1991). In Molecular Biology of Hepatitis B Virus, McLachlan, A. (ed.) (1991). CRC Press Inc., Boca Raton, Florida, p. 307-322.
3. Imai, M., Yanase, Y., Nojiri, T., Miyakawa, Y. and Miyumi, M. (1979). *Gastroenterology* **76**, 242-247.
4. Heerman, K.H., Kruse, F., Seifer, M. and Gerlich, W.H. (1987). *Intervirology* **28**, 14-25.
5. Machida, A., Kishimoto, S., Ohnuma, H., Miyamoto, H., Baba, K., Oda, K., Nakamura, T., Miyakawa, Y. and Mayumi, M. (1983). *Gastroenterology* **85**, 268-274.
6. Krone, B., Lenz, A., and Heerman, K.H. (1990). *Hepatology* **11**, 1050-1056.
7. Persing, D.H., Varmus, H.E. and Ganem, D. (1985). *Proc. Natl. Acad. Sci. USA* **82**, 3440-3444.
8. Pontisso, P., Petit, M.A. Bankowski, M.J. and Peeples, M.E. (1989). *J. Virol.* **63**, 1981-1988.
9. Milich, D.R., McNamara, M.K., McLachlan, A., Thornton, G.B. and Chisari F.V. (1985). *Proc. Natl. Acad. Sci. USA* **82**, 8168-8172.
10. Milich, D.R., Thornton, G.B., Neurath A.R., Kent, S.B. Michel, M.L., Tiollais, P. and Chisari, F.V. (1985). *Science* **228**, 1195-1199.
11. Budkowska, A., Dubreuil, P., Capel, F. and Pillot, J. (1986). *Hepatology* **6**, 360-368.
12. Milich, D.R., McLachlan, A., Chisari, F.V., Nakamura, T. and Thornton, G.B. (1986). *J. Immnunol.* **173**, 2703-2710.
13. Okamoto, H. Usuda, S., Imai, M., Tachibana, K., Tanaka, E., Kumakura, T., Itahashi, M., Takai, E., Tsuda, F., Makamura, T., Miyakawa, Y. and Mayumi, M. (1986). *Hepatology* **6**, 354-359.
14. Pontisso, P., Schiavon, E., Fraiese, A., Pornaro, E., Realdi, and Alberti, A. (1986). J. *Hepatol.* **3**, 393-398.
15. Meisel, H., Sominskaya, I., Pumpens, P., Pushko, P., Borisova, G., Deepen, Lu, X., Spiller, G.H., Kruger, D.H., Grens, E. and Gerlich, W.H. (1994). *Intervirology* **37**, 330-339.
16. Thawaranantha, D., Balachandra, K., Watanaseree, J., Boonchird, C., Warachit, P., Bhumisawasdi and Pantuwatana, S. (1994). *J. Sci. Soc. Thailand* **20**, 115-124.
17. Balachandra, K., Thawaranantha, D., Watanaseree, J., Techapaiboonying, D., Warachit, P., Bhumisawasdi, J. And Pantuwatana, S. (1994). *J. of the Government Pharmaceutical Organization* **20**, 9-19.
18. Birnboim, H.C. and Doly, J. (1979). *Nucleic Acids Res.* **7**, 1513-1523.
19. Holmes, A., Hick, J.B. and Fink, G.R. (1981). *Anal. Biochem.* **114**, 193-197.
20. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1982). Molecular cloning A laboratory manual. 2nd edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
21. Sanger F., Nicklen, S. and Coulon, A.R. (1977). *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
22. Southern, E.W. (1975). *J. Mol. Biol.* **98**, 503-517.
23. Ono, Y., Onda, H., Sasada, R., Igarashi, K., Sugino, Y. and Nishioka, Y. (1983). *Nuceic Acids Res.* **11**, 1747-1757.
24. Fujiyama, A., Miyanohara, A., Nozaki, C., Yoneyama, T., Ohtoma, N., Matsubara, K. (1983). *Nucleic Acids Res.* **11**, 4601-4610.
25. Okamoto, H., Tsuda, F., Sakugawa, H., Sastrosoewignjo, R., Imai, M., Miyakawa, Y., and Mayumi, M. (1998). *J. Gen. Virol.* **69**, 2575-2583.
26. Raney, A.K. and McLachlan, A. (1991). In *Molecular Biology of Hepatitis B Virus*, McLachlan, A. (ed.) (1991). CRC Press Inc., Boca Raton, Florida, p. 1-37.

27. Kobayashi, M. and Koike, K. (1984). *Gene* **30**, 227-232.

28. Steward, M.W., Partidos, C.D., D'Mello, F. and Howard, C.R. (1993). *Vaccine* **11**, 1405-1414.

29. Stirk, H.J., Thornton, J.M. and Howard, C.R. (1992). *Intervirology* **33**, 146-158.

30. Ashton-Rickardt, P.G. and Murray, K. (1989). *J. Med. Virol.* **29**, 204-214.

31. Carman, W.F., Zanetti, A.R., Karayiannis, P., Waters, J., Manzillo, G., Tanzi, E., Zuckerman, A.J. and Thomas, H.C. (1990). *Lancet* **336**, 325-329.

32. Harrison, T.J., Hopes, E.A., Oon, C.J., Zanetti, A.R. and Zuckerman, A.J. (1991). *J. Hepatol.* **13** (suppl. 4), 105-107.

33. Fujii, H., Moriyama, K., Sakamoto, N., Kondo, T., Yasuda, K., Hiraizumi, Y., Yamazami, M., Sakaki, Y., Okochi, K. and Brechot, C. (1992). *Biochem. Biophy. Res. Commun.* **184**, 1152-1157.

34. Okamoto, H., Yano, K., Nozaki, Y., Matsui, A., Miyazaki, H., Yamamoto, K., Tsuda, F., Machida, A. and Mishiro, S. (1992). *Pediatr. Res.* **32**, 264-268.

35. Yamamoto, K., Horikita, M., Tsuda, F., Itoh, K., Akahane, Y., Yotsumoto, S., Okamoto, H., Miyakawa, Y. and Mayumi, M. (1994). *J. Virol.* **68**, 2671-2676.

36. Okamoto, H., Imai, M., Tsuda, F., Tanaka, T., Miyakawa, Y. and Mayumi, M. (1987). *J. Virol.* **61**, 3030-3034.

37. Okamoto, H., Omi, S., Wang, Y., Itoh, Y., Tsuda, F., Tanaka, T., Akahane, Y., Miyakawa, Y. and Mayumi, M. (1989). *Mol. Immuno.* **26**, 197-205.

38. Norder, H., Hammas, B., Lee, S.D., Courouce', A.-M., Mashashwar, I.K., and Madnius, L.O. (1993). *J. Gen. Virol.* **74**, 1341-1348.

39. Norder, H., Hammas, B., Lofdahl, S., Courouce', A.M. and Magnius, L.O. (1992). *J. Gen. Virol.* **73**, 1201-1208.

40. Neurath, A.R. and Kent, S.B.H. (1998). *Adv. Virus Res.* **34**, 65-142.

41. Mimms, L.T., Floreani, M., Tyner, J., Whitters, E., Rosenlof, R., Wray, L., Goetze, A., Sarin, V. and Edle, K. (1990). *Virology* **176**, 604-619.

บทคัดย่อ

ดีเอ็นเอของไวรัสตับอักเสบมีสายพันธุ์ *adr* จาก Dane particle ซึ่งแยกได้จากเลือดที่บริจาคให้แก่สภากาชาดไทย ได้ถูกโคลนเข้าพลาสมิด pBluescript ลำดับเบสทั้งหมดของยีโนมได้ถูกวิเคราะห์ และพบว่ามีจำนวนเบส 3,215 สายคู่ ที่มี Open reading frame สำหรับสร้างเอนติเจนส่วนผิว (HBsAg) และ Pre-HBsAg (Pre-S) โปรตีน P เอนติเจนส่วนแกน (HBsAg) และโปรตีน X โดยมีจำนวนกรดอะมิโน 226 และ 174, 843, 212 และ 154 ตัว ตามลำดับ ดีเอ็นเอที่โคลนได้นี้มีคุณสมบัติที่จำเพาะของไวรัสตับอักเสบบี กล่าวคือ มีลำดับเบสที่เป็น Direct repeat ที่ประกอบด้วย 11 เบส (DR1) และ Direct repeat ที่มีลำดับเบสสั้น ๆ อีก 2 แบบ (DR2 และ DR3)

จากการวิเคราะห์ลำดับกรดอะมิโนของเอนติเจนส่วนผิวซึ่งถอดระหัสมาจากดีเอ็นเอ พบกรดอะมิโน 9 ตัว ที่ตำแหน่ง 139-CTKPSDGNC-147 ซึ่งเป็นลักษณะจำเพาะของ Antigenic determinant *a* นอกจากนี้ยังพบกรดอะมิโน Cys-124, Cys-147 และ Pro-142 ที่มีส่วนช่วยให้เกิด Antigenicity โดยสมบูรณ์ จากการเปรียบเทียบลำดับกรดอะมิโนของเอนติเจนส่วนผิว พบว่าดีเอ็นเอของไวรัสที่โคลนได้จัดอยู่ใน Group C และมีกรดอะมิโน Val-177 และ Pro-178 ซึ่งบ่งชี้ว่าเป็น *q* Subtype determinant นอกจากนี้จากการวิเคราะห์ลำดับกรดอะมิโนของส่วน Pre-S2 พบว่ามีกรดอะมิโนที่ตำแหน่ง 131-LLDPRVRGLYFAG-144 ที่มีผู้เสนอว่าเป็น Immunodominant

Note : The DNA sequence has been submitted to the GenBank and its accession number is AF068756