Cloning and Nucleotide Sequence analysis of Pre-S1 Surface Antigen Region of Hepatitis B Virus

NAAZ ABBAS, SAMINA ABBAS AND TALAT AFROZE
PCSIR Labs. Complex, Peshawar (NA) and National Centre of Excellence in Molecular Biology, University of the Punjab, Lahore (SA, TA)

Abstract.- Nucleotide sequence of the Pre-S1 region of the surface antigen of hepatitis B virus (HBV) from local population has been determined. Viral DNA extracted from blood plasma of two blood samples NA007 and NA013 from Services Hospital, Lahore was amplified as 1.1 Kb and 0.6 Kb fragments using two different sets of primers (NHBF1 and NHBR1; B170AS and B2833S). The amplified fragment was subcloned in pGEM-T vector. Pre-S1 region of the two samples was sequenced by non radioactive dideoxy chain termination method. Pre-S1 region nucleotide sequence of both samples were found homologous with the Pre-S1 region of reported subtype “adr4”. Sequence Alignments showed that NA007 and NA013 were 96% to 100% identical with various adr subtypes (genotype C) and had 9.8% to 23.4% nucleotide variations with other subtypes/genotypes.

Key words: Hepatitis B virus, PreS surface antigen, PCR amplification.

INTRODUCTION

Hepatitis B virus is the smallest known DNA virus that can infect man. It is serious health problem with more than 500 million carriers. It is the major cause of liver cancer and 62% of all reported cancers are caused by HBV (Wai and Fontana, 2004; Raffa et al., 2007). IARC (1994) reported South Asia as the area of intermediate endemicity (2-7%), while reports from Pakistan shows 10-30% endemicity (Malik and Akhtar, 1995; Malik et al., 1995). Hepatitis B virus has a complex genomic organization. It has four overlapping reading frames S, C, P and X, which encode seven polypeptides or viral proteins. Three envelope proteins viz., major (HBsAg), middle (PreS2 + HBsAg) and large proteins (PreS1+PreS2+HBsAg) are found in the serum and can induce protective antibodies (Neurath et al., 1989; Nunez et al., 2001; Madalinski et al., 2004; Abbas et al., 2007). Pre-S region of surface gene is highly immunogenic (Milich et al., 1985; Madalinski et al., 2004) and is most hypervariable part of the HBV genome. Immunogenic response of Pre-S region is able to protect from HBV infection (Milich, 1989; Kazaks et al., 2004). The Pre-S region has five known epitopes (Kuroki et al., 1990; Mimms et al., 1990) and two hepatocellular binding sites (Pontisso et al., 1989). Brind et al. (1997) have also provided in vivo evidence for reinfection of the Liver by HBV particles lacking PreS envelope protein expression.

Based on sequence heterogeneity, hepatitis virus has eight genotypes A-H that are geographically distributed (Okamoto et al., 1988; Norder et al., 1993; Stuyver et al., 2000; Arauz-Ring et al., 2002; Ong et al., 2005). Significant heterogeneity (greater than 10%; reviewed in IARC monograph, 1994) in the envelope proteins of HBV, reducing the efficacy of passive immunity reagents and also allows carriers of Asian HBV isolates to go undetected with current serological detection kits because most of these kits are developed with European and American HBV isolates. This heterogeneity also reduces the efficacy of vaccines (developed with non Asian HBV isolates) for Asian population. So to find out the variation among the Pre-S1 surface antigen region the Pre-S1 region of Pakistani HBV isolates was sequenced and compared with other reported sequences.

MATERIALS AND METHODS

Patient history

Two clinically diagnosed HBsAg positive patients blood samples were collected from Services Hospital, Lahore. Patient designated NA007 was female (15Y) while NA013 (55Y) was male.
According to the medical record of the hospital NA007, admitted for endocarditis, belonged to Lahore while NA013 with severe condition of blood vomiting was from Faisalabad.

The HBV DNA was amplified by PCR using external set of primers. PCR amplified product was run on 0.8% agarose gel and 1.1 kb band was eluted by Wizard PCR purification system or Sigma’s Nucleiclean Kit. The gel purified DNA was cloned in pGEM-T vector, specially designed for the cloning of PCR product because this vector has T overhangs while Taq polymerase adds A overhangs at 3’ end of PCR product. The plasmid DNA containing 1.1 kb insert was isolated by minipreps and confirmed by restriction digestion and PCR amplification.

**Sampling and extraction of HBV DNA**

Blood samples were collected from clinically diagnosed HBV positive patients from Services Hospital, Lahore, Pakistan. EDTA was added to a final concentration of 7.0 mM and plasma was extracted by centrifugation. Viral DNA was extracted from the blood plasma using the method of Persing et al. (1993). 10 µl of extracted DNA was used for PCR.

**PCR amplification**

Two sets of primers were used. The first pair of primers NHBF1 (GTC CCC TAG AAG AAG AAC TCC CTC) and NHBR1 (CCC TAG AAA ATT GAG AGA AGT CCA) was selected from the conserved region of HBV genome near S gene. A region of about 1.12 kb from nucleotide No. 2362 to 283 was amplified by using this set of external primer (Abbas, 2001). Second set of primers B2833S (GGG TCA CCA TAT TCT TGG) and B170AS (GTC CTA GGA ATC CTG ATG) was selected from a published source (Thiers et al., 1988) and used for sequencing of Pre S region of HBV by dideoxy chain termination method (Sanger et al., 1977).

The denaturation was done at 94°C for 1 minute, annealing was done at 55 or 50°C for 1 minute, and extension at 72°C for 2 minutes for 20-25 cycles.

Total volume of PCR reaction was 25 µl and consisted of 20 µl of master mix (9 mM Tris-Cl pH8.3, 45mM KCl, 2.25mM MgCl2, 180mM of each dNTP and 0.9µM of each primer), 3 µl of HBV DNA and 2 µl (1 unit) of Taq DNA polymerase.

**Sub cloning**

The PCR product of external primers was run on 0.8% low melting agarose gel and the bands of interest was purified from the gel slice through Wizard DNA purification system (Promega/Madison, WI, USA) or Sigma’s Nucleiclean Kit. This gel purified DNA was cloned in pGEM-T vector, especially designed for cloning PCR products. The following ligation reaction was set up: Ligase 10X buffer, 1 µl; PGEM-T vector, 1 µl (50 ng); PCR product, 3 µl (30 ng of 1.1 kb), and T4 DNA ligase, 1 µl (3 units). Later water was added to make the volume 10 ml. The contents were mixed by finger flicking, centrifuged briefly and then incubated at 4°C overnight. 2 µl of ligation reaction was added to an eppendorf tube containing 50 µl of competent cells, incubated on ice for 20 minutes and then heat shocked at 42°C for 2 minutes. Later 1 ml of LB medium was added and incubated at 37°C for 1 hour in incubator shaker. Later serial dilutions were prepared and plated onto LB-Amp./IPTG/X-Gal plates.

**DNA sequencing**

Non-radioactive dideoxy sequencing was carried out according to the manufacturer’s instructions (Silver sequence kit; Promega/Madison, WI, USA). The sequencing was performed to characterize the pre-S1 region of two Pakistanian HBV isolates. The sequencing reactions were resolved on a 0.2 mm thin 6% polyacrylamide gel (PAGE) (Sambrook et al., 1989). After resolving the gel was silver stained according to the manufacturer’s instructions (Silver sequence kit; Promega/Madison, WI, USA). The silver stained image of the gel was stored as computer file by a video imaging system and the DNA sequence was read manually and stored as a computer file.

**RESULTS**

Figure 1 shows PCR product of ten clones
NUCLEOTIDE SEQUENCE ANALYSIS OF PRE-S1 SURFACE ANTIGEN OF HBV

from HBV isolates # NA007 and NA013 each, having 1.1Kb inserts.

Fig. 1. PCR confirmation of NA007 (Panel A) and NA013 clones (Panel B) having 1.1Kb (upper half) and 0.6 kb inserts (lower half), using specific primers in each case.
The agarose gel in panel A and B consists of two halves. The upper half of the gel shows PCR products of NHBF1 & NHBR1 primers. Lane 1 shows the λ HindIII size marker, lane2 is the negative PCR control, lane3 has the positive PCR control and lanes 4 to 13 contain PCR products of ten clones derived from Pakistani HBV isolate # NA007. The panel B has λ HindIII size marker in lane 1 and HBV positive PCR control in lane 2. Lanes 3 to 13 contain PCR products of ten clones derived from Pakistani HBV isolate # NA013. The lower half of the gel is loaded with PCR products using nested primers B170AS and B2833S. Negative PCR control next to the λ HindIII size marker is clear and all ten clones from Pakistani HBV isolates # NA007 and NA013 show a 0.6 Kb DNA fragment which confirms the identity of the cloned HBV DNA.

Amplified product containing PreS region was sequenced in both sense and antisense directions using one of the internal set of primer (B170AS and B2833S). Nucleotide sequence (357bp) of the PreS1 region of NA007 and NA013 was read manually from the silver sequencing gel (sequence gels not shown). HBV PreS1 sequence was fed into PC/GENE program to get its amino acid sequence (Fig. 2).

Sequence variation analysis

PreS1 sequence from samples NA007, and NA013 were found similar to adr4 after homology/ variation analysis through PCGENE. Sequence alignment through the NCBI Blast search facility revealed that Pakistani HBV PreS1 sequence showed 96.6% with ayr, 90.2% with adw2, 86.8Adw, 82.4% with ayw, 79.4% with ayw4. Both of these sequences varies with other subtypes follows the order adw4 > ayw4 > ayw > adw > adw2 (Table I). The variations due to vaccine were excluded in these cases, as the patients were not vaccinated. There may be other reasons such as
persistent of HBV infection in chronic state. These mutations may be correlated to cirrhotic state, as these are common in two patients with liver cirrhosis.

```
10     20     30     40
|      |      |      |
ATG   GGA   GGT   TGG   TCT
Gly   Gly   Trp   Ser   Ser

50     60     70     80     90
|      |      |      |
MET   Gly   Trp   Ser   Ser   Lys
Pro   Arg   Gly   MET   Gly   Thr   Asn

100    110    120    130
|      |      |      |
CTT   TCT   GTT   CCC   AAT   CCT   CTG   GGA   TTA   TTC
Leu   Ser   Val   Pro   Asn   Pro   Leu   Gly   Phe   Phe

140    150    160    170    180
|      |      |      |      |
GAC   CCT   GGC   TGC   GGA   GCC   AAC   TCA   CAC   AAT   CCC   GAT   CAC   GAG   AAT   CAG   GTC   TGC   GAC   TTC   TTC   CCC   GAT   CAC   GAG   AAT   CAG   GTC
Asp   Pro   Ala   Phe   Gly   Ala   Asn   Ser   His   Asn   Pro   Asp   Trp   Asp   Phe

190    200    210    220
|      |      |      |
AAG   Pro   Asn   Lys   Asp   His   Trp   Pro   Glu   Ala   Asn   Gin   Val   Gly   Ala

230    240    250    260    270
|      |      |      |      |
GGG   TGG   AGG   TGC   CAG   CCA   GAG   CTA   GCC   CCA   CAC   GCC   GGT   CTT   TTG
Gly   Ala   Phe   Gly   Phe   Thr   Pro   Pro   His   Gly   Gly   Leu   Leu

280    290    300    310
|      |      |      |
GGG   TGG   AGG   TGC   CAG   CCA   GAG   CTA   GCC   CCA   CAC   GCC   GGT   CTT   TTG
Gly   Ala   Phe   Gly   Phe   Thr   Pro   Pro   His   Gly   Gly   Leu   Leu

320    330    340    350
|      |      |      |
GGG   TGG   AGG   TGC   CAG   CCA   GAG   CTA   GCC   CCA   CAC   GCC   GGT   CTT   TTG
Gly   Ala   Phe   Gly   Phe   Thr   Pro   Pro   His   Gly   Gly   Leu   Leu

360    370    380    390    400
|      |      |      |      |
ACT   CCC   ATC   TCT   CCA   CCT   CTA   AGA   GAC   AGT   CAT   CCT   CAG   GCC
Thr   Pro   Ile   Ser   Pro   Ser   Leu   Arg   Ser   His   Pro   Leu   Gin   Ala

Fig. 2. DNA and amino acid sequences of PreS1 region of NA007 and NA013.

Table I.- Comparison of sequence with other subtypes/genotypes

<table>
<thead>
<tr>
<th>Subtypes (geno)</th>
<th>References</th>
<th>Homology</th>
<th>Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADR4(C)</td>
<td>Fujiyama et al. (1983)</td>
<td>100%</td>
<td>0.0%</td>
</tr>
<tr>
<td>AYR(C)</td>
<td>Okamoto et al. (1986)</td>
<td>96.6%</td>
<td>3.4%</td>
</tr>
<tr>
<td>ADRQ(C)</td>
<td>Nordet al. (1994)</td>
<td>95.5%</td>
<td>4.5%</td>
</tr>
<tr>
<td>Adv2(A)</td>
<td>Valenzuela et al. (1979)</td>
<td>90.2%</td>
<td>9.8%</td>
</tr>
<tr>
<td>Adv(B)</td>
<td>Okamoto et al. (1988)</td>
<td>86.8%</td>
<td>13.2%</td>
</tr>
<tr>
<td>AYW(D)</td>
<td>Tong et al. (1990)</td>
<td>82.4%</td>
<td>17.6%</td>
</tr>
<tr>
<td>AYW(D)</td>
<td>Lai et al. (1991)</td>
<td>82.7%</td>
<td>17.3%</td>
</tr>
<tr>
<td>AYW4(E)</td>
<td>Nordet al. (1994)</td>
<td>79.4%</td>
<td>20.6%</td>
</tr>
<tr>
<td>AYW4(F)</td>
<td>Nordet al. (1994)</td>
<td>76.5%</td>
<td>23.5%</td>
</tr>
</tbody>
</table>

Interest in the PreS1 region was due to its hypervariability, having point mutations (Lauder et al., 1993). Amino acid sequence encoded by the PreS1 region is found highly immunogenic (Okamoto et al., 1985). Of the total mutations occurring in the HBV genome, 50% occurs in the hypervariable regions, including PreS1. It is reported that some of the mutations affect the clinical course of HBV infection (Tong et al., 2005).

REFERENCES


Principles and applications. American Society for Microbiology, Washington D.C., USA.


(Received 15 December 2007, revised 4 April 2008)