Molecular basis of hepatitis B virus serotype variations within the four major subtypes

Heléne Norder, Anne-Marie Couroucé and Lars O. Magnius

Department of Virology, The National Bacteriological Laboratory, S-105 21 Stockholm, Sweden and Institut National de Transfusion Sanguine, F-75739 Paris, France

Amino acid residues 101 to 180 of hepatitis B surface antigen (HBsAg) were predicted by sequencing the corresponding part of the S gene of hepatitis B virus (HBV) DNA in 46 HBsAg-positive sera, which had been subtyped by immunodiffusion with respect to d/y, w/r, w1 to w4 and q. The sequences of the nine different HBV serotypes defined by these specificities were found to be homogeneous proving that they represent consistent variations of HBV at the genomic level. Residue 127 was found to be important as were Pro, Thr and Leu for w/r, w1 to w2, w3 and w4, respectively. Five residues were found to differ between ayw1 and ayw2. These were at positions 134 (Phe instead of Tyr), 143 (Thr instead of Ser), 159 (Ala instead of Gly), 161 (Tyr instead of Phe) and 168 (Val instead of Ala). However, all these residues were shared by ayw1 and adw2, implying that Arg122 was also important for w1 expression. All genomes expressing r, apart from one ayr strain, had an Ile126, which might explain the pseudo-allelism of w1 to w4 in relation to r, since this substitution might influence the w epitope. There were two regions where adw4q- and adrq differed from all the q+ subtypes. These were located at residues 158 and 159, and at residues 177 and 178, where both the q- subtypes had amino acid substitutions in adjacent positions. The mapping of the epitopes defining these antigenic specificities will help to link information on the world-wide distribution of HBsAg subtypes to future molecular epidemiology with regard to HBV.

Introduction

Hepatitis B surface antigen (HBsAg) is the small surface protein of hepatitis B virus (HBV) and is the predominant protein of the 20 nm small spherical particles representing circulating excess surface protein. The 226 amino acids of HBsAg form a hydrophobic protein with two moderately hydrophilic regions, one at residues 30 to 79, which is internal, and another at residues 99 to 168, which is exposed on the surface of the HBsAg particle (Lerner et al., 1981; Howard et al., 1988; Gerlich & Heermann, 1991). The latter region, rich in cysteines and prolines, is considered to encompass the a determinant common to all HBV strains, and is presumed to be important for the protective immune response to HBV (Bhatnagar et al., 1982; Dreesman et al., 1982; Gerin et al., 1983; Brown et al., 1984; Waters et al., 1992). The a determinant is partially sensitive to reduction and alkylation (Imai et al., 1974), which indicates that disulphide bonds are important in maintaining this determinant.

The occurrence of different serotypes of HBsAg reflecting the genetic variability of HBV is well documented. The common determinant a and two pairs of mutually exclusive determinants, d/y (Le Bouvier, 1971) and w/r (Bancroft et al., 1972), enables the distinction of four major subtypes of HBsAg, adw, adr, ayw and ayr. Additional determinants designated subdeterminants of w (w1 to w4) have allowed the definition of four serotypes of ayw (ayw1, ayw2, ayw3 and ayw4) and two serotypes of adw, i.e. adw2 and adw4 (Couroucé et al., 1976). The determinants w1 to w4 were once designated subdeterminants of a, since they were considered to be variations of the major a determinant (Soulier & Couroucé-Pauty, 1973; Couroucé-Pauty & Soulier, 1974). The q determinant, originally described as present on all subtypes apart from adw4 (Magnius et al., 1975), was later also found to be absent from adr strains in the Pacific region, thus defining adr as either q+ or q- (Couroucé-Pauty et al., 1978).

A classification of HBV genomes into six groups, A to F, based on the degree of similarity in the nucleotide sequence has been suggested (Okamoto et al., 1988; Norder et al., 1992). Strains specifying adw are found in groups A, B, C and F, and those specifying ayw in groups A, B, D and E (Sastrosawignjo et al., 1991; Norder et al., 1992). Strains specifying r have so far only been found in group C (Okamoto et al., 1988).
The part of the S gene that corresponds to the hydrophilic domains of HBsAg shows point mutations associated with subtype variations (Gerin et al., 1983; Okamoto et al., 1987; Ashton-Rickardt & Murray, 1989). The important allelic substitutions associated with d/y and w/r expression are found at residues 122 and 160, respectively, and at both these sites the subtypic changes are mediated by a shift from Lys to Arg (Okamoto et al., 1987). The molecular basis has not been allocated for w1 to w4 (Couroucé-Pauty & Soulier, 1976), x (Le Bouvier, 1971), q (Magnius et al., 1975) or g (Shorey, 1976), nor has it been clarified whether the observed serological variations within the four major subtypes might be linked to the genomic groups of HBV.

In a study on 12 HBV genomes of known subtypes, amino acid substitutions were found which might explain some of the serological variations of HBsAg (Norder et al., 1992). So far, however, no systematic studies have been conducted to correlate these to multiple amino acid sequences of defined subtypes. Therefore, another 34 S genes were sequenced from HBV DNA-positive sera that had been subtyped with respect to d/y, w/r, w1 to w4 and q.

**Methods**

**Sera.** The 46 sera used as the source of HBV DNA were all positive for HBsAg of known subtype, and are described in Table 1. The sera had been subtyped by immunodiffusion at the Institut National de Transfusion Sanguine, Paris, France, as described previously (Couroucé-Pauty & Soulier, 1974). Two HBV genomes from sera Tar and Man, which had given aberrant results on subtyping, were also sequenced. Tar had been found to give the same reaction in immunodiffusion with adw4 and adr, but had given a reaction of partial identity with adw2 when tested against the rabbit antiserum R182, workshop no. 14, which exhibits anti-d and another reactivity shared by ayw1 and adw2 (Couroucé-Pauty & Soulier, 1974). Man is derived from a child infected by HBV in 1978 at a haemodialysis unit with several carriers of HBV adw4 and ayw3 subtypes. This sample was found to be negative for d, y, q, w1 to w4 and x at subtyping.

**PCR amplification and sequencing.** Two oligonucleotide primers were used to amplify a fragment corresponding to amino acids -8 to 198 of the S gene: hep3 at positions 738 to 750 is identical to MD03 of Larzul et al. (1988) and hep33 at positions 131 to 146 is 5' AGGACTGGGGAC-CCTG 3'. PCR was performed as described previously (Norder et al., 1990). The amplification products from three tubes in each amplification run were pooled and extracted once with chloroform. The oligonucleotides and excess dNTPs were removed from the amplified DNA product by using Magic PCR Preps (Promega). The sequencing reaction was performed as described previously (Norder et al., 1992) with hep3 as the sequencing primer.

**Table 1. Identification and geographic origin of 46 sera containing HBsAg of specified subtype that were used as source for HBV DNA for S gene sequencing**

<table>
<thead>
<tr>
<th>Subtype</th>
<th>Donor code</th>
<th>Workshop no.</th>
<th>Geographic origin</th>
<th>Subtype</th>
<th>Donor code</th>
<th>Workshop no.</th>
<th>Geographic origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>ayw1 PHD</td>
<td>23</td>
<td>Vietnam</td>
<td>ayw1 VoN</td>
<td>82</td>
<td>Vietnam</td>
<td>ayw1 Pon</td>
<td>Vietnam</td>
</tr>
<tr>
<td>ayw1 Bel</td>
<td>Vietnam</td>
<td>ayw1 Ngu</td>
<td>Vietnam</td>
<td>adw2 Rio</td>
<td>France</td>
<td>ayw1 Bat</td>
<td>Sub-Saharan Africa</td>
</tr>
<tr>
<td>ayw2 Ren</td>
<td>24</td>
<td>France</td>
<td>ayw2 Wai</td>
<td>France</td>
<td>adw4q- Gal</td>
<td>France</td>
<td>ayw2 Hac</td>
</tr>
<tr>
<td>ayw3 Mau</td>
<td>85</td>
<td>France</td>
<td>ayw3 Gay</td>
<td>86</td>
<td>France</td>
<td>ayw3 Gag</td>
<td>France</td>
</tr>
<tr>
<td>ayw3 Kli</td>
<td>France</td>
<td>ayw3 Flo</td>
<td>France</td>
<td>adrq Waf</td>
<td>Polynesia</td>
<td>ayw3 Nar</td>
<td>France</td>
</tr>
<tr>
<td>ayw4 Tah</td>
<td>Nigeria</td>
<td>ayw4 Tah</td>
<td>Nigeria</td>
<td>indet Man</td>
<td>France</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* West African contact.
† Indeterminate.
Results and Discussion

The amino acid sequences covering residues 101 to 180 of HBsAg of 46 different HBV strains are shown in Fig. 1. The different serotypes were found to be homogeneous, showing that these serotypes represent consistent variations of HBV at the genomic level.

The only variation observed between African ayw1 and adw2 in genomic group A, as well as between Vietnamese ayw1 and adw2 in genomic group B, resided at position 122, confirming that substitution at this position alone explains the d/y variation.

Five residues were found to differ between ayw1 and ayw2. These were residues 134 (Phe instead of Tyr), 143 (Ser instead of Thr), 160 (Ser instead of Tyr), 161 (Ser instead of Tyr), and 180 (Ser instead of Thr).
(Thr instead of Ser), 159 (Ala instead of Gly), 161 (Tyr instead of Phe) and 168 (Val instead of Ala). However, aywl and adw2 had all these substitutions in common, and apart from the d/y site at residue 122, an identical amino acid sequence was found for African aywl and common adw2, as well as for Vietnamese aywl and Sru, serotyped as normal adw2. Positive reactions for w1 are, however, shown by some ayr samples such as PVC (Couroucé et al., 1976). Therefore, w1 recognition is apparently dependent on the Arg122 that renders the region, compatible with the loss of a common reactivity have a pronounced influence on the antigenicity in this position 127 might, by removing a beta-turn, most likely critical site for the serological variations between serotype. Our finding that amino acid residue 127 is the also obvious, since these are the only residues in common for ayr4 and adw4 that are not shared by any other serotype. Our finding that amino acid residue 127 is the critical site for the serological variations between w1/w2 in relation to w3 and w4 also has implications with regard to the nature of the x determinant (Le Bouvier, 1971) and the g determinant (Shorey, 1976). The Pro to Thr shift in ayr3, and the Pro to Leu shift in ayr4 and adw4 at position 127 might, by removing a beta-turn, most likely have a pronounced influence on the antigenicity in this region, compatible with the loss of a common reactivity for aywl, ayw2 and adw2. Interestingly, Pro127 is shared by all HBsAg categories exhibiting x as detected by reagent CH. 12-40448 (Couroucé et al., 1976) and g. The conclusion that x/g, w3 and w4 are mutually exclusive (Couroucé et al., 1976) and thus behave in an allelic manner corroborates our results.

The allocation described herein of the w subdeterminants to positions 127 and 134 within the major a determinant is supported by the finding that monoclonal antibodies with the ability to bind to a cyclized peptide covering residues 124 to 147 have also been found to possess relative discriminatory properties with regard to the w subdeterminants (Waters et al., 1992). This indicates that the reactivities behind the 'w' specificities are rather variations within a as originally suggested (Soulier & Couroucé-Pauty, 1973) and since all w strains had a Lys160, the w1 to w4 determinants are apparently not alleles of r. The pseudoallelic relation between w1 to w4 and r was one reason for changing the designation from subdeterminants of a to subdeterminants of w. The non-reactivity of adr and most ayr strains with antisera against subdeterminants of w might be explained by the fact that Ile126, present in most of the group C strains (Onuhma et al., 1990), is likely to abolish the w subdeterminants. However, the PVC ayr strain, which we have sequenced, did not express an Ile126 and did, as mentioned, exhibit w1 reactivity also.

Regarding q, the amino acid substitutions abolishing q in adw4q- and adrq- were found to differ, since the two q- subtypes had no amino acid substitutions in common, when compared to the q+ ones. However, there were two regions, residues 158 to 159, and residues 177 to 178, where the q- subtypes shared unique amino acid substitutions at adjacent positions. If the substitutions at the former region would abolish q, it is necessary that Gly159 in ayw2, ayw3 and ayw4, instead of Ala159 in the other q+ subtypes, would have no influence on the expression of q.

The two sera with indeterminate subtypes both had genomes with unique changes in the region of the S gene sequenced in this study. The HBsAg sequence of Tar was found to differ from the normal adw2, genomic group A, by having Ser instead of Thr at residue 143. Therefore, this residue seems important for the recognition of a common specificity for adw2 and aywl by the antisera R182. In the HBsAg sequence of Man, the Pro120 to Gln120 shift not only rendered this sample not typable for the d/y polymorphism at residue 122, but also for the expression of w4 of this apparently mutant adw4q- strain.

The allocation of determinants described herein to specific amino acid residues does not imply that their epitopes depend on these residues alone, but that the indicated residues are essential parts of the epitopes, whether linear or conformation-dependent, together with residues shared with the heterologous subtypes, which at present cannot be identified.

The concept that there are at least nine different HBV serotypes is now widely accepted (Couroucé et al., 1976; Wands et al., 1984; Ben-Porath et al., 1985; Swenson et al., 1991). The molecular mapping of the subdeterminants defining these serotypes is important for several reasons. Apart from its apparent value in elucidating the specificity of diagnostic procedures in HBV serodiagnosis, it also provides an opportunity to use previous information on the world-wide distribution on HBsAg subtypes (Couroucé-Pauty et al., 1983) as background information for directing future research with regard to HBV molecular epidemiology based on S gene sequencing.

Berit Hammas is acknowledged for skilful technical assistance. This work was supported by the Swedish Medical Research Council, grant no. B93-16X-10385-01.

References

predicted from the nucleotide sequence of the hepatitis B virus genome elicit antibodies reactive with the native envelope protein of Dane particles. *Proceedings of the National Academy of Sciences, U.S.A.* 78, 3403-3407.


(Received 1 June 1992; Accepted 24 August 1992)