## Discontinuous epitopes of hepatitis B surface antigen derived from a filamentous phage peptide library

(biopanning/envelope protein/monoclonal antibody/protein structure/mimotope/surface antigen)

Y.-C. JACK CHEN\*, KATHY DELBROOK, CHRIS DEALWIS, LARRY MIMMS<sup>†</sup>, ISA K. MUSHAHWAR, AND WLODEK MANDECKI

Molecular Biology Laboratory, Viral Discovery Group, Diagnostics Division, Abbott Laboratories, 1401 Sheridan Road, North Chicago, IL 60064-4000

Communicated by Emmanuel Margoliash, University of Illinois, Chicago, IL, October 2, 1995

ABSTRACT The structure of the small hepatitis B virus surface antigen (HBsAg) was investigated by epitope mapping of four anti-HBsAg monoclonal antibodies (mAbs). Amino acid sequences of epitopes were derived from affinity-enrichment experiments (biopanning) using a filamentous phage peptide library. The library consists of 10<sup>9</sup> different clones bearing a 30-residue peptide fused to gene III. Sequence homologies between peptides obtained from panning the library against the antibodies and the native HBsAg sequence allowed for precise description of the binding regions. Three of four mAbs were found to bind to distinct discontinuous epitopes between amino acid residues 101 and 207 of HBsAg. The fourth mAb was demonstrated to bind to residues 121-124. The sequence data are supported by ELISA assays demonstrating the binding of the HBsAg-specific peptides on filamentous phage to mAbs. The sequence data were used to map the surface of HBsAg and to derive a topological model for the  $\alpha$ -carbon trace of the 101–207 region of HBsAg. The approach should be useful for other proteins for which the crystal structure is not available but a representative set of mAbs can be obtained.

Elucidation of epitope structure is an integral part of studying antigen–antibody interactions. Peptide libraries on filamentous phage have recently been used for rapid epitope mapping of antibodies. The procedure entails affinity screening the library against the antibody, isolating the phage that bind, and determining the peptide sequence of the binding phage through DNA sequencing of the relevant portion of the phage genome. By using this procedure, a large number of peptides can be screened for binding to an antibody in a relatively short time (1, 2).

A filamentous phage in a typical library carries on its surface 3 to 5 random peptides, 6–38 residues long, fused to the gene *III* protein. The peptide library used in this paper contains  $10^9$  different phages, each displaying a 30-residue peptide fused to the gene *III* protein (3). The library was originally applied to select peptides that bind to monoclonal antibodies (mAbs) raised against human immunodeficiency virus gp120 or hepatitis C virus core proteins. In this study, we applied the phage library approach to mapping the surface antigen of the hepatitis B virus (HBV). The antigen is the diagnostic marker for the HBV. The virus causes major endemic illness throughout the world and is associated with a greatly increased frequency of primary hepatoma, a major cancer in the world (4).

The envelope of the hepatitis B virion consists of three proteins, termed small, middle, and large HBV surface proteins, and their glycosylated derivatives. The small HBV surface antigen (HBsAg) is the major component of the envelope of the virion and is mainly present in the plasma of HBV-infected individuals as 22-nm spherical particles composed of about 100 HBsAg monomers each (5). The small HBsAg is a 226-residue protein and bears determinants important for the induction of a protective humoral immune response in man. Although the three-dimensional structure of HBsAg is not known, structure–function studies indicate that its central core (residues 99–169) is involved in binding to antibodies against small HBsAg. This region contains eight cysteine residues thought to form disulfide bonds since most epitopes of the small HBsAg are lost if the 22-nm particles are treated with a reducing agent (6). The HBsAg can be glyco-sylated at Asp-146 (7). Experiments utilizing peptide libraries on filamentous phage showed that many residues within the 115–135 region interact with antibodies from human serum (8) or mouse mAb against HBsAg (9).

The objective of this study was to characterize HBV epitopes by investigating the binding properties of four mAbs raised against the HBsAg. The four mAbs were chosen because they represent four distinct, nonoverlapping epitopes within the "a" group-specific determinant located on the S gene product, and all four react with high affinity to all known HBV subtypes. Three of four mAbs, namely H5, H35, and H53 (10, 11), were thought to bind to unknown discontinuous epitopes since they did not cross-react with denatured HBsAg samples in Western blotting experiments. The recognition sequences of these antibodies were unknown since they could not be mapped by using synthetic peptides. The fourth antibody, H166, was known to interact with a continuous epitope (11). We report here the sequences of epitopes recognized by these antibodies and propose a partial map of the surface of HBsAg and a topological model for its structure.

## **MATERIALS AND METHODS**

**General.** mAbs H5, H35, H53, and H166 (10, 11) were obtained from Abbott Monoclonal Antibody Development. The immunogen was human plasma-derived HBsAg particles, subtype adw2, ayw3, or ayw2 (11). Rabbit anti-alkaline phosphatase and sheep anti-M13 were purchased from 5 Prime–3 Prime, Inc. Donkey anti-sheep horseradish peroxidase and goat anti-rabbit horseradish peroxidase conjugates were obtained from Jackson ImmunoResearch Labs, Inc. Recombinant HBsAg (rHBsAg) was prepared as described in ref. 11. The fNG1 phage library and the panning protocol have been described (3).

**Competition Assays.** Inhibition constants were determined by competitive ELISAs. Assays were performed by mixing the desired concentration of rHBsAg with a fixed concentration of phage (H5,  $2.75 \times 10^9$ ; H35,  $2.3 \times 10^{11}$ ; H53.3,  $4.76 \times 10^9$ ;

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HBsAg, hepatitis B virus surface antigen; rHBsAg, recombinant HBsAg; HBV, hepatitis B virus; mAb, monoclonal antibody.

<sup>\*</sup>To whom reprint requests should be addressed.

<sup>&</sup>lt;sup>†</sup>Present address: GenProbe, Inc., 9880 Campus Point Drive, San Diego, CA 92121-1598.

H53.6,  $6.88 \times 10^{9}$ ; H53.7,  $9.04 \times 10^{9}$ ; H53.16,  $9.3 \times 10^{9}$ ; H35.7,  $1.62 \times 10^{10}$ ) and incubating the mixture at 25°C for 1 h. The equilibrated phage/antigen solution was added to wells of a microtiter plate coated with the desired antibody and allowed to incubate at 37°C for 1 h. Residual phage binding was detected colorimetrically by incubating with a sheep anti-M13 antibody and a goat anti-sheep antibody conjugated to horse-radish peroxidase, followed by colorimetric detection by using *o*-phenylenediamine.

## **RESULTS AND DISCUSSION**

The fNG1 phage library was panned against each of four anti-HBsAg mAbs, H166, H35, H53, and H5 (10). Several clones from the second round of each biopanning were sequenced, and their binding properties were analyzed by competitive ELISA.

H166. A total of 16 different sequences of 22 clones isolated from panning against H166 are shown in Fig. 1A. A -CXTCpeptide motif spanning residues 121-124 is apparent (X = K or R, single-letter code). Thr at position 123 is always conserved. Clone H166.15 contains two potential binding motifs, -CKTC- and -CQTC-. The diversity of amino acids surrounding the tetramer in the selected sequences suggests that residues adjacent to the determinant do not contribute to binding. Trypsin digestion of HBsAg destroys the ability of antibodies to bind the epitope, suggesting that the binding site straddles the -CXTC- sequence. Studies with a synthetic peptide corresponding to region 115-129 that binds an antibody with similar reactivity to that of H166 showed that the epitope was preserved upon alkylation of the peptide. Binding, however, was completely destroyed upon reductive alkylation of the peptide (12). We conclude that the epitope is a loop formed by residues 121-124 held together by a disulfide bond, consistent with previous epitope mapping studies (7, 12, 13).

H5. Three different clones were isolated from panning against antibody H5, and their sequences are listed in Fig. 1*B*. Clone H5.2 contains two cysteines and bears similarity to residues 158–167 of HBsAg. Clone H5.4 shares sequence similarity with three regions of HBsAg. The sequence between Gln-6 and Leu-11 contains a match of 4 of 6 amino acids with residues 101–106 of HBsAg. Additional regions of similarity exist at 19-QGN-21 and at 22-WEW-24, which map to residues 129–131 and 163–165 of HBsAg, respectively. All the H5.4 phage bound to immobilized antibody, and specific binding could be competed with rHBsAg, as shown in Fig. 2*A* (IC<sub>50</sub> of 20  $\mu$ g of rHBsAg per ml, i.e., 800 nM as monomer rHBsAg). No sequence similarities could be detected for clone H5.3.

H35. Sequencing of the 24 clones binding to H35 revealed only two different clones, one of them being represented 23 times (Fig. 1*C*). Both clones contain two cysteines 14 or 15 residues apart preceded by an -RAR- motif. Clone 35.1 was isolated only once and bears sequence similarity to the 166– 175 region of HBsAg. Whole 35.1 phage particles competed weakly against rHBsAg for binding to immobilized H35 (data not shown). Clone 35.2 was the predominant clone selected. It bears a measure of sequence similarity to the 122–130 region of HBsAg. Unlike 35.1, however, 35.2 binds much more strongly to the H35 antibody (Fig. 2*B*) and competes directly with free rHBsAg for binding to the antibody (IC<sub>50</sub> of 33  $\mu$ g of rHBsAg per ml, i.e., 1.3  $\mu$ M).

**H53.** Four different peptides isolated from panning against mAb H53 are listed in Fig. 1*D*. The clone H53.16 bears similarity (5 of 12 residues) to the 187–198 HBsAg region. In clone H53.6, a similarity of three of five consecutive residues with residues 196–200 of HBsAg is observed. The sequence is bound by two cysteines in the selected peptide but not in the native sequence. The two cysteines may form a disulfide bond resulting in presentation of the residues between them in a rigid conformation similar to the one they might adopt in

HBsAg. Clone H53.7 exhibits almost an exact match with residues 112–117 of HBsAg. Interestingly, this sequence is 70–80 residues in the linear sequence away from the epitope suggested by the similarities shown by the other two clones. Finally, clone H53.3 bears similarity (match of 5 of 11 residues) to the 197–207 region of HBsAg. Clone H53.3 also contains a stretch of residues (Gly-7 to Thr-12) that bear a match of 4 of 5 residues to the HBsAg sequence from the region 112–116. These data suggest that H53 recognizes a discontinuous epitope composed of residues from two different regions of HBsAg, namely 112–117 and 187–207.

HBsAg was observed to compete against each of the different phage clones for binding to H53 (Fig. 2 C and D). All the graphs are characterized by sigmoidal curves with a narrow dynamic range. Clone 53.16 gives the highest signal in the ELISAs. This contrasts with clone 53.7 which, despite having a higher degree of sequence similarity, appears to bind more weakly than 53.16. Clone 53.3 possesses attributes of both sequences, and it gives a lower signal than 53.16. This suggests that the 187–207 region is more important for binding than the 112–117 region.

Biopanning against the four mAbs gave a set of short sequences similar to various regions of HBsAg (Fig. 1). Many residues from these sequences, and especially the residues indicated in boldface in Fig. 1, are expected to interact with the mAbs employed here. The number of residues identified as forming the epitope and the number of discontinuous regions within the epitope (H5, 14 residues and 3 regions; H35, 13 residues and 2 regions; H53, 18 residues and 2 regions; H166, 4 residues and 1 region) is consistent with the results of the crystallographic analysis of five antigen–antibody complexes (three with lysozyme and two with neuraminidase). These studies show that the antibody makes direct contacts with 14-22 amino acid residues of the antigen and that 2–5 short linear polypeptide segments constitute the binding region on the antigen (14).

A Model for the Surface of HBsAg. The sequence information from biopanning was used to construct a partial map of the surface of the antigen. The analysis was based on the following premises: (i) the majority of residues identified from biopanning to interact with mAbs should be displayed on the surface of the protein, (ii) residues identified by biopanning to form a single discontinuous epitope should be in spatial proximity to each other, and (iii) different epitopes that share the same residue(s) must also be in propinquity.

The schematic representation of the surface is given in Fig. 3, where data from panning experiments in this paper, other papers [regions I and II (9)], and additional biochemical evidence on the surface exposure of the 140–146 region [region III (16)] are compiled. Circles represent areas of HBsAg recognized by different mAbs. The discontinuity of three epitopes is indicated by splitting the circles with a broken line. The circles overlap one another if the corresponding mAbs recognize the same residue(s), as evidenced by the sequences from phage display.

The central region of the surface is formed by residues 115–129 since this region is recognized by five of seven antibodies analyzed here, namely I, II, H35, H53, and H166. mAb H166 recognizes the 121–124 loop, while the epitope for mAb H35 shares the 122–129 region with antibody I. Antibodies H53 and II share regions 112–115 and 127–129, respectively, with antibody I. The discontinuous epitope of the H5 antibody (residues 101–106 and 158–167) overlaps the H35 epitope. Six of seven mAb-recognition sites are linked; that is, they partially or fully overlap one another. Region III is the only unlinked area in Fig. 3. It is shown in the proximity of area II because only a short spacer (residues 136–139) separates the two areas. The proximity of several regions does not require that epitopes outlined in Fig. 3 be formed within a single monomer of HBsAg; different segments in Fig. 3 can originate

H166       MCRESNPLIC CVTC GFRVRQTVGESDGGS       H5.2       2x       EQARTCPMQAVGCMASTLCRSVVSGNVR         .1       MCRESNPLICTNSTGRQAVA CRTC F       4x       4x </th <th>A</th> <th></th> <th></th> <th>В</th> <th></th> <th></th> <th></th> <th></th>	A			В				
.1       MCRRESNPLIG       CUPUE       P=0.008       FAXYLMEWASV       adw2         .2       VUNNERSPATLECTINSTORGROUNA CERPC F       4x       130       101       100       100       adw2         .3       RYSGVVGNAVSEGERLINGLSSS       CUPUE       GEVGLMTRPOVEMNA       QGM       adw2         .4       CARC       GEVGLMTRPOVEMNA       QGM       adw2         .5       GAGQVERLERAND CERPC GEVGROGONGORSONGER       2x       RXSGVVGNAVSEGERLINGLSSS       CUPUE       WEWAS       all         .7       TITINASGLH, CERPC AFKEL       2x       2x       8       SLDRWPEHLATMGNELOMTRQ CERPC WASSERT       P=0.035       QGMLPV       WEWAS       all         .10       VORNNDTEMME CERPC WASSEGAUGEORS CERPC VIGSOLVAGEOPY CARC NERVAN       E       H5.3       2x       LTEVVISRSADCREDVIGECGOGHRGGCE         .11       RIPHKRYNQYDGMEWARAGODES CENC CINC V       LCRC VISHQERTVKGQOVTGTRICOCC       101       106       155         .13       NUKRPVVNLRESENG CARLESONG CRAC CINC V       LCRC VISHQERTVKGQOVTGTRICOCC       101       106       107         .14       WERPOINGEGALSONG CINC VISHQE MARCESONG       ENDITION       101       106       106       106         .15       VISHOVERAREVENDANCYCORELETY CARC VISHOVERALING CINC VISHON VISHON ALL PROLI	H166			<b>H5.2</b> 2x	EQ <b>FAK</b> T <u>C</u> PMQA	VKGG <b>WAS</b> TL <u>C</u> RS\	<i>J</i> YSGNVR	
2       VNWRGPSATLEGTNSNTGRRQQAVA CRTC F       4x       158       167         3       RYSGVVGNAVSEGERLNGLSS GYTC LOWR       CRTC GEVGLMTRPGVRMNA       QGN adw2         4       GAGQVERLEAKDP CRTC GGSWRGEPPWM       P.0.035       QGMLPY       WEWAS all         5       DERQIQROEFMVRISSERDAW CRTC GYTC LLOWR       2x       P.0.035       QGMLPY       WEWAS all         7       TITNSASGLHF CRTC MKSGCPAVAGKQDE 2x       2x       P.0.035       QGMLPY       WEWAS all         10       WORMWORKAGCOWS CRTC LAPGY       Y       WEWAS       all         11       RHPHRVPQTDGRGAGPGNS CRTC LAPGY       HS epitope:       101       106       165         13       L       CKTC VRSNQEPETYCGATO       ANTO VRSNQEPETYCGATO       Y       WEWAS       all         13       RHSRVPUNCPSCAGAPS CATC GYGGVE       D       101       106       130       158       167         14       WINNEPGIRQCPAASVGKWURLAGI CRTC V       VNNEPGIRQCPAASVGKWURLAGI CRTC TST       DYQOMLPYCPQGNWAFAKYLMEWASVC adw2       D         145       121       124       121       166       175       166       175         153       132       130       166       175       ayw3       156       207         1	.1	MQRKESNPNLG CVTC GFRVRQTVGESDGGS		P=0.008	FAKYLWE	<b>WAS</b> V		adw2
.3       RYSGVVGNAVSEGERLNGLSSS CVTC LGWR       130         .4       CRTC GEVGLMTRFQVRMNA       QGN adw2         .5       GAGQVERLREAKDP CRTC GESGRWRGEPFWM       2x         .6       DERQIQROEPMVENSERDAWR CRTC AFKEL       2x         .7       TITNSASCLH CKTC WINGGGAVAGKODE       2x         .8       SLDRWPENLATMCNRLIGMTRQ CKTC VOSTL       2x         .10       VORWNDTEWR CKTC MISSE       2x         .11       RHPHKUNQUTGENGRAGGDWS CKTC LRPGY       2x         .12       PKRQISMERNLQVTOGENT CKTC VISSTL       2x         .15       LCKTC VRSOGENTVKGDQVTGTRICQTCC       101       106       158         .15       LCKTC VRSOGENTVKOGDQVTGTRICQTCC       101       106       158       167         .19       DKNEPUVULREESCRLSR CATC GVGGVE       V       VORMANSTEWR BARK       2w       101       106       130       158       167         .19       DVINEPEIGCAPAASOKKVVRLAACI CKTC V       V       VORMANSTEWR BARK       2w       101       106       130       158       167         .24       SVLRQAAQFGRFELYVRREGN CRALTG KR       V       V       VMMWASOK       adw2       101       167       166       175       166       175       166       175       166	.2	VNWRGPSATLEGTNSNTGRRGQAVA CRTC F	4x		158	167		
.4       CRTC GEVGLMTRBGVENDAA       OGN       adw2         .5       GGGQUFELRERADP CRTC GGSNERGEPFWM       2x	.3	RYSGVVGNAVSEGERLNGLSSS CVTC LGWR			100	130		
.5       GAGQVERLERARD CHTC GGSRWRGEFYMM         .6       DERQIQRQEPMYRNSERDAMR CHTC AFKEL       2x         .7       TITNSASGLHF CKTC WKNSGOPAVAGKQDE       2x         .8       SLDRWFEHLATMGNRLGMTRQ CKTC VGSTL       2x         .10       VQWRNDTEMMR CKTC MLSE       101       106       165         .11       RHPHRKVRQYDGMEGAGGDWS CKTC LRPGY       101       106       165         .12       PKRQISMERWLQVTQGEEVTP CATC NPMVA       LCKTC VRHQERTVKGQQVTGTRICOTC       101       101       106       167         .15       LCKTC VRHQERTVKGQQVTGTRICOTC       0       VQWINDTEMMARKONKUVRLAGE       101       106       167         .16       WDKRPVVWLRFEESQRJSR CATC GVGGVE       V       VGMLVYFKRSVPDVTGSQGAPP CRTC TST       101       106       130       158       167         .10       I** is R or K)       121       124       124       126       127       198         H35.1       RASGARARCEHRSGLSLSWOPSECSDSRTT       F20.0       GSSTTS       63×       GRKTEKYSDGWTSMHSSLGDIQSYWT       116       166       196       200         H35.2       23×       SPVSYGEWRARYCTNGGQOTVQRADRSCW       ayw3       112       117       112       112       112       112       112       112       <	.4	CRTC GEVGLMTRPGVRMNA				OGN		adw2
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	.5	GAGQVERLREAKDP CRTC GGSRWRGEPFWM		H5.4 2x	RVAKG <b>OG</b> PLRV	YLTORRK <b>OGN</b> AS	NEWEEFI	uunb
.7       TITNSASGLHF CKTC WKNSGGPAVAGKQDE 2x       101 106 165         .8       SLDRWPEHLATMCNRLGMTRQ CKTC WGSTL 2x       2x         .10       WORRWNDTEMMR CKTC MLSE       101 106 165         .11       RHPIKKURQYDGMRGAGGBWS CKTC LRPGY       4         .12       PKRJISMERLUQTYQGEET P CATC NFWVA       H5.3 2x       LTEVVISRSADCRFRDVITGECCGMHRGCF         .15       L CKTC VRSHQETVKGDQVTGTRICOTC       101 106 130 158 167         .16       WDKRPVVLRFESQRLSR CATC GVGGVE       D         .19       NUNEPGIRQGPAASVGMKVVRLAGI CKTC V       DYQGMLPVCPQGNWAFAKYLWEWASVR adw2         .20       GMKIVVFPKRSVPDVTGSQGAPP CRTC TST      WAFAKYLWEWASVR adw2         .24       SVLRQAAQFGNFELYVRREON CRALTGC MR       H65WAFAKYLWEWAGEEESFGYWLVNWQEMM         H166 epitope:       C       H53.16       5x GSVPLSRKEVWAGEEESFGYWLVNWQEM         H35.1       RASGARARCEHRSGLSLSWQPSECSDSRTT       P=0.00017       ASARTSWLSL       ayw3         166       175       ayw3       122 130       166 175       gSSTT       ayw2 or ayw3         122       130       166 175       ayw3       122 117       187 207         H35 epitope:       122       130       166 175       197 207         H35       FQRTGTTPAQGTSM EWASARFSWLSLLV       ayw	.6	DERQIQRQEPMVRNSERDAMR CRTC AFKEL	2x	P=0.035	OGMLPV		NEWAS	all
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	.7	TITNSASGLHF CKTC WKNSGGPAVAGKQDE	2x	1-01000	101 10		165	
.10       VOWRNNDTEWMR CKTC MLSE         .11       RHPHKRVRQYDGMRGAGGDWS CKTC LRPGY         .12       PKRQISMERMLQVTQGEVTP CATC NPWVA         .15       L CKTC VRSHQENTVKGQVTGTRICQTC         .16       WDKRPVWLRFEESQRLSR CATC GVGGVE         .19       NVNEPGIRQEPAASVGMKVVLAGI CKTC V         .20       GMKIVVFPKRSVPDVTGSQGAPP CKTC TST         .24       SVLRQAAQFGNFELYVRREGN CRALTGC MR         H166 epitope:       C+TC all         (*** is R or K)       121 124         C       H53.16         H35.1       RASGARARCEHRSGLSLSWQPSEQSDSRTT         P=0.00017       ASARFSWLSL         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         123       146         124       150         GSSTT       ayw2 or ayw3         123       166         124       130         1	.8	SLDRWPEHLATMGNRLGMTRQ CKTC VGSTL	2x		101 10	0	105	
.11       RHPHRVPRQYDQMGRGGGGWS CKTC LRPGY         .12       PKRQISMERWLQVTQGEEVTP CATC NPWVA         .15       L CKTC VRSHQERTVKGDQVTGTRICQTC         .16       WDKRPVVWLRFEESQRLSR CATC GVGGVE         .19       NVNEPGIRQGPASVGWKVVRLAGI CKTC V         .20       GKTVVFFKRSVPDVTGSQGAP CRTC TST         .24       SVLRQAAQFGNFELYVRRGN CRALTGC MR         H166 epitope:       C+TC all         (*** is R or K)       121 124         C       MS1.6         H35.1       RASGARARCEHRSGLSLSWQPSECSDSRTT         P=0.00017       ASARFSWLSL         ajw3       166         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         123       121         124       130         125       197	.10	VQWRWNDTEWMR CKTC MLSE		H5.3 2x	LTEVVISESAD	CRERDVITGECCO	WHRGCE	
.12       PKRQISMERWLQVTQGEEVTP CATC NEWVA         .15       L CKTC VRSHQERTVKGQQVTGTRICQTC         .16       WDKRPVVWLRFEESQLSR CATC GVGGVE         .19       NVNEPGIRQCPAASVGKKVVRLAGI CKTC V         .20       GMKIVVFPKRSVPDVTGSQGAPP CRTC TST         .24       SVLRQAAQFORFELYVREON CRALTES MR         H166 epitope:       C+TC all         (*** is R or K)       121 124         C       H35.1         RASGARARCEHRSGLSLSWQPSECSDSRTT         P=0.00017       ASARFSWLSL         166       175         H35.2       23x         SPVSYGEWRARYCTNGGQOTVQRRADRSCW         P=0.0041       RTCTTPAQGT         122       130         122       130         122       130         135       PCRTCTTPAQGTSM         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         122       130         135       POLO045         GSSTT       ayw2 or ayw3	.11	RHPHKRVRQYDGMRGAGGDWS CKTC LRPGY		<b></b>		<u>o</u> ni no vi i od <u>oo</u>		
.15 L CKTC VRSHQERTVKGDQVTGTRICQTC .16 WDKRPVURIRFEESQRLSR CATC GVGGVE 19 NVNEPGIRQGPAASVGMKVVRLAGI CKTC V .20 GMKIVVFERSVPDVTGSQGAPP CRTC TST .24 SVLRQAAQFGNFELYVREGN CRALTGC MR H166 epitope: C+TC all ('+* is R or K) 121 124 C H35.1 RASGARARCEHRSGLSLSWQPSECSDSRTT P=0.00017 ASARFSWLSL ayw3 166 1 <sup>75</sup> H35.2 23x SPVSYGEWRARYCTNGGQOTVQRADRSCW P=0.0041 RTCTTPAQGT ayw3 122 130 166 175 PCRTCTTPAQGTSM EWASARFSWLSLLV ayw3 H35 epitope: 122 130 166 175 PCRTCTTPAQGTSM EWASARFSWLSLLV ayw3 H35 epitope: 124 117 187 207 H53 epitope: 125 117 207 H53 epitope: 126 117 187 H53 epitope: 127 187 H53 epitope: 128 117 187 H53 epitope: 129 117 187 H53 epitope: 120 117 187 H53 epitope: H54 epitope: H55 epi	.12	PKRQISMERWLQVTQGEEVTP CATC NPWVA		H5 epitope	•			
$ \begin{array}{c} 16 & WDKRPVWWLRFEESQRLSR CATC GVGGVE \\ 19 NVNEPGIRQGPAASVGWKVVRLAGI CKTC V \\ 20 & GKIVVFPKRSVPDVTGSQGAPP CFTC TST \\ 24 & SVLRQAAQFGNFELYVRREGN CRALTGC MR \\ \hline H166 epitope: C+TC all \\ (*+* is R or K) 121 124 \\ \hline \\ \hline \\ H35.1 & RASGARARCEHRSGLSLSWQPSECSDSRTT \\ P=0.00017 & ASARFSWLSL \\ 166 & 175 \\ \hline \\ H35.2 23x & SPVSYGEWRARYCTNGGQOTVQRRADRSCW \\ P=0.0041 & RTCTTPAQGT ayw3 \\ 122 & 130 \\ \hline \\ H35 epitope: \\ 122 & 130 \\ \hline \\ H53 epitope: \\ 121 & 117 \\ \hline \\ H53 epitope: \\ 122 & 117 \\ \hline \\ H53 epitope: \\ 121 & 117 \\ \hline \\ H53 epitope: \\ 122 & 117 \\ \hline \\ H53 epitope: \\ 124 & 117 \\ \hline \\ H53 epitope: \\ 124 & 117 \\ \hline \\ H53 epitope: \\ 124 & 117 \\ \hline \\ H53 epitope: \\ 124 & 117 \\ \hline \\ H53 epitope: \\ 124 & 117 \\ \hline \\ H53 epitope: \\ 124 & 117 \\ \hline \\ H53 epitope: \\ 124 & 117 \\ \hline \\ H53 epitope: \\ 124 & 117 \\ \hline \\ H53 epitope: \\ 124 & 117 \\ \hline \\ H53 epitope: \\ 125 & 117 \\ \hline \\ H53 epitope: \\ 126 & 117 \\ \hline \\ H53 epitope: \\ 127 & 117 \\ \hline \\ H53 epitope: \\ 127 & 117 \\ \hline \\ H53 epitope: \\ 127 & 117 \\ \hline \\ H53 epitope: \\ 127 & 117 \\ \hline \\ H53 epitope: \\ 127 & 117 \\ \hline \\ H53 epitope: \\ 127 & 117 \\ \hline \\ H53 epitope: \\ 127 & 117 \\ \hline \\ H53 epitope: \\ H53 epitopi$	.15	L CKTC VRSHQERTVKGDQVTGT	RI <b>CQTC</b>	101	106 130	158 1	167	
19NVNEPGIRGCPAAS/GWKVVRLAGI CKTC VDYGMLEVCPGMWAFAKYLWWASVRadw220GMKIVVFPKRSVPDVTGSQGAPP CRTC TSTD.24SVLRQAAQFGNFELYVREGN CRALTGC MRH166 epitope: C+TC all $(***$ is R or K) 121 124H53.165xGSVPLSRKEVWAGEEESFGYWLVNWQEMM $P=0.14$ CH15.1 $(***$ is R or K) 121 124H53.63xGRKTEKYSDGWTSMHSAEVCTQWNMSYQMI $P=0.00017$ H35.1 $P=0.00017$ $166$ ASARFSWLSL $122$ ayw3H53.63xGRKTEKYSDGWTSMHSAEVCTQWNMSYQMI $P=2.0$ H35.2 $122$ 23x $122$ SPVSYGEWRARYCTNGGQGTVQRADRSCW $P=0.0041$ H53.7 $122$ 2xGVELGKRANRGGSTTSWHGSSLGDIQSYWT $P=0.0045$ H53.7 $0SSTT$ $112$ ayw2 or ayw3H35 $P_{CRTCTTPAQGTSM}$ ayw3112 $117$ 117 $112$ 117 $112$ ayw2 or ayw3H35 $P_{CRTCTTPAQGTSM}$ Iff $122$ ayw3Iff $122$ ayw3Iff $122$ ayw3H35 $P_{CRTCTTPAQGTSM}$ Iff $122$ ayw3Iff $122$ ayw3Iff $122$ ayw3H35 $P_{CRTCTTPAQGTSM}$ Iff $122$ ayw3Iff $122$ ayw3Iff $122$ ayw3H35 $P_{CRTCTTPAQGTSM}$ Iff $122$ ayw3Iff $122$ ayw3Iff $122$ ayw2 or ayw3H35 $P_{CRTCTTPAQGTSM}$ Iff $122$ ayw3Iff $122$ ayw3Iff $122$ ayw3H35 $P_{CRTCTTPAQGTSM}$ Iff $122$ ayw3Iff $122$ ayw3Iff $122$ ayw2 or ayw3H35 <br< td=""><td>.16</td><td>WDKRPVVWLRFEESQRLSR CATC GVGGVE</td><td></td><td></td><td></td><td></td><td></td><td></td></br<>	.16	WDKRPVVWLRFEESQRLSR CATC GVGGVE						
20 GMKIVVFFKSVPDVTGSQGAPP CRTC TST 24 SVLRQAAQFGNFELVVRREGN CRALTGC MR H166 epitope: C+TC all (*+* is R or K) 121 124 C H35.1 RASGARARCEHRSGLSLSWQPSECSDSRTT P=0.00017 ASARFSWLSL ayw3 166 $175H35.2 23x SPVSYGEWRARYCTNGGQOTVQRRADRSCWP=0.0041$ RTCTTPAQGT ayw3 122 $130$ $166$ $175H35 epitope:122$ $130$ $166$ $175PCRTCTTPAQGTSM EWASARFSWLSLLV ayw3 H35 epitope: 122 130 166 175PCRTCTTPAQGTSM EWASARFSWLSLLV ayw3 112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207122 117 187 207122 117 187 207122 117 187 207123 112 117 187 207124 117 187 207125 112 117 187 207125 112 117 187 207125 112 117 187 207125 112 117 187 207125 112 117 187 207125 175 175 175 112 117 187 207112 117 187 207125 175 175 175 175 175 175 175 175 175 175 175 112 117 187 207112 117 187 207 112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 207112 117 187 120 175 112 117 112 117 112 117 112 117 112 117 112 117 112 117 112 117 1$	.19	NVNEPGIRQGPAASVGWKVVRLAGI CKTC V		DYQGM	LPV <u>C</u> P <b>QGN</b> .	WAFAKYLWEWA	SVR	adw2
.24 SVLRQAAQFGNFELYVRREGN CRALTGC MR H166 epitope: C+TC all ('+*' is R or K) 121 124 C H35.1 RASGARARCEHRSGLSLSWQPSECSDSRTT P=0.00017 ASARFSWLSL ayw3 166 175 P=0.0041 RTCTTPAQGT ayw3 122 130 H35 epitope: 122 130 H36 epitope: 122 130 H37 207 H38 epitope: 122 130 H38 epitope: 122 130 H39 epitope: 122 130	.20	GMKIVVFPKRSVPDVTGSQGAPP CRTC TST		_				
H166 epitope: C+TC allH53.165xGSVPLSRRKEVWAGEEESFGYWLUNWQEMM SPTVWL-SAIWMM all $187$ allCH53.65xGSVPLSRRKEVWAGEEESFGYWLUNWQEMM SPTVWL-SAIWMM all $187$ allFGN or K)121 124H53.65xGSVPLSRRKEVWAGEEESFGYWLUNWQEMM SPTVWL-SAIWMM $121$ CH53.65xGSVPLSRRKEVWAGEEESFGYWLUNWQEMM SPTVWL-SAIWMM $187$ CH53.63xGRTENGESLSLWQPSECSDSRTT $P=0.00017$ ASARFSWLSL $196$ ayw3H53.223xSPVSYGEWRARYCTNGGQGTVQRRADRSCW $P=0.0041$ RTCTTPAQGT $122$ ayw3H53.72xGVELGKRANRGGSTTSWHGSSLGDIQSYWT $P=0.0045$ GSSTTS $GSSTTS$ ayw2 or ayw3H35.223xSPVSYGEWRARYCTNGQOTVQRRADRSCW $P=0.0041$ RTCTTPAQGT $1122$ ayw31122130166175H53.310xRGSTT $1122$ ayw2 or ayw31122130166175PCRTCTTPAQGTayw3122130166 <td>.24</td> <td>SVLRQAAQFGNFELYVRREGN CRALTGC MR</td> <td></td> <td>D</td> <td></td> <td></td> <td></td> <td></td>	.24	SVLRQAAQFGNFELYVRREGN CRALTGC MR		D				
H166 epitope: C+TC all $P=0.14$ SPTVWL-SAIWMM all(*+* is R or K)121124 $P=0.14$ $Ie7$ <				Н53.16	5x GSVPLSRR	KEVWAGEEE <b>S</b> FGY	YWLVNWQE	M
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		H166 epitope: C+TC all		P=0.14		SPT	WIL-SAIW	<b>M</b> all
$ \begin{array}{c} \textbf{C} \\ \textbf{H35.1} \\ P=0.00017 \\ \textbf{ASARF}SWLSL \\ 166 \\ 175 \\ \textbf{H35.2} 23x \\ P=0.0041 \\ P=0.0041 \\ \textbf{RTCTTPAQGT} \\ 122 \\ 130 \\ \textbf{H35.epitope:} \\ \textbf{H33.epitope:} \\ \textbf{H33.epitope:} \\ \textbf{H33.epitope:} \\ \textbf{H33.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H35.epitope:} \\ 112 \\ 117 \\ 197 \\ 207 \\ \textbf{H35.epitope:} \\ 112 \\ 117 \\ 187 \\ 207 \\ \textbf{H35.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 187 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 \\ 197 \\ 207 \\ \textbf{H53.epitope:} \\ 122 \\ 117 $		("+" is R or K) 121 124				187	t	198
CH53.63xGRKTEKYSDGWTSMHSAEVCTOWNMSYCMIH35.1RASGARARCEHRSGLSLSWQPSECSDSRTT $P=2.0$ WMMWYall $P=0.00017$ ASARFSWLSLayw3 $196$ $200$ H35.2 $23x$ SPVSYGEWRARYCTNGGQGTVQRADRSCW $P=0.0045$ GSSTTS $ayw3$ $P=0.0041$ RTCTTPAQGT $ayw3$ $112$ $117$ $122$ $130$ $166$ $175$ $GSSTT$ $ayw2$ or $ayw3$ H35 epitope: $122$ $130$ $166$ $175$ $P=0.0035$ $MMWYWGPSLYS$ $all$ $122$ $130$ $166$ $175$ $P=0.0035$ $MMWYWGPSLYS$ $all$ $122$ $130$ $166$ $175$ $P=0.0035$ $MMWYWGPSLYS$ $all$ $122$ $130$ $166$ $175$ $197$ $207$ $112$ $117$ $197$ $207$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$ $112$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$ $112$ $117$ $187$ $207$								
P=0.00017RASGARARCEHRSGLSLSWQPSECSDSRTTP=2.0WMMWYallP=0.00017ASARFSWLSLayw3196200166175H53.72xGVELGKRANRGGSTSWHGSSLGDLQSYWTH35.223xSPVSYGEWRARYCTNGGQQCTVQRRADRSCW $P=0.0045$ GSSTTSayw2 or ayw3P=0.0041RTCTTPAQGTayw3112116122130166175GSSTTayw2 or ayw3H35 epitope:122130166175197207122130166175197207112117122130166175197207112117122130166175197207112117122130166175197207112117122130166175197207112117122130166175197207112117122130166175197207112117122130166175197207112117187207122130166175112117187207112117187207122130166175112117187207112114114114114114114114115123124125124125125124124124<	C			Н53.6	3x GRKTEKYS	DGWTSMHSAEVC7	rq <b>wnm</b> s <b>y</b> <u>c</u> n	<b>1</b> I
H35.1RASGARACELERSGLESDS/UPSE_SDS/RTT196200 $P=0.00017$ ASARFSWLSLayw3196200166175H53.72xGVELGKRANRGGSTTS/HGSSLGDIQSY/WTH35.223xSPVSYGEWRARYCTNGGQGT/QRRADRSCW $P=0.0045$ GSSTTSayw2 or ayw3 $P=0.0041$ RTCTTPAQGTayw3112117122130166175197207H35 epitope:H53.310xRGGL/MRGGSSDTRL/MGWQ-SSSIYSFQARGS122130166175197207PCRTCTTPAQGTSMEWASARFSWLSLLVayw3H53 epitope:122130166175197207PCRTCTTPAQGTSMEWASARFSWLSLLVayw3H53 epitope:112117187207112117187207112117187207112117187207112117187207112117187207112117187207112117187207112117187207112117187207112117187207112117187207112117187207113112117187114115207115117187207116112117187117112114114 <td colspan="3"></td> <td>P=2.0</td> <td></td> <td></td> <td>WMMWY</td> <td>all</td>				P=2.0			WMMWY	all
$P=0.00017$ $rac{ASARF}{SWLSL}$ $i66$ $175$ $H53.7$ $P=0.0045$ $P=0.0041$ $rac{CTTPAQGT}{122}$ $130$ $rac{CTTPAQGT}{122}$ $130$ $rac{CTTPAQGT}{122}$ $130$ $rac{CTTPAQGT}{122}$ $rac{CTTPAQGTSM}{122}$ $rac{CTTPACTTPACTTPACCTSM}{122}$ $rac{CTTPACTTPACTTPACTTPACTTPACTTPACTTSC}{122}$ $rac{CTTPACTTPACTTPACTTPACTTPACTTSC}{122}$ $rac{CTTPACTTPACTTPACTTPACTTPACTTPACTTSC}{122}$ $rac{CTTPACTTPACTTPACTTPACTTPACTTPACTTTPACTTPA$	<b>n</b> 35.	I RASGARARCERRSGISISWOPSECSDSRTT					196 200	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	P=0.	00017 ASARFSWLSL ayw3						
$\begin{array}{cccccccccccccccccccccccccccccccccccc$		166 175		<b>H53.7</b> 2x	GVELGKRANRG	G <b>STTS</b> WHGSSLGI	DIQSYWT	
H35.2 23x       SPVSYGEWRARYCTNGGQQTVQRRADRSCW       112       117         P=0.0041       RTCTTPAQGT       ayw3       112       116         122       130       GSSTT       ayw2 or ayw3         H35 epitope:       H53.3       10x       RGGLMRGGSSDTRLMGWQ-SSSIYSFQARGS         122       130       166       175       Pe0.0035       MMWYWGPSLYS       all         122       130       166       175       197       207       197       207         PCRTCTTPAQGTSM       EWASARFSWLSLLV       ayw3       H53 epitope:       112       117       187       207         I12       117       187       207       112       117       187       207         I12       117       187       207       112       117       187       207				P=0.0045	G	SSTTS	ayw	2 or ayw3
P=0.0041         RTCTTPAQGT         ayw3         112 116           122 130         GSSTT         ayw2 or ayw3           H35 epitope:         H53.3 10x         RGGLMRGGSSDTRLMGWQ-SSSIYSFQARGS           122 130         166 175         Pe0.0035         MMWYWGPSLYS         all           PCRTCTTPAQGTSM         EWASARFSWLSLLV         ayw3         H53 epitope:         197 207           112 117         187         207         112 117         187         207           IDESSTTSTG         GLSPTVWI,SVIWMWYWGPSLYSUL avw2 or avw3         112 117         187         207	нз5.	2 Z3X SPVSYGEWRARYCTNGGQGTVQRRADRSCW			11	2 1'17		
122     130     GSSTT     ayw2 or ayw3       H35 epitope:     H53.3     10x     RGGLMRGGSSDTRLMGWQ-SSSIYSFQARGS       122     130     166     175     197     207       PCRTCTTPAQGTSM     EWASARFSWLSLLV     ayw3     H53 epitope:     112     117     187     207       112     117     187     207     112     117     187     207	P=0.	0041 RTCTTPAQGT ayw3			112 1	.16		
H35 epitope:       H53.3 10x       RGGLMRGGSSDTRLMGWQ-SSSIYSFQARGS         122       130       166       175       197       207         PCRTCTTPAQGTSM       EWASARFSWLSLLV       ayw3       H53 epitope:       112       117       187       207         I12       117       187       207       112       117       187       207         I12       117       187       207       112       117       187       207		122 130			GSSI	Ť	ayw	2 or ayw3
H35 epitope:       P=0.0035       MMWYWGPSLYS       all         122       130       166       175       197       207         PCRTCTTPAQGTSM       EWASARFSWLSLLV       ayw3       H53 epitope:       112       117       187       207         I12       117       187       207       198       199       112       117       187       207				<b>H53.3</b> 10x	RGGLMRG <b>GSS</b> I	TRLMGWQ-SSSI	<b>TS</b> FQARGS	
122 130 166 175 P <u>C</u> RT <u>C</u> TTPAQGTSM EWASARFSWLSLLV ayw3 H53 epitope: 112 117 187 207 IPCSSTTSTG GLSPTVWI.SVIWMWYWGPSI.YSIL avw2 or avw1	H35	epitope:		P=0.0035		MMWYWGPSLY	TS	all
P <u>C</u> RT <u>C</u> TTPAQGTSM EWASARFSWLSLLV ayw3 H53 epitope: 112 117 187 207 UPCSSTTSTG GLSPTVWI.SVIWMMWYWGPSI.Y51L avw2 or avw1		122 130 166 175				197 2	207	
112 117 187 207 IDESSTTATE CLASTING SVIWMMYYWGPSI.YSTL avw2 or avw1		P <u>C</u> RT <u>C</u> TTPAQGTSM EWASARFSWLSLLV ayw3		H53 epitop	e:			
IDCSSTTSTC CLSPTVMT,SVTWMMVTWCPSI,YSTI, avv2 or avv2				112 117	187		207	
				IP <b>GSSTTS</b> TG	GLSPTVWI	SVI <b>WMMWYW</b> GP <b>S</b> I	L <b>YS</b> IL avw	2 or avw

FIG. 1. Peptide sequences from phage display. (A, B, C, and D) Results of biopanning against mAbs H166, H5, H35, and H53, respectively. The peptide sequence is preceded by the clone name and the number of identical clones sequenced (if more than one). Residues identified as matching the HBsAg sequence are highlighted in boldface; the corresponding sequence of the HBsAg is also given and numbered as in the sequence of the mature form of HBsAg. The HBsAg sequence from one of three possible immunogens, subtype adw2, ayw2, or ayw3, is presented (sequences are from the Swiss-Prot 30 protein sequence database). A descriptor "all" is used if sequences from all three subtypes match the sequence that is highlighted in boldface. Cysteine residues are underlined. The lower part of each panel contains the predicted epitope sequence for the antibody used in panning that is a composite of sequences matching the HBsAg sequence in that panel. No similar sequences were identified in clone H5.3 of *B*. The similarities between sequences were identified by the manual analysis of dot matrix diagrams (GENEWORKS 2.3) (*B*, *C*, and *D*) or by visual inspection (*A*). Numerical values for *P* indicate an estimated frequency of obtaining the sequence match (indicated in boldface) when comparing two random sequences 130 and 30 amino acids long. The following equation gives the frequency *P* of finding similar sequences in two random sequences having the length of *m* and *n* residues, respectively. It is assumed that 20 different residues are allowed with equal probability at each position of the chain. A match of *k* residues out of a block of *l* residues is sought in both sequences. An approximate formula for *P* is

$$P = \binom{l}{k} \cdot \frac{(m-l+1)\cdot(n-l+1)}{20^k} \cdot \left(\frac{k}{l}\right)^{1.4}.$$
[1]

An estimate for the frequency of two nearby similar regions within a pair of two sequences is

$$P = \frac{(a+1)(b+1)}{mn} P_1 P_2,$$
 [2]

where  $P_1$  and  $P_2$  are frequencies of observing two similar regions as calculated from Eq. 1, and a and b are distances between the similar regions measured within the peptide and within the sequence to which the peptide is compared, respectively, and a << m, b << n. Eqs. 1 and 2 were derived from basic combinatorics formulas; in addition, the second term in Eq. 1,  $(k/l)^{1.4}$ , is obtained from a Monte Carlo computation of frequencies of sequence matches. These formulas were applied to estimate the probabilities of occurrences of similarities shown in B, C, and D. Parameter values are m = 130 and n = 30. As an example, for H5.2, two matches observed of k = 3 out of l = 3 residues separated by a = 4 residues in one sequence and within the length of the other, P = 0.0080. Relative high numerical values of P were observed for two sequences, H53.6 and H53.16. The justification for the significance of the alignment is based on the similarity to other H53-derived sequences and the high content of aromatic residues (Trp and Tyr) and Met in the similar regions.

from two or more monomers while maintaining the consistency with the phage display data and the surface model presented here.

An attempt has been made to construct an  $\alpha$ -carbon trace for the polypeptide chain of the 101–207 region that is consistent with the model shown in Fig. 3. The main assumption is that the observed epitopes are formed within a single monomer of HBsAg. We took advantage of knowing the assignment of one disulfide bridge (Cys-121 Cys-124; this paper, as well as ref. 12), and a high likelihood of another assignment (Cys-139 Cys-147; ref. 16). Although the model presented in Fig. 4 shows two other disulfide bonds (Cys-107 Cys-138 and Cys-137 Cys-149) consistent with the full oxidation of eight cysteine residues in the antigenic region of HBsAg (4), the topological properties of the model would not be drastically changed if the latter four cysteine residues were not involved in forming disulfide bridges or formed disulfide bridges with other monomer(s) within the HBsAg particle.

The trace presented in Fig. 4 was derived on the basis of the presumption that the dominant feature of the surface is an



FIG. 2. Inhibition of phage binding to anti-HBsAg mAbs H5, H35, and H53 (*A*, *B*, *C*, and *D*, respectively) by HBsAg. The ability to block phage binding was determined by preincubating dilutions of HBsAg with a constant amount of phage. For H5 (*A*) and H35 (*B*), results of individual experiments are plotted. For H53, plotted values are results of an average of triplicate experiments: (*C*) 53.16,  $\diamond$ ; 53.3,  $\Box$ ; and 53.3 (second replicate),  $\bigcirc$ ; and (*D*) 53.6,  $\Box$ ; 53.7,  $\diamond$ ; and 35.7,  $\bigcirc$ .

extended conformation of the chain. We believe this to be so for two reasons. First, 55 of a total of 75 residues in the 101–175 region (73%) are found to be on the surface (this paper and ref. 16). The only residues thought to be buried in this region are 107–111, 136–139, and 147–157. Second, surface elements in envelope proteins of spherical small RNA viruses (to which HBV is thought to be structurally related) are in extended conformation (typically a  $\beta$ -sheet) in either the open form (bacteriophage MS2; Protein Data Bank entry 1MS2) or in the jelly roll topology (such as in the satellite tobacco necrosis virus; Protein Data Bank entry 2STV).

The extended parts of the polypeptide chain are formed by the residues defining the epitopes in the 101-175 region (Fig. 1), except for region 121-124, which forms a turn anchored by a disulfide bridge. Region 136-139 is mostly buried, with its three cysteine residues being part of a network of disulfide bridges with other cysteine residues within HBsAg. Residues 150-157 have not been identified as forming any epitope. In the model, they are buried and link loop 139-147 with an extended region starting at residue 158.

Two  $\alpha$ -helices in the model are proposed on the basis of the secondary structure assignment (18) and a pattern of residues in sequences from biopanning. Persson and Argos (18) predicted four helices in HBsAg, namely 6-33, 75-103, 173-193, and 202-222. We found that several residues in the 187-207 region lie on the surface of HBsAg (Fig. 1D), and the pattern of residues in contact with the H53 antibody is consistent with a helix formed by residues 176-195, followed by a turn at 196-201 (see sequence of clone H53.6 in Fig. 1D, where a conserved sequence is flanked by two cysteine residues, suggesting a turn) and another helix at 202-207+, in agreement with Persson and Argos (18). The proposed helices are not consistent with an earlier topological model for the HBsAg (19) in which residues 160-184 and 189-210 form transmembrane  $\alpha$ -helices and the middle section (185–188) is buried inside the HBsAg particle.

The model does not provide structure for residues 1-100 and 208-226; these regions form the hydrophobic core of the antigen and support the surface-exposed fragments of the



FIG. 3. Representation of the surface of the antigen. Circles correspond to regions of similarity between the peptide identified from phage display experiments (Fig. 1) and the antigen sequence. The circles are labeled with an antibody name (H5, H35, H53, or H166), and the numbers inside the circle identify residues forming the epitope. Regions labeled I and II were identified as epitopes by Motti *et al.* (9), and region III is the binding site for the mAb RFHB7 (15). Circles are shown as overlapping if epitopes recognized by two different antibodies overlap. Different regions in the diagram may correspond to protein fragments coming from either one or possibly more than one antigen monomer.

chain which are in an extended conformation. A high percentage (45–50%) of  $\alpha$ -helix in HBsAg, as revealed by circular dichroism studies (6, 13), implies that these two regions are mostly  $\alpha$ -helical.

Several point mutations in the HBV surface antigen have been described. The following mutations affect the recognition properties of the surface antigen by one or more antibodies: D144Q, G145A, K160N, P120Q, T/I126N, K141E, T126S-T131N-M133T triple mutant (reviewed in ref. 20), and G145R (15). The mutations listed are predicted in the model to be on the surface of the antigen and, as such, can directly influence antibody binding, consistent with the experimental evidence.



FIG. 4. A plausible topological model for the three-dimensional structure of the antigen. An oval ribbon representation of the  $\alpha$ -carbon trace is shown. The initial model based on the distance constraints from phage library data was modified by replacing the loops with the closest resembling structures chosen from the o structural data base (17). The model shown was prepared by using the GRAMPS graphic system and RAYT program. Residue numbers for small HBsAg are given.

The approach presented in the paper can be summarized as obtaining distance constraints from biopanning experiments with peptide libraries on filamentous phage and using the distance constraints to derive a model for the surface of a protein or for the polypeptide chain trace. The approach is conceptually analogous to deriving a three-dimensional structure of a protein from nuclear Overhauser effect determinations in NMR, although the number of constraints from biopanning is much less than in NMR, and their accuracy is much lower (which also prevented us from using distance geometry programs to build the model). To further improve the accuracy of the model, it would be desirable to increase the number of antibodies used in biopanning experiments and the number of clones investigated and sequenced. The approach could be especially useful to study structural properties and to derive peptide mimics for those systems where the structure of a protein cannot be easily determined by crystallography or NMR, as is the case for HBsAg.

We thank Joan Tyner for providing mAbs and Craig Carlson for providing surface antigen. We are grateful to Bob Carlson for programming and T. J. O'Donnell for computer ray tracing. We thank the members of the Protein Engineering group for helpful discussions and Michael Klass for support.

- 1. Parmley, S. F. & Smith, G. P. (1988) Gene 73, 305-318.
- 2. Scott, J. K. & Smith, G. P. (1990) Science 249, 386-390.
- Grihalde, N. D., Chen, Y.-C. J., Golden, A., Gubbins, E. & Mandecki, W. (1995) Gene 166, 187-195.
- Hollinger, F. B. (1990) in *Virology*, eds. Fields, B. N. & Knippe, D. M. (Raven, New York), 2nd Ed., Vol. 2, pp. 2171–2236.
- Mangold, C. M. T. & Streeck, R. E. (1993) J. Virol. 67, 4588– 4597.
- Antoni, B. A., Rodriguez-Crespo, I., Gomez-Gutierrez, J., Nieto, M., Peterson, D. & Gavilanez, F. (1994) Eur. J. Biochem. 222, 121–127.
- Peterson, D. L., Nath, N. & Gavilanez, F. (1981) J. Biol. Chem. 257, 10414-10420.
- Folgori, A., Tafi, R., Meola, A., Felici, F., Galfre, G., Cortese, R., Monaci, P. & Nicosia, A. (1994) *EMBO J.* 13, 2236–2243.
- Motti, C. M., Nuzzo, M., Meola, A., Galfre, G., Felici, F., Cortese, R., Nicosia, A. & Monaci, P. (1994) Gene 146, 191–198.
- Peterson, D. L., Paul, D. A., Lam, J., Tribby, I. I. E. & Achord, D. (1984) J. Immunol. 132, 920–927.
- Mimms, L., Goetze, A., Swanson, S., Floreani, M., Edwards, B., Macioszek, J., Okasinski, G. & Kiang, W. (1989) J. Virol. Methods 25, 211–232.
- Ohnuma, H., Takai, E., Machida, A., Tsuda, F., Okamoto, H., Tanaka, T., Naito, M., Munekata, E., Miki, K., Miyakawa, Y. & Mayumi, M. (1990) J. Immunol. 145, 2265-2271.
- 13. Guererro, E., Gavilanes, F. & Peterson, D. L. (1988) in Viral Hepatitis and Viral Disease, ed. Zuckerman, A. J. (Liss, New York).
- 14. Padlan, E. A. (1991) in *Structure of Antigens*, ed. Van Regenmortel, M. H. V. (CRC, Boca Raton, FL), pp. 29-42.
- Carman, W. F. (1994) in Viral Hepatitis and Liver Disease, eds. Nishioka, K., Suzuki, H., Mishiro, S. & Oda, T. (Springer, New York), pp. 243-247.
- Howard, C. R., Stirk, H. J., Brown, S. E. & Steward, M. W. (1988) in Viral Hepatitis and Liver Disease, ed. Zuckerman, A. J. (Liss, New York), pp. 1094-1101.
- 17. Jones, T. A., Cowan, S., Zou, J.-Y. & Kjeldgaard, M. (1991) Acta Crystallogr. A 47, 110-119.
- 18. Persson, B. & Argos, P. (1994) J. Mol. Biol. 237, 182-193.
- Stirk, H. J., Thornton, J. M. & Howard, C. R. (1992) Intervirology 33, 148–158.
- Dawson, G. J., Mimms, L. T. & Lesniewski, R. R. (1993) in *Current Hepatology*, ed. Gittnick, G. (Wiley, New York), Vol. 14, pp. 63-116.