Cloning, expression, isolation and characterization of the pre-S domains of hepatitis B surface antigen, devoid of the S protein

Sue DELOS, Maria Teresa VILLAR, Peisheng HU and Darrell L. PETERSON

Department of Biochemistry and Molecular Biophysics, Virginia Commonwealth University, Richmond, VA 23298, U.S.A.

The 'pre-S' parts of the envelope protein of hepatitis B virus (HBV) have been proposed to be involved in the infection of hepatocytes by HBV. In order to facilitate the study of these processes, we have developed an expression system to allow the production and purification of large quantities of the pre-S protein. To obtain a protein containing all of the pre-S sequence and only this sequence, mutations were introduced into the HBV(*ayw*) genome to create an *Nde*I restriction site at the initial ATG of the large surface protein gene. Also, stop codons and a *Bgl*II restriction site were introduced after the last codon of pre-S2. This fragment was then cloned into the high-expression vector pET-3A. A protein of the expected M_r was expressed at a level of up to 10% of the total soluble protein in HMS174 (DE3) cells, as judged by SDS/PAGE. A rapid purification method has been developed for this protein. The protein retains the polyalbumin-binding activity ascribed to the pre-S2 sequence, and is recognized by both polyclonal and monoclonal antibodies directed against pre-S determinants. Gel filtration chromatography demonstrates that the protein is monomeric and globular, and c.d. spectroscopy indicates that β -sheet is the major periodic structure.

INTRODUCTION

The hepatitis B surface antigen (HBsAg) is synthesized as a nested set of three proteins [1–4]. The bulk of the protein occurs as the short form, designated S, which is 226 amino acids long [2]. The two additional proteins result from *N*-terminal extensions of the S protein. The middle-length protein, M, contains an additional 55 amino acids. These are designated pre-S2 [1]. The large surface antigen protein, L, contains the M protein plus an additional 109 or 118 amino acids, depending on the antigen subtype. These amino acids are designated pre-S1 [3,4]. Although HBsAg has a macromolecular lipoprotein structure, only the S portions of proteins are closely associated with the lipid, whereas the pre-S sequences are hydrophilic and have been shown to exist entirely external to the particle lipid [5]. The removal of the pre-S does not affect the particle structure of HBsAg or the immunochemical properties of the S sequences within the HBsAg particle, suggesting that the pre-S and S sequences exist as independent domains [5]. Although both pre-S1 and pre-S2 contain potential glycosylation sites, published data are conflicting concerning whether either domain is glycosylated within the L protein [4,6], and the only post-translational modification of this sequence that has been clearly demonstrated is that the *N*-terminus of the L protein is myristoylated [7].

The pre-S domains of HBsAg have been implicated in a number of functions. The pre-S protein has important antigenic sites for both B cells and T cells [8–11]. It contains sequences that have been shown to bind specifically to hepatocytes [12–14] and others which bind to polymerized human serum albumin (pHSA) [6,15,16]. These binding functions are thought to be important in the initial phase of viral infection [6,12–16]. Studies of the detailed structural basis for these activities is complicated by the fact that the L and M proteins are present in small and highly variable amounts [4]. Moreover, in all cases, the S protein is in great excess over the pre-S-containing M and L proteins, which precludes many physical and chemical studies of the structure of the pre-S domains. For these reasons we have developed methods to allow the synthesis and isolation of the pure pre-S domains

devoid of S protein. The physical and immunochemical properties of this pre-S protein are reported here.

MATERIALS AND METHODS

Enzymes and reagents

Restriction endonucleases and T4 ligase (400 units/ml) were purchased from New England BioLabs, Beverly, MA, U.S.A. Isopropyl β -D-thiogalactopyranoside (IPTG) and culture medium components were obtained from Sigma Chemical Co., St. Louis, MO, U.S.A. All other chemicals were from Fisher Chemical Co., Columbia, MD, U.S.A.

Vectors

pKSV10 (Pharmacia–LKB Biotechnology, Piscataway, NJ, U.S.A.) containing the complete hepatitis B virus (HBV) surface antigen gene (subtype *ayw*) had previously been prepared in our laboratory (D. L. Peterson, unpublished work). The pET expression system was obtained from W. Studier, Brookhaven National Laboratory, Upton, NY, U.S.A. [17]. All plasmids were isolated from cell cultures by the alkaline lysis method followed by dye-buoyant density-equilibrium centrifugation [18].

Synthesis of the pre-S gene

The full-length pre-S1S2 gene, devoid of S sequences, was prepared by mutation and synthesis using the polymerase chain reaction as follows. Primers were synthesized at the Nucleic Acid Synthesis and Analysis Core Facility, Virginia Commonwealth University, and were purified by electrophoresis on a 20% polyacrylamide gel containing 8 M-urea. The primers used in the reaction were CAAGATCTACcatATGGGGCAGAATCTTTCCAC for the 5' end and CATGaGATcTTCTaCtaGTTcAGCGCAGGGTCCCCAA for the 3' end (the bases in lower case letters denote the mismatches used). The reaction was performed in a total volume of 100 μ l, and contained 50 mM-KCl, 10 mM-Tris, pH 8.3, 1 mM-MgCl₂, 0.05% gelatin, 0.2 mM-dNTPs, 2 μ M primers and 1 ng of HBsAg DNA. After an initial

Abbreviations used: HBV, hepatitis B virus; HBsAg, hepatitis B surface antigen; S, M and L, short, medium and long forms of HBsAg; IPTG, isopropyl β -D-galactopyranoside; (p)HSA, (polymerized) human serum albumin; OPD, *o*-phenylenediamine.

incubation at 94 °C for 3 min, 2.5 units of Taq polymerase were added, and 30 reaction cycles were run. The cycle consisted of 2 min at 94 °C, 2 min at 37 °C and 3 min at 72 °C. The final reaction mixture was extracted with phenol/chloroform (1:1, v/v), then stored at -20 °C until used.

Genetic construction

The DNA obtained from the amplification reaction was digested with *NdeI* and *BglII*, and the pET-3A vector was cut with *NdeI* and *BamHI*. The resultant DNAs with compatible ends were then mixed, ligated and used to transform HMS174 cells. Plasmid-containing cells were obtained by selection on ampicillin-containing medium. Colonies containing the correct insert were identified by restriction enzyme analysis of plasmid DNA prepared from mini-lysates [19]. Plasmid DNA from cells containing the correct insert was purified and then transfected into HMS174 (DE3) cells. These cells contain T7 polymerase under the control of an IPTG-inducible promoter, lacUV5, as a permanent phage lysogen [17]. Cells were screened for the expression of the pre-S protein by first cultivating them on medium containing 0.5 mM-IPTG. Whole cell mini-lysates were prepared by suspension of the cells obtained from 1.5 ml of culture in SDS sample buffer and boiling for 2 min. Samples were analysed by immunoblotting of the proteins separated by SDS/PAGE using anti-pre-S antibodies, as described below. Alternatively, cells were screened for the production of the pre-S protein using a polymerized HSA-binding assay.

DNA sequencing

DNA sequencing was performed by the dideoxy chain-termination method of Sanger *et al.* [19] using the Sequenase reagents (US Biochemicals) according to the manufacturer's directions.

pHSA-binding assay for the pre-S protein

One activity ascribed to the pre-S domain of the HBsAg is the binding to polymerized serum albumin [16]. To assay for this activity, pHSA was prepared according to previously described methods [6]. Polystyrene beads of 0.635 cm diam. ($\frac{1}{4}$ inch) (Abbott Laboratories) were coated with pHSA by incubation with this protein (400 µg/ml) in sodium carbonate buffer, pH 9.5, for 18 h at room temperature. The beads were then washed with water and stored at 4 °C. Cells were disrupted by sonication and samples to be tested were diluted in phosphate-buffered saline containing 0.1 % Tween 20, and 0.2 ml was added to an AUSRIA tray (Abbott Laboratories, North Chicago, IL, U.S.A.). A pHSA-coated bead was added to each well, and the tray was then incubated at 37 °C for 1 h. The beads were washed, and then peroxidase-labelled anti-pre-S antibody, diluted in phosphate-buffered saline containing 10 % normal human serum, was added. The tray was incubated for an additional 1 h at 37 °C. Following extensive washing with saline, the beads were transferred to tubes and the peroxidase substrate [*o*-phenylenediamine (OPD) substrate solution from Abbott Laboratories] was added. The reaction was stopped after 30 min or when the colour was judged to be sufficiently developed. The reaction was monitored at 492 nm with an Abbott Quantum II e.l.i.s.a. reader.

Purification of the pre-S protein

Overnight cultures of the HMS174 (DE3) cells containing the correct plasmid were transferred to fresh medium (5 %, v/v) and grown to an A_{600} of 0.6. IPTG was then added to a final concentration of 0.5 mM. Cells were incubated for an additional 3 h, then harvested by centrifugation. The cell pellet was

resuspended in cold water containing 50 mM-EDTA, 10 mM-iodoacetate and 1 mM-phenylmethanesulphonyl fluoride. All subsequent procedures were performed at 4 °C, and all buffers contained 0.02 % NaN₃. The cells were disrupted by a single passage through a French press at 138 000 kPa (20 000 lb/in²). Cell debris was removed by centrifugation at 25 000 *g* for 5 min. The supernatant was then diluted until the conductivity was < 1500 µS, and applied to a CM-Sephadex column equilibrated with 10 mM-sodium phosphate buffer, pH 6.8. Unbound proteins were removed by washing the column with the same buffer until the A_{280} returned to baseline values. The pre-S protein was then eluted with 100 mM-sodium phosphate buffer, pH 6.8. The protein was concentrated by ammonium sulphate precipitation by addition of solid ammonium sulphate (245 g/litre). To separate full-length pre-S protein from that which had been proteolytically cleaved before and during the purification procedure, the protein was chromatographed on a Bio-gel P-10 column (4 cm × 110 cm), equilibrated with 50 mM-sodium phosphate buffer, pH 6.8.

SDS/PAGE and immunoblotting

Analytical PAGE was performed on a Bio-Rad slab gel apparatus with a 15 % polyacrylamide gel; a 4 % stacking gel was used. The gel and buffer formulations were those of O'Farrell [20]. For immunoblotting, the protein was transferred to nitrocellulose according to the method of Burnette [21]. The nitrocellulose was treated with 5 % BSA in 10 mM-Tris/HCl, pH 7.5, containing 154 mM-NaCl. After 30 min this blocking solution was removed and anti-pre-S antibody of desired specificity was added in the same Tris/HCl buffer containing 0.2 % Tween 20, 3 % BSA, 0.1 % NaN₃ and 0.1 mM-EDTA. Following overnight incubation, the nitrocellulose filter was washed extensively with 10 mM-Tris/HCl, pH 7.5, containing 0.5 M-NaCl. Peroxidase-labelled goat anti-rabbit immunoglobins in 10 mM-Tris/HCl, pH 7.5, containing 10 % fetal calf serum, was then added. Following a 2 h incubation at room temperature, the filter was extensively washed with 0.9 % NaCl, then developed by the addition of 1-chloronaphthol and H₂O₂ (peroxidase substrate solution from Kirkegaard & Perry Laboratories, Gaithersburg, MD, U.S.A.).

Protein analysis

The N-terminal sequence of the protein was determined using an Applied Biosystems model 470A sequencer, equipped with an on-line phenylthiohydantoin-amino acid analyser. The C-terminal sequence was determined by amino acid analysis following digestion with carboxypeptidase Y as previously described [20]. Amino acid analyses were performed with a Durrum MBF amino acid analyser using *o*-phthaldehyde as the detection reagent. Tryptophan was determined spectrophotometrically [22]. C.d. spectra were obtained using a Jasco J-500C spectropolarimeter equipped with a thermostatically controlled cell. The M_r of the protein was estimated by chromatography on a 2.5 cm × 100 cm Bio-Gel P-100 column, by comparison with standard proteins.

RESULTS

Construction and expression of the protein

Fig. 1 shows the strategy chosen for the creation of a clone capable of directing the synthesis of the full-length pre-S amino acid sequence, but devoid of the S domains, of HBsAg/*ayw*. Mismatched primers corresponding to the 5' and 3' ends of the pre-S sequences of the HBV S gene were used to synthesize the pre-S DNA sequence shown, by polymerase chain reaction

Fig. 1. Construction of the pre-S expression vector

The pre-S region of the HBV genome was synthesized by a polymerase chain reaction, yielding a DNA with the sequences at the 5' and 3' ends as shown. This DNA was cloned into the *NdeI*/*BamHI*-cleaved pET-3A vector to yield the pET-3A/pre-S1S2 vector. The amino acid sequence of the protein that would be produced from this vector is also shown. Synthetic peptides corresponding to the underlined sections of this protein sequence were prepared and used to immunize rabbits in order to generate antibodies against specific regions of the pre-S protein.

mutagenesis and amplification. The final product had *NdeI* and *BglII* restriction sites as well as termination codons following the coding triplet for the final amino acid of the pre-S2 domain. Also shown are the relevant aspects of the pET-3A vector. The pre-S gene was inserted into the *NdeI*/*BamHI* site of this vector as shown. The expected final protein sequence is also shown in Fig. 1. This sequence exactly matches the pre-S sequence of the *ayw* antigen, with no additional amino acids [23].

A single fragment with the expected length of 512 bp was obtained from the reaction, and this product gave the expected two fragments upon digestion with *EcoRI* (results not shown) [23].

This DNA was cloned into the pET-3A vector, transformed into HMS174 cells, and ampicillin-resistant colonies were obtained. *EcoRI* digestion, which should yield fragments of 4220 and 655 bp, was used to identify potentially correct clones [17,23]. Several such clones were obtained and screened for the production of the pre-S protein. The DNA encoding the pre-S protein from one of these clones was completely sequenced and shown to be as expected (results not shown). This clone was used for all subsequent studies.

Fig. 2. SDS/PAGE of fractions obtained during purification of the pre-S protein

Lanes A-F are Coomassie-Blue-stained proteins; lane G and H were transferred to nitrocellulose and stained with anti-pre-S antisera. Lane A is a mixture of M_r standards. Lanes B and C are bacteria containing or lacking the pET-3A/pre-S1S2 vector respectively after induction with IPTG. Lane D is the eluate from the CM-Sephadex column, lane E is the protein after ammonium sulphate precipitation and lane F is the high- M_r fraction obtained from Bio-Gel P-10. Lanes G and H are the immunoblotted samples corresponding to lanes B and C respectively.

In order to express the pre-S protein, plasmid DNA from colonies containing the correct construct was purified and transfected into HMS174 (DE3) cells. Cells were grown, induced and harvested as described in the Materials and methods section. When the cells were lysed directly in sample buffer for SDS/PAGE and immunoblotting, the results shown in Fig. 2 were obtained. A protein having the expected M_r for the pre-S protein (M_r 17363) was observed after induction of pET-3A/pre-S1S2-containing cells (Fig. 2, lane B) but not in control cells containing just the pET-3A vector (Fig. 2, lane C). This protein was shown to be the expected pre-S protein by immunoblotting with rabbit antiserum made against a peptide corresponding to residues 1-20 of the pre-S sequence as shown in Fig. 1. The new protein was readily stained with this antiserum (Fig. 2, lane G), whereas control cells lacking the pET-3A/pre-S1S2 vector were not stained (Fig. 2, lane H). This protein represents approx. 10% of the soluble proteins, as measured by densitometric scanning of the stained gel.

Purification of the pre-S protein

Since a relatively large amount of pre-S protein is made, only 1-litre cultures were necessary to allow the purification of substantial quantities of the protein. From 2 litres of culture, approx. 6 g of cells were obtained. The cells were suspended in 30 ml of water containing proteinase inhibitors and disrupted, as described in the Materials and methods section. Fig. 2 shows an SDS/PAGE gel of samples taken from the various purification steps. When the cells were disrupted directly in the SDS sample buffer, only a single molecular species of pre-S protein was observed having the M_r expected for the full-length protein (Fig. 2, lane B). However, in subsequent steps, additional lower- M_r species were observed (Fig. 2, lanes D and E). These must result from proteolytic cleavage following disruption. The amount of cleavage could be decreased, although not eliminated, by inclusion of proteinase inhibitors, keeping the solution cold and accelerating the purification process (results not shown). We have not attempted to identify the specific proteinases involved but rather have added several well-known proteinase inhibitors. EDTA was included to inhibit metalloendopeptidases, iodoacetate was included to inhibit cysteine proteinases (pre-S contains no cysteine residues), and phenylmethanesulphonyl fluoride was included to inhibit serine proteinases. The cells were disrupted in cold water to allow direct application of the clarified

lysate to the CM-Sephadex column, which was run at 4 °C. The pre-S protein readily bound to CM-Sephadex at pH 6.8, whereas virtually all other proteins did not. Therefore the protein was eluted in a nearly pure form with 100 mM-phosphate buffer, pH 6.8 (Fig. 2, lane D). The protein was concentrated and some remaining contaminants were removed by precipitation of the protein with ammonium sulphate (Fig. 2, lane E). It is clear from the SDS/PAGE, and the results of immunoblotting studies presented later, that the final product is pure pre-S protein, although it is heterogeneous in molecular size. The full-length protein was subsequently separated from the proteolysed protein by gel filtration on Bio-Gel P-10 (Fig. 2, lane F). The typical yield of pure pre-S protein was 30–50 mg/litre of cell culture, and the current procedure requires about 2 h from the time of disruption of the cells.

M_r determination

In order to establish the M_r of the pre-S protein under native conditions, the protein was chromatographed on a column of Bio-Gel P-100 which had been calibrated with known protein standards. The pre-S protein was eluted at a position consistent with it being a monomeric protein having an apparent M_r of about 23000. This is somewhat higher than the value of 17400 calculated from the amino acid sequence; however, it is clearly not sufficiently high as to indicate a multimeric protein. From the V_e/V_0 values obtained by gel filtration, we have calculated a frictional coefficient of 1.1, which is consistent with a globular protein having the shape of a prolate ellipsoid with an axial ratio of approx. 1:3 [19].

Amino acid analysis and primary structural analysis of the protein

The amino acid composition of the protein obtained by amino acid analysis of the full-length protein obtained from the Bio-Gel P-10 column (Fig. 2, lane F) was in very good agreement with that expected based on the DNA sequence [23]. The pre-S protein shown in Fig. 2 (lane E) was subjected to 10 cycles of automatic Edman degradation, and a single sequence was observed. The *N*-terminal residue was found to be glycine, indicating that the *N*-terminal methionine had been removed, after synthesis, by the *Escherichia coli* methionine aminopeptidase [24]. The subsequent steps were as expected from the DNA sequence and are shown in Fig. 1. Since this sample was a mixture of the various forms (in terms of length) of the pre-S protein, the presence of a single *N*-terminal residue indicates that only the C-terminus of the protein has been proteolysed.

When the full-length protein (Fig. 2, lane F) was subjected to carboxypeptidase Y digestion, the C-terminal sequence Ala-Leu-Asn was determined, consistent with that expected from the DNA sequence.

Immunoblotting studies of the pre-S protein

Although the pre-S protein observed in cells that were lysed directly in SDS sample buffer appeared to be of a single M_r species corresponding to the full-length protein, the protein that was obtained when the cells were lysed in buffer was heterogeneous in size, indicating that some of the protein was proteolysed. Therefore we have used anti-pre-S antibodies of known specificity to more clearly localize the sites of proteolysis. The antibodies used were prepared by immunizing rabbits with synthetic peptides corresponding to selected regions of the pre-S1 or pre-S2 sequences [5]. The specific peptide antigens which were used are indicated in Fig. 1. In addition, a monoclonal anti-pre-S1 antibody, obtained from Abbott Laboratories, was used. Fig. 3 shows the results of immunoblotting the pre-S protein with

Fig. 3. Immunoblotting of the pre-S proteins with anti-pre-S antisera of defined specificity

The pre-S protein used was that shown in lane E of Fig. 2. Lane A was stained with a monoclonal anti-pre-S1; lane B was stained with anti-pre-S-(1–20); lane C was stained with anti-pre-S-(23–46); lane D was stained with anti-pre-S-(109–133); and lane E was stained with anti-pre-S-(146–163) (see Fig. 1).

three different anti-pre-S1 (lanes A–C) and two different anti-pre-S2 (lanes D and E) antibodies. Since each protein observed on the Coomassie-Blue-stained SDS/PAGE gel was also stained by the monoclonal anti-pre-S1 antibody (lane A), it is clear that these proteins are the expected pre-S protein. Moreover, they all react with antibodies directed against peptides corresponding to the *N*-terminus of pre-S1 (lanes B and C), whereas only the higher- M_r bands were stained with anti-pre-S2 (lanes D and E) antibodies. Therefore the lower- M_r bands lack at least part of the pre-S2 sequence, and the heterogeneity must arise from proteolysis at the C-terminus of the protein. These results are also consistent with the results of Edman degradation presented above.

Polymerized albumin activity of the pre-S protein

The e.l.i.s.a. assay for pHSA binding was performed on cell lysates and samples obtained from the various steps of purification. Table 1 shows the results of such assays on the initial cell lysates and on the final purified protein. The cells containing the pET-3A/pre-S1S2 vector had a very high titre of pHSA-binding activity which was lacking in the control cells. Moreover, this activity copurified with the pre-S protein, clearly demonstrating

Table 1. pHSA-binding activity of the pre-S protein

The protein concentration was calculated by assuming a molar absorption coefficient of $1 \text{ M}^{-1} \cdot \text{cm}^{-1}$ for a 1 mg/ml solution.

Fraction assayed	Protein concn. (mg/ml)	Activity	
		Dilution	A_{492}
Cell lysate*	120	1:1000	1.754
		1:10000	1.048
		1:100000	0.511
Control cells†	120	1:1000	0.024
		1:10000	0.000
		1:100000	0.014
Purified pre-S‡	1.87	1:1000	1.293
		1:10000	0.549
		1:100000	0.524

* Cells containing the pET-3A/pre-S1S2 vector.

† Cells containing only the pET-3A vector.

‡ Purified pre-S protein corresponding to lane E of Fig. 3.

Fig. 4. C.d. spectra of the pre-S protein

The pre-S protein was present at 0.1 mg/ml in 0.05 M-sodium phosphate buffer, pH 6.8. A thermostatically controlled cell was used. The continuous line is the spectrum obtained at 4 °C; the broken line is the spectrum obtained at 37 °C.

that the pre-S protein has the pHSA-binding activity described for the HBsAg and HBV [6,12,14,16]. This assay served as our initial method to evaluate procedures for the purification of the protein, although the final purification method used does not require such assays in order to be able to locate the protein.

C.d. studies of the pre-S protein

The purified pre-S protein is soluble to at least 4–5 mg/ml when kept at 4 °C. However, when allowed to warm up to room temperature, solutions of the protein at concentrations above about 1 mg/ml become cloudy and a precipitate forms. Upon cooling back to 4 °C, the protein dissolves completely. This suggests that the protein undergoes a reversible conformational change which alters its solubility. Therefore, to examine the conformation of the protein, c.d. spectra were recorded. Fig. 4 shows the c.d. spectra of the protein at 4 °C and at 37 °C. In each case the buffer blank has been subtracted from the spectrum. It is evident that the spectra differ. Moreover, this difference was found to be reversible upon lowering the temperature back to 4 °C (results not shown). When the secondary structure was calculated from the spectra using the algorithm of Yang *et al.* [25], the following values were obtained. At 4 °C the protein has 65% β -sheet, 30% random coil and 5% β -turn; no helical structure was observed. At 37 °C the protein has 60% β -sheet, 35% random coil and 5% β -turn. Thus the observed change in dichroicity is consistent with the conversion of about 5% of the β -sheet structure into random coil. Predictions based only on the amino acid sequence and using the method of Garnier *et al.* [26] yield 48.4% β -sheet, 23.3% β -turn, 43% random coil and 1.8% α -helix.

DISCUSSION

The function of the pre-S sequences of the HBsAg in the biology and immunology of the HBV is only beginning to be understood. These sequences have been implicated in tissue- and species-specific binding of the virus to hepatocytes [13,14], they contain highly antigenic T and B cell epitopes which may play an important role in recovery from viral infection or protection from infection [8–10], and they may also be involved in the association of the core particles with the viral envelope during maturation, since virions are found to be rich in the pre-S protein [1,3,19]. The ability to produce and isolate the pre-S protein should enable the investigation of these diverse functions.

The pre-S sequences have been shown to exist completely

external to the HBsAg particle, and presumably to the HBV as well [1,3,5,7,16,19,27]. Thus antibodies to peptides representing segments along the full pre-S sequence bind HBsAg, suggesting that these are external [15]. Furthermore, tryptic digestion has been shown to remove the entire pre-S domain from the particle, and this removal of the pre-S protein does not appear to affect the structure of the remaining HBsAg [15]. These results suggest that the pre-S exists as an independent domain (or domains) and may not depend upon the rest of the S protein for its structural integrity. Therefore pre-S might be expected to retain its native structure and functions in the absence of the S protein and the normal HBsAg-associated lipids.

The results presented here demonstrate that the pre-S protein can be expressed and isolated in a pure form. The isolated protein exists as a monomer and is judged to be a globular protein by its elution from a gel filtration column. The predominant secondary structure is β -sheet, with no measurable α -helix content as determined by c.d. spectroscopy. The results obtained are in reasonable agreement with those predicted from the amino acid sequence. Interestingly, the protein appears to undergo a reversible conformational change upon warming which correlates with an apparent change in solubility. Whether this has any relevance to the proposed function of the protein as a receptor-binding protein is unknown. However, other viral envelope proteins that are involved with internalization of the virus after initial attachment have been shown to undergo conformational changes (reviewed by Marsh & Helenius [27]).

The pre-S protein is susceptible to proteolysis during purification. However, the sites of proteolysis appear to be restricted to the pre-S2 domain. These results are very similar to those obtained in studies of the proteolytic susceptibility of the pre-S domains of naturally occurring HBsAg [5]. In those previous studies we demonstrated that the pre-S2 domain was most susceptible to digestion with a variety of proteinases, including trypsin, chymotrypsin, clostripain, and *Staphylococcus aureus* proteinase, whereas the pre-S1 was relatively more resistant to such attack. These observations further support the conclusion that the pre-S protein produced from the cloned gene accurately mimics the pre-S domains of the natural HBsAg protein, and suggest that, in both, the pre-S1 domain is likely to be more compact, thereby having fewer accessible bonds, whereas the pre-S2 domain is a more open structure. Perhaps it is the pre-S2 protein existing in an extended open conformation that results in the protein having an apparent M_r greater than expected upon gel filtration.

The pre-S domains of natural HBsAg are post-translationally modified. In the M protein, the pre-S2 is acetylated and has both *N*-linked and *O*-linked carbohydrates (D. L. Peterson, unpublished work). In the L protein the *N*-terminal methionine residue is removed and the exposed glycyl residue is myristoylated [7]; however, there are conflicting reports concerning whether the pre-S domains are glycosylated in the L protein [28,29]. In the pre-S protein described here, the only post-translational modification which occurs is the removal of the *N*-terminal methionine. *E. coli* do not myristoylate or acetylate the protein, nor do they glycosylate it. Thus, if the pre-S protein is glycosylated in the natural HBsAg [29], then the cloned material differs by lacking this modification. However, if the pre-S domains are not glycosylated [28], then the pre-S protein accurately reflects this structure. Since the *N*-terminal methionine is removed by the *E. coli*, it would be relatively easy to myristoylate the protein chemically. Thus the role of the myristoyl group in the maintenance of the structure and function of the pre-S could be investigated.

In summary, we have demonstrated that the pre-S protein can be produced and purified in large amounts. The pure protein

retains the antigenic properties of the natural pre-S domains as well as its pHSA-binding activity. It exists as a monomer in solution and is apparently a globular protein. It is expected to be a good model for the study of the structural basis for the various functions of pre-S.

This work was supported in part by grant AI15955 from the National Institutes of Health, and by the A. D. Williams Fund, Virginia Commonwealth University.

REFERENCES

1. Neurath, A. R., Kent, S. B. H., Strick, N., Taylor, P. & Stevens, C. E. (1985) *Nature (London)* **315**, 154–156
2. Peterson, D. L. (1981) *J. Biol. Chem.* **256**, 6975–6983
3. Pfaff, E., Klinkert, M. Q., Theilman, L. & Schaller, H. (1986) *Virology* **148**, 15–22
4. Welch, R. A., Jones, K. R. & Macrina, F. L. (1979) *Plasmid* **2**, 261–268
5. Hu, P. & Peterson, D. L. (1988) in *Viral Hepatitis and Liver Disease* (Zuckerman, A. J., ed.), pp. 318–322, Alan R. Liss, New York
6. Machida, A., Kishimoto, S. & Ohnuma, H. (1983) *Gastroenterology* **85**, 268–274
7. Persing, D. H., Varmus, H. E. & Ganem, D. (1987) *J. Virol.* **61**, 1672–1677
8. Milich, D. R., McLachlan, A., Chisari, F. V., Kent, S. B. H. & Thornton, G. B. (1986) *J. Immunol.* **137**, 315–322
9. Milich, D. R., McLachlan, A., Thornton, G. B. & Chisari, F. V. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 8168–8172
10. Petit, M., Capel, F., Riottot, M. M., Dauguet, C. & Pillot, J. (1987) *J. Gen. Virol.* **68**, 2759–2767
11. Petit, M., Maillard, P., Capel, F. & Pillot, J. (1986) *Mol. Immunol.* **23**, 511–523
12. Imai, M., Yanase, Y. & Nojiri, T. (1979) *Gastroenterology* **76**, 242–247
13. Neurath, A. R., Kent, S. B. H., Strick, N. & Parker, K. (1986) *Cell* **46**, 429–436
14. Pontisso, P., Petit, M., Braukowski, M. J. & Peebles, M. (1989) *J. Virol.* **63**, 1981–1988
15. Hu, P., Carithers, R. L., Fiorenza, V. & Peterson, D. L. (1987) *J. Virol. Methods* **16**, 97–114
16. Okamoto, H., Imai, M. & Usuda, J. (1985) *J. Immunol.* **134**, 1212–1216
17. Rosenberg, A. H., Lade, B. N., Chui, D., Lin, S., Dun, J. N. & Studier, W. (1987) *Gene* **56**, 125–135
18. Macrina, F. L., Tobian, J. A., Jones, K. R., Evans, R. P. & Clewell, D. B. (1982) *Gene* **19**, 345–353
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
20. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021
21. Burnette, W. (1981) *Anal. Biochem.* **112**, 195–203
22. Edelhoch, H. (1967) *Biochemistry* **6**, 1948–1954
23. Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. & Charnay, P. (1979) *Nature (London)* **281**, 646–650
24. Beu-Bassat, A., Bauer, K., Chang, S., Myambo, K., Boosman, A. & Chang, S. (1987) *J. Bacteriol.* **169**, 751–757
25. Yang, J. T., Wu, C. C. & Martinez, H. M. (1986) *Methods Enzymol.* **130**, 208–269
26. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120
27. Marsh, M. & Helenius, A. (1989) *Adv. Virus Res.* **36**, 107–151
28. Heerman, K. H., Goldman, U., Schwartz, W., Seyffarth, T., Baumgarten, H. & Gerlich, W. (1984) *J. Virol.* **52**, 396–402
29. Wong, D. T., Nath, N. & Sninsky, J. J. (1985) *J. Virol.* **55**, 223–231

Received 3 September 1990/28 December 1990; accepted 8 January 1991