JOURNAL OF VIROLOGY, July 1990, p. 3319–3330
0022-538X/90/073319-12$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 64, No. 7

Hepatitis B Virus Nucleocapsid Assembly: Primary Structure Requirements in the Core Protein

FRANK BIRNBAUM AND MICHAEL NASSAL*

Zentrum für Molekulare Biologie Heidelberg, Universität Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, Federal Republic of Germany

Received 28 December 1989/Accepted 28 March 1990

As a step toward understanding the assembly of the hepatitis B virus (HBV) nucleocapsid at a molecular level, we sought to define the primary sequence requirements for assembly of the HBV core protein. This protein can self assemble upon expression in *Escherichia coli*. Applying this system to a series of C-terminally truncated core protein variants, we mapped the C-terminal limit for assembly to the region between amino acid residues 139 and 144. The size of this domain agrees well with the minimum length of RNA virus capsid proteins that fold into an eight-stranded β-barrel structure. The entire Arg-rich C-terminal domain of the HBV core protein is not necessary for assembly. However, the nucleic acid content of particles formed by assembly-competent core protein variants correlates with the presence or absence of this region, as does particle stability. The nucleic acid found in the particles is RNA, between about 100 to some 3,000 nucleotides in length. In particles formed by the full-length protein, the core protein mRNA appears to be enriched over other, cellular RNAs. These data indicate that protein-protein interactions provided by the core protein domain from the N terminus to the region around amino acid 144 are the major factor in HBV capsid assembly, which proceeds without the need for substantial amounts of nucleic acid. The presence of the basic C terminus, however, greatly enhances encapsidation of nucleic acid and appears to make an important contribution to capsid stability via protein-nucleic acid interactions. The observation of low but detectable levels of nucleic acid in particles formed by core protein variants lacking the Arg-rich C terminus suggests the presence of a second nucleic acid-binding motif in the first 144 amino acids of the core protein. Based on these findings, the potential importance of the C-terminal core protein region during assembly in vivo into authentic, replication-competent nucleocapsids is discussed.

The core gene of hepatitis B virus (HBV) (for a review, see reference 7) gives rise to several distinctly different proteins. p21c, consisting of 183 amino acids (5), is the primary translation product which assembles intracytoplasmically into viral capsids (serologically defined as HBcAg). Alternative usage of the initiation codon of the pre-C region, a short in-phase open reading frame directly preceding the C gene, leads to a pre-C/C fusion protein which is sorted into the host cell secretory pathway (21, 29) and, after several processing steps, is found in the serum of HBV-infected individuals as a nonparticulate protein of about 17 kilodaltons (HBeAg) (15). Despite their markedly different antigenic and physical properties, these proteins share a large stretch of identical primary sequence from amino acid positions 1 to 149 (36). In addition, a third, membrane-associated core gene product of unknown primary sequence, which exhibits HBe antigenicity, has recently been described (30).

One of the most striking features in the amino acid sequence of HBV and related mammalian hepadnavirus core proteins is the presence at the C terminus of an extremely Arg-rich region (see Fig. 1B) which, on this basis, has been postulated to have a nucleic acid-binding function (22); nucleic acid binding to C-gene products, possibly independent of this sequence, has been reported after immobilization of the protein on nitrocellulose membranes (17, 23). A structural model (1) for p21, based on a sequence comparison with several spherical RNA viruses of known three-dimensional structure and on serological data about accessible parts in hepatitis B core antigen (HBcAg) (28) suggests a close relationship of the HBV core protein with the capsid proteins of RNA viruses. Many of these contain an amino-terminal basic region which appears to be involved in correct assembly via interaction with nucleic acid (33). HBV cores show the same icosahedral architecture as many RNA viruses, with triangulation number $T = 3$ (C. von Bonsdorff, M. Cyrklaff, F. Birnbaum, M. Nassal, and S. H. Fuller, unpublished data).

Interestingly, HBeAg, the second major core gene product, lacks almost exactly the Arg-rich C-terminal region (see Fig. 1B), which could explain the different properties of this protein, especially its nonparticulate nature. Immunological analysis of HBeAg (28) has indicated that the entire N-terminal sequence, including the region around amino acid 140, is important for full e-antigenicity. Likewise, proteolytic digestion of p21 (3, 8; R. Dalseg, Ph.D. thesis, University of Heidelberg, Heidelberg, Federal Republic of Germany 1990) suggests that the first 140 to 150 amino acids form a tightly folded, almost protease-resistant domain, whereas the basic C terminus is easily clipped off.

On the basis of these data, we chose to study the influence of the C-terminal region on capsid assembly by constructing a series of C-terminal deletion mutants, ending with amino acids R-164, V-149, P-144, I-139, and P-138 (see Fig. 1B). In this series the positively charged tail would be successively removed, and the overall length of some of the proteins would be limited beyond the minimal size (about 150 amino acids [26]) expected for a capsid protein that folds into the eight-stranded β-barrel structure commonly observed for RNA virus capsid proteins. To simplify the analysis, we constructed mutant C genes encoding only core protein-

* Corresponding author.

derived amino acids, since the influence of the newly created terminal regions in previously described core protein fusion variants (2, 18; H. Müller, diploma thesis, University of Heidelberg, 1986) is not easily interpreted. Since HBV and HBV capsids cannot be obtained in large quantity from the serum or liver of infected patients or from transfected animal cells, we took advantage of the fact that the full-length HBV core protein can be efficiently expressed in *Escherichia coli* (19, 22) and purified on the basis of its self-assembly capability, which is retained even in the absence of other viral components. We expected that stable deletion mutants could also be obtained in preparative amounts from this expression system and thus be available for analyzing the contributions of different regions in the core protein to capsid assembly. The various core protein derivatives described here also served as starting material for a structural analysis of the HBV core particle by electron-microscopic methods. The results of this investigation will be separately reported.

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* NF1 (K-12ΔH1Δ*trp* [24]) harboring a defective λ bacteriophage genome providing the temperature-sensitive λ *c*I857 repressor and *E. coli* 537 [C600(pcI587) (25)] were used for expression experiments. Bacteria were grown at 28°C in standard I medium (E. Merck AG, Darmstadt, Federal Republic of Germany) in the presence of 50 μg of ampicillin per ml and 25 μg of kanamycin per ml when appropriate. All expression plasmids were based on plasmid pPLC4-1 (19), which contains a synthetic HBV C gene under control of the phage λ p_L promoter.

Chemicals and enzymes. Chemicals for DNA synthesis were obtained from Applied Biosystems, Weiterstadt, Federal Republic of Germany, and all reagents for sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) were obtained from Serva, Heidelberg, Federal Republic of Germany. Agarose and low-gelling-temperature agarose were purchased from FMC Corp., Marine Colloids Div., Rockland, Maine. T4 DNA ligase and restriction enzymes were obtained from Boehringer GmbH, Mannheim, Federal Republic of Germany, and used as recommended by the manufacturer.

Synthesis of oligodeoxynucleotides. All oligonucleotides were synthesized on an Applied Biosystems 380B synthesizer and purified as previously described (20).

Construction of deletion mutants. Core protein variant 1-149 was generated by replacing the *Sal*I (nucleotide [nt] 420)-*Hin*dIII (nt 555) restriction fragment in plasmid pPLC4-1 (19) by a synthetic oligonucleotide duplex deleting all codons after position 149. The mutant C gene 1-139 was constructed similarly by replacement of the *Bsm*I (nt 405)-*Sal*I (nt 420) restriction fragment with a synthetic duplex introducing a stop codon (TAG) at codon 140. Variants 1-138, 1-144, and 1-164 were created by changing codons I-139 (ATC to TAG), E-145 (GAG to TAG), and R-165 (AGA to TGA) into stop codons via the mismatched-primer method (38) by using an authentic HBV genome (5) cloned in M13mp11 as the template. As mutagenic primers, synthetic 18-mers were used which, except for the desired nucleotide exchanges, corresponded to the authentic HBV sequence. Mutant C genes generated in M13 were transferred as *Taq*I (nt 8)-*Ava*I (nt 533) fragments into plasmid pPLC4-1 digested with *Cla*I and *Ava*I (Fig. 1A). All mutations were confirmed by direct sequencing of the expression plasmids with Sequenase (United States Biochemical Corp., Cleveland, Ohio) and appropriately located constituent oligonucleotides of the synthetic C gene (19) as sequencing primers.

Expression and purification of mutant C-gene products. *E. coli* NF1 or 537 cells were transformed with the pPLC plasmids carrying individual mutant C genes. Overnight cultures grown at 28°C were diluted and induced by shifting the temperature to 42°C. After 2 h at this temperature, cell lysates were obtained as previously described (19). Briefly, cells were lysed with lysozyme–Triton X-100–EDTA. After the lysates had been clarified by low-speed centrifugation, core proteins were precipitated with ammonium sulfate (40% saturation), and the resulting pellets were washed with 30% saturated ammonium sulfate in 50 mM morpholinepropanesulfonic acid (MOPS) buffer (pH 7.0). The pellets were suspended in MOPS buffer and subjected to sedimentation in a 10 to 60% sucrose gradient (SW28 rotor, 25,000 rpm, 16 h, 4°C). Fractions were monitored for core proteins by direct electrophoresis of aliquots on 0.1% SDS–15% polyacrylamide gels (12). Peak fractions were pooled, dialyzed overnight against 50 mM MOPS buffer (pH 7), and subjected to sedimentation on a similar gradient (SW40 rotor, 25,000 rpm, 2 h, 4°C) (see Fig. 2). Thereafter, the material was usually nearly homogeneous. For purification of core proteins obtained from cultures larger than 1 liter, a third sedimentation under similar conditions removed most of the contaminating *E. coli* material. The variant core proteins were designated according to the position of their C-terminal amino acid (e.g., 1-164 is variant 164). For an approximate determination of their sedimentation coefficients, a mixture of the various particles was analyzed on a linear 5 to 20% sucrose gradient by using ovalbumin (3.5S), thyroglobulin (19.4S), and coliphage fr (82S) as markers.

Agarose gel electrophoresis of core particles and RNA. Aliquots of sucrose density gradient fractions, or purified particles, were analyzed on 1 or 1.4% agarose gels in TAE buffer (16) containing 0.5 μg of ethidium bromide per ml. Nucleic acid was detected by illumination with UV light. Protein was stained with Coomassie blue R (Serva). Immunological detection of HBV core protein was performed after capillary transfer from the agarose gel to nitrocellulose (Schleicher & Schüll, Dassel, Federal Republic of Germany) with the core-specific monoclonal antibody c-312 (28). Bound antibody was determined by using alkaline phosphatase-coupled anti-mouse immunoglobulin G and 5-bromo-4-chloro-3-indolyl phosphate–*p*-Nitro-Blue Tetrazolium chloride (Biomol, Mannheim, Federal Republic of Germany) as the substrate. *E. coli* total RNA was isolated from uninduced or induced cultures by the hot-phenol method (16).

For nuclease protection analysis, particles purified by twofold sedimentation on sucrose gradients (10 to 15 μg of protein) were mixed with 6 μg of *E. coli* 16S plus 23S RNA (Pharmacia, Freiburg, Federal Republic of Germany) and 10 μg of a 5.5-kilobase plasmid in a total volume of 100 μl. One third was incubated with 1 μg of RNase A for 45 min at room temperature; one third was incubated with 1 U of DNase, after addition of $MgCl_2$ to 10 mM and dithiothreitol (DTT) to 1 mM, and incubated for 45 min at 37°C. Reaction products were analyzed on a nondenaturing 1.2% agarose gel. To digest all nonprotected RNA, the gel was soaked overnight in TE buffer (16) containing 20 μg of RNase A per ml.

Nucleic acid encapsidated in core particles was isolated by incubating the desired amount of sucrose gradient-purified cores at a concentration of about 0.1 to 0.2 mg/ml with 10 μg of proteinase K per ml in the presence of 1% SDS for 30 min

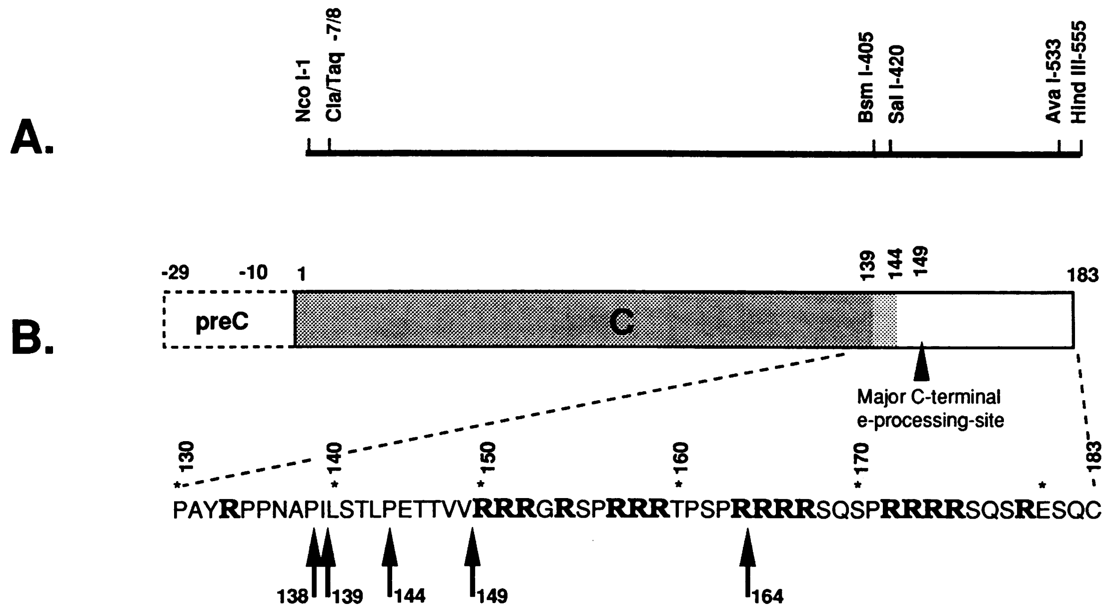

FIG. 1. Schematic view of the HBV C gene and the core protein. (A) Restriction sites in the C gene (19) relevant for the construction of 3'-terminal deletions are indicated. Numbers refer to the nucleotide position in the HBV C gene and follow the numbering system of Pasek et al. (22). (B) Numbers on top refer to amino acid positions in the core protein; the dashed box represents the 29 amino acids encoded by the pre-C region which is required for HBeAg secretion (21, 29); the major C-terminal HBeAg processing site is indicated at amino acid position 149 (36). The shaded area represents the domain essential for assembly proficiency, the C-terminal border of which (lighter shading) has been confined to the sequence between amino acid positions 140 to 144 (see text). The region from amino acid positions 130 to 183, including the Arg-rich tail, is shown in detail by using the single-letter amino acid code; Arg residues are in boldface type. The upward arrows indicate the location of the C-terminal amino acid in the truncated variants.

at 37°C. Subsequently, proteins were removed by extraction with phenol, and nucleic acid was precipitated with sodium acetate-ethanol. Pellets were washed in 70% ethanol and dissolved in diethylpyrocarbonate-treated water. Northern (RNA) blotting (10) was performed with two of the oligonucleotides constituting the synthetic C gene (no. 8, antisense orientation; no. 17, sense orientation [19]) as probes. The oligonucleotide with antisense orientation differs at 27 of 50 positions from the corresponding region in the woodchuck hepatitis B virus (WHV) C gene. Hybridizations were performed at room temperature in hybridization mix containing 50% formamide, followed by several washes at room temperature with 2×, 1×, and 0.5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) (10). RNA size markers were obtained from GIBCO BRL, Eggenstein, Federal Republic of Germany.

Stability of particles. Aliquots from sucrose gradient-purified core preparations (approximately 0.1 to 0.2 mg/ml) were incubated in 1% SDS–100 mM DTT (final concentration) for 15 min at room temperature, then serially diluted (1:5; starting at 1 μg/100 μl per well) into phosphate-buffered saline containing 2% bovine serum albumin–0.05% Nonidet P-40, and immediately analyzed for HBc or HBe antigenicity by enzyme-linked immunosorbent assay (ELISA) (see below).

Particle stability toward high pH was analyzed by diluting aliquots of particles from variants 144, 149, and 165 and the full-length protein obtained after two sucrose gradients (5 to 10 μg of core protein) into 1 ml of 50 mM MOPS buffer (pH 7). To this solution was added 1 M NaOH in an amount which had been shown to be sufficient to produce a pH 10 or 12 in the identical buffer without protein. After incubation at 37°C for 3 h, 1 ml of the reaction mixture was loaded onto a sucrose step gradient consisting of 200 μl of 10% (wt/wt) sucrose and 200 μl of 60% (wt/wt) sucrose and centrifuged in a Beckman TLS-55 rotor at 55,000 rpm for 1 h at 20°C. Gradients were fractionated from the top, and individual fractions were analyzed for core protein by SDS-PAGE and Western immunoblotting with rabbit anti-c/e antiserum.

ELISA. An HBcAg- or HBeAg-specific ELISA was set up that uses a polyclonal rabbit anti-c/e antiserum bound to the ELISA plate which recognizes all known forms of C-gene products. After incubation with the antigen, HBc-specific (c-275 or c-312 [28]) or HBe-specific (monoclonal antibody 03 or 158; M. Noah, unpublished results) monoclonal antibody was used to sandwich the antigen. Bound mouse monoclonal antibody was detected by peroxidase-coupled rabbit anti-mouse immunoglobulin G with 2,2'-azino-bis (3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS; Boehringer) as the chromogenic substrate.

RESULTS

Expression of HBV core protein variants in *E. coli*; primary sequence requirements for assembly. The constructs encoding the variously truncated core proteins were obtained by introducing stop codons at the positions following the desired C-terminal amino acid residue, either by restriction fragment replacement in the synthetic (19) C gene (*Bsm*I to *Sal*I for variant 139) or by use of mismatched primers (38) on a complete HBV genome cloned into M13mp11 as template (for variants 138, 144, and 164). For variant 149, the fragment between the *Sal*I and *Hin*dIII sites in the synthetic C gene was replaced by a synthetic duplex such that the coding region from codons 150 to 183 was deleted (Fig. 1A). The expected C-terminal sequences in the core protein variants are indicated in the lower part of Fig. 1B.

For expression in *E. coli*, the mutated C genes were

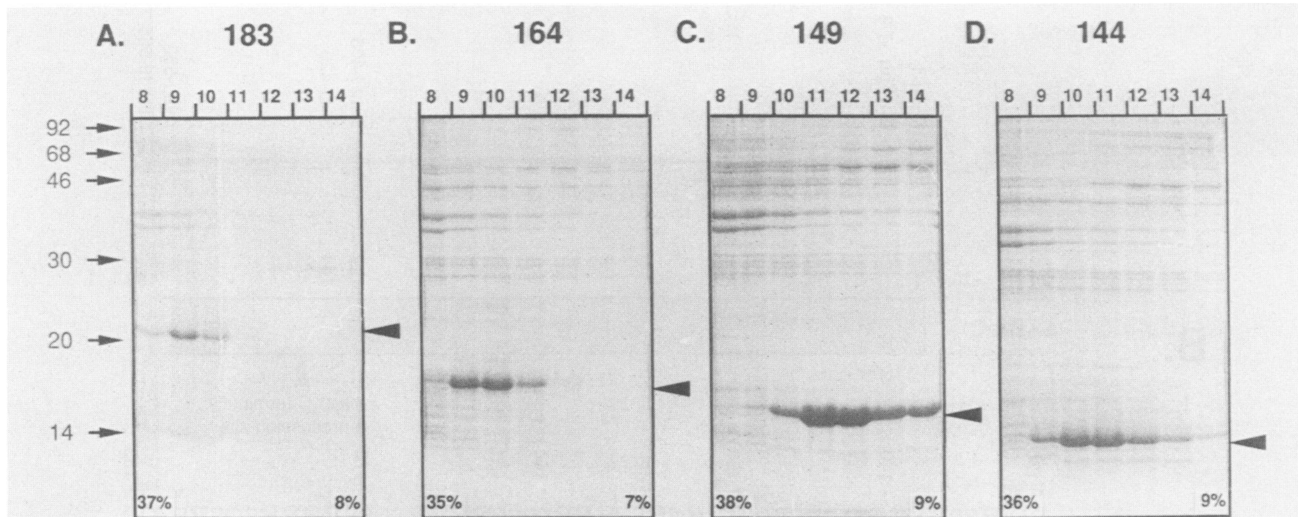

FIG. 2. Partial purification of particle-forming full-length and C-terminally truncated HBV core proteins expressed in *E. coli*. Core proteins 183 (A), 164 (B), 149 (C), and 144 (D), partially purified by ammonium sulfate precipitation and sedimentation on a first sucrose gradient (see Materials and Methods), were subjected to a second centrifugation on a 10 to 60% sucrose gradient (SW28 rotor, 2 h, 27,000 rpm, 4°C). Gradients were fractionated from the bottom (fraction 1) to the top (fraction 20). Aliquots were analyzed by SDS-PAGE. Only the region of the gradients where particles would be expected is shown. Numbers on top of each gel indicate individual fractions; numbers on the left show the positions in the gel of marker proteins of the indicated sizes (in kilodaltons). Numbers in the lower corners of each gel refer to the sucrose concentrations in the first and last fractions shown. The arrowheads highlight the positions of individual core protein variants. The decrease in sedimentation velocity visible from panels A to D was confirmed by determining the approximate S values of the individual particles (see Results).

inserted into plasmid pPLC4-1, which had been previously shown to efficiently direct the synthesis of the full-length core protein (19). Crude lysates from induced *E. coli* cultures harboring the corresponding expression plasmids were analyzed by SDS-PAGE and Western blotting. All proteins were synthesized in amounts visible by Coomassie blue staining and showed the expected differences in electrophoretic mobility (data not shown). It should be noted that variant 149 was expressed reproducibly better than any of the other variants or the full-length protein.

To investigate the influence of the C-terminal truncations on particle formation, the proteins were partially purified by ammonium sulfate precipitation, and subsequently subjected to sedimentation on sucrose gradients (see Materials and Methods). Variants containing 144 or more N-terminal amino acids sedimented, at least in part, with a velocity indicative of their particulate nature (Fig. 2). Direct evidence for particle formation was obtained by electron microscopic analysis of the fast-sedimenting peak fractions (von Bonsdorff et al., unpublished). Core proteins 144 and 149 did not yield clearly resolved peaks of particulate and nonparticulate material, but rather showed a broad distribution indicating an intrinsic instability of the particles formed by these mutant proteins (data not shown).

In contrast, when the lysates from *E. coli* cells producing variants 138 (Fig. 3) or 139 (data not shown) were fractionated by sedimentation and analyzed by Western blotting, the proteins were found exclusively in the upper region of the gradient, indicating that no significant numbers of particles had been formed. When ELISA was used to monitor core protein in individual fractions, however, small amounts of material exhibiting HBc antigenicity could be detected in the

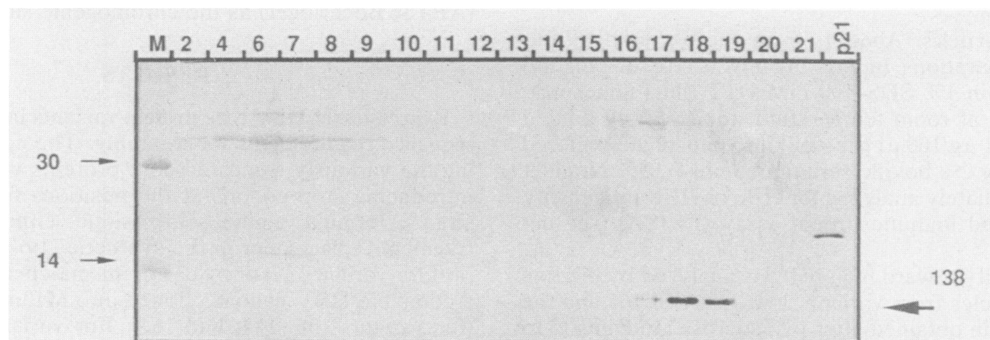

FIG. 3. Core protein variant 138 does not form stable particles. The lysate from an *E. coli* culture expressing variant 138 was loaded on a 10 to 60% sucrose gradient (see Materials and Methods) and fractionated as described in the legend to Fig. 2. Aliquots were analyzed by immunoblotting with a polyclonal rabbit anti-c/e antiserum. The only strongly reacting bands, besides several weakly stained background bands, are variant 138 in fractions 17 to 19 near the top of the gradient and the full-length core protein (p21) included as marker. M, Molecular mass standards; the 14- and 30-kilodalton marker proteins are unspecifically stained during immunoblotting. Core protein variant 139, in a separate experiment, was found at the corresponding region in the gradient.

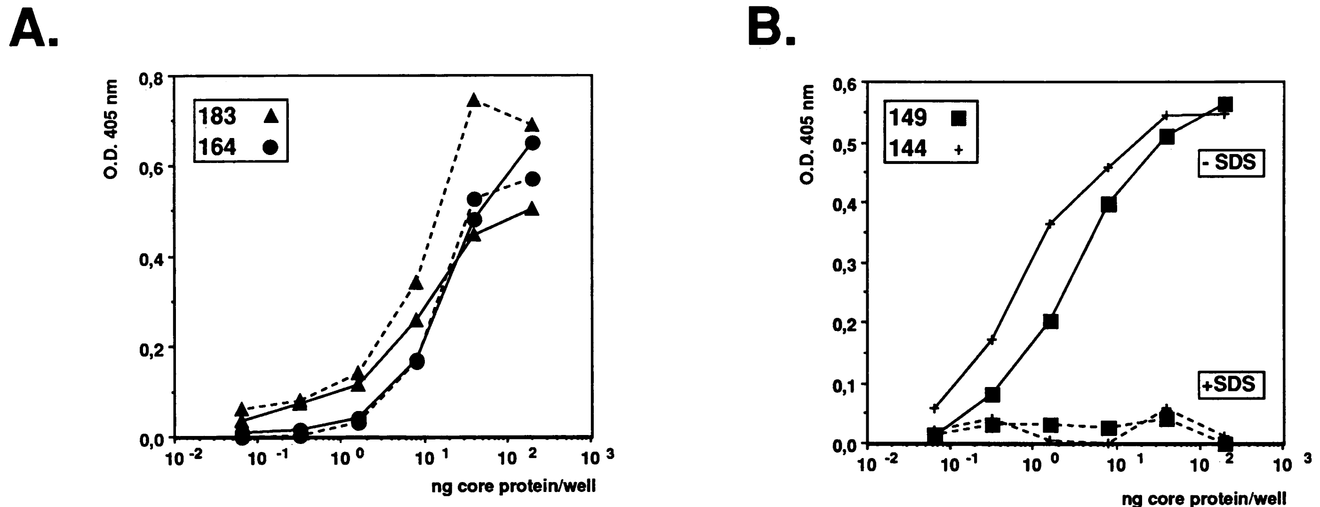

FIG. 4. HBc antigenicity of particles from full-length and C-terminally truncated core protein variants is differentially influenced by treatment with SDS-DTT. Equal amounts of purified particles from full-length and variant core proteins were bound to immobilized polyclonal rabbit anti-c/e antiserum either directly (———) or after treatment with 0.1% SDS–100 mM DTT (--------), and their antigenicities were analyzed by ELISA with HBcAg-specific mouse monoclonal antibody c-275 (28) as described in Materials and Methods. (A) Particles from full-length core protein and variant 164 retain their c antigenicity upon SDS-DTT treatment. (B) Particles from variants 149 and 144 lose c antigenicity completely. The same result was obtained with HBc-specific monoclonal antibody c-312 (28). O.D., Optical density.

lower fractions of the gradient. The amount and purity of these samples were insufficient for direct electron-microscopic observation. The stable particles formed by variants containing at least 144 amino acids, in contrast, could be analyzed by cryoelectron microscopy, and were shown to consist of 180 subunits, like those formed by the full-length protein (von Bonsdorff et al., unpublished).

Stability of particles formed by core protein variants. HBV nucleocapsids exhibit a distinctly different antigenicity (HBcAg) from HBeAg, the second, nonparticulate gene product of the C gene. Since HBcAg can be converted, by denaturation, to "HBeAg" (14), the presence or absence of HBc- or HBe-specific epitopes may be used as a measure of particle integrity. A sandwich ELISA based on an immobilized polyclonal antiserum directed against both HBcAg and HBeAg and different soluble mouse monoclonal antibodies with either HBcAg or HBeAg specificities was used to determine the antigenic properties of the various particles. The full-length protein, isolated from *E. coli* as described above, gave a response which was 2 to 3 orders of magnitude higher with HBc-specific monoclonal antibodies c-312 and c-275 (28) than with HBe-specific monoclonal antibodies 158 and 03, indicating that we had isolated largely intact particles. Upon treatment with 0.1% SDS–100 mM DTT, a corresponding increase in the HBe response was detected (data not shown), in agreement with earlier reports (14). Variant 164 showed similar behavior, whereas variants 149 and 144 gave comparable HBc and HBe responses in their native form. Such dual antigenicity has recently been described for core protein derivatives similar to variant 144, which contained foreign sequences at their N and C termini (18, 34). Analysis of the HBc response upon treatment with SDS-DTT revealed a distinct difference between the full-length protein and variant 164 on the one hand and the shorter variants on the other. Both longer proteins retained their HBc antigenicity after exposure to SDS-DTT, whereas variants 149 and 144 lost it completely (Fig. 4A). These data demonstrate that epitopes recognized by the HBc-specific antibodies are able to survive partial denaturation in particles from the longer but not from the further truncated core proteins, thereby suggesting that particle stability toward SDS is correlated with the length of the core protein and/or the nucleic acid content of the particles (see below).

Direct proof for the differential stabilities of the particles toward a different denaturing agent was obtained by high-pH treatment of a mixture of purified particles from the full-length protein and its truncated variants and subsequent analysis of their sedimentation behavior. The reaction mixture was layered on top of two-step sucrose gradients consisting of a 10 and a 60% sucrose cushion. After being subjected to pH 10 conditions, a large fraction of variants 144 and 149 did not sediment through the 10% sucrose layer into the 60% cushion under the given centrifugation conditions. Full-length protein and variant 164, in contrast, were found exclusively in the bottom fractions of the gradient (Fig. 5A). After exposure to pH 12, variants 144 and 149 were confined almost exclusively to the upper regions of the gradient; a substantial fraction of variant 164 was converted into slow-sedimenting material, whereas particles from the full-length protein were still stable even under these harsh conditions (Fig. 5B). As controls, aliquots of the individual particle mixtures before high-pH treatment were subjected to sedimentation in identical step gradients run in parallel (data not shown). As noted above, particles from variants 144 and 149 appeared inherently unstable, in that even without exposure to high pH, a fraction of these proteins were not fast sedimenting; high-pH treatment markedly increased this fraction. It should be noted that all protein samples in these experiments had been obtained as the fastest-sedimenting, i.e., particulate, peaks on the preparative sucrose gradients. These results indicate again that particle stability correlates with the presence or absence of the basic C-terminal tail.

The approximate sedimentation coefficients of the particles formed by the assembly-proficient core protein variants were determined by sedimentation of a mixture of purified particles on a linear sucrose gradient with ovalbumin, thyroglobulin and coliphage fr as markers. Variants 144 and 149

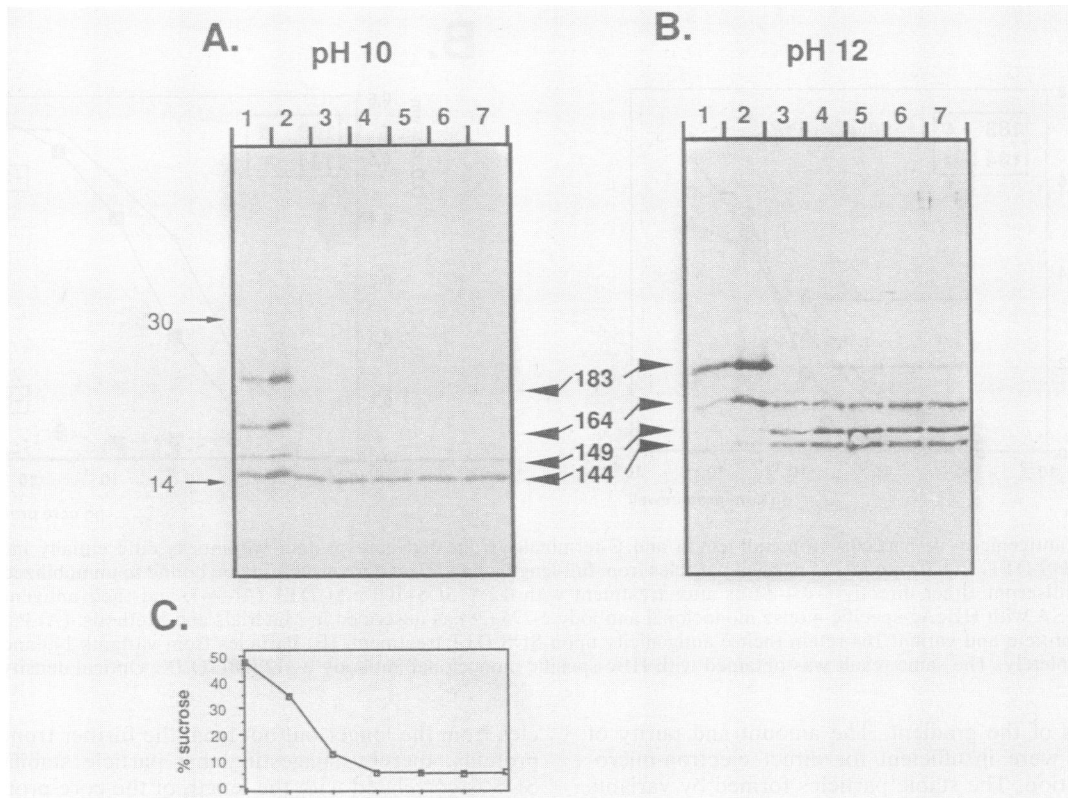

FIG. 5. Particle stability at high pH correlates with the presence or absence of the Arg-rich C terminus. A mixture of purified particles from full-length and assembly-proficient core protein variants was adjusted to pH 10 or 12 as described in Materials and Methods. Reaction products were subjected to sedimentation on a two-step sucrose gradient. Core proteins were detected by immunoblotting. (A) Sedimentation behavior of particle mixture after pH 10 treatment. (B) Particle mixture after pH 12 treatment. (C) Sucrose concentration profile in the seven fractions of the gradient shown in panel A (fraction 1, top; fraction 7, bottom). The same profile was obtained for the gradient shown in panel B. Arrowheads indicate the positions of the full-length and variant core proteins in the gel. Note that after pH 10 treatment, a large proportion of core protein variants 144 and 149 does not sediment into the more highly concentrated sucrose layers, whereas the full-length protein is found exclusively in the bottom fractions even after exposure to pH 12. As controls, aliquots of the particle mixtures before exposure to high pH were run in parallel gradients (results not shown).

sedimented at 55S to 60S, in agreement with an empirical formula derived for the sedimentation behavior of ordinary proteins (37). Variant 164 cosedimented with coliphage fr, allowing its sedimentation coefficient to be determined rather exactly as 82S; the approximate s value for the full-length protein was about 110S. These two s values are distinctly higher than would be calculated for pure protein aggregates of 180 subunits, and they are due to the nucleic acid encapsidated in these particles (see below).

Nucleic acid content of particles from full-length and variant core proteins. Sucrose gradient fractions were also analyzed by gel electrophoresis in ethidium bromide-containing agarose gels to determine whether the particles isolated from *E. coli* contained nucleic acid. In addition to the expected smear of differently sized breakdown products of cellular nucleic acids generated during the sonication of the cell lysates, a distinct band of ethidium bromide staining material was seen in fractions of the gradient which also contained core protein as detected on protein gels. To prove that this band indeed represented core particles, the agarose gel was blotted onto nitrocellulose and the membrane was stained for core protein by a procedure identical to a conventional Western blot with monoclonal antibody c-312 (28), which is directed against the core protein. This blot showed clearly that ethidium bromide and immunostaining detected the identical band on the gel, indicating that the intercalating dye bound to nucleic acid which is either strongly associated with the core particles or possibly encapsidated during assembly (see below). Qualitatively the same results were obtained with all assembly-proficient core protein variants.

The availability of the agarose gel assay allowed us to estimate the relative nucleic acid content of the different particles by using a dilution series of particles from the full-length protein for calibration (Fig. 6A). As a control, the same gel was stained for protein with Coomassie blue (Fig. 6B). Whereas particles from variant 164 showed a similar or slightly decreased nucleic acid/protein ratio compared with those formed by the full-length core protein, the nucleic acid content of the truncated core protein particles appeared to be 10- to 30-fold lower (Fig. 6A, compare lanes 7 and 8 with lanes 3 and 4). The nucleic acid/protein ratio in the different particles could be approximated by using the classical Warburg-Christian method (13). The measured ratios of A_{280}/A_{260} (0.63 and 0.71 for two independent preparations of the full-length protein, and 1.38 for variant 149) would indicate nucleic acid percentages of about 10 to 15% for the full-length protein and 0.9% for variant 149. Although a more accurate determination would require knowledge of the individual extinction coefficients of the protein and the

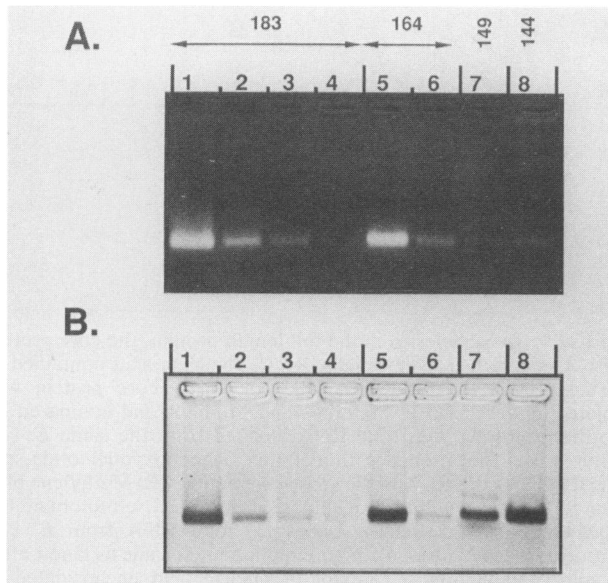

FIG. 6. Nucleic acid content of core particles correlates with the presence or absence of the positively charged C terminus. Aliquots of purified full-length and truncated core proteins were analyzed on a 1.0% agarose gel containing ethidium bromide; after detection of nucleic acid (A) the gel was stained for protein with Coomassie blue (B). Lanes: 1 to 4, serial dilution corresponding to approximately 10, 3, 1, and 0.3 μg of full-length protein per lane; 5 and 6, 10 and 1 μg, respectively, of variant 164; 7, 10 μg of core protein 149; 8, 10 μg of core protein 144. Note that the amount of nucleic acid in lanes 7 and 8 corresponds approximately to that in lanes 3 and 4; i.e., the nucleic acid/protein ratio in particles from variants 144 and 149 is 10- to 30-fold lower than in those from the full-length core protein or variant 1-164. Note also that all particles show a very similar mobility in the gel.

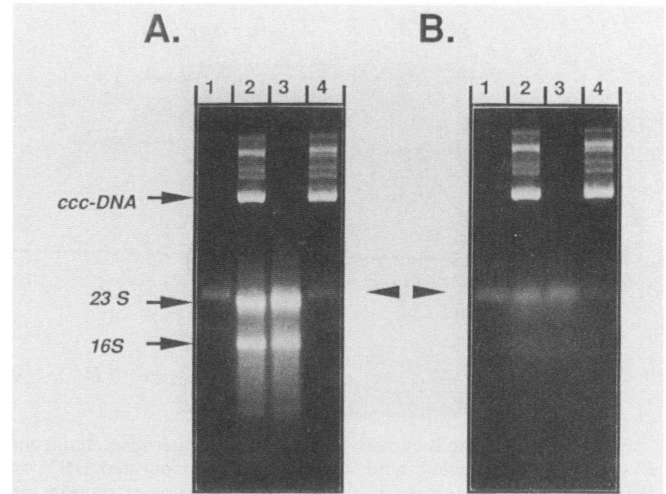

FIG. 7. Nucleic acid associated with core particles is encapsidated. Core particles from the full-length protein were mixed with *E. coli* 23S and 16S rRNA and a 5.5-kilobase plasmid DNA and then incubated with either RNase A or DNase I as described in Materials and Methods. Reaction products were analyzed on a 1.2% nondenaturing agarose gel containing ethidium bromide. The arrows indicate the positions of 16S and 23S rRNA; ccc-DNA indicates the position of the covalently closed circular form of the plasmid DNA; arrowheads point to the band corresponding to core particles. (A) Lanes: 1, purified core particles; 2, mixture of core particles, 16S plus 23S rRNA, and plasmid, no enzyme added; 3, as in lane 2 after DNase digestion; 4, as in lane 2 after RNase treatment. (B) Same gel as in panel A, but after overnight incubation in RNase A-containing buffer. Note that almost all of the added rRNA has disappeared, whereas the core particle-associated nucleic acid is still visible. The slightly higher intensity of the core particle-related band in lanes 2 and 3 compared with lane 1 is due to some residual rRNA.

nucleic acid, these values agree reasonably well with the staining pattern shown in Fig. 6.

These results show that the amount of nucleic acid associated with or encapsidated in the particles depends largely on the presence of at least part of the positively charged C-terminal region. Small amounts of nucleic acid, however, are also found with the particles formed by further truncated core protein variants.

The observation of a very similar mobility of particles from full-length and C-terminally truncated core proteins during agarose gel electrophoresis (Fig. 6) suggests that this property is not markedly influenced by the Arg-rich C terminus. Given a similar diameter for all particles (von Bonsdorff et al., unpublished), this result indicates that the positive charges of the C terminus do not contribute in a major way to the surface charge of the particle.

The nucleic acid associated with core particles produced in *E. coli* is protected from degradation by nuclease and is therefore encapsidated. To determine whether the nucleic acid associated with the particulate core proteins is encapsidated, purified particles from the full-length protein were mixed with excess 16S and 23S RNA from *E. coli* and with plasmid DNA and then treated with RNase and DNase. The reaction products were analyzed in the agarose gel assay (Fig. 7) described above. Whereas the exogenous nucleic acids were digested in the expected way, the nucleic acid associated with the core protein was protected from the enzymes (a slight decrease in ethidium staining was observed for the RNase-treated samples). We conclude that this nucleic acid is encapsidated inside the core particles. The same results were obtained with particles from core protein variants 164 and 149 (data not shown).

Core particles produced in *E. coli* contain RNA. The nature of the encapsidated nucleic acid was determined by digesting the capsid protein with proteinase K and subsequent phenol extraction. When the liberated nucleic acid was analyzed on a formaldehyde-containing agarose gel (Fig. 8, lane 1), a rather broad distribution of ethidium bromide-stainable material was seen with a mobility ranging from about 100 to about 3,000 nt (compared with the RNA size markers M1 and M2). However, within the smear a prominent band of about 830 nt and two larger species (around 1,600 and 2,000 nt) were clearly visible. The nucleic acid is sensitive only to RNase (Fig. 8, lane 4), but not to DNase (lane 3), and therefore represents RNA. In several independent experiments, we have found no evidence for packaging of DNA, although the isolated core protein, when immobilized on nitrocellulose membranes, binds better to DNA than to RNA (T. W. Hatton and D. N. Standring, personal communication). Since the mRNA transcribed from our expression plasmid pPLC4-1 is expected to have a size of about 840 nt (a 246-nt 5′ leader comprising λ p_L and MS2 replicase gene sequences followed by the 560-nt synthetic C gene and about 40 nt of a synthetic transcription termination signal), we suspected this band to represent the core gene mRNA. An apparently even higher enrichment of a single RNA of about the same size was seen when the nucleic acid isolated from WHV cores expressed in *E. coli* was analyzed on the same

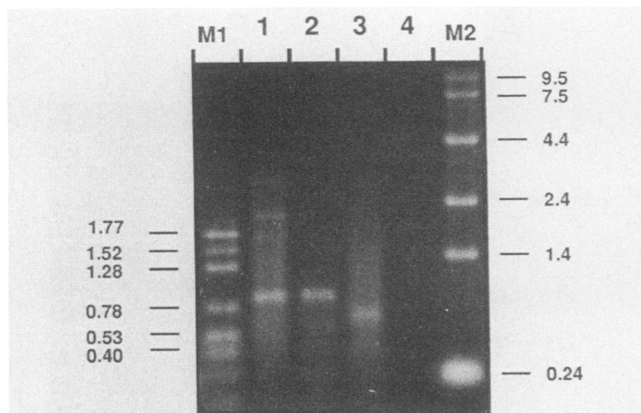

FIG. 8. Nucleic acid encapsidated in core particles isolated from *E. coli* is RNA. Nucleic acid was isolated from purified HBV or WHV cores by proteinase K digestion and phenol extraction, as described in Materials and Methods, and analyzed on a 1.2% denaturing agarose gel. Lanes: M1, RNA size standards of the lengths indicated on the left of the gel; M2, RNA size standards as indicated on the right of the gel; 1, nucleic acid isolated from HBV cores; 2, nucleic acid isolated from WHV cores; 3, same nucleic acid as in lane 1, but treated with DNase; 4, same nucleic acid as in lane 1, but treated with RNase. The apparent mobility shift of the major encapsidated nucleic acid after DNase treatment (lane 3) might be due to some minor RNase contamination.

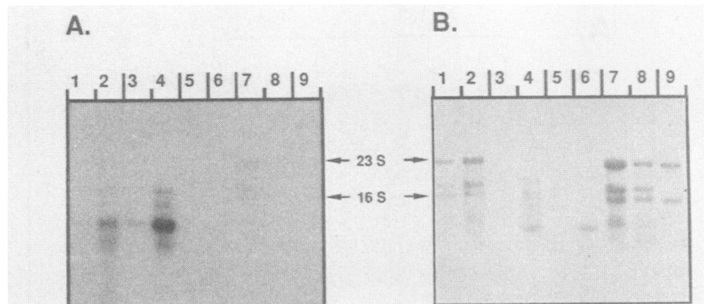

FIG. 9. In particles from the full-length protein, the core protein mRNA is enriched over cellular RNAs. Nucleic acid contained in particles from the full-length HBV or WHV core protein was isolated as described in Materials and Methods and compared by Northern blotting with total RNA isolated from the same *E. coli* cultures. (A) Blot probed with a 50-mer oligodeoxynucleotide specific for HBV core RNA in the sense orientation. (B) Methylene blue stain for nucleic acid of nitrocellulose membrane used to obtain the Northern blot in panel A. Lanes: 1, total RNA from *E. coli* harboring plasmid pPLC4-1 before induction; 2, same as lane 1 after 30 min of induction; 3, aliquot of nucleic acid encapsidated in approximately 2 μg of HBV core protein; 4, as in lane 3, but with 20 μg of core protein; 5, RNA size markers; 6, aliquot of nucleic acid contained in approximately 20 μg of WHV core protein; 7, total RNA from an induced *E. coli* culture harboring a plasmid corresponding to pPLC4-1, in which the HBV C gene is substituted by the WHV C gene; 8, as in lane 7, but before induction; 9, *E. coli* 23S and 16S rRNA. When an oligonucleotide of the opposite orientation was used as the probe, no specific hybridization could be detected.

gel (Fig. 8, lane 2). The expression plasmid, except that it harbors the WHV rather than the HBV C gene (M. Nassal, unpublished results), is otherwise identical with pPLC4-1, and an RNA of similar length should be produced upon induction.

The C-gene mRNA is enriched over other RNA species in core particles. To investigate whether the major individual RNA species inside the core particles was indeed the core protein mRNA, the RNA encapsidated in HBV cores was analyzed on a Northern blot with a synthetic oligonucleotide (19) as probe which is complementary to the HBV core protein mRNA. This oligonucleotide differs at 27 of 50 positions from the WHc sequence. As controls, we used the RNA isolated from WHV cores expressed in *E. coli* and total RNA from induced and uninduced *E. coli* cultures harboring the corresponding HBc and WHc expression plasmids.

With this probe, a very similar pattern of RNAs was detected in the nucleic acid isolated from HBV core particles (Fig. 9A, lanes 3 and 4) and in the total RNA from the induced culture harboring the HBV core protein expression vector (Fig. 9A, lane 2). However, the uninduced culture (lane 1) and all samples from the WHV core protein-expressing bacteria (lanes 6 to 8) were negative. Subsequent staining of the nitrocellulose membrane for total nucleic acid with methylene blue (Fig. 9B) showed that it was indeed the major individual RNA species encapsidated in the HBV cores (Fig. 9B, lane 4) that hybridized to the probe. Specificity of hybridization is demonstrated by the fact that the nucleic acid isolated from the WHV cores (Fig. 9B, lane 6), although present in similar amounts to the nucleic acid from the HBV cores (Fig. 9B, lane 4), did not hybridize to the HBV-specific probe. An oligonucleotide probe recognizing the HBV core gene sequence in the antisense orientation did not detect any band, whereas a WHc-specific probe hybridized to the RNA from WHV cores (data not shown). These data demonstrate that (i) the bands recognized by the "sense" probe depend on the presence of the correct expression plasmid and on the induction of transcription (i.e., they are not of *E. coli* origin) and (ii) the pattern of core-specific RNA packaged in the core particles is very similar to that found in the total RNA from the induced culture harboring the correct plasmid. We conclude that the major single species of nucleic acid found inside the bacterially expressed HBV cores is the core mRNA.

DISCUSSION

The assembly of viral nucleocapsids is still only poorly understood. The HBV core particle provides an attractive model to study capsid assembly, for several reasons: (i) hepadnaviruses are enveloped, in contrast to most other viruses whose assembly has been studied so far; (ii) the hepadnavirus core is one of a few capsids which, depending on the stage in the virus life cycle, can contain RNA (the viral pregenome) and DNA (the partially double-stranded DNA genome in mature virions); and (iii) hepadnavirus cores are built from a single species of capsid protein which, although not easily available from the native source, is accessible in preparative amounts by heterologous expression and does not require the presence of other viral components for assembly.

Since the purpose of any virus capsid is to provide a protective container for the viral genome, capsid proteins are expected to exhibit some potential for direct (basic or other polar amino acid side chains) or indirect (encapsidated polyamines) interaction with nucleic acid (4). This interaction may be important in nucleation of capsid assembly, but could also serve a structural purpose in defining the architecture (i.e., arrangement of subunits) of the capsid (4, 33). Moreover, a mechanism ensuring selective encapsidation of

the genomic nucleic acid, and possibly further viral components, is required for the formation of replication-competent viral nucleocapsids.

One very prominent feature in the sequence of the HBV and other hepadnavirus core proteins is the presence of an extremely basic, Arg-rich region comprising the C-terminal part of the protein. On the basis of its unusual composition, this sequence has been postulated to possess nucleic acid-binding capability. The sequence, in fact, contains several copies of the motif SPRR, which is also found in some histone proteins (35), where it appears to provide a rather tight binding to nucleic acid.

To investigate the influence of this region on capsid assembly without any perturbation by foreign sequences fused to the core protein (18, 19), we constructed mutant C genes encoding solely HBV core protein-derived sequences. All of the variant genes were well expressed in *E. coli*; therefore, we could analyze the proteins for their capability to self assemble and could compare the stability and nucleic acid contents of the particles formed by the variant proteins with those from the full-length protein.

Our data show that (i) the C terminus of the HBV core protein from amino acid 144 on, and therefore the entire Arg-rich region, is dispensable for core assembly, whereas truncations beyond amino acid 140 are not tolerated; (ii) the presence or absence of the Arg-rich tail does not markedly influence the electrophoretic mobility of the particles; (iii) the bulk of nucleic acid binding in *E. coli*, i.e., in the absence of the correct genomic RNA and HBV P-gene products, is correlated with the presence of at least part of the highly basic C terminus; (iv) the packaged nucleic acid is RNA, ranging from about 100 to 3,000 nt in length (in packaging, the full-length core protein seems to exhibit some selectivity for its own mRNA); (v) particles formed by core protein variants completely lacking the basic C terminus still contain nucleic acid, although the level is about 20-fold lower than in particles from full-length protein or variant 164; and (vi) particle stability correlates with the presence or absence of the Arg-rich region and with the nucleic acid content of the particle.

Primary sequence requirements for assembly: structural implications. The results described above indicate that the assembly capability of the HBV core protein resides in its first 140 to 144 amino acids (the shaded area in Fig. 1B); that is, all the structural elements required for productive intersubunit contacts in the complete particle are provided by this sequence, and the entire basic C-terminal tail is dispensable. A structural feature important for correct monomer folding and/or stable intersubunit contacts is apparently present between amino acids 140 and 144. The further truncated core protein variants may still have some potential for assembly, since on sucrose density gradients they exhibited a fast-sedimenting HBcAg reactive peak, in addition to the bulk of protein sedimenting at low velocity. However, owing to the small amount of material and the presence of excess proteinaceous contaminants, this could not be substantiated by electron-microscopic observation.

As will be separately described (von Bonsdorff et al., unpublished), all stable particles showed clear $T = 3$ symmetry; that is, like particles from the wild-type core protein, they consist of 180 subunits. This contrasts markedly with many RNA plant virus capsids of known structure, which commonly have a basic N-terminal sequence that appears to be essential for assembly into the higher-order 180-subunit structure; enzymatic removal of this sequence yields truncated proteins forming empty $T = 1$ shells. That the presence of the basic terminus in the HBV core protein leads to encapsidation of markedly more RNA allows us to infer that, regardless of binding specificity, a substantial part of this sequence is available in the inside of the particle. Proteolysis data (3, 8; our unpublished data) suggest that part of the sequence between amino acids 140 and 150 might be exposed on the particle surface; this implies that the rest of the C terminus must be threaded from the outside to the inside of the particle in a way that does not interfere with the packing of subunits. However, it might impose a slight structural change on the particle which could be responsible for the differing antigenicities observed for cores from full-length protein or variant 164 (high HBc antigenicity; low HBe antigenicity) and the further truncated variants (high HBc and HBe antigenicity) as observed in this study and (for truncated C-terminal fusion proteins) by others (18). We cannot exclude, however, that the observed HBe antigenicity is due to the presence of incomplete particles or even smaller intermediates, since the particles from variants 144 and 149 appeared inherently unstable.

If there is a structural difference between assembled particles of the longer and shorter core proteins, it does not seem to be dramatic, as indicated by our finding that all stable particles exhibited almost identical electrophoretic mobility. Mobility in agarose gel electrophoresis is, under otherwise identical conditions, a function of particle radius and surface charge (31, 32). Since the radii of the particles are very similar (von Bonsdorff et al., unpublished), surface charge should not be dramatically different; basically, the same surface residues seem to be exposed to the solvent in particles from full-length and truncated core protein subunits. If the Arg-rich region were located at the particle surface, it would be expected to have drastic influence on surface charge and, therefore, on electrophoretic mobility.

The loss of assembly capability occurring after removal of the amino acids between positions 139 and 144 is compatible, in terms of overall size requirements, with the recently postulated model for the folding of the core protein (1), which predicts a similar eight-stranded β-barrel structure as the one commonly found for RNA virus capsid proteins. There, the minimal length for correct folding into such a structure is estimated to be about 150 amino acid residues (26). For exhibiting full HBe antigenicity, amino acid residues from the N-terminal region as well as from around position 138 are required (28), suggesting that this sequence (Fig. 1B, shaded area) is sufficient for folding of the peptide chain into a stable, tightly folded domain. HBcAg and HBeAg differ in structure, as shown by their distinct antigenicities; however, they may share a similarly folded core with identical structure elements. It seems likely, therefore, that in the nonassembling variants 138 and 139, a structural element of importance in intersubunit contacts in the particle rather than in folding of the protein monomer is lacking. We have no direct evidence, however, whether these variant proteins can fold properly.

Variants 144 and 149 efficiently assembled into particles under the conditions in the overexpressing *E. coli* cell. This indicates that the apparent assembly deficiency of authentic p17e-HBeAg (consisting of amino acids -10 to 149) cannot be solely due to the lack of the Arg-rich region. Alternative explanations have to be invoked which might include an inhibitory influence of the -10 to -1 peptide remaining after cleavage of the first 19 amino acids of the pre-C region (8); more probably, however, the effect is related to the different biosynthetic pathways of HBcAg and HBeAg.

Functional role of the Arg-rich C terminus. As shown

above, the basic C terminus of the core protein is not required for correct assembly of HBV capsids; however, much more RNA is encapsidated if this sequence is present, confirming the expected affinity of the poly-Arg region for nucleic acid.

Our data suggest that in addition to selectivity for RNA over DNA, at least the full-length protein encapsidates its own mRNA better than other, cellular RNAs. Although one tends to overestimate the proportion of a single band within a smear of numerous less prominent bands, this seems to indicate a certain degree of specific recognition between the core protein and HBV-related RNA. Similar observations have been made by others (V. Bruss and W. H. Gerlich, personal communication). On the other hand, no evidence for packaging selectivity has been obtained so far from a corresponding analysis of the RNA contained in duck hepatitis B virus core particles expressed in *E. coli* (R. Lenhoff, P. Smith, M. Yu, and J. Summers, personal communication). The available literature contains conflicting reports about whether nucleic acid binding by the core protein occurs in a sequence-specific manner. Our data may, at first glance, support the idea that during assembly of a replication-competent core particle in an infected hepatocyte, the correct (i.e., genomic) RNA is encapsidated by sequence-specific recognition of this RNA by the core protein. Recent data from this laboratory on RNA packaging into HBV core particles in animal cells, however, do not support the notion that the apparent packaging selectivity observed in *E. coli* is due to a direct sequence specific recognition process between core protein and HBV RNA. First, the HBV P gene product(s) appears to be essential for efficient packaging of the genomic RNA (R. Bartenschlager and H. Schaller, (submitted for publication); the P gene is, however, absent in our heterologous expression system. Second, an essential part of the major *cis*-acting element on the genomic RNA mediating correct packaging (i.e., the packaging signal) resides in the pre-C region (M. Niepmann, R. Bartenschlager, and H. Schaller, submitted for publication), which is again absent from the constructs used for expression in *E. coli*.

How then, in the *E. coli* expression system, could the core mRNA be found enriched in the core particles? Although we cannot exclude a contribution of sequences present in the C-gene mRNA, we favor the model that utilizing the general, unspecific affinity of the core protein for nucleic acid in a cotranslational or *cis*-acting manner would also lead to preferential encapsidation of the core mRNA: during or after emergence of the protein chain from the translating ribosome, this RNA would be one of the spatially closest nucleic acids for the core protein to bind to. Completion of capsids around such a core protein-RNA nucleation complex would then be driven by the relatively high concentration of core protein in the overexpressing *E. coli* cell. Considering that during natural HBV infection the core protein is likely to be translated from the genomic RNA, a similar *cis*-acting process could contribute in vivo to efficient assembly of replication-competent capsids by increasing the local concentration of core protein subunits at the correct site of assembly, i.e., around the genomic RNA.

Although some of these interpretations regarding the assembly process are speculative, it is clear from our results that the Arg-rich C terminus of the core protein plays a crucial role in providing the high stability to the HBV core particle that was noticed earlier (18). The high stability of particles from full-length protein or, to some lesser extent, from variant 164, is lost after removal of the C-terminal basic amino acid residues, and it parallels the much lower contents of nucleic acid in the latter particles. We therefore conclude that in addition to protein-protein interactions, those between protein and nucleic acid contribute importantly to the remarkable stability of the HBV capsid. This raises the question of how, in natural infection, such a stable capsid could disassemble and uncoat the viral genome. In view of the data presented above, a modification eliminating, or at least lowering, the affinity of the core protein for nucleic acid should be one important component in the uncoating mechanism. One obvious candidate for such modification would be phosphorylation; not only do authentic core particles carry an associated kinase activity able to phosphorylate Ser residues in the C-terminal region of the protein (9), but also phosphorylation seems to interfere with DNA binding of proteins containing the same SPRR motif that occurs in the mammalian hepadnavirus core proteins (35).

Implications for assembly of replication-competent HBV nucleocapsids. As outlined above, the *E. coli* expression system lacks several of the components essential for the formation of a replication-competent HBV nucleocapsid, but it still allows us to draw a number of conclusions regarding assembly in vivo. The most obvious difference from the authentic situation, in addition to the lack of the genomic RNA and P-gene product(s), is the high concentration of core protein in the overexpressing *E. coli* cell. Since 180 subunits of the core protein must come together to form the complete core particle, high concentrations will drive the process in the direction of the assembled capsid. Therefore, we are confident that the border for assembly we have mapped to the amino acids between residues 139 and 144 reflects a real limit inherent to the properties of this sequence. This domain must provide the principal protein-protein interactions which hold together the capsid. The reported existence of empty core particles in preparations from human liver (11, 27) suggests that assembly based on protein-protein interactions alone also occurs in infected hepatocytes, but its efficiency would be expected to depend on the actual concentration of core protein in the cell.

Adding to the capsid-forming domain, the Arg-rich C-terminal tail as a general nucleic acid-binding region could, in the presence of a suitable nucleic acid, drive the equilibrium between separate intermediates and assembled cores to the side of the particles: enhanced particle stability correlates with nucleic acid content of the core and is most probably due to cooperativity between protein-protein and protein-nucleic acid interactions. Furthermore, the affinity of the protein for nucleic acid will enhance its local concentration and thereby promote assembly. If binding of the core protein to the nucleic acid is cotranslationally controlled, it may also contribute to packaging specificity.

For an HBV capsid to be replication competent, however, it must also encapsidate the viral P-gene product(s). To prevent exclusion of this essential component, assembly based on interactions between core protein subunits alone, or core protein and RNA, as seen in the overexpressing *E. coli* cell, should not proceed with the same efficiency in an HBV-infected hepatocyte. Considering the apparent requirement for P-gene product(s) in packaging the genomic RNA in animal cells, the formation of a binary (genomic RNA plus P-gene product[s]) or ternary (genomic RNA plus P-gene product[s] plus core protein) assembly-nucleation complex offers the most likely explanation, if this complex has a higher affinity for additional core protein subunits than do any of the P-gene product-independent assembly intermediates. Together with stringent regulation of the concentra-

tions of all reaction components in the HBV-infected cell, this would allow for the economic production of complete, replication-competent viral nucleocapsids.

As the manuscript was being completed, an independent study was reported which, by using one core protein variant containing short stretches of foreign sequence at its N terminus and at its engineered C terminus at amino acid position 144, also arrived at the conclusion that the first 144 amino acids are sufficient for assembly (6). Whereas nucleic acid binding was analyzed by blotting techniques, no data were presented on the amount or nature of nucleic acid encapsidated in the particles.

ACKNOWLEDGMENTS

We are grateful to H. Schaller for continuous support and many helpful discussions throughout the course of this work, and we gratefully acknowledge the excellent technical assistance of A. Rieger. We thank M. Noah, Behringwerke, Marburg, Federal Republic of Germany, for kindly providing the monoclonal antibodies used in this study and S. Fuller, European Molecular Biology Laboratory, Heidelberg, for critically reading the manuscript.

This work was supported by the Bundesminsterium für Forschung und Technologie (grant no. BCT318-5).

LITERATURE CITED

1. **Argos, P., and S. D. Fuller.** 1988. A model for the hepatitis B virus core protein: predicition of antigenic sites and relationship to RNA virus capsid proteins. EMBO J. **7**:819–824.
2. **Borisova, G. P., Y. V. Kalis, P. M. Pushko, V. V. Tsibinogin, V. Y. Loseva, V. P. Ose, E. I. Stankevich, A. Y. Dreimane, D. Y. Sniker, E. E. Grinshtein, P. P. Pumpen, and E. Y. Gren.** 1988. Genetically engineered mutants of the core antigen of human hepatitis B virus, retaining the ability for native autoassembly. Dokl. Biochem. **298**:52–55.
3. **Bruss, V., and W. H. Gerlich.** 1988. Formation of transmembranous hepatitis B e-antigen by cotranslational *in vitro* processing of the viral precore protein. Virology **163**:268–275.
4. **Chen, Z., C. Stauffacher, Y. Li, T. Schmidt, W. Bomu, G. Kramer, M. Shanks, G. Lomonosoff, and J. E. Johnson.** 1989. Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution. Science **245**:154–159.
5. **Galibert, F., E. Mandart, F. Fitoussi, P. Tiollais, and P. Charnay.** 1979. Nucleotide sequence of the hepatitis B virus genome (subtype ayw) cloned in *E. coli*. Nature (London) **281**:646–650.
6. **Gallina, A., F. Bonelli, L. Zentilin, G. Rindi, M. Muttini, and G. Milanesi.** 1989. A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acid. J. Virol. **63**: 4645–4652.
7. **Ganem, D., and H. E. Varmus.** 1987. The molecular biology of hepatitis B viruses. Annu. Rev. Biochem. **56**:651–693.
8. **Garcia, P. D., J.-H. Ou, W. J. Rutter, and P. Walter.** 1988. Targeting of the hepatitis B virus precore protein to the endoplasmic reticulum membrane: after signal peptide cleavage translocation can be aborted and the product released into the cytoplasm. J. Cell Biol. **106**:1093–1104.
9. **Gerlich, W. H., U. Goldmann, R. Müller, W. Stibbe, and W. Wolff.** 1982. Specificity and localization of the hepatitis B virus-associated protein kinase. J. Virol. **42**:761–766.
10. **Greenberg, M. E.** 1987. Preparation and analysis of RNA, p. 4.9.1–4.9.4. *In* F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.), Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
11. **Kaplan, P. M., E. C. Ford, R. H. Purcell, and J. L. Gerin.** 1976. Demonstration of subpopulations of Dane particles. J. Virol. **17**:885–893.
12. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
13. **Layne, E.** 1957. Spectrophotometric and turbidimetric methods for measuring proteins. Methods Enzymol. **3**:447–454.
14. **McKay, P., J. Lees, and K. Murray.** 1981. The conversion of hepatitis B core antigen synthesized in E. coli into e antigen. J. Med. Virol. **8**:237–243.
15. **Magnius, L. O., and J. A. Espmark.** 1972. New specificities in Australia antigen positive sera distinct from Le Bouvier determinants. J. Immunol. **109**:1017–1021.
16. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1983. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Matsuda, K., S. Satoh, and H. Ohori.** 1988. DNA-binding activity of hepatitis B e antigen polypeptide lacking the protaminelike sequence of nucleocapsid protein of human hepatitis B virus. J. Virol. **62**:3517–3521.
18. **Milich, D. R., A. McLachlan, S. Stahl, P. Wingfield, G. B. Thornton, J. L. Hughes, and J. E. Jones.** 1988. Comparative immunogenicity of hepatitis B virus core and e antigens. J. Immunol. **141**:3617–3624.
19. **Nassal, M.** 1988. Total chemical synthesis of a gene for hepatitis B virus core protein and its functional characterization. Gene **66**:279–294.
20. **Nassal, M., T. Mogi, S. S. Karnik, and H. G. Khorana.** 1987. Structure-function studies on bacteriorhodopsin. III. Total synthesis of a gene for bacterio-opsin and its expression in Escherichia coli. J. Biol. Chem. **262**:9264–9270.
21. **Ou, J.-H., O. Laub, and W. J. Rutter.** 1986. Hepatitis B virus gene function: the precore region targets the core antigen to cellular membranes and causes the secretion of the e antigen. Proc. Natl. Acad. Sci. USA **83**:1578–1582.
22. **Pasek, M., T. Goto, W. Gilbert, B. Zink, H. Schaller, P. MacKay, G. Leadbetter, and K. Murray.** 1979. Hepatitis B virus genes and their expression in *E. coli*. Nature (London) **282**: 575–579.
23. **Petit, M.-A., and J. Pillot.** 1985. HBc and HBe antigenicity and DNA-binding activity of major core protein p22 in hepatitis B virus core particles isolated from the cytoplasm of human liver cells. J. Virol. **53**:543–553.
24. **Remaut, E., P. Stanssens, and W. Fiers.** 1981. Plasmid vectors for high efficiency expression controlled by the pL promoter of coliphage λ. Gene **15**:81–93.
25. **Remaut, E., H. Tsao, and W. Fiers.** 1983. Improved plasmid vectors with a thermoinducible expression and temperature-regulated runaway replication. Gene **22**:103–113.
26. **Rossmann, M. G., and J. E. Johnson.** 1989. Icosahedral RNA virus structure. Annu. Rev. Biochem. **58**:533–573.
27. **Sakamoto, Y., G. Yamada, M. Mizuno, T. Nishihara, S. Kinoyama, T. Kobayashi, T. Takahashi, and H. Nagashima.** 1983. Full and empty particles of hepatitis B virus in hepatocytes from patients with HBsAg-positive chronic active hepatitis. Lab. Invest. **48**:678–682.
28. **Salfeld, J., E. Pfaff, M. Noah, and H. Schaller.** 1989. Antigenic determinants and functional domains in core- and e-antigen from hepatitis B virus. J. Virol. **63**:798–808.
29. **Schlicht, H.-J., J. Salfeld, and H. Schaller.** 1987. The duck hepatitis B virus pre-C region encodes a signal sequence which is essential for the synthesis and secretion of processed core proteins but not for virus formation. J. Virol. **61**:3701–3709.
30. **Schlicht, H.-J., and H. Schaller.** 1989. The secretory core protein of human hepatitis B virus is expressed on the cell surface. J. Virol. **63**:5399–5404.
31. **Serwer, P.** 1986. Use of gel electrophoresis to characterize multimolecular aggregates. Methods Enzymol. **130**:116–132.
32. **Serwer, P., R. A. Easom, S. J. Hayes, and M. S. Olson.** 1989. Rapid detection and characterization of multimolecular cellular constituents by two-dimensional agarose gel electrophoresis. Trends Biochem. Sci. **14**:4–7.
33. **Sorger, P. K., P. G. Stockley, and S. C. Harrison.** 1986. Structure and assembly of turnip crinkle virus. II. Mechanism of reassembly in vitro. J. Mol. Biol. **191**:639–658.
34. **Stahl, S. J., and K. Murray.** 1989. Immunogenicity of peptide fusions to hepatitis B virus core antigen. Proc. Natl. Acad. Sci. USA **86**:6283–6287.
35. **Suzuki, M.** 1989. SPKK, a new nucleic acid-binding unit of protein found in histone. EMBO J. **8**:797–804.

36. **Takahashi, K., A. Machida, G. Funatsu, M. Nomura, S. Usuda, S. Aoyagi, K. Tachibana, H. Miyamoto, M. Imai, T. Nakamura, Y. Miyakawa, and M. Mayumi.** 1983. Immunochemical structure of hepatitis B e antigen in the serum. J. Immunol. **130:** 2903–2907.
37. **Younger, B. D.** 1984. Measurement of sedimentation coefficients and computer simulation of rate-zonal separations, p. 127–159. *In* D. Rickwood (ed.), Centrifugation—a practical approach. IRL Press, Oxford.
38. **Zoller, M., and M. Smith.** 1983. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. Methods Enzymol. **100:**468–500.