The EMBO Journal vol.13 no.5 pp.1048 – 1057, 1994

A dramatic shift in the transmembrane topology of a viral envelope glycoprotein accompanies hepatitis B viral morphogenesis

Philomena Ostapchuk, Patrick Hearing and Don Ganem[1]

Department of Microbiology, Health Sciences Center, State University of New York, Stony Brook, NY 11794 and [1]Howard Hughes Medical Institute and Department of Microbiology and Immunology, Box 0502, University of California, San Francisco, CA 94143, USA
[1]Correspondence

Communicated by P.Tiollais

The envelope of hepatitis B virus contains three related glycoproteins (termed L, M and S) produced by alternative translation initiation in a single coding region. The smallest of these, the S protein, is a 24 kDa glycoprotein with multiple transmembrane domains. The M and L proteins contain the entire S domain at their C-termini, but harbor at their N-termini additional (preS) domains of 55 or 174 amino acids, respectively. Most of these preS residues are displayed on the surface of mature virions and hence would be expected to be translocated into the endoplasmic reticulum (ER) lumen during biosynthesis. Using a coupled, *in vitro* translation/translocation system we now demonstrate that, contrary to expectation, virtually all preS residues of the L protein are cytoplasmically disposed in the initial translocation product. This includes some preS sequences which in the M protein are indeed translocated into the ER lumen. Since preS sequences are found on the external surface of the virion envelope, our results indicate that during or following budding a dramatic reorganization of either the envelope proteins or the lipid bilayer (or both components) must occur to allow surface display of these sequences. These findings imply that some membrane budding events can have remarkable and previously unsuspected topological consequences.

Key words: hepatitis B virus/membrane proteins/viral assembly

Introduction

The assembly and budding reactions of enveloped animal viruses have been extensively studied not only for their intrinsic virologic interest but also as model systems for understanding the more complex membrane recognition and budding events that occur in eukaryotic cells. The budded virus particle can be thought of as a highly specialized vesicle whose cargo, the nucleocapsid, is surrounded by a patch of extensively derivatized cellular membrane. Viruses have several properties which commend them for study in this context. They encode only a few polypeptides, an even smaller subset of which is involved in these morphogenetic processes. Moreover, the product of the budding event, the virion, is usually efficiently secreted from the cell, greatly simplifying its purification and characterization. Structural analyses of mature virions can provide great insight into the molecular interactions occurring during the budding process (see, for example, Fuller, 1987).

In most viral systems, the participating envelope glycoproteins are simple integral membrane proteins with a single transmembrane domain spanning a conventional lipid bilayer. The budding reaction is thought to be triggered by interaction of the cytoplasmic nucleocapsid with the transmembrane envelope protein; this event promotes lateral interactions between the envelope polypeptides, greatly concentrating them and typically excluding other (host) chains in the process. In the final budded particle, the envelope proteins, although densely packed, retain the transmembrane orientation that they acquired during their biosynthesis (reviewed in Simons and Garoff, 1980, and Stephens and Compans, 1988).

Here we report on a remarkable variation on this theme which occurs during the envelopment of the hepatitis B virus (HBV). In HBV virions, the 3.2 kb viral DNA genome is located within a nucleocapsid core composed of 180 subunits of the 21 kDa C protein; surrounding this core is an envelope containing three related viral glycoproteins (termed L, M and S) (reviewed in Ganem and Varmus, 1987). These three proteins are encoded by alternative translational initiation within a single open reading frame (see Figure 1A). The 24 kDa S protein is initiated at the innermost AUG codon and is the most abundantly expressed HBV surface glycoprotein. A current view of its transmembrane structure is shown in Figure 1B. Both its N- and C-termini are known to project into the endoplasmic reticulum (ER) lumen; it is felt that the molecule may have up to four transmembrane domains, though only the two most N-terminal ones have been directly shown to participate in membrane interactions (Eble *et al.*, 1986, 1987; Gavilanes *et al.*, 1982; Guerrero *et al.*, 1988). This molecule is unique among viral envelope proteins in that it does not require nucleocapsid interactions to be released from the cell: when expressed singly in uninfected cells, multimers of S are capable of aggregating in the ER bilayer and budding into the lumen as a 20 nm subviral lipoprotein particle (Liu *et al.*, 1982; Simon *et al.*, 1988; Huovila *et al.*, 1992).

However, S protein expression is not sufficient for the envelopment of HBV nucleocapsids; expression of the upstream (or 'preS') coding region is also required for the assembly of mature virions. Two less abundant proteins, termed M and L, are initiated at in-frame AUG codons in this region, 55 and 174 codons (respectively) 5′ to the S AUG. (By convention, the M initiator divides the preS region into two subregions, preS1 and preS2; see Figure 1.) These proteins are found in only small quantities on the 20 nm subviral particles but are greatly enriched on preparations of mature virions (Heermann *et al.*, 1984). Genetic studies indicate that in addition to S protein the expression of L (but not M) protein is required for viral assembly: mutations ablating L expression fail to assemble and release mature virions (Bruss and Ganem, 1991; Ueda *et al.*, 1991).

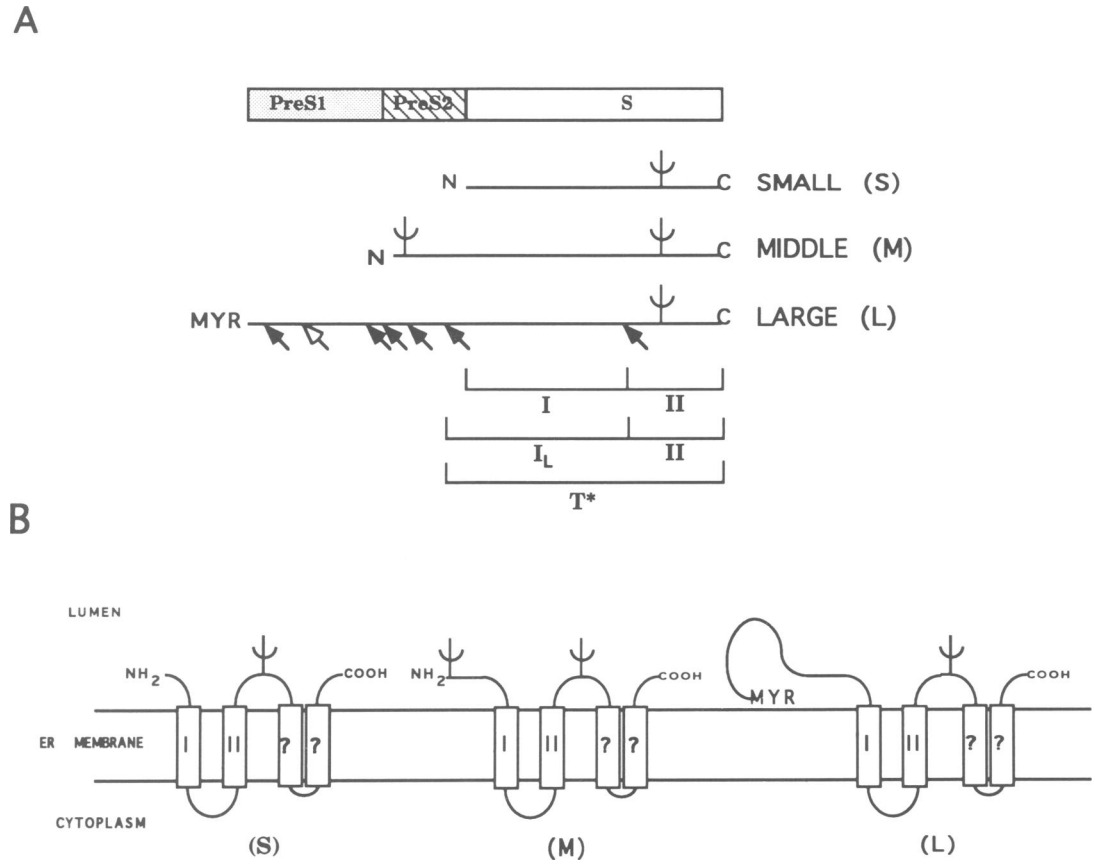

Fig. 1. Structural relationships and proposed topologies of small (S), middle (M) and large (L) surface antigen proteins in the ER membrane. (**A**) S, M and L proteins (shown as line drawings) are translated from a common open reading frame of HBV (indicated by the rectangle) by the use of three in-frame initiation codons at the N-termini of preS1, preS2 and S. All three proteins share in common the S domain. The 55 amino acid preS2 domain is common to M and L, and L has an additional N-terminal domain, preS1. Sites where asparagine-linked glycosylation occurs are indicated by forked symbols. A myristic acid group (MYR) is amide-linked to glycine at the N-terminus of L. Filled arrows indicate the site of trypsin cleavage common to all three proteins in S, as well as sites within preS1 and preS2 (sites within the preS domains separated by fewer than four amino acids are indicated by a single arrow). The open arrow indicates a potential trypsin cleavage site within preS1 that is not cleavable (see text). Trypsin digestion of untranslocated S yields proteolytic products I and II and the derivation of those products are indicated by brackets. Also indicated by brackets are the significant proteolytic products of untranslocated L (I_L and II) as well as the trypsin product, T*, produced by digestion of translocated L (as discussed in the text). (**B**) Proposed topologies of S, M and L are illustrated. The lumenal and cytoplasmic sides of the endoplasmic reticulum (ER) membrane are indicated. Note that the N- and C-terminal ends of the proteins are lumenally disposed as well as the entire preS domains of both M and L. Extensive analyses of the topology of S have confirmed the orientation demonstrated and defined the two membrane spanning domains (I and II) (Eble *et al.*, 1986, 1987). The proposed membrane spanning domains marked with a '?' are based on model building but have not been examined experimentally (Guerrero *et al.*, 1988). The S domain within M is similarly disposed and the preS2 domain is translocated via the uncleaved signal sequence of membrane spanning domain I (Eble *et al.*, 1990). The proposed lumenal disposition of the preS domains of L was largely based on inferences drawn from analyses of subviral and viral particles.

Several lines of evidence demonstrate that in both virions and subviral particles most preS determinants are on the particle surface. Monoclonal antibodies specific for preS2 and preS1 determinants can precipitate both subviral and viral particles (Heermann *et al.*, 1984; Kuroki *et al.*, 1990; Mimms *et al.*, 1990; this work), and preS determinants on particles are highly accessible to attack by exogenous proteases (Heermann *et al.*, 1984). According to conventional views of membrane topogenesis and viral budding, these domains should therefore be translocated into the ER lumen during their biosynthesis (see Figure 1 for diagram). In earlier work we demonstrated that this is indeed so for the preS2 domain of the M protein (Eble *et al.*, 1990). When M protein is translated *in vitro* in the presence of microsomal membranes, its preS2 domain undergoes N-linked glycosylation and is preferentially protected from exogenous trypsin; presumably, this translocation is effected by the downstream signal sequence encoded at the N-terminus of S. Here we report that similar studies of the L protein reveal a strikingly different result: in this molecule, neither preS1 nor preS2 sequences are translocated into the vesicle lumen, under conditions in which the S sequences of the same protein are correctly disposed. Our results rationalize the requirement for L protein in viral assembly, since only in this protein would preS determinants be available for interactions with cytoplasmic nucleocapsids. More importantly, they indicate that HBV morphogenesis must involve a dramatic reorganization of either the envelope proteins, the lipid bilayer, or both components in order to allow surface display of preS1.

Results

Translocation of L chains in vitro reveals an unexpected topology

To examine the transmembrane disposition of the L protein we employed a coupled *in vitro* translation/translocation system programmed with synthetic transcripts of the L

protein coding region. Translation was carried out in a rabbit reticulocyte lysate supplemented with [^{35}S]methionine and microsomal vesicles from dog pancreas. Membrane translocation of newly synthesized chains can be detected by (i) the addition of N-linked carbohydrates to appropriate glycosylation sites (Asn-X-Ser/Thr) in translocated domains; or (ii) the protection of translocated domains from digestion by exogenous proteases. Glycosylation is detected by its ability to retard the migration of the corresponding protein on SDS−PAGE; where necessary, proof that retardation is due to glycosylation is supplied by parallel translation in the presence of tripeptide competitors of N-linked glycosylation.

When L mRNA is translated in this system, a doublet of bands at 39 and 42 kDa appears, corresponding to the unglycosylated and glycosylated forms of L protein, respectively (Figure 2, lane 1). These bands are identical to the products of L expression in transfected cells *in vivo*, with glycosylation of the translocated chains occurring at a single site within the S domain (Figure 6A and data not shown; see Figure 1A for diagram). S chains translated in parallel (lane 2) were likewise translocated and glycosylated at this same position (Eble *et al.*, 1986). We next asked if the disposition of the S domain in the L protein is similar to that in the 24 kDa S protein itself. To do this, both translocated preparations were exposed to exogenous proteinase K, which is known to cleave correctly translocated S chains only in the cytoplasmic loop between transmembrane domains I and II, resulting in protection of a C-terminal fragment termed Ro (Eble *et al.*, 1986; see diagram in Figure 2). As shown in lanes 7 and 8, this signature fragment was efficiently protected in both L and S chains following translocation; when the translocation products were exposed to detergent prior to proteolysis (lanes 4 and 5), the Ro bands were degraded, confirming that protection was due to membrane sequestration rather than to intrinsic protease-resistance of the chains. As expected, translocated yeast prepro-alpha factor, which is fully secreted into the vesicle lumen, remained fully resistant to exogenous proteinase K unless vesicles were disrupted with detergent (lanes 3, 6 and 9). These data indicate that preS sequences do not drastically alter the transmembrane disposition of the S domain within L, specifically those S sequences distal to the first cytoplasmic loop (see diagram in Figure 2).

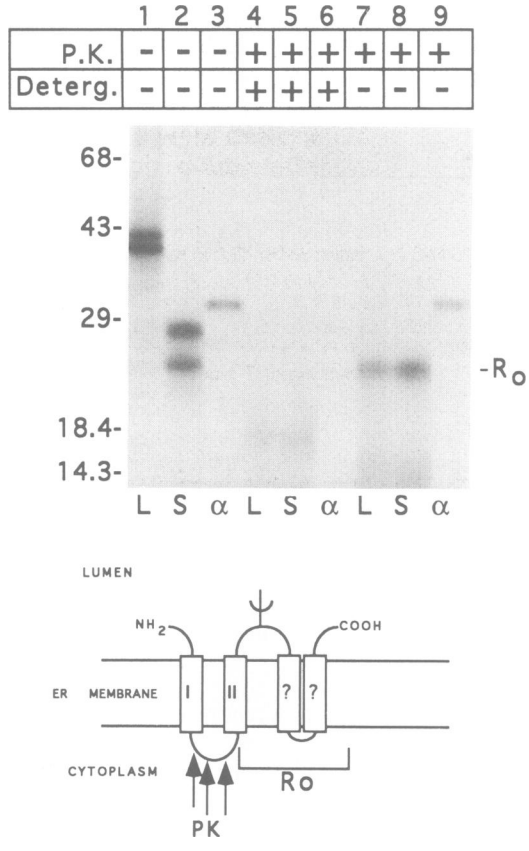

Fig. 2. Analysis of the topology of L protein using proteinase K. *In vitro* synthesized RNAs for S, L and prepro-α factor (α) were translated in reticulocyte lysates in the presence of dog pancreas microsomes and ^{35}S Translabel. Following translation, samples were incubated in either the presence (+) or absence (−) of proteinase K (P.K.) and Triton X-100 (Deterg.) as indicated in the key and described in Materials and methods. ^{35}S-labeled proteolytic products were examined by SDS−PAGE and autoradiography. Lanes 1, 4 and 7 are results with L RNA in the translations; lanes 2, 5 and 8 are results with S RNA; lanes 3, 6 and 9 are results with prepro-alpha factor RNA. Numbers to the left of lane 1 are mol. wt standards expressed in kDa. Ro indicates the previously defined proteinase K-resistant fragment of translocated S protein (Eble *et al.*, 1986). The illustration below shows the topology of S. The region of S digested by proteinase K is indicated by filled arrows and the extent of the resulting Ro product is indicated in brackets.

The current model for the transmembrane disposition of preS sequences within L envisions that the entire preS region is disposed lumenally [Guerrero *et al.* (1988); see Figure 1B]. This model predicts that proteinase K digestion of translocated L should produce an N-terminal fragment of ~21 kDa (comprising preS1, preS2 and S transmembrane domain I). A fragment of this size would be expected to comigrate with the Ro fragment in SDS−PAGE. If, however, glycosylation of translocated L chains is prevented, the Ro fragment resulting from proteinase K treatment of such chains should shift out of this region and readily reveal the predicted 21 kDa preS fragment just described. When this experiment was performed, however, no additional 21 kDa band was observed (data not shown), implying that proteolysis at sites within the preS1 and/or preS2 domains must have occurred. This finding is inconsistent with the model of Figure 1B.

As an independent test of that model we next examined the accessibility of the translocated L product to trypsin. Figure 1A schematically depicts the location of all predicted trypsin cleavage sites in the L chain. Fortuitously the S domain has only one cleavable trypsin site, and this site is known to be translocated into the vesicle lumen in the S protein (see Figure 1B and Eble *et al.*, 1986). By contrast, the preS domain is rich in potential trypsin cleavage sites. If the model of Figure 1 applies, then translocated L and M chains should be fully resistant to exogenous trypsin. We have previously shown that some translocated M chains are indeed trypsin-resistant (Eble *et al.*, 1990). However, when L chains are examined in a similar fashion the results are remarkably different (Figure 3). Prior to proteolysis, the expected doublet of 39−42 kDa is observed (lane 1). But after trypsin exposure, no full-length protected chains are observed (lane 3). This finding was highly reproducible, and even prolonged overexposure of autoradiograms from multiple experiments consistently failed to reveal fully-protected chains (data not shown). Under identical conditions, translocated (glycosylated) S chains are fully protected while untranslocated (unglycosylated) chains are digested (compare lanes 4 and 6).

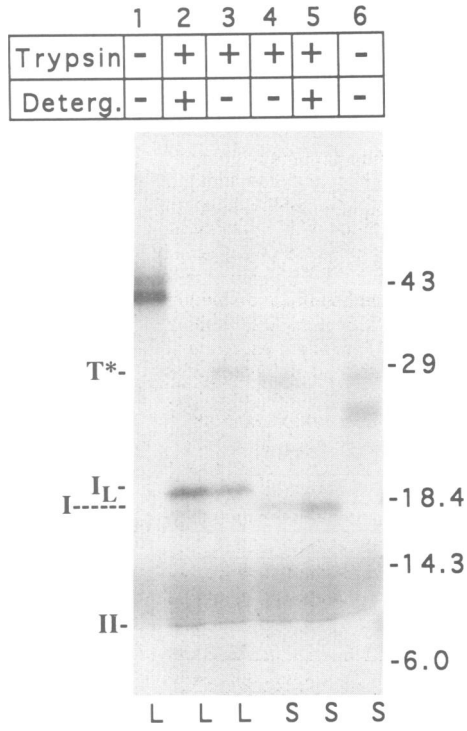

Fig. 3. Translocated L protein is not fully protected from trypsin digestion. *In vitro* synthesized RNAs for L (lanes 1, 2 and 3) and S (lanes 4, 5 and 6) were translated in reticulocyte lysates in the presence of dog pancreas microsomes and ^{35}S Translabel. Following translation, samples were incubated in the presence (+) or absence (−) of trypsin and Triton X-100 (Deterg.) as indicated in the table and as described in Materials and methods. Proteolytic products were examined by SDS−PAGE and autoradiography. Numbers at the right of lane 6 are mol. wt standards expressed in kDa. I, I_L and II indicate limit products of digestion of untranslocated chains of L and S as described in Figure 1. T* indicates the tryspin-resistant product unique to translocated L protein (illustrated in Figure 1).

Comparison of the trypsin cleavage products of L with those of S is instructive. Untranslocated S chains should be cleaved into two fragments, I and II (see Figure 1A for map). These products are observed (lanes 4 and 5), though fragment II is largely obscured by the huge excess of unlabeled rabbit globin chains in this region of the gel (data not shown). In addition to fragment II, the untranslocated L chains are expected to generate numerous small preS fragments and a larger fragment, I_L, which contains all of fragment I plus the sequences between the S initiator methionine and the most distal trypsin site in preS2 (see Figure 1A). This expected product is also observed (lanes 2 and 3). Most importantly, the translocated L chains should produce a protected fragment larger than the translocated S protein, corresponding to the protection of the entire S domain plus whatever portion of preS is protected. In lane 3, such a band is indeed detected, here labeled T*, but it is only slightly larger than the translocated S protein, suggesting that trypsin must have cleaved at a very distal site within preS2. These data indicate that at least the distal preS2 region must be cytoplasmically disposed in the initial biosynthetic product of L expression and are not compatible with the conventional view of L topogenesis (which we note was derived by inference from the structure of the virion rather than by experiment).

Mapping the cytoplasmic preS domains of translocated L protein

To map the extent of the cytoplasmically disposed preS sequences in L, we examined the pattern of protected fragments generated from mutant L chains in which trypsin sites had been sequentially inactivated. The mutants we examined are summarized in Figure 4A. The 10 trypsin sites in the preS region of the wild-type (WT) L protein (top line) are shown as open circles and numbered sequentially from the N-terminus. In the series of mutants shown below the WT, mutationally inactivated trypsin sites are indicated by black circles; in all cases, the mutant sites substituted glutamine for the arginine or lysine cleavage target. Closely clustered trypsin sites were inactivated together. To validate that the mutations had indeed inactivated the corresponding trypsin sites, each mutant was translated *in vitro* without microsomes and analyzed by trypsin digestion (not shown); in all cases save one, the predicted proteolytic products were observed. The exception was in mutant RQ4, in which cleavage did not occur at site 4, but rather at the cluster of sites 1−3. We infer from this that local folding around site 4 renders this site uncleavable.

Each mutant was then translocated *in vitro* and the products examined for glycosylation and trypsin protection. Results of the glycosylation analysis are presented in Figure 5. Not unexpectedly, each of the mutants is successfully glycosylated at a single site; results to be presented below indicate that this site is, as expected, within S (see Figure 6B). The retarded electrophoretic mobilities of several of the mutants (Figure 5, lanes 7−12) correlate with the cumulative elimination of charged residues in the preS domain, resulting in aberrant migration in SDS−PAGE.

Trypsin cleavage of the mutant translocation products is presented in Figure 6A. As seen previously for the WT L chain, no full-length mutant L chains are observed after trypsin digestion (lanes 1−8). [For mutants RQ4 and KQ5 (lanes 5−7), protected products of nearly full length are observed, but electrophoresis of these species alongside their uncleaved counterparts on higher-resolution gel systems (data not shown) indicates that they, too, have been shortened, presumably by cleavage within the cluster of trypsin sites 1−3 (see Figure 4A).] Instead, two series of prominent bands are observed that increase in size in a stepwise fashion with each mutant (Figure 6A, lanes 1−8). The origin of these bands is depicted schematically in Figure 4. For clarity, in Figure 6 the T* species are labeled with a downward pointing arrow next to each band and the I_L species with an upward pointing arrow at each band. The T* series represents the products of trypsin cleavage of translocated chains, while the I_L series derives from cleavage of untranslocated chains. Two lines of evidence support this view. First, the T* fragments disappear when membrane vesicles are solubilized with detergent prior to trypsin digestion, indicating that they derive from L chains that have at least partly translocated into the vesicle; by contrast, the I_L bands are completely unaffected by this treatment, indicating that they represent limit digest products of untranslocated chains (Figure 6A, lanes 9−16). Second, the T* fragments are glycosylated while the I_L fragments are not: when translation is carried out in the presence of glycosylation inhibitor, only the T* bands are shifted in mobility (Figure 6B). (Given that the N-termini of these fragments must be in preS, this places the glycosylation site

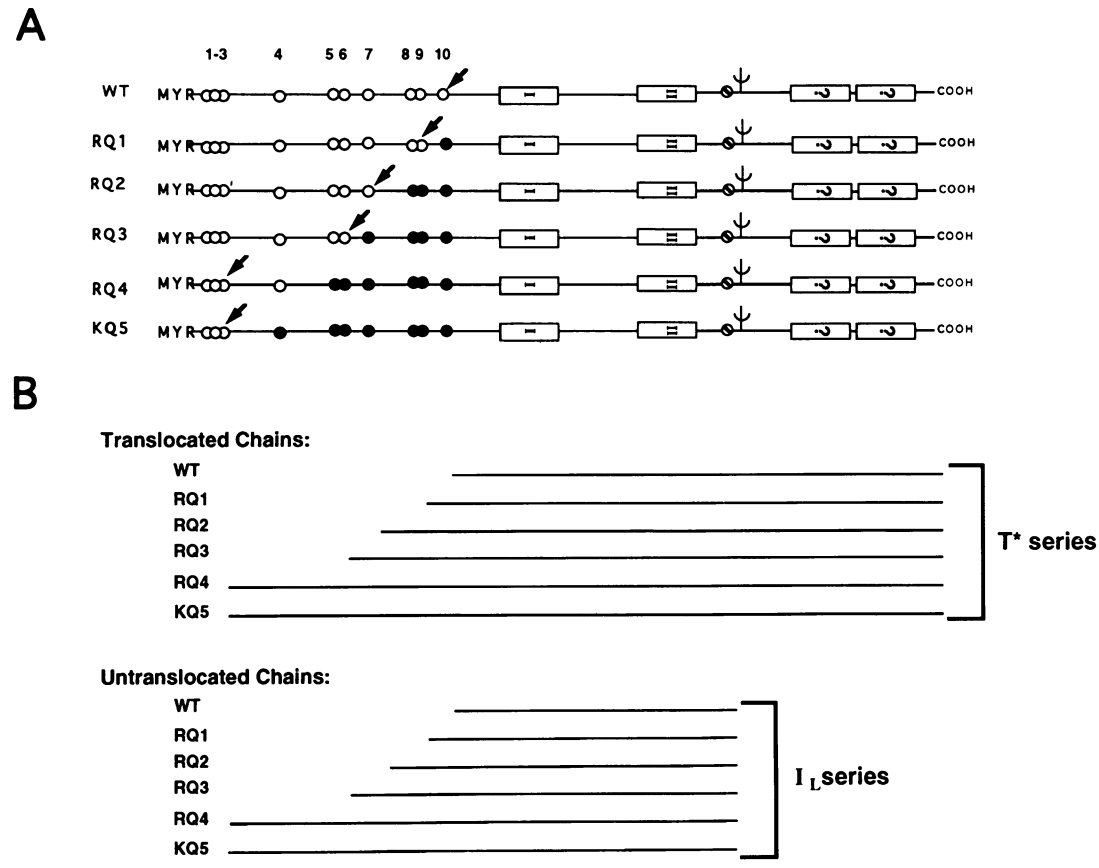

Fig. 4. Schematic depiction of mutations ablating preS trypsin sites and the resulting tryptic fragments. (**A**) A line diagram of L (WT) is shown with features previously described in Figure 1. Also shown are the 10 potential trypsin cleavage sites (open circles, numbered 1–10) within the preS domains of L. The circle with a slash indicates the trypsin cleavage site within the S domain that is protected upon translocation. Below WT L are illustrations of L mutants, RQ1–4 and KQ5, that contain arginine to glutamine, or lysine to glutamine changes (indicated by filled circles) introduced by site-directed mutagenesis as described in Materials and methods. Arrows indicate the site of trypsin cleavage of the translocated L chains that result in the generation of the T* proteolytic products depicted in panel B and demonstrated experimentally in Figure 6A and B. (**B**) The diagnostic trypsin products produced from the translocated (T* series) and untranslocated (I_L series) chains are indicated. The lines show the extent of T* and the I_L products for the corresponding WT and mutant L chains.

in our translocated L chains within the S domain, exactly as is found *in vivo*.)

The stepwise increase in size in the T* fragments with progressive elimination of preS trypsin sites indicates that these fragments contain increasing amounts of preS sequences. The molecular weights of the fragments indicate that they contain the entire S domain plus increasing amounts of preS–i.e. that in each mutant the protected translocated T* band derives from trypsin cleavage at the first cleavage site upstream of the most 5′ mutation (indicated schematically by the arrows in Figure 4). Taken together, this means that each trypsin site indicated by an arrow in Figure 4A must reside on the cytoplasmic face of the vesicle in the initial translocation product. Since these sites span virtually the entire preS region, we conclude that almost all preS sequences are cytoplasmically disposed in newly synthesized L chains (see Figure 10 for model).

We were concerned that the substitution of glutamines for the numerous arginines and lysines in our mutants might have rendered the mutant preS regions incompetent for translocation across the vesicle membrane. To exclude this possibility we carried out two control experiments. First, the three trypsin site mutations in the preS2 region were introduced into the M protein. As noted earlier, we had previously shown that preS2 sequences in the M protein can be translocated into the vesicle lumen under the influence of downstream S signals (Eble *et al.*, 1990). Because the preS2 and S regions each contain a glycosylation site, the resulting, fully translocated product can be doubly glycosylated. Accordingly, we translated the WT and mutant M proteins in the presence of microsomes and examined the products for glycosylated derivatives by SDS–PAGE. As shown in Figure 7 (lanes 3 and 4), both proteins generated a series of three bands corresponding to the unglycosylated, mono- and di-glycosylated isoforms of M; glycosylation competitor peptide experiments (not shown) confirmed that the more slowly migrating bands indeed represent glycosylation products. The relative proportions of the three species in lanes 3 and 4 indicates that the mutant preS2 domain is translocated with essentially WT efficiency.

In an independent experiment we asked if the preS domain in our most extensively substituted mutant (KQ5) could be translocated under the influence of a heterologous signal sequence. We therefore constructed mutant SSKQ5, which bears the preprolactin signal sequence fused to the N-terminus of KQ5 (one extra amino acid of preprolactin has been added, and the L initiator methionine was replaced with leucine as a result of the introduction of a restriction site for cloning). When translated in the presence of microsomes, three bands are seen (Figure 7, lane 2). The lowest of these

bands comigrates with the SSKQ5 product made in the absence of microsomes (not shown); the upper two bands disappear when translocation occurs in the presence of the peptide competitor of glycosylation (lane 1), indicating that the protein has become multiply glycosylated. The most highly glycosylated form was fully resistant to trypsin digestion (not shown), indicating that its preS region had been translocated. Moreover, close inspection of the gel reveals that in the presence of the glycosylation inhibitor, an additional band appears that migrates more quickly than the untranslocated product and at the same mobility as the KQ5 product. This band likely corresponds to L chains whose signal peptides have been cleaved off by signal peptidase. We conclude that the preS mutations we introduced pose no barrier to the successful translocation of the chain.

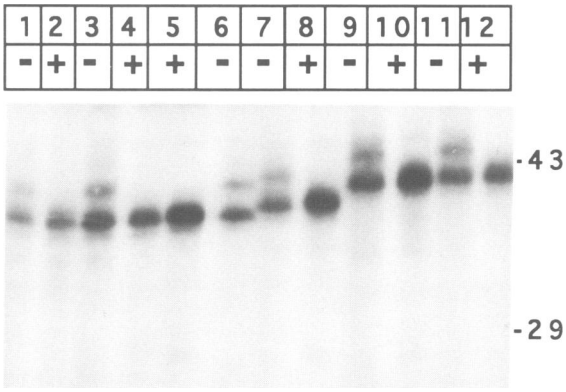

Fig. 5. The mutated L proteins are translocated and glycosylated upon synthesis in the *in vitro* translation/translocation system. The following RNAs were synthesized *in vitro* and translated in reticulocyte lysates containing dog pancreas microsomes, ^{35}S Translabel and either in the presence (+) or absence (−) of glycosylation acceptor peptide: lanes 1 and 2, L; lanes 3 and 4, mutant RQ1; lanes 5 and 6, mutant RQ2; lanes 7 and 8, mutant RQ3; lanes 9 and 10, mutant RQ4; and lanes 11 and 12, mutant KQ5. Labeled proteins were examined by SDS−PAGE and autoradiography. At right are indicated the positions of mol. wt standards of 43 and 29 kDa.

The preS region is devoid of signal sequences

The inability of the preS region of L to be translocated *in vitro* suggested that this region may be devoid of functional signal sequences. In accord with this, sequence inspection in this region discloses no stretch of hydrophobic residues typical of such elements. To explore this issue directly, we constructed the plasmid preS1+2-glycoglobin, in which the entire preS region is fused in-frame to a chimpanzee α-globin coding sequence bearing an engineered glycosylation site. Translation of synthetic mRNA for this construct in the presence of microsomes revealed no newly glycosylated products (Figure 8, compare lanes 1 and 3). Moreover, when the products synthesized in the presence of membrane vesicles were treated with exogenous proteinase K, no protected fragments were observed (lane 4). Thus, the preS domain appears to lack functional topogenic signals.

The preS domain of L is on the external surface of the virion

Given the unexpected topology of the L protein we felt it important to confirm that preS domains of L truly are exposed on the surface of the virion. Virions were obtained by transfection of HepG2 cells with overlength plasmid DNAs encoding either WT virus or a mutant genome (M−) in which the M protein AUG had been inactivated by conversion to AUC (Bruss and Ganem, 1991). Virus particles were harvested from the culture medium and their complement of surface epitopes was determined by immunoprecipitation with a variety of polyclonal and monoclonal antibodies. Virions were detected by examining the immune precipitates for endogenous HBV DNA

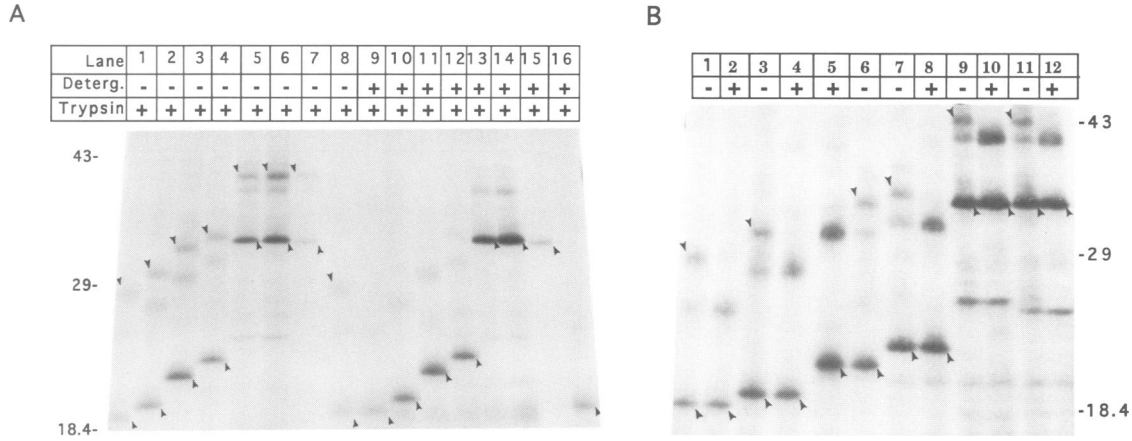

Fig. 6. The entire preS domain of L protein is cytoplasmically disposed. (**A**) *In vitro* synthesized RNAs for L (lanes 1, 8, 9 and 16) and the mutants of L (lanes 2 and 10, RQ1; lanes 3 and 11, RQ2; lanes 4 and 12, RQ3; lanes 5 and 13, an RQ4 mutant; lanes 6 and 14, another independent RQ4 mutant; lanes 7 and 15, KQ5) were translated *in vitro* in the presence of dog pancreas microsomes and ^{35}S Translabel. Following translation, samples were incubated with trypsin either in the presence (+) or absence (−) of Triton X-100 (Deterg.) as indicated in the key. Proteolytic products were examined by SDS−PAGE and autoradiography. Downward pointing arrows mark the T* proteolytic products of the translocated chains and upward pointing arrows mark the I_L proteolytic products of the untranslocated chains (see Figure 4). The numbers to the left of lane 1 indicate the mol. wt standards in kDa. (**B**) The trypsin-resistant fragments, T*, of L and the mutants of L are glycosylated. Synthetic RNAs were translated in the *in vitro* translation/translocation system as described in (A) except that glycosylation acceptor peptide was either present (+) or absent (−) from the reactions as indicated in the key. Following translation, samples were incubated with trypsin and proteolytic products were examined by SDS−PAGE and autoradiography. Lanes 1 and 2, L; lanes 3 and 4, RQ1; lanes 5 and 6, RQ2; lanes 7 and 8, RQ3; lanes 9 and 10, RQ4; lanes 11 and 12, KQ5. Downward pointing arrows mark the T* products and upward pointing arrows mark the I_L products. Numbers to the right of lane 12 show positions of mol. wt standards in kDa.

polymerase activity, which assays for the ability of the virion polymerase to elongate HBV plus strands on a template of genomic minus strands. Precipitates were labeled with [^{32}P]dATP and the DNA product extracted and analyzed by agarose gel electrophoresis. In Figure 9 (lanes 1–6) virions from WT HBV infection are examined. As expected, no virion DNA is precipitated with any anti-envelope antibody if the envelope is first removed with Triton X-100 (lanes 1–3). However, virions are efficiently precipitated by polyclonal anti-S (lanes 4) or monoclonal anti-preS1 (lanes 5) or preS2 (lanes 6). Identical results are seen with M– virus (lanes 7–12); the reactivity of this virus with an anti-preS2 mAb (lanes 9) is particularly noteworthy, since in the absence of M protein this region of preS2 is only represented in L chains. Control experiments with anti-capsid antibodies (Bruss and Ganem, 1991) or irrelevant mAbs (not shown) excluded nonspecific precipitation of virions. These results provide direct evidence that both preS1 and preS2 epitopes of L are surface-exposed on the mature, budded virion.

In accord with this, recent studies of others indicate that in authentic HBV virions purified from human serum up to 50% of the L chains contain externally disposed preS1 determinants, as judged by accessibility to protease (V.Bruss, L.Xuanyong, W.Gerlich and R.Thomssen, personal communication). The remaining 50% have preS1 regions that are trypsin-resistant; this would be expected if molecules oriented as in Figure 10 were also incorporated into virions without any change in their topology (see below).

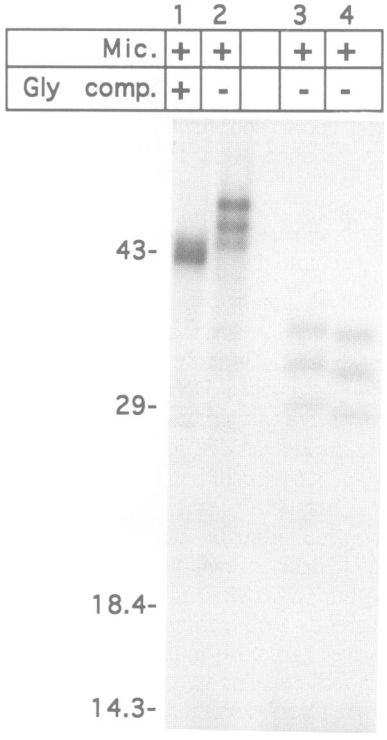

Fig. 7. PreS1 and preS2 domains with the arginine and lysine to glutamine changes can be translocated. Synthetic RNAs for SSKQ5 (L mutant KQ5 engineered to contain the signal sequence from preprolactin) were translated in reticulocyte lysates containing ^{35}S Translabel and dog pancreas microsomes. Proteins were analyzed by SDS–PAGE followed by autoradiography. Shown are the results of translations done in the presence (+, lane 1) or absence (–, lane 2) of glycosylation acceptor peptide (Gly comp.). Lanes 3 and 4 show the results of products obtained in the translation/translocation reactions programmed with either M RNA (lane 3) or RNA for a mutant M protein engineered to contain the arginine to glutamine changes in the preS2 domain (lane 4). Numbers to the left of lane 1 show the positions of the mol. wt standards in kDa.

Discussion

These results indicate that the initial product of L protein biosynthesis has a transmembrane orientation radically

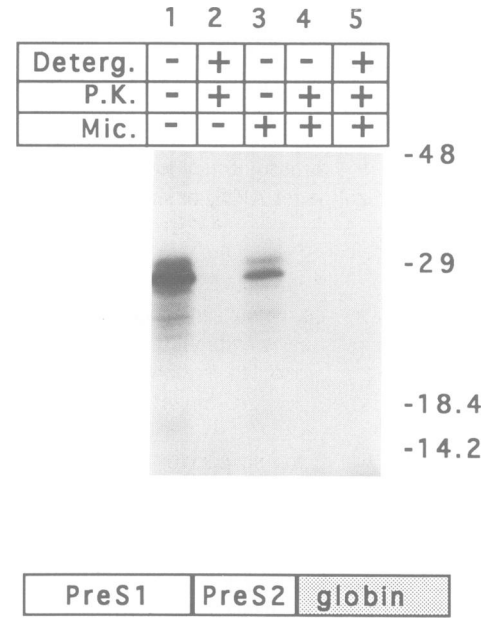

Fig. 8. The preS domains do not contain topogenic sequences. The fusion protein, preS1+2 glycoglobin contains the preS1 and preS2 domains fused to a modified α-globin containing an engineered glycosylation site. *In vitro* synthesized RNAs for this protein were translated in reticulocyte lysates in the presence of ^{35}S Translabel. Reactions were done either in the presence (+) or absence (–) of dog pancreas microsomes (Mic.); following translation, samples were incubated in either the presence (+) or absence (–) of proteinase K (P.K.) and Triton X-100 (Deterg.) as indicated in the table. Numbers to the right of lane 5 show the positions of mol. wt standards expressed in kDa.

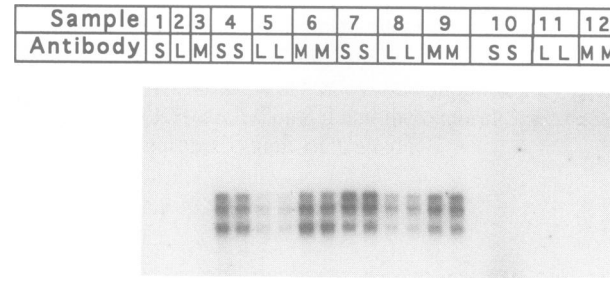

Fig. 9. Epitopes within preS1 and preS2 of L protein are on the surface of the viral particle. Viruses were immunoprecipitated from cell culture media of HepG2 cells transfected with either a plasmid with an overlength copy of WT DNA (lanes 1–6) or a plasmid, M–, bearing a mutation (AUC) in the initiator AUG of M (lanes 7–12). Immunoprecipitated viruses were detected by endogenously labeling the viral DNAs (as described in Materials and methods). Linear and relaxed circular forms of the labeled DNAs were analyzed by agarose gel electrophoresis. The following antibodies were used: S, polyclonal sera raised to surface antigen particles; M, mAb specific for the preS2 domain; L, mAb specific for the preS1 domain. Control samples in lanes 1–3, and 10–12 were pretreated with Triton X-100 and DTT to remove the surface antigens from the viral particle prior to immunoprecipitation. All samples are shown in duplicate except samples 1–3.

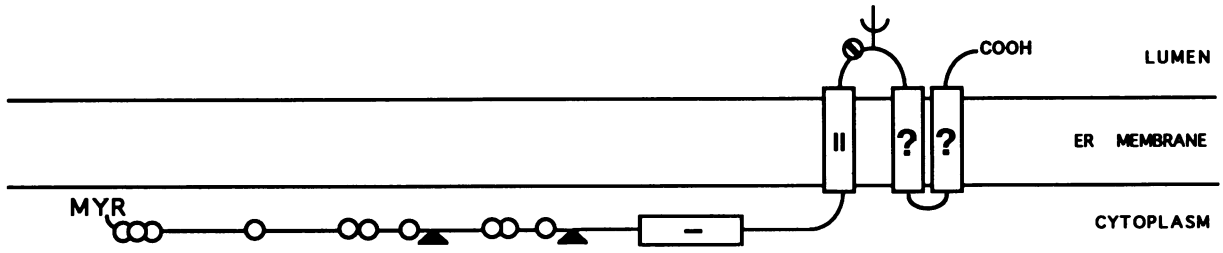

Fig. 10. Model for the initial transmembrane topology of L protein. The lumenal and cytoplasmic sides of the membrane are indicated. All of preS1 and preS2 are cytoplasmically disposed. Transmembrane domain I is presented as cytoplasmically disposed for reasons described in the text. MYR, N-terminal myristate group; open circles, preS trypsin sites; slashed circle, S trypsin site; triangles, M and S AUG initiator codons; ψ, glycosylation site.

different from that predicted by models of its topology in the mature virion. Although most of its S domain appears to be conventionally disposed, its preS1 and preS2 sequences appear to be largely or exclusively cytoplasmic. Strikingly, the same sequences in preS2 that are efficiently transported into the ER lumen in the related M protein are unable to enter the ER when additional preS1 sequences are added to their N-termini. This is not because the preS1 sequences themselves cannot be transposed: when a heterologous signal sequence from prolactin is supplied at the N-terminus of preS1 (in mutant SSKQ5), these sequences are readily delivered to the ER lumen (Figure 7).

A model for the initial transmembrane orientation of L protein is presented in Figure 10. We note that the exact disposition of transmembrane domain I of the S protein is uncertain at present. We have drawn it as cytoplasmic in this model because the distance (nine amino acids) between the N-terminus of S (known to be lumenal in S chains) and the most distal trypsin site in preS2 (known to be cytoplasmic in L chains) is insufficient to allow another transmembrane passage. Further studies of the fine structure of the protein will be required to resolve this issue. In any case, the proteinase K protection data of Figure 2 indicate that transmembrane domain II (and more distal sequences) of S are correctly disposed. This is not surprising, since the topogenic sequences in the transmembrane domain II normally function in an internal position and would not be expected to be disturbed by the preS addition. Consistent with this, earlier experiments with globin−S fusion proteins showed that transmembrane domain II could correctly orient the remainder of S even when the upstream sequences were replaced entirely by globin domains (Eble *et al.*, 1987).

The structure of the initial transmembrane form of L was determined entirely from *in vitro* translocation experiments, but there is little doubt that the results reflect the situation *in vivo*. First, our results are consistent with every experimental fact about L topology that has been generated *in vivo*. The most compelling of these concerns the disposition of glycosylation sites. Only two glycosylation sites are used in the HBV envelope proteins: one in S and one near the N-terminus of preS2. In particles purified from infected human plasma (or from transfected human cells), both sites are glycosylated in the M protein but only the S site is glycosylated in L. Earlier models for L topology (see Figure 1) could not explain this discrepancy, but this pattern of glycosylation site use is exactly what would be predicted from our results. Importantly, when the preS region of L is experimentally directed into the ER lumen (e.g. by a cleavable heterologous signal sequence) it acquires up to two extra glycosylations (Figure 7); thus, the preS sites are not intrinsically inaccessible to the glycosylation machinery in full length L chains. Since such extra glycosylations are never found on virion-derived L chains (Heermann *et al.*, 1984), this argues strongly that virions are not derived from a minor subpopulation of translocated chains that has escaped our detection.

Additionally, we note that the N-terminus of L is myristoylated, a co-translational modification that is known to occur in the cytosol; a cytoplasmic localization for preS1 readily accommodates this observation as well. Thirdly, the L protein topology we observe is generated under conditions that reproduce the correct topologies of the S protein and the yeast α factor. Finally, and most directly, recent studies confirm that in transfected COS cells *in vivo* the preS1 regions of all newly synthesized L chains are cytoplasmically disposed (i.e. trypsin-sensitive), exactly as described herein (V.Bruss, L.Xuanyong, W.Gerlich and R.Thomssen, personal communication).

Our results can also rationalize the requirement for L but not M protein in HBV virion assembly (Bruss and Ganem, 1991). HBV nucleocapsids form in the cytoplasm; once they have completed reverse transcription they are competent for interaction with the viral envelope glycoproteins on intracellular membranes. This is thought to lead to a budding reaction that generates enveloped virions within the lumen of the secretory pathway, following which virions are exported from the cell by constitutive vesicular transport. The data presented here allow formulation of a simple and attractive model for the requirement for L proteins in assembly. The model posits that preS sequences are required for nucleocapsid interactions that trigger budding; if so, then only in the L protein will these sequences be exposed to cytoplasmic nucleocapsids. Recent genetic studies of V.Bruss and Thomsen (personal communication) are consistent with this view and suggest candidate preS sequences for this role. They found that L proteins with deletions of up to 85% of their N-terminal preS1 sequences remain competent for assembly *in vivo*, but that deletions of the sequences closer to the preS1/preS2 interface ablate assembly. Importantly, in every assembly-incompetent mutant the remaining preS sequences were translocated into the ER lumen (as judged by their glycosylation pattern), in keeping with the view that sequestration of these sequences within the ER does not allow virion envelopment.

Finally, our results raise a major mechanistic question: how do cytoplasmically disposed sequences ultimately come to reside on the external surface of the budded virion? We can envision two possibilities: either the preS domain is post-

translationally translocated across the bilayer following core interactions, or the bilayer itself is attenuated or reorganized during budding to expose regions of preS. Unfortunately, present knowledge of the structure of the HBV virion is insufficient to discriminate rigorously between these possibilities. Although electron micrographs of intact HBV virions (Dane *et al.*, 1970) reveal a dense outer coat that contains the viral surface glycoproteins, these negative staining studies reveal little about the molecular organization of the envelope. There have been no chemical or structural analyses of the lipid composition of HBV virions, in part because all HBV preparations also contain a 1000-fold excess of subviral lipoprotein particles whose presence would greatly distort the resulting profile. Interestingly, however, the structure of these subviral particles may furnish clues to the nature of the morphogenetic events in virion budding. These S-containing structures are known to derive from budding or extrusion from pre-Golgi membranes (Simon *et al.*, 1988; Huovila *et al.*, 1992), yet EM studies of them (Patzer *et al.*, 1986) show no unit membrane structure, and their buoyant densities (1.2−1.22 g/ml) are substantially higher than those of normal cellular membranes. It has previously been proposed that lipid reorganization may occur during their biogenesis (Simon *et al.*, 1988); if so, then the same process may also be at work in virion envelopment.

Irrespective of the mechanism by which this topologic feat is achieved, our results indicate that at least in certain situations cellular machinery can reorganize the topology of membrane proteins in surprising and remarkable ways. Although the viral system we report here is obviously a specialized example of this, it raises the intriguing possibility that such rearrangements may be found in other biological settings as well.

Materials and methods

Plasmids

The following SP6-based plasmids were used as templates in *in vitro* transcription reactions: pSP24H for synthesis of S protein RNA; pSP45Hm for L protein RNA, and pSP31H for M protein RNA. The construction of pSP24H, pSP31H and pSP45H has been previously described (Eble *et al.*, 1986, 1990; Persing *et al.*, 1986). In order to facilitate the construction of the RQ and KQ mutants, the *Xba*I site in the polylinker region of pSP45H was eliminated by digestion with *Stu*I and *Hin*dIII, followed by repair with Klenow and ligation generating the plasmid, pSP45Hm. The RQ and KQ mutations were introduced into the L protein coding region DNA of pSP45Hm using site-directed mutagenesis methods (Kunkel, 1985). Oligonucleotides containing the following mutations were used to introduce the arginine or lysine to glutamine changes in L protein: for RQ1, codon 167 was changed from AGG to CAG; for RQ2, in addition to the RQ1 mutation, codon 135 was changed from AGA to CAG and codon 137, AGG to CAG; for RQ3 in addition to the RQ1 and 2 mutations, codon 113 was changed from AGA to CAA; for RQ4, in addition to mutations RQ1−3, changes were made at codons 99 (CGG to CAG) and 103 (AGG to CAG); for KQ5, in addition to RQ1−4 mutations, codon 49 was changed from AAG to CAA. An additional conservative mutation was introduced: isoleucine (codon 48, ATC) was changed to valine (GTC) because the sequence of the oligonucleotide used for mutagenesis was based on the published sequence (Valenzuela *et al.*, 1981) which predicted a valine at position 48. The original plasmid pSP45H contains an isoleucine at position 48 as confirmed by subsequent dideoxynucleotide sequencing. The presence of mutations was verified by dideoxynucleotide sequencing. pRQ2/31Hm expressed M protein containing the arginine to glutamine changes in the preS2 domain. This plasmid is a derivative of pSP31H; the *Spe*I−*Pvu*II restriction fragment of pSP31H was replaced with the *Spe*I−*Pvu*II fragment of pSP45Hm in order to eliminate restriction sites in the polylinker region; followed by the replacement of the *Bam*HI−*Xba*I restriction fragment with *Bam*HI−*Xba*I of the RQ4 plasmid. The presence of the mutations was confirmed by restriction analysis, since the presence of the RQ mutations results in two additional *Bst*NI sites. SSKQ5 encodes a fusion protein of the L mutant, KQ5, with a heterologous signal sequence from prolactin. It was generated in two steps. A plasmid containing the coding sequences for preprolactin was partially digested with *Bsa*I followed by complete digestion with *Bam*HI. The resulting plasmid of interest contains DNA encoding the signal sequence of prolactin plus one additional codon. This was ligated to a PCR-generated fragment of L spanning the N-terminus to the *Bam*HI site in the preS coding region of HBV. The oligonucleotide used for PCR amplification introduced a *Bsa*I site at the N-terminus of KQ5 and resulted in a change of the initiator methionine to a leucine. The DNA sequence of the *Bsa*I−*Bam*HI PCR amplified fragment was confirmed by dideoxnucleotide sequencing. The resulting plasmid was digested with *Bam*HI and *Sma*I and ligated with the *Bam*HI−*Pvu*II fragment of KQ5 to yield the final plasmid SSKQ5.

preS1+2 glycoglobin was constructed by recombining a *Pvu*II−*Eco*RI fragment of HBV DNA (containing preS1 and some preS2 sequences under SP6 promoter control) with a *Pvu*II−*Eco*RI fragment of plasmid PS2/gG (Eble *et al.*, 1990) to generate a plasmid containing both preS1 and preS2 fused in-frame to a chimp α-globin cDNA bearing an engineered glycosylation site.

Both pHBV1.5 and M− plasmids bear overlength copies of HBV adw. The initiator methionine codon of M protein was changed to AUC in the M− virus-bearing plasmid. The construction of these plasmids has been previously described (Bruss and Ganem, 1991)

In vitro transcription, translation and translocation

Templates for transcription were linearized by digestion with *Hpa*I except for preS1+2 glycoglobin, which was linearized with *Pvu*II. Transcription of plasmids was done as described previously (Melton *et al.*, 1984) using SP6 polymerase (Promega, Madison, WI). Synthetic RNAs were stored at −70°C in 10 μl aliquots. Translations were done at 30°C for 1 h using micrococcal nuclease-treated reticulocyte lysates (Promega, Madison, WI), following the protocol recommended by the manufacturer. 10 μl reactions contained 1 μl of RNA and 0.8 μl ^{35}S Translabel (ICN, Irvine, CA). Dog pancreas microsomes used in the translocation assays were either purchased from Promega (Madison, WI) or prepared as described previously (Eble *et al.*, 1986). 4−6 OD units of prepared microsomes were added. Where indicated, the tripeptide glycosylation competitor (acetylated Asn-Tyr-Thr) was added to a final concentration of 1.0 mM.

Proteolysis reactions

Following translations, reactions were cooled on ice and either CaCl to 10 mM final concentration or tetracaine to 3 mM final concentration were added. To 5 μl of translation reaction, proteinase K or trypsin was added to a final concentration of 400 μg/ml and the reactions were incubated for 1 h on ice. Triton X-100 was added to a final concentration of 0.7%. All reactions were brought to the same volume (7 μl) using the appropriate buffers. Proteolysis was stopped by transferring 5 μl samples to 50 μl of boiling 1% SDS and 0.1 M Tris pH 9 using a micropipet tip dipped in 0.5 mM phenylmethylsulfonylfluoride in dimethyl sulfoxide. Samples were analyzed by SDS−PAGE using either 20% or 15% polyacrylamide. Polyacrylamide gels were prepared for autoradiography using Amplify (Amersham, Arlington Heights, IL).

Virus production, immunoprecipitation, and endogenous polymerase reaction

Transfections of HepG2 cells with pHBV1.5 and the plasmid bearing the M− virus were done using the calcium phosphate precipitation method (Wigler *et al.*, 1977). Three days post-transfection, virus was immunoprecipitated from media. The following antibodies were used for the immunoprecipitations: goat polyclonal anti-hepatitis B surface antigen (Dako Corp., Santa Barbara, CA); mAb 116-109 (Kuroki *et al.*, 1990; Mimms *et al.*, 1990), which recognizes an epitope in preS1; and mAb 50-80, which recognizes an epitope in preS2 (these mAbs were the generous gift of L.Mimms, Abbott Labs, Abbott Park, IL). Endogenous polymerase reactions were done with the immunoprecipitated virus using methods previously described (Bruss and Ganem, 1991).

Acknowledgements

We thank Greg Shelness for providing the tripeptide glycosylation competitor, V.R.Lingappa for the prolactin plasmid and Volker Bruss for permission to cite unpublished work. This work was initiated by P.Ostapchuk in the laboratory of D.Ganem and completed in the laboratory of P.Hearing. The work was supported by grants from the National Institutes of Health to D.G. and P.H.

References

Bruss,V. and Ganem,D. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 1059–1063.

Dane,D.S., Cameron,C.H. and Briggs,M. (1970) *Lancet*, **i**, 695–698.

Eble,B.E., Lingappa,V.R. and Ganem,D. (1986) *Mol. Cell. Biol.*, **6**, 1454–1463.

Eble,B.E., Macrae,D.R., Lingappa,V.R. and Ganem,D. (1987) *Mol. Cell. Biol.*, **7**, 3591–3601.

Eble,B.E., Lingappa,V.R. and Ganem,D. (1990) *J. Virol.*, **64**, 1414–1419.

Fuller,S.D. (1987) *Cell*, **48**, 923–934.

Ganem,D. and Varmus,H.E. (1987) *Annu. Rev. Biochem.*, **56**, 651–693.

Gavilanes,F., Gonzalez-Ros,J.M. and Peterson,D.L. (1982) *J. Biol. Chem.*, **257**, 7770–7777.

Guerrero,E., Gavilanes,F. and Peterson,D.L. (1988) In A.J.Zuckerman (ed.), *Viral Hepatitis and Liver Disease*. Alan R. Liss, Inc., New York, pp. 606–613.

Heermann,K.H., Goldmann,U., Schwartz,W., Seyffarth,T., Baumgarten,H. and Gerlich,W.H. (1984) *J. Virol.*, **52**, 396–402.

Huovila,A.-P.J., Eder,A.M. and Fuller,S.D. (1992) *J. Cell Biol.*, **118**, 1305–1320.

Kunkel,T.A. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 488–492.

Kuroki,K., Floreani,M., Mimms,L.T. and Ganem,D. (1990) *Virology*, **176**, 620–624.

Liu,C., Yansura,D. and Levinson,A. (1982) *DNA*, **1**, 213–221.

Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.

Mimms,L.T., Floreani,M., Tyner,J., Whitters,E., Rosenlof,R., Wray,L., Goetze,A., Sarin,V. and Eble,K. (1990) *Virology*, **176**, 604–619.

Patzer,E., Nakamura,G., Simonsen,C., Levinson,A. and Brands,R. (1986) *J. Virol.*, **58**, 884–892.

Persing,D.H., Varmus,H. and Ganem,D. (1986) *Science*, **234**, 1388–1391.

Simon,K., Lingappa,V. and Ganem,D. (1988) *J. Cell. Biol.*, **107**, 2163–2168.

Simons,K. and Garoff,H. (1980) *J. Gen. Virol.*, **50**, 1–21.

Stephens,E.B. and Compans,R.W. (1988) *Annu. Rev. Microbiol.*, **42**, 489–516.

Ueda,K., Tsurimoto,T. and Matsubara,K. (1991) *J. Virol.*, **65**, 3521–3529.

Valenzuela,P., Gray,P., Quiroza,M., Zaldivar,J., Goodman,H.M. and Rutter,W.J. (1981) In B.Fields, R.Jaenisch and C.F.Fox (eds), *Animal Virus Genetics*. Academic Press, Inc., New York, pp. 57–70.

Wigler,M., Silverstein,S., Lee,L.S., Pellicer,A., Cheng,Y.C. and Axel,R. (1977) *Cell*, **11**, 223–232.

Received on September 27, 1993; revised on December 10, 1993