Prediction of a putative fusion peptide in the S protein of hepatitis B virus

Ignacio Rodríguez-Crespo, Julián Gómez-Gutiérrez, Marta Nieto, Darrell L. Peterson and Francisco Gavilanes

Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Químicas, Universidad Complutense, 28040 Madrid, Spain and Department of Biochemistry and Molecular Biophysics, Virginia Commonwealth University, Richmond, Virginia 23298, U.S.A.

Sequence analysis of the S protein of hepatitis B virus (HBV) reveals a stretch of 23 hydrophobic amino acids in the amino-terminal region which shows a high degree of similarity with known fusogenic peptides from other viruses. Additionally, this sequence also appears to be highly conserved within the hepadnavirus family. Taken together, the different criteria used in this work suggest fusogenic activity in the amino-terminal region of the S protein of the envelope of HBV.

None of the proteins of the envelope of hepatitis B virus (HBV), the prototype member of the hepadnavirus family, has yet been found to be involved in the fusion of the viral and cellular membranes which is necessary to allow the virus genome to enter the cytoplasm and carry out subsequent infective events. In fact, the hepadnavirus family is one of the few families of known enveloped viruses whose fusion proteins have not been identified yet. However, there are a few recent reports regarding the initial infective steps of a hepadnavirus, the related virus isolated from ducks, DHBV (Offensperger et al., 1991; Rigg & Schaller, 1992). However, the results of these reports are contradictory and, obviously, further investigation needs to be done in order to establish definitive conclusions.

In order to search for a putative fusogenic moiety in the envelope proteins we have scanned the sequences of these proteins and found a hydrophobic stretch in the amino terminus of the S protein that can be tentatively assigned as a fusion peptide. It consists of a segment of 23 hydrophobic amino acids interrupted by an Arg residue at position 24 (Fig. 1). Several features of this region of the S protein have led us to the conclusion that it might have a fusogenic role in the virus life cycle. Although there is no general motif for all fusion peptides studied to date, there are certainly several special characteristics that can be identified (Glushakova et al., 1990; White, 1990). In a first analysis it can be observed that this sequence is long enough to form a fusion peptide, it is far from glycosylation sites (Peterson et al., 1982), it shows a predominantly extracytoplasmic location, with the N terminus outside and the C terminus inside the viral membrane (Guerrero et al., 1988), and its hydrophobic index is higher than 0.5 according to the scale of Kyte & Doolittle (1982). All these characteristics have been previously assigned to other fusogenic viral peptides.

One of the most important features of a fusion peptide is its high degree of sequence conservation within the virus family (White, 1990). Fig. 1 shows the sequence comparison of this peptide from different subtypes and from other viruses within the hepadnaviruses. It can be observed that it is highly conserved within the different subtypes of S proteins and also within the hepadnavirus family. In addition, most of the unmatched amino acids are conservative substitutions. Perhaps the most outstanding conclusion that can be reached from sequence analysis is the fact that this portion of the S protein of HBV also shares significant similarities with fusogenic peptides from both the retrovirus and paramyxovirus families (Fig. 1b). This circumstance is of very special interest, since sequence homology between different virus families is seldom found (Gallaher, 1987; White, 1990). In fact, in this sequence is the tripeptide Phe-X-Gly, which is present in all paramyxoviruses and also in some retroviruses and which has been directly implied in fusogenic processes. This tripeptide and homologous small peptides are known to inhibit fusion and plaque formation by paramyxoviruses (Kelsey et al., 1990; Richardson et al., 1980). Although this tripeptide is not conserved among the hepadnavirus family, the Phe residue is substituted by a Leu residue [woodchuck hepatitis virus (WHV) and ground squirrel hepatitis virus (GSHV)] or by an Ile residue [DHBV and heron HBV (HHBV)]. These two substitutions seem to be conservative enough to maintain a fusogenic activity. However, the substitution of the Gly residue at position 10 by an Ala residue in DHBV and HHBV is of interest.
since this substitution in the fusion peptide of simian immunodeficiency virus (SIV) has been proved to enhance the fusogenic activity of the virus by sixfold (Bosch et al., 1989).

The observed sequence resemblance to those of some retroviruses is also of interest. Several studies indicate that the hydrophobic amino terminus of the transmembrane glycoprotein gp41 of human immunodeficiency virus (HIV) and gp32 of SIV are primarily responsible for the membrane fusion events involved with virus infection and syncytium formation (Gallaher, 1987; Kowalski et al., 1987). If we use a helical wheel chart to compare the sequence of the S protein of HBV and that of gp41 of HIV (Fig. 2), a similar distribution of residues on both sides of the helix can be observed. Most of the bulkier, more hydrophobic amino acids fall on one face of the helix, and most of the smaller, apolar amino acids, including the majority of the alanine and glycine residues, fall on the opposite face of the helix. In addition, both the amino-terminal sequences of HBV and HIV are known to behave in a similar way when these proteins are cloned and expressed in Escherichia coli. It has been reported that these hydrophobic tails must be deleted both in the S protein of HBV and in gp41 of HIV in order to enable the expression of the entire proteins in this heterologous expression system (Edman et al., 1981; Sisk et al., 1992).

Considering that the S envelope proteins of the hepadnavirus family share greater sequence homology among each other than the M and L proteins (those

---

Fig. 1. Sequence comparison of the amino terminus region of the S protein of different subtypes of HBV (a) and of the same sequence with other members of the hepadnavirus family as well as with fusogenic peptides of some retroviruses and paramyxoviruses (b). Sequence identity with HBV (ayw) is shown in boxes. References for the sequences are: HBV subtypes (Okamoto et al., 1986); WHV and DHBV (Mandart et al., 1984); GSHV (Seeger et al., 1984); HHBV (Sprengel et al., 1988); HIV-1, HIV-2 and SIV (Gallaher, 1987); respiratory syncytial virus (RSV), measles virus (MV), simian virus 5 (SV5) and Sendai virus (SV) (Morrison & Portner, 1991).

---

Fig. 2. Helical wheel charts of the amino terminus sequence, residues 3 to 20 of the S protein of HBV (a) and of gp41 of HIV (b). Hydrophobic residues are labelled in black.
which contain the pre-S sequences), it seems reasonable to assume that a highly conserved sequence of the S protein might be involved in the initial infective step. Moreover, it has been reported that the pre-S1 region of the L protein has a direct role in the interaction with a membrane receptor(s) on the surface of the hepatocyte. The pre-S2 region of the M protein is also suspected to interact with the hepatocyte by means of an indirect association with human polymeric serum albumin (Pontisso et al., 1989). This is in accordance with a putative role of the pre-S domains in terms of distinct receptor recognition and hepatotropism among different animals, and S proteins performing a subsequent fusogenic process. In summary, the amino-terminal region of the S protein might have a fusogenic role in the virus life cycle and, then, might be considered as an important target in the design of immunological strategies for the prevention of HBV infection in vivo.

This research was supported by Grants from the DGICYT (Spain) (PB89-090) and from the National Institutes of Health (U.S.A.) (AI 15955). The authors would like to thank the BioMolecular Engineering Research Center at Boston University for supplying the software necessary for the helical wheel charts.

References


(Received 2 July 1993; Accepted 19 October 1993)