Conservation of Glycoprotein H (gH) in Herpesviruses: Nucleotide Sequence of the gH Gene from Herpesvirus Saimiri

By U. A. GOMPELS,*† M. A. CRAXTON AND R. W. HONESS
Division of Virology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.

(Accepted 4 July 1988)

SUMMARY

We present the nucleotide sequence of the glycoprotein H (gH) gene of herpesvirus saimiri (HVS), a representative of the T lymphotropic herpesviruses of New World monkeys, and compare the predicted amino acid sequence with sequences of homologous proteins from four human herpesviruses. The HVS gH gene is located within a block of genes encoding products conserved in all herpesvirus subgroups as represented by the human herpesviruses herpes simplex virus, varicella-zoster virus, cytomegalovirus and Epstein–Barr virus. In agreement with the biological grouping of HVS as a lymphotropic gammaherpesvirus, its gH amino acid sequence shows greatest similarity to that of the B lymphotropic Epstein–Barr virus, although the nucleotide sequences of their respective gH genes show little similarity given different G+C compositions of 31% and 54%. The similarity observed between the gH amino acid sequences of the two representative gammaherpesviruses is greater than that between the two human alphaherpesviruses varicella-zoster virus and herpes simplex virus. The members of the gH family range in size from 706 to 743 amino acid residues for the beta- and gammaherpesviruses, to 838 to 841 for the alphaherpesviruses, giving non-glycosylated precursors with Mr values of 78322 to 93651. The difference in size is due to heterogeneity in the poorly conserved N-terminal regions of the larger alphaherpesviruses compared to the smaller beta- and gammaherpesvirus molecules. Greatest similarity is observed in the C-terminal halves of the proteins including residues surrounding four conserved cysteine residues, a conserved N-linked glycosylation site (within the sequence NGTV) 13 to 18 residues proximal to the membrane-spanning sequences, and a short cytoplasmic domain of seven or eight residues for the beta- and gammaherpesviruses’ gH. Thus, the representatives of all subgroups of herpesviruses, including those with a non-human host, encode gH homologues. Together with the observation that gH of these viruses are major targets for virus neutralization by antibody, this suggests that this glycoprotein family is essential among all herpesviruses and represents a major component involved in herpesvirus infectivity.

INTRODUCTION

Glycoproteins H and B (gH and gB) are the only virion glycoproteins conserved in members of all three subgroups of herpesviruses (Gompels & Minson, 1986; McGeoch & Davison, 1986; Davison & Taylor, 1987; Cranage et al., 1986, 1988) as represented by the diverse human herpesviruses: herpes simplex virus (HSV) and varicella zoster virus (VZV) for the neurotropic alphaherpesviruses, cytomegalovirus (CMV) for the slow-growing, salivary gland inclusion body agents or betaherpesviruses, and Epstein–Barr virus for the lymphocyte-immortalizing

*Present address: Division of Virology, Department of Pathology, Tennis Court Road, Cambridge CB2 1QP, U.K.

0000-8399 © 1988 SGM
gammaherpesviruses [classification as described in detail by Roizman (1982), Honess (1984) and Honess & Watson (1977)]. Both these glycoproteins function in mediating virus infectivity but in the most characterized herpesvirus, HSV, it is gH rather than gB that acts as a major target for complement-independent antibody-mediated neutralization of virus infectivity and inhibition of cell-to-cell spread (Buckmaster et al., 1984; Para et al., 1985; Gompels & Minson, 1986). These characteristics have also been observed for the gH homologue in VZV, gpIII (Grose et al., 1983; Forghani et al., 1984; Keller et al., 1984; Davison et al., 1986; Montalvo & Grose, 1986; Keller et al., 1987). Consistent with the results shown with antibodies are the studies with an HSV conditional lethal mutant, tsQ26 (Weller et al., 1983), which has a point mutation in the gH gene (Desai et al., 1988). The mutation appears to affect the virus infectivity; intracellular virus retaining gH remains infectious while extracellular virus, devoid of gH, is uninfected (Desai et al., 1988). Although gH of both HSV and VZV are important for virus entry, they appear to be relatively less abundant and immunogenic than other glycoproteins (Buckmaster et al., 1984; Keller et al., 1984). Of 88 monoclonal antibodies (MAbs) directed towards HSV glycoproteins, as summarized by Para et al. (1985), only four were specific for gH (M₁, 110K) (Showalter et al., 1981), although another has since been described (Buckmaster et al., 1984; Gompels & Minson, 1986). Similarly, of 43 MAbs isolated that were specific for VZV glycoproteins, as summarized by Davison et al. (1986), only four were specific for VZV gH (gpIII) and another four have since been described (Montalvo & Grose, 1986). However, these antibodies to gH in VZV and HSV can neutralize virus infectivity without complement and at high titres, and only the non-conserved glycoprotein D of HSV serves as a similar target.

Although less is known about gH homologues in the beta- and gammaherpesviruses, results of studies with antibodies parallel those of the alphaherpesviruses. Of at least 90 MAbs described (Pereira et al., 1982; Rasmussen et al., 1984, 1985; Kari et al., 1986; Cranage et al., 1986, 1988) against CMV glycoproteins only four have been shown to neutralize CMV infectivity without complement. Three of these are specific for the CMV homologue of gH (Rasmussen et al., 1984; Cranage et al., 1988), a component of gCIII which is one of the three major glycoprotein complexes in CMV (Gretch et al., 1988). In EBV there are two major targets for neutralization, gp350/220 and gp85 (Thorley-Lawson & Poodry, 1982; Strnad et al., 1982). Of 42 MAbs raised against EBV membrane antigens only six neutralize virus infectivity (Hoffman et al., 1980; Thorley-Lawson & Geilinger, 1980; Mueller-Lantzch et al., 1981; Qualtiere et al., 1982; Strnad et al., 1982; Sairenji et al., 1985; Yokochi et al., 1986; Balachandran & Hutt-Fletcher, 1986); three are specific for gp350/220 (Hoffman et al., 1980; Thorley-Lawson & Geilinger, 1980; Sairenji et al., 1985), two for gp85 (Strnad et al., 1982) and one human MAb for a 93K glycoprotein (Koizumi et al., 1986). The gp350/220 mediates entry by binding to a cellular receptor for C3d (Nemerow et al., 1987; Thorley-Lawson & Poodry, 1982), while gp85 is the EBV equivalent of gH (Oba & Hutt-Fletcher, 1988; Heineman et al., 1988), which is encoded by open reading frame (ORF) BXLF2 (McGeoch & Davison, 1986; Baer et al., 1984). Thus, with the available evidence, the gH homologue in all these human herpesviruses appears to be important for mediating virus infectivity.

We have been studying herpesvirus saimiri (HVS) of the squirrel monkey which represents a subdivision of gammaherpesviruses, which are T cell-tropic and have an A + T-rich genome composition as opposed to the B cell-tropic G + C-rich EBV. The identification of a gH homologue for such a virus would support the hypothesis that gH performs an essential function, and would assist in the delineation of conserved amino acid sequences which may be functionally important. Although HVS also has a genome structure quite distinct from EBV and encodes a number of proteins not found in other herpesviruses (Stamminger et al., 1987; Desrosiers et al., 1985; Honess et al., 1986; Trimble et al., 1988), it shares with EBV a similar overall genetic organization of conserved genes (Gompels et al., 1988) quite different from that of the alpha- or betaherpesviruses (Davison & Taylor, 1987; Kouzarides et al., 1987). By short sequence analyses we have previously identified the location of the HVS gH gene within a block of genes conserved amongst all human herpesviruses (Gompels et al., 1988). Here we present the complete nucleotide sequence of the HVS gH gene and compare its encoded amino acid sequence to the previously described human herpesvirus gH homologues.
Nucleotide sequence of HVS gH gene

METHODS

Recombinant plasmids. The EcoRI B fragment of the HVS strain 11 (Falk et al., 1972) genome was cloned in the plasmid vector pACYC184 as described by Knust et al. (1983).

M13 subcloning and DNA sequencing. The PstI M 3-9 kb subfragment of the EcoRI B plasmid (Gompels et al., 1988) was purified and the nucleotide sequence determined by shotgun dideoxynucleotide sequencing (Sanger et al., 1977) using methods detailed by Bankier & Barrell (1983). Briefly, the DNA was sheared by sonication (Deininger, 1983), randomly subcloned into M13 mpl8 vectors (Messing, 1983), the sequence was determined using [35S]dATP (Amersham) as a radioactive label and the products were fractionated by Tris-borate-EDTA buffer gradient electrophoresis (Biggin et al., 1983). The 5' end of the gene was determined by nucleotide sequencing of the overlapping 4.8 kb MspI I subfragment of EcoRI B (Gompels et al., 1988).

Computer analyses of sequence data. Sequence data were compiled and analysed on a Dec 20-60 computer (Digital Equipment) using the computer programs of Staden (1982a, 1984, 1986) as described previously (Gompels et al., 1988). Protein similarities were searched for in the PIR (Protein Identification Resource) sequence database (Chen & Barker, 1985) using the procedure of Wilbur & Lipman (1983) as implemented by R. Mott on the Molecular Genetic Sequencing (MGS) set of programs at the National Institute for Medical Research (NIMR), London and by running FASTP (Lipman & Pearson, 1985) on the VAX computer (Digital Equipment) at the Laboratory for Molecular Biology, Cambridge. Dot matrix comparisons were performed using a program similar to DIAGON of Staden (1982b) as implemented by P. Gillett on MGS at the NIMR. Multiple alignments were produced by analyses of all possible double alignments made by the MGS Wilbur & Lipman program. The multiple alignment was optimized by the introduction of gaps using the FORMAT option of MGS.

RESULTS AND DISCUSSION

DNA sequence analysis

By determining the nucleotide sequence of short fragments (300 bp average length) of HVS DNA from selected regions of the genome, we had previously located a number of HVS genes homologous to those of EBV, VZV and HSV (Gompels et al., 1988). Amino acid sequence similarity to residues 36 to 85 of EBV gH (BXLF2) were encoded by sequence at the left end of the 3-4 kb PstI M fragment of EcoRI B with the genome in prototype orientation (i.e. with the 160K protein-coding gene at the right). This gene was shown to be flanked by other genes conserved among the herpesviruses, which additionally have a conserved overall organization in a gene block (Davison & Taylor, 1987; Gompels et al., 1988). Of those encoding products with known function, the genes flanking the HVS gH gene were those encoding the HVS thymidine kinase, adjacent at the left end, and the HVS major capsid protein, separated by one gene from the 3' end (U. A. Gompels, M. A. Craxton & R. W. Honess, unpublished results; Gompels et al., 1988). In this report, we have derived the complete nucleotide sequence of the HVS gH gene by shotgun sequence analyses of the PstI M and the overlapping 4-8 kb MspI I subfragments of EcoRI B from HVS strain 11. The MspI I sequence overlaps the left end of the PstI M fragment and terminates 2815 bp from the PstI site.

In Fig. 1, the nucleotide sequence of 2-46 kb of HVS DNA is shown, which includes the ORF encoding HVS gH from positions 256 to 2406. There are several consensus sequences for promotion and termination of transcription flanking this ORF: two possible TATA boxes at positions 8 and 105 (Bucher & Trifonov, 1986), four upstream simian virus 40 (SV40) 'core' enhancer sequences (Weihat et al., 1983) and a 3' polyadenylation signal, AATAAA, followed by a G + T-rich signal proposed to be involved in efficient 3' termination (Birnstiel et al., 1985; McLauchlan et al., 1985; Zhang & Cole, 1987) (Fig. 1).

The 717 amino acid sequence encoded by the HVS gH gene shows characteristics predictive of a membrane-spanning glycoprotein; these include hydrophobic signal and transmembrane sequences, a positively charged cytoplasmic anchor sequence, and the presence of N-linked glycosylation sites (Schlesinger & Schlesinger, 1987; Kornfeld & Kornfeld, 1985; Walter et al., 1984; Eisenberg, 1984). In Fig. 2, a hydropathic plot indicates two regions that may be located within a lipid bilayer. As shown by Kyte & Doolittle (1982) using their values for hydrophobicity of each amino acid, if the hydropathy averages greater than 1-6 across a span length of 19 there is a high probability that this region will be membrane-traversing. Two such regions exist within the HVS gH sequence, one at the amino terminus and one towards the carboxy terminus (Fig. 2). The region at the N terminus could serve as a signal sequence for cotranslational translocation of
Fig. 1. Nucleotide sequence and deduced amino acid sequence of HVS gH. Restriction sites for PvuII, PstI and XbaI are marked on the nucleotide sequence. Underlined sequences 5' of the initiating methionine codon are SV40 core enhancer sequences in either orientation; dotted sequences indicate region.

Fig. 2. Hydropathy plot of HVS gH amino acid sequence using hydrophobicity scale of Kyte & Doolittle (1982). Scores were averaged within a sliding window of 19. Sequences with a high probability to be membrane-spanning have a score greater than +1.6; this cut-off is marked by the upper line. Sequences at the N- and C-terminal ends are above this cut-off.
Fig. 3. Dot matrix analyses of relation of HVS gH to the gH homologues of the human herpesviruses. In each comparison a K-tuple of 41 was used and a dot marked where there was an average score greater than 10.8 per residue. A score above 10 denotes an acceptable amino acid substitution as deduced from studies of 71 families of closely related proteins. This weight matrix used is Staden's modification (Staden, 1982b) of the relatedness odds matrix determined by Dayhoff (1969; Dayhoff et al., 1983). Complete sequences are compared from the N-terminal methionines, starting at the bottom left-hand corner.

the protein across the endoplasmic reticulum. It is similar in length and composition to other described eukaryotic signal sequences (McGeoch, 1985; von Heijne, 1985). A signal sequence cleavage site may be predicted after Ala 16, as the positions of amino acids in this 16 residue sequence score highest with a weight matrix determined by von Heijne (1986) from evaluations of amino acid counts in each position of known eukaryotic signal sequences. Cleavage at this site would result in a protein of 701 residues with a predicted Mr of 80859. The second extended hydrophobic region towards the C terminus may serve as a transmembrane sequence of 21 residues (Fig. 1). This region is also identified using the 'normalized' hydrophobicity scale of Eisenberg (1984). The eight remaining residues, including one positive charge at Lys 711, may act as a cytoplasmic anchor sequence. The body of the protein from Leu 17 to Arg 688 contains eight potential acceptor sites for N-linked glycosylation (Fig. 1). Thus, the HVS gH sequence has the common features associated with membrane-spanning glycoproteins.

Similarity of the HVS gH gene to homologous genes from human herpesviruses

Using the FASTP program (Lipman & Pearson, 1985) to search for amino acid sequence similarities within the PIR database plus a library of all available herpesvirus sequences, the only significant matches detected were between the HVS gH and the gH of HSV-1 and its homologues in VZV, CMV and EBV. Scores obtained were 118, 118, 305 and 776, respectively, compared to a mean score of 28 [standard deviation (s.d.) 7], and for the 10 next best selected a mean score of 54 (s.d. 12). By dot matrix analyses, the HVS product is clearly homologous with the EBV gH but also shows limited similarity to all the gH glycoproteins in the C-terminal portions of the sequences (Fig. 3). The similarity between the HVS and EBV products is more than that shown between VZV and HSV gH homologues (Fig. 3; see also McGeoch & Davison, 1986). Furthermore, by this method the CMV gH shows more similarity to the HVS gH sequence than to the homologues in the other human herpesviruses (Fig. 4).

In Fig. 5, the homology between the HVS gH and EBV gH proteins is shown in more detail by an alignment of the two amino acid sequences. The overall similarity is 36% while the C-terminal halves show 47%. Of the nine cysteines conserved overall, six out of six are conserved in the C-terminal halves, whereas three out of seven possible residues are in the N-terminal halves. Most of the sequence similarity is clustered around the six conserved cysteines towards the C terminus. The organization of the two sequences appear similar. Using the rules of von
Heijne (1986), both sequences have predicted signal sequences of 16 and 17 residues, which may be cleaved to give proteins of $M_r$ 80859 and 76403, respectively. All five sites for $N$-linked glycosylation in EBV gH (Strnad et al., 1983) are conserved in position in HVS gH. Two of these sites are also conserved in sequence, one towards the N terminus (WREANVTE) and one towards the C terminus, 11 to 13 residues before the proposed transmembrane sequences, in the motif NGTV. Both transmembrane sequences are followed by a short (eight amino acids) positively charged sequence which may serve as the 'stop transfer' or cytoplasmic anchor sequence.

By the dot matrix analyses (Fig. 3 and 4) differences and similarities between the N- or C-terminal portions of the gH proteins appear to distinguish the virus subgroups. In the C-terminal region the CMV gH appears to be more closely related to the HVS gH than to the homologues in the other human herpesviruses (Fig. 4), but in the N-terminal domain there seems to be conservation only between gH homologues from the same subgroup: HVS and EBV of the gammaherpesviruses, VZV and HSV type 1 (HSV-1) of the alphaherpesviruses. Dot matrix analyses, using the Dayhoff weight matrix for similar amino acids (Dayhoff, 1969; Dayhoff et al., 1983), have been shown to be one of the most sensitive methods for showing similarities between distantly related proteins (Feng et al., 1985; Staden 1982b). By this method, the HVS, EBV and CMV gH homologues can be grouped separately from those of VZV and HSV.

**Conserved domains in the gH homologues**

There are three regions which show similarity between all the gH homologues, one possible N-terminal domain and two regions in the C-terminal half of the proteins (Fig. 6a, b, c). Multiple alignments of the HVS, EBV and CMV gH with the VZV and HSV gH indicate that the size difference between these groups originate at the amino termini (Fig. 6a). The total numbers of amino acid residues of the HVS, EBV and CMV gH are 717, 707 and 743 respectively, whereas those of the HSV-1 and VZV gH are 838 and 841. The N-terminal sequence of the HVS, EBV and CMV gH can be aligned within the VZV and HSV-1 gH sequence starting at residue positions 77 and 104 respectively (Fig. 6a). The limited similarity between the 88 to 127 amino acid residues aligned here may be explained by expansion or contraction of gH genes to give rise to encoded N-terminal signal sequences of HVS, CMV and EBV gH derived internally from the alphaherpesvirus gH sequence.

The second two regions with similarity between all the gH homologues are in the C-terminal halves of the proteins. In comparisons between any two herpesvirus gH this region also shows the greatest similarity. As shown in comparisons between the EBV and HVS gH homologues...
Nucleotide sequence of HVS gH gene

MT I DVLFLM I LELCDQDLPKPR I MKPPAERKLKLNGYNTTLIEFDDGVSFMNLHV 60/HVS
NOLLCEFLULLWEGARLSVEUKLHLDIEGASHYTIPLN 40/EBV

TK I IETHIPDEL I LLELAERAAENTPELUNLTLLKRSDTYPETNHU1HPOGHSYRALCYPWYSY
TELARNKUGPSLEARHANEDLASNLHAYKLVIKTSGTGLIREPUDPVLPASEGSM

TQDQEDEAK--TTGQLMHDGPISTKTLNEVFDFDVQUYHFQTVETAYVLCHIGATTP 178
QVQASKWMPUGISGLSPACMLSAFPELQKFLYVITMPLATHTPSYVQALCHSVARYL

AVPTISCHT----PNYL----FUSU-EFTKDSLTLLFHSHLYPLKGIYVNDIEGAS
SINGDKFQSYGATSKFLNGTVKRGKDEUMUSLFUGKTDLPSLGGPSFLSSLTSAQ 160

MDUSFLUFSVTDFGKHNSEFSDIFLUKCEFIEETPFLIKQDNEMFTIEIDGCNI 290
SGDYSLUUTTFVUHYNLHMYFUPMHLKOMPSRAVATITASRYAVYULQKLLUNKEN- 279

MN I I HNP KTLFRAKSHA WFLUIDSL---ATQOHILLMCAMYMSELEFLKMLMSECFEFF
EPECQELTLTFEVSURSHULKHCYLAFLAKSFELIULDIICGCYRAT 408

EEFP FY TI ETL A RASQ AL HPM UXI TSLSDQTMLSL FRA SHS HMUS ARI SE I IDL
UKQNGQSYGLELARLNLNLEKNEGHLTHTEKEQERVALATUVPKGAVYGSGLIGATSU 399

ISHIVITYA S-YTHM LTSKE AEKML LDAVI ULDHNMKHKHTUKQDO LHPY-ULSS SMS T I 
LLSAYWALPFPOLHTYRATLFGS HUA RE-ALM HT IQOGPM LALV Y L STI CQ SI 

I GQMLH FQGOKDUVETFSPC YLSHMNF DFTIEKLIETFPOSSL1IQAEK I NLGTNGF Q 526
IGEV LA EDLG TESGL----FSCP YLSMTDLAKLNSAPQ EALDQAASVHUDLG 

TLMN-AMHTSLE I LP I KCI KSLSTD I ILSI PLKH NYV UST-KPUPMKSIYQUSEFLK
ALSLEERDA九龙PKVCCICR-OKULNIPLIUNTFI I SDEVRAGSLYESTT YLS

TSHA S -UNO CKP YGGSARRQH IQV YITYUP PRGCYPPSSUYSDEQGFOSSHNY 643
SSFLSPUMHCSCQAO VAPQPK 1 OPHET Q TQ K SICFQGSFALLSYDEKEGLETTVY 634

TDTVQENLTFENSSPFDDGKMLHILHVILAMAGTUIERGAAYARLUNGFI II U I N U F E FL
TSEOUQMLS- SNYFDOFLMHLVYLLTTNGTUMEAGLYVEERAMULALLLYFAFA 

UGLYL LVLKFLUYLT (HVS gH aa 1 to 717)
LGFLFKMTCWFL (EBV gH aa 1 to 706)

Fig. 5. Alignment of HVS gH with EBV gH. The alignment was optimized from the output of FASTP
(Lipman & Pearson, 1985) using a K-tuple of 1, after a search of the PIR database. Identical and similar
residues are marked with an asterisk according to the following match set: L = V = I; R = K; D = E;
S = T; Y = F. Potential signal and membrane-spanning regions are underlined, as well as five
conserved sites for ?v:linked glycosylation. Conserved cysteines are marked by filled circles. The
coordinates for each sequence are as marked; HVS gH is the upper sequence, EBV gH the lower
sequence.

(Fig. 5), regions of similarity centre around six conserved cysteines and a conserved
glycosylation site 11 to 15 residues before conserved hydrophobic sequences which may serve as
transmembrane domains. In HSV-1, two of the conserved cysteines are not present (Fig. 3c), but
it is not known whether this is peculiar to the two laboratory strains analysed, HFEM and strain
17 (Gompels & Minson, 1986; McGeoch & Davison, 1986). On the basis of overall sequence
similarity in this region, HVS, EBV and CMV gH can again be grouped separately from the
VZV and HSV gH (Fig. 6b and c). The potential transmembrane sequences also show similarity
within the HVS, EBV, CMV gH group rather than to those of VZV and HSV. Similarities in the
first group include Phe, Tyr, Leu and Ile residues, whereas the second group favours Ala or Gly.
Fig. 6. Three regions of similarity between the herpesvirus gH amino acid sequences: (a) N-terminal, (b) and (c) C-terminal. Identical and similar residues are boxed (match set as for Fig. 5), gaps introduced to maximize the alignment are marked with a dash. Asterisks mark beginning and ends of sequences. The coordinates of each region are given. Conserved potential membrane-spanning regions are marked by open circles.
This distinction continues with the characteristically short cytoplasmic anchor sequences. Of all the herpesvirus glycoproteins described only the gH homologues have a short anchor sequence of less than 15 residues; however, the HVS, EBV and CMV gH homologues have only seven or eight residues in this domain whereas the HSV and VZV gH have 14 or 15 residues. These differences in potential membrane-spanning and 'stop transfer' sequences may reflect distinctions in membrane interactions with different cell types.

As the strongest sequence conservation exists around the conserved C-terminal cysteines in the herpesvirus gH homologues, a common structure involving disulphide bonding is suggested. It is possible that this structural feature would be part of a domain with conserved function. As described in the introduction, the role of this glycoprotein in herpesvirus infection appears to be essential. MAbs directed against individual gH homologues are potent neutralizers, independent of complement, of the infectivity of the respective herpesvirus. In addition antibodies to the alphaherpesvirus HSV and VZV gH prevent cell-to-cell spread of virus and inhibit cell fusion induced by syncytial strains. It seems likely that the conserved C-terminal domain of the gH homologues plays a critical role in the infectivity of herpesviruses.

We are grateful to A. Bankier and B. Barrell, Laboratory of Molecular Biology, Cambridge, for access to the CMV gH sequence before publication, and for use of computer facilities.

REFERENCES


(Received 18 April 1988)