JOURNAL OF VIROLOGY, Dec. 1985, p. 807–813
0022-538X/85/120807-07$02.00/0
Copyright © 1985, American Society for Microbiology
Vol. 56, No. 3

Epstein-Barr Virus Genome May Encode a Protein Showing Significant Amino Acid and Predicted Secondary Structure Homology with Glycoprotein B of Herpes Simplex Virus 1

PHILIP E. PELLETT,[1] MARK D. BIGGIN,[2]† BART BARRELL,[2] AND BERNARD ROIZMAN[1]*

Marjorie B. Kovler Viral Oncology Laboratories, The University of Chicago, Chicago, Illinois 60637,[1] and The Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, United Kingdom[2]

Received 27 June 1985/Accepted 28 August 1985

We report significant sequence and predicted secondary structure homology between the herpes simplex virus 1 glycoprotein B (gB) and a protein predicted to be encoded by the BALF4 reading frame of Epstein-Barr virus (EBV). Homology was detectable at the DNA level and was highly significant at the protein level and when evolutionary substitution frequencies of amino acids in related proteins were taken into account. Hydropathic analyses predicted that the two proteins possess conserved N-terminal and C-terminal hydrophobic domains. The N-terminal hydrophobic domains share features in common with known cleavable membrane insertion signal sequences. The amino acid sequences of the C-terminal hydrophobic domains predict three adjacent membrane-spanning segments as had been previously predicted for gB. In an alignment of the two amino acid sequences, 247 of 903 gB residues had a matched pair in the BALF4 sequence, and 247 of 854 BALF4 residues were found to have a matched pair in the gB sequence. In addition, all 10 cysteine residues located outside the predicted signal sequence of both proteins were conserved, as were four predicted N-linked glycosylation sites. In all, 43% of the residues in the aligned sequences are predicted to possess equivalent secondary structures. gB is a virion envelope glycoprotein required for virus entry into cells. The domain of gB determining the rate of entry into cells has been mapped; the predicted structure of this domain in gB and the predicted EBV protein are almost identical. Similarly, the cytoplasmic domain of gB postulated to interact with submembrane proteins was also nearly identical in predicted structure to that of the EBV protein. These results suggest that EBV encodes a protein similar in structure and function to the herpes simplex virus 1 gB.

The herpesviruses constitute a family of related viruses characterized by a core containing double-stranded DNA enclosed in a rigid icosadeltahedral capsid surrounded by proteins (the tegument) and a membrane (the envelope). Herpesviruses have been isolated from most vertebrates examined for the presence of these viruses. In addition to the morphology of the virion, the members of the herpesvirus family share a similar reproductive cycle and ability to remain latent in their hosts. The herpesviruses differ, however, in the size of their genome, host range, and in other biologic properties (27).

Of the five herpesviruses that infect humans (herpes simplex virus 1 [HSV-1], herpes simplex virus 2 [HSV-2], Epstein-Barr virus [EBV], cytomegalovirus, and varicella-zoster virus), only HSV-1 and HSV-2 are closely related as measured by DNA homology (17) and viable recombinant progeny (22). Because HSV-1 and HSV-2 grow readily and yield high-titer progeny in a variety of cell species and nonhuman hosts, they have been extensively characterized genetically and biochemically. The studies on the function of EBV gene products have lagged largely because the poor growth of EBV in cell culture has precluded the isolation of conditionally lethal mutants.

One approach toward the identification of EBV genes and elucidation of their function is based on the assumption that although EBV and HSV diverged significantly from a common ancestor, gene products performing similar functions may retain some amino acid sequence homology and secondary structure features. This approach has led to the discovery that several HSV-1 genes have DNA and protein sequence homologs in EBV. These include the ribonucleotide reductase (12), the DNA polymerase (1), and proteins of unknown function (6; unpublished data). In this paper, we report that an open reading frame of the EBV genome designated as BALF4 (EBV coordinates 159,322 to 156,752 [1]) may encode a protein sharing significant amino acid sequence homology and many predicted structural features with the glycoprotein B (gB) specified by HSV-1.

Relevant to this report are the following properties of the HSV-1 gB (for a review, see reference 31). gB is one of the five major glycoproteins specified by HSV, a major target of the host immune response and the only HSV glycoprotein whose gene has yielded well-characterized, conditionally lethal and nonlethal mutations. Analyses of these mutants has led to the conclusion that gB is involved in the fusion of the virion envelope to the cellular membrane and in the fusion of cells (Syn^+ phenotype). Another class of well-characterized mutants exhibit altered reactivity to neutralizing antibody directed against gB (14, 18). A population of the HSV-1 gB extracted from virions is a dimeric protein (29). Structural predictions (24) for gB based on genetic, biochemical, immunological, and DNA sequencing studies indicated that gB is a 903-amino-acid protein consisting of a N-terminal domain projecting into the extracellular environment containing six potential sites for N-linked glycosylation and mutations affecting reactivity with neutralizing antibody, a hydrophobic domain consisting of three membrane-spanning segments, and a positively charged domain carrying the Syn^- mutations and projecting into the cytoplasmic side of the membrane.

* Corresponding author.
† Present address: Ludwig Institute for Cancer Research, Cambridge CB2 2QH, United Kingdom.

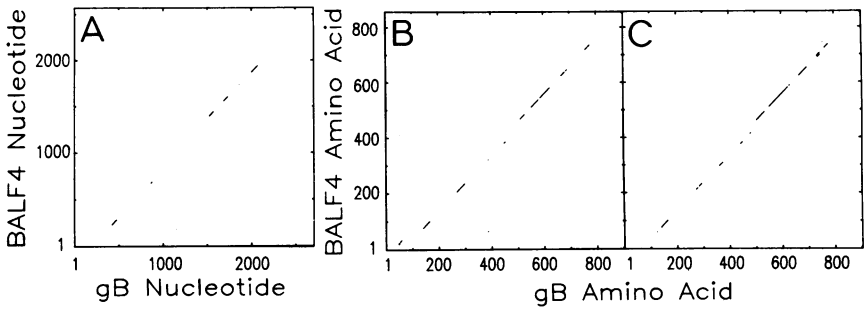

FIG. 1. Dot homology matrix plots at the DNA and protein levels between HSV-1 gB and EBV BALF4. In all three plots gB is represented on the horizontal axis, and BALF4 is represented on the vertical axis. (A) DNA sequence comparison between the coding regions of gB and BALF4. A dot is plotted if a minimum of 34 nucleotides within a 57-nucleotide window match. (B) Protein sequence comparison between gB and BALF4. A dot is plotted if 8 amino acids within a 19-residue-wide window match. (C) Comparison of evolutionary relatedness between gB and BALF4. A dot is plotted if a score of evolutionary relatedness (see Materials and Methods) summed over a 19-amino-acid window is equal to or greater than 230.

In this report, we extend these predictions to the protein encoded by the BALF4 reading frame of EBV and conclude that gB and BALF4 have arisen from a common progenitor gene and share functional domains.

MATERIALS AND METHODS

DNA sequences. The nucleotide sequences of the HSV-1 gB gene and of the BALF4 reading frame of EBV have been previously reported (2–4, 24). The sequence of the HSV-1 strain F gB (24) was used in all analyses reported here.

Computer analysis of the sequences. The homology was originally found by using the LIBSEQ program of A. D. McLachlan (unpublished data). A dot matrix analysis program with an algorithm similar to the program DIAGON (32) was used for graphic comparison of the two sequences. Comparisons based on evolutionary substitutions were done as described by Dayhoff (7), with the amino acid substitution values of Staden (32). This comparison is based on the assignment of values for specific substitutions of one amino acid by another amino acid, as deduced from substitution frequencies among related sequences. The conservation of some amino acids receives a higher value (e.g., a value of 27 is assigned for a conserved Trp) than does conservation of other amino acids (e.g., a conserved Ala receives a value of 12). The difference in assigned values between a Trp and an Ala conservation reflects the relative ease with which Ala can be substituted evolutionarily by other amino acids and, conversely, the relative difficulty of substituting for Trp. The extremes of values assigned by the substitution matrix for heterologous amino acid replacements are represented by a value of 2 for a Cys-to-Trp substitution and 17 for a Phe-to-Tyr substitution.

The initial alignment of the two sequences was done by using the program of Wilbur and Lipman (34), and adjustment of the alignment to take into account predicted structural features was done with the aid of the "scissors and glue" alignment program obtained from Hugo Martinez at the University of California at San Francisco. Hydropathic analysis was done as determined by the procedure of Kyte and Doolittle (19), and amphipathic analysis was done as described by Eisenberg (8, 9).

RESULTS

Comparison of nucleotide and predicted amino acid sequences of HSV-1 gB and of EBV BALF4. The diagonal in the plot of Fig. 1A indicates the homologous domains in the coding regions of gB and BALF4. The most striking homology was exhibited by a region of approximately 500 nucleotides encoding a portion of the C-terminal half of the molecules, but some similarity was present in other domains of the coding regions.

A comparison of the amino acid sequences of gB and BALF4 is shown in Fig. 1B. In this plot, the similarity of the two sequences was much more apparent inasmuch as it negates the effect of evolutionary randomization of codon usage on nucleotide sequence homology. These results indicated that the regions of homology at the DNA level were a subset of the homologous regions seen in the amino acid comparison and that the C-terminal half of the HSV-1 strain F gB is the most conserved domain of the gene. If the evolutionary relatedness of amino acids is taken into account, an even higher degree of relatedness is seen (Fig. 1C).

Hydropathic analyses of the proteins predicted by the DNA sequences of HSV-1 gB and EBV BALF4. Confirmation of the evolutionary relatedness of the coding sequences of HSV-1 gB and EBV BALF4 emerged from analyses of the predicted structures of the proteins. Figure 2 presents the hydropathic profiles of gB and BALF4 constructed as described by Kyte and Doolittle (19). The most striking features seen in the plots are the N-terminal hydrophobic maxima and the three hydrophobic peaks near the C terminus of both molecules.

It was previously predicted that the N-terminal hydrophobic region of gB could serve as a cleavable signal sequence, because the amino acid sequence within this region was in agreement with known features of such sequences (24). A similar analysis of the BALF4 hydrophobic region indicates that it also has the features of a cleavable signal sequence (25, 33). Specifically, the hydrophobic core sequence is preceded by a region containing positively charged residues (25) and is followed by a sequence that fits the cannonical cleavage recognition site (25, 33) which is near a predicted beta turn (33). In addition, the hydrophobic core sequence of both molecules contains a Leu at core position 2 and a Val-Val pair at core positions 7 and 8, features that have been found to be conserved in eucaryotic and procaryotic signal sequences (25). The cleavage of the gB signal sequence was predicted to occur after the Ala_{29} (24), whereas the BALF4 signal sequence is predicted to be cleaved after Ala_{22}.

The hydrophobic domain near the C terminus of HSV-1 gB was previously predicted to constitute a three-segment membrane-spanning domain of the glycoprotein (24). Although sequence conservation between HSV-1 gB and EBV

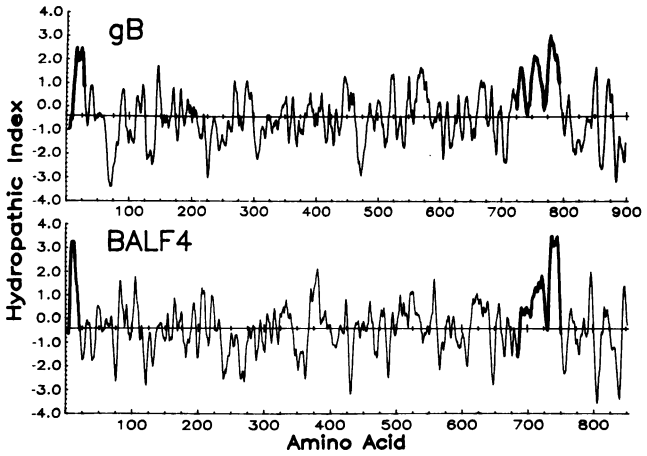

FIG. 2. Hydropathic analysis of HSV-1 gB and EBV BALF4. The hydropathic profiles of both proteins were determined by using the algorithm of Kyte and Doolittle (19) with a moving window of 7 residues. The subsequent profile was smoothed for plotting by taking its average in a moving 3-residue-wide window. The *x* axis of the plots is drawn at the average hydropathicity value deduced by Kyte and Doolittle so that points above the line are of above-average hydrophobicity and those below the line are of above-average hydrophilicity. The portions of each molecule predicted to function as the signal sequence and membrane-spanning anchor sequence are highlighted by heavier lines at the N-terminal end (left) and C-terminal end, respectively.

BALF4 in this region is less extensive than elsewhere, the presence of three hydrophobic maxima near the carboxyl terminus of BALF4 (Fig. 2), of width sufficient for the amino acids in the segments to span the lipid bilayer, indicated that BALF4 may possess a structure similar to that predicted for gB. Hydrophobic moment analyses (Fig. 3) and helical wheel analyses (30; data not shown) of the three hydrophobic segments indicate that the predicted membrane-spanning helices of both proteins have the features expected of membrane-bound helices, which interact with other membrane-bound protein structures in that the segments are simultaneously amphipathic and hydrophobic, i.e., one face of each helix contains smaller residues and polar residues which could allow interactions with other protein segments, while the opposite side of the helix is composed of nonpolar and relatively larger residues which would interact with the nonpolar lipid portion of the membrane. Analyses of crystallographic data bases indicate that inter- and intrahelical hydrogen bonding and salt bridges within hydrophobic regions occur frequently and significantly stabilize the structures (13, 26). Theoretical thermodynamic considerations also argue that charged residues may be stably incorporated into membranes (10, 15). These considerations lead us to predict that the membrane-spanning region of both gB and the protein encoded by BALF4 are composed of three helical membrane-spanning segments which cluster together such that the exterior surface of the three-helix cluster is nonpolar and is able to interact stably with the lipid environment of the membrane, whereas the interior, helix contact portion of the structure is stabilized by hydrogen bonding and salt bridges. It is possible that other proteins interact with this intramembrane structure.

Alignment of HSV-1 gB and EBV BALF4 predicted amino acid sequences. The hydropathic analysis and structural predictions for the hydrophobic regions provided the basis for the amino acid alignment shown in Fig. 4. At the N termini, the signal sequences were aligned first by the last positively charged residue before the hydrophobic core, which brought the conserved features within the core into alignment, and then a two-residue gap was introduced into BALF4 to align the predicted signal peptidase cleavage sites. The domains immediately downstream from the signal cleavage site show two instances of divergence which could reflect either a deletion in BALF4 or an insertion in gB. The homologous alignment begins again at gB position 108 and BALF4 position 44. In the remainder of the alignment, there is one other region in which a large gap containing a small

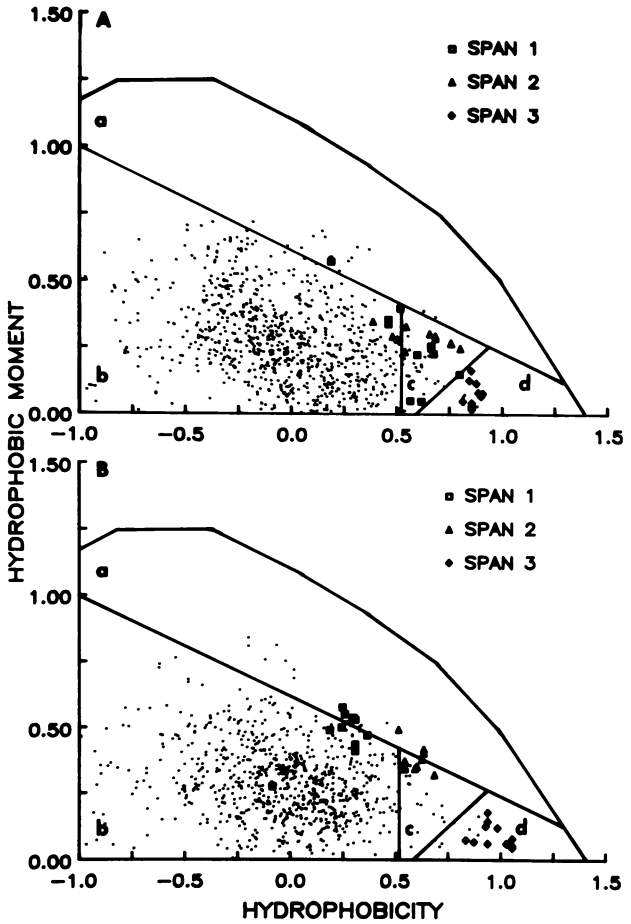

FIG. 3. Helical hydrophobic moment analyses of HSV-1 gB and EBV BALF4. (A) Helical hydrophobic moment plot for HSV-1 gB. (B) Helical hydrophobic plot for EBV BALF4. The plots were constructed as described in the procedure of Eisenberg (8, 9) by using his normalized consensus scale of hydrophobicity (8) and calculating the hydrophobicity per residue (*x* axis) and the hydrophobic moment per residue (*y* axis) for 11 residue segments. The points indicated with squares are from the 11-residue segments totally within the predicted membrane-spanning segment 1 (Fig. 4); the triangles and squares represent 11-residue segments totally within spans 2 and 3, respectively. The portion of the plots labeled a is that portion of the plot indicative of surface-seeking α-helices, the portions labeled b, c, and d indicate globular protein character, multimeric transmembrane character, and monomeric transmembrane character, respectively (8). The points indicated by dots are those determined for every 11-residue segment in the proteins. This makes the implicit and otherwise unwarranted assumption that every residue in the two proteins is in an α-helical conformation but serves as a reference for the nonrandom clustering of the points representing the predicted membrane-spanning segments.

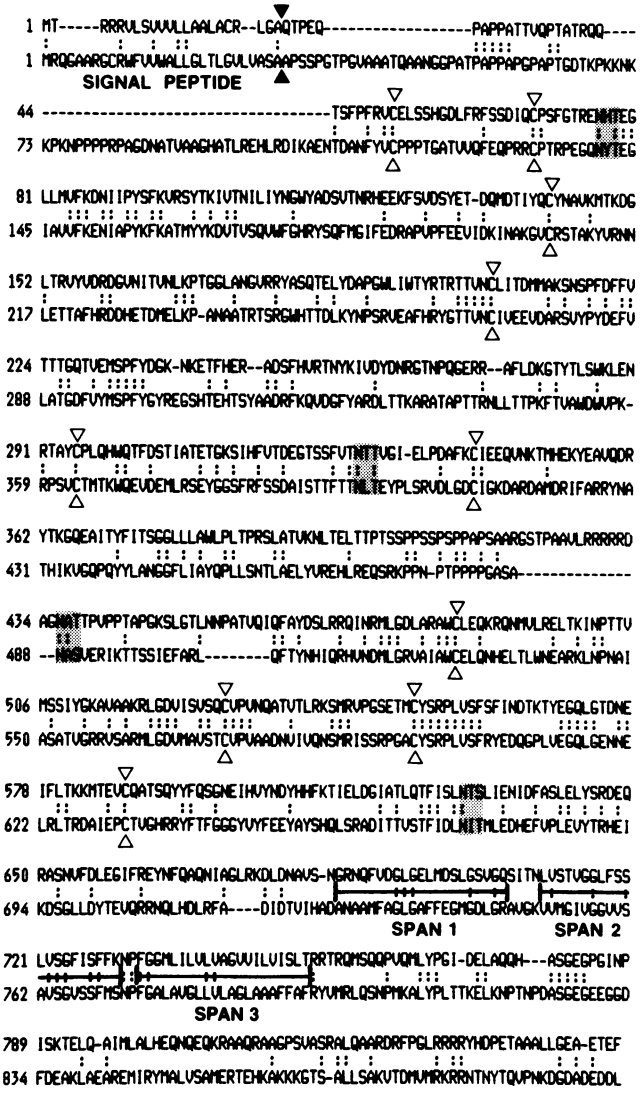

FIG. 4. Alignment of the amino acid sequences of HSV-1 strain F gB and EBV BALF4. Gaps inserted in the sequences to account for the difference in sequence length and to increase the number of aligned residues (–) and residues which are equivalent in the aligned sequences (:) are shown. Other residues indicated include aligned potential N-linked glycosylation sites (▒), aligned Cys residues (△), the predicted site of cleavage of the signal peptides (▲) and the predicted membrane-spanning segments (—). BALF4 is the upper sequence.

homologous patch must be introduced into one sequence (gB) to allow alignment of the remainder of the sequence. In the alignment shown in Fig. 4, 247 of the 903 gB residues have a matched pair in the BALF4 sequence.

The most striking feature of Fig. 4 is the alignment of all Cys except the one which occurs in each signal sequence. This observation suggests that the constraints imposed by Cys-Cys disulfide bonds are important determinants of the tertiary structure of the proteins and that their structures are similar, notwithstanding the considerable divergence in their sequences. The predicted similarity in the tertiary structure is reinforced by the observation that four of six potential N-linked glycosylation sites of gB and four of nine potential BALF4 glycosylation sites are aligned (Fig. 4). The alignment suggests that their location on the surface of the molecule may be important for either correct processing or the function of the molecules.

The obvious relatedness of the two sequences is reinforced by the presence of a few stretches of identical amino acid sequence; the longest is a predicted stretch of 9 amino acids, beginning at amino acid 595 of gB and 455 of BALF4. This region occurs in the middle of the portion of the molecules with the highest degree of sequence identity.

A quantitative assessment of an alignment similar to that shown in Fig. 4, made with the program of Lipman and Pearson (20), which takes into account the evolutionary substitution matrix used in Fig. 1, scored the relationship between the two sequences as very highly significant, with a Z value (20) of >100; a Z value of >10 is taken as significant.

Comparison of the predicted secondary structures of HSV-1 gB and EBV BALF4. Previously, the secondary structure of gB was deduced from hydropathic, helical wheel, and empirically based, secondary structure predictive analyses (24). The same procedures were then applied to the BALF4 amino acid sequence, and the results of the analyses were drawn so as to represent the aligned portions of the sequences in approximately equivalent positions of the diagrams (Fig. 5). Although currently available secondary structure predictive schemes often fail to anticipate the structure derived from X-ray crystallographic analyses (16, 23) and such predictions should therefore not be taken as the definitive structure, comparisons of the secondary structure predictions for related proteins can be useful for identifying regions of similar structural character in much the same way that the evolutionary substitution data for amino acids (7) has been used to determine possible relatedness of amino acid sequences. The techniques used here predict that a significant fraction (43%) of the aligned domains of the molecules shown in Fig. 4 have equivalent secondary structure. The diagrammatic representations show that there are several domains of nearly identical predicted secondary structure. In particular, the portion of the molecules represented as residing on the underside of the membrane has almost identical predicted secondary structure in both molecules, notwithstanding extensive amino acid sequence divergence (Fig. 4).

DISCUSSION

We report here considerable sequence and predicted structural similarity between the HSV-1 gB and the predicted product of the BALF4 reading frame of EBV. As could be predicted from the difference in the base composition of the two genomes, the sequence similarity was least extensive at the nucleotide level (Fig. 1A), strong at amino acid level (Fig. 1B), and even stronger when the comparison took into account evolutionary amino acid substitutions calculated as described by Dayhoff (7) (Fig. 1C).

Structural considerations. The specific structural features shared by the predicted products of gB and BALF4 may be summarized as follows.

(i) The N-terminal domains of both proteins share structural features common to cleavable signal sequences for membrane translocation (25, 33). Thus, in both proteins a hydrophobic core sequence is preceded by a region containing positively charged residues and is followed by a sequence which fits the canonical cleavage recognition site and is near a predicted beta turn. In addition, the hydrophobic core sequence of both molecules contains a Leu at core position 2 and a Val-Val pair at core positions 7 and 8, features which have been found to be conserved in eucaryotic and procaryotic signal sequences (25). On the basis of their structure, the signal sequences of gB and BALF4 are

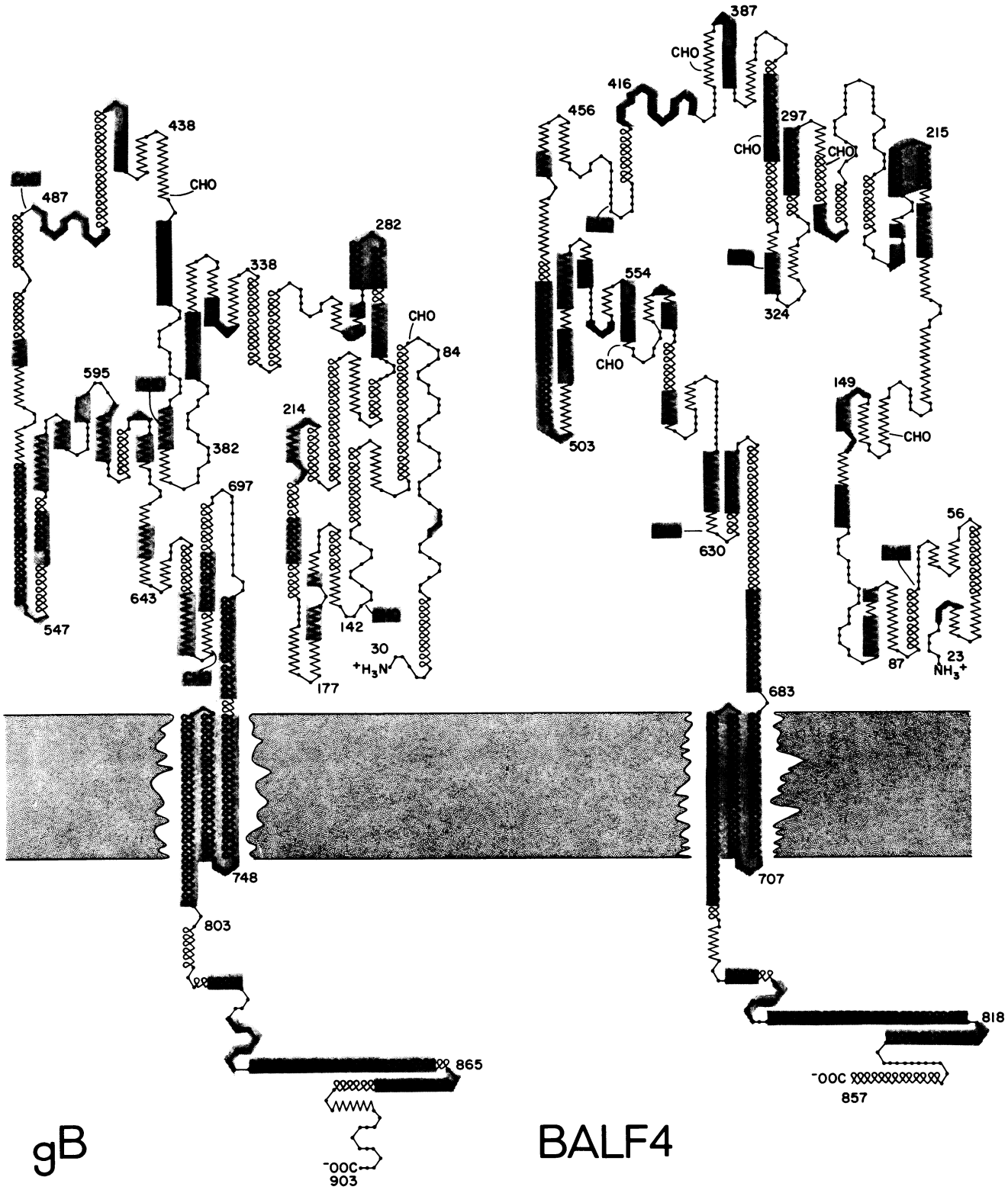

FIG. 5. Predicted secondary structures of gB and BALF4. The secondary structures of the proteins were predicted by using a combination of hydropathic (19), helical wheel (30), and empirically based secondary structure analyses (5, 11). Predicted alpha helices are represented by ⌇⌇, predicted beta sheets are represented by ⩚, beta turns are represented by 2- to 4-residue changes in chain direction, and random coil regions are represented by ⟿. Regions of identical predicted secondary structure are highlighted by shading. Sites with potential for N-linked glycosylation are indicated by −CHO, and those glycosylation sites which are conserved in the aligned sequences are highlighted with shading. The predicted membrane-spanning region is shown passing through a membrane, with the cytoplasmic or virion interior side being at the bottom of the picture. The N-terminal residue in the drawings is the first residue predicted to follow the signal peptidase cleavage site. The signal sequences are not shown in this diagram.

predicted to be cleaved after the Ala at amino acids 29 and 22, respectively.

(ii) The largest fraction of both molecules is predicted to project into the extracellular environment. The degree of structural conservation of this domain of the molecule is likely to be higher than that demonstrated by primary sequence or secondary structure comparison. This conclusion is based on the observation that every cysteine and four N-glycosylation sites in the mature proteins are readily aligned and together are likely to impose dominant constraints on the secondary and tertiary structures of these molecules.

(iii) A domain located near the carboxy terminus of both molecules is predicted to form three interacting membrane-spanning segments. The model structure for both proteins consists of three helical membrane-spanning segments which cluster together such that the exterior surface of the three-helix cluster is nonpolar and is able to interact stably with the lipid environment of the membrane, whereas the interior, helix contact portion of the structure is stabilized by hydrogen bonding and salt bridges. It remains a possibility that other proteins interact with this intramembrane structure.

(iv) The C-terminal domains of the two proteins are predicted to project into the cytoplasm; these domains exhibit the most striking example of predicted structure conservation, notwithstanding relatively little sequence homology.

Functional considerations. HSV-1 and EBV differ in host cell range. The limited human host cell range of EBV reflects at least in part the availability of receptors for adsorption and entry of EBV. Since gB very likely interacts with the constituents of host cell membranes during adsorption and entry into susceptible cells (21, 29), it could be argued that the BALF4 gene product has similar functions and that some of the differences in sequence and structure reflect not only genetic drift but also differences in the structure of host cell membranes. It could also be, however, that conserved sequences reflect domains essential for the secondary structure of the molecule. While little can be inferred from divergent structures, the conserved structures serve as powerful indicators of domains which might be responsible for the function of the molecules.

It has been postulated that gB is part of a multiprotein complex which determines the structure of infected cell membranes and the social behavior of infected cells (28). The C-terminal portion of the gB molecule is the domain which would be expected to interact with virion tegument proteins and possibly with other membrane proteins to affect the social behavior of infected cells inasmuch as the Syn^- mutations in gB have been mapped in that location (4, 24), and one *syn* lesion was explained by a substitution of amino acid 857 (4). This amino acid is located in the long predicted alpha helix ($gB_{836-864}$ and $BALF4_{888-816}$). A portion of this region comprising gB residues 841 to 858, when drawn as a helical wheel or analyzed by using hydrophobic moments (data not shown), appears to have an amphipathic character in that there is a hydrophobic and a hydrophilic face. It is tempting to speculate that this portion of the structure is involved in specific interactions with another protein.

Another well-conserved portion of the two molecules is found between gB residues 524 and 573 and BALF4 residues 480 and 529. This portion of the molecule may be involved in cell entry, because gB amino acid 552 is found to be altered in fast entry strains relative to slower entering strains (4). This function may place this portion of the molecule under selective pressure to maintain a membrane fusion function and, out of necessity, to conserve the structure of this portion of the molecule.

This paper reports comparisons of the proteins predicted by sequences of two herpesviruses only. It is likely that as more genes homologous to gB and BALF4 are identified and more common domains become known, it might be possible to define more precisely the conserved functional domains of this family of essential herpesvirus glycoproteins.

ACKNOWLEDGMENTS

We thank Garret Toomey for computer programming assistance.

The studies at the University of Chicago were aided by Public Health Service grants CA-08494 and CA-19264 from the National Cancer Institute and by American Cancer Society grant MV-2R to the University of Chicago. P.E.P. is a predoctoral trainee supported by Public Health Service training grant T32 PHS AI 07182.

LITERATURE CITED

1. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deininger, P. J. Farrell, T. G. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Seguin, P. S. Tuffnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**:207–211.
2. **Bankier, A. T., P. L. Deininger, P. J. Farrell, and B. G. Barrell.** 1983. Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr Virus. Mol. Biol. Med. **1**:21–45.
3. **Bzik, D. J., B. A. Fox, N. A. DeLuca, and S. Person.** 1984. Nucleotide sequence specifying the glycoprotein gene, gB, of herpes simplex virus type 1. Virology **133**:301–314.
4. **Bzik, D. J., B. A. Fox, N. A. DeLuca, and S. Person.** 1984. Nucleotide sequence of a region of the herpes simplex virus type 1 gB glycoprotein gene: mutations affecting rate of virus entry and cell fusion. Virology **137**:185–190.
5. **Chou, P. Y., and G. D. Fasman.** 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Protein Chem. **47**:45–148.
6. **Costa, R. H., K. G. Draper, T. J. Kelly, and E. K. Wagner.** 1985. An unusual spliced herpes simplex virus type 1 transcript with sequence homology to Epstein-Barr virus DNA. J. Virol. **54**:317–328.
7. **Dayhoff, M. O.** 1969. Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, Md.
8. **Eisenberg, D.** 1984. Three-dimensional structure of membrane and surface proteins. Annu. Rev. Biochem. **53**:595–623.
9. **Eisenberg, D., R. M. Weiss, and T. C. Terwilliger.** 1982. The helical hydrophobic moment: a measure of the amphipathicity of a helix. Nature (London) **299**:371–374.
10. **Engleman, D. M., and T. A. Steitz.** 1981. The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. Cell **23**:411–422.
11. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implication of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
12. **Gibson, T., P. Stockwell, M. Ginsburg, and B. Barrell.** 1984. Homology between two EBV early genes and HSV ribonucleotide reductase and 38K genes. Nucleic Acids Res. **12**: 5087–5099.
13. **Gray, T. M., and B. W. Matthews.** 1984. Intrahelical hydrogen bonding of serine, threonine and cysteine residues within α helices and its relevance to membrane proteins. J. Mol. Biol. **175**:75–81.
14. **Holland, T. C., S. D. Marlin, M. Levine, and J. Glorioso.** 1983. Antigenic variants of herpes simplex virus selected with glycoprotein-specific monoclonal antibodies. J. Virol. **45**:672–682.
15. **Honig, B. H., and W. L. Hubbell.** 1984. Stability of "salt bridges" in membrane proteins. Proc. Natl. Acad. Sci. USA **81**:5412–5416.
16. **Kabsch, W., and C. Sander.** 1983. How good are predictions of protein secondary structure? FEBS Lett. **155**:179–182.
17. **Kieff, E., B. Hoyer, S. Bachenheimer, and B. Roizman.** 1972.

Genetic relatedness of type 1 and type 2 herpes simplex viruses. J. Virol. **9**:738–745.
18. **Kousoulas, K. G., P. E. Pellett, L. Pereira, and B. Roizman.** 1984. Mutations affecting conformation or sequence of neutralizing epitopes identified by reactivity of viable plaques segregated from *syn* and *ts* domains of HSV-1(F) gB gene. Virology **135**:379–394.
19. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
20. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
21. **Little, S. P., J. T. Jofre, R. J. Courtney, and P. A. Schaffer.** 1981. A virion-associated glycoprotein essential for infectivity of herpes simplex virus type 1. Virology **115**:149–160.
22. **Morse, L. S., L. Pereira, B. Roizman, and P. A. Schaffer.** 1978. Anatomy of herpes simplex virus (HSV) DNA. X. Mapping of viral genes by analysis of polypeptides and functions specified by HSV-1 × HSV-2 recombinants. J. Virol. **26**:389–410.
23. **Nishikawa, K.** 1983. Assessment of secondary-structure prediction of proteins: comparison of computerized Chou-Fasman method with others. Biochim. Biophys. Acta **748**:285–299.
24. **Pellett, P. E., K. G. Kousoulas, L. Pereira, and B. Roizman.** 1985. The anatomy of the Herpes simplex virus 1 strain F glycoprotein B gene: primary sequence and predicted protein structure of the wild type and of monoclonal antibody-resistant mutants. J. Virol. **53**:243–253.
25. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391–409.
26. **Rashin, A. A., and B. Honig.** 1984. On the environment of ionizable groups in globular proteins. J. Mol. Biol. **173**:515–521.
27. **Roizman, B., L. E. Carmichael, F. Deinhardt, G. de The, A. J. Nahmias, W. Plowright, F. Rapp, P. Sheldrick, M. Takahashi, and K. Wolf.** 1981. Herpesviridae: definition, provisional nomenclature and taxonomy. Intervirology **16**:201–217.
28. **Ruyechan, W. T., L. S. Morse, D. M. Knipe, and B. Roizman.** 1979. Molecular genetics of herpes simplex virus. II. Mapping of the major viral glycoproteins and of the genetic loci specifying the social behavior of infected cells. J. Virol. **29**:677–697.
29. **Sarmiento, M., M. Haffey, and P. G. Spear.** 1979. Membrane proteins specified by herpes simplex viruses. III. Role of glycoprotein VP7(B_2) in virion infectivity. J. Virol. **29**: 1149–1158.
30. **Schiffer, M., and A. B. Edmundson.** 1967. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. Biophys. J. **7**:121–135.
31. **Spear, P. G.** 1984. Glycoproteins specified by herpes simplex viruses, p. 315–356. *In* B. Roizman (ed.), The herpesviruses. Plenum Publishing Corp., New York.
32. **Staden, R.** 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. Nucleic Acids Res. **10**:2951–2961.
33. **von Heijne, G.** 1984. How signal sequences maintain sequence specificity. J. Mol. Biol. **173**:243–351.
34. **Wilbur, W. J., and D. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.