Mapping of the Major Glycoprotein Gene of Human Cytomegalovirus

By MICHAEL MACH, URSULA UTZ AND BERNHARD FLECKENSTEIN*
Institut für Klinische Virologie der Universität Erlangen-Nürnberg, Loschgestrasse 7, D-8520 Erlangen, F.R.G.

(Accepted 7 April 1986)

SUMMARY

The gene coding for the most abundant glycoprotein (gp58) of human cytomegalovirus (HCMV), strain AD169, was physically mapped on the viral genome. A monospecific rabbit antiserum against gp58 was used to screen a cDNA library that was constructed from poly(A)+ RNA of HCMV-infected cells in the prokaryotic expression vector λgt11. A cDNA clone was identified which synthesized part of the glycoprotein. It allowed localization of the coding region within the right terminal sequence of the HindIII-F fragment between map coordinates 0.344 and 0.380 of HCMV virion DNA.

Human cytomegalovirus (HCMV) is a herpesvirus capable of causing a variety of diseases in man that can arise either from primary infection or from the reactivation of latent virus during periods of immunosuppression. The genome codes for approximately 30 structural proteins. The glycosylated surface polypeptides of HCMV from both virions and infected cells have been analysed by using polyvalent hyperimmune sera and monoclonal antibodies (Pereira et al., 1982; Nowak et al., 1984). Extensive studies of other herpesviruses such as Epstein–Barr virus (EBV) and herpes simplex virus (HSV) have demonstrated that virion glycoproteins are important for determining the host’s immune response to virus infection (Thorley-Lawson & Geilinger, 1980; Spear, 1980). To date, however, the nature of the HCMV-encoded protein(s) recognized by neutralizing antibodies remains poorly defined. Several authors described neutralizing antibodies which react with multiple glycoproteins found within HCMV-infected cells and purified virions (Pereira et al., 1984; Nowak et al., 1984; Britt, 1984; Law et al., 1985; Rasmussen et al., 1985). Most of the monoclonal antibodies precipitate a set of at least three glycoproteins from infected cells with estimated mol. wt. of 130000 (130K), 95K to 115K and 58K (glycoprotein A or gA complex) (Pereira et al., 1984). The authors suggested that the 58K protein is derived from a 95K unglycosylated precursor via a 130K glycosylated intermediate. In Western blot analyses with purified virions, only the 58K protein is recognized by a monoclonal antibody which, in immunoprecipitations, also reacts with the 95K and 130K protein complexes (Law et al., 1985). Although the published data still leave open several possible pathways for the maturation process of the gA complex, it is clear that the primary unglycosylated precursor is a polypeptide of about 95K, and by far the most abundant glycoprotein of extracellular virions is the polypeptide of 58K apparent mol. wt. Physical mapping and structural analysis of the gA gene will help to elucidate the biochemistry of the maturation process, and will also allow expression of the dominant viral surface protein for new diagnostic procedures and immunoprevention of disease by a subunit vaccine. This paper describes the identification of the gene coding for the 58K glycoprotein by prokaryotic expression vector cloning and use of a monospecific antiserum.

Total RNA was extracted from HCMV AD169-infected human foreskin fibroblasts (HFF) 96 to 120 h post-infection using the guanidinium isothiocyanate method described by Chirgwin et al. (1979). Poly(A)-containing [poly(A)+] RNA was isolated by two cycles on an oligo(dT)-
Fig. 1. SDS-PAGE and immunoassay of viral proteins and hybrid protein accumulation. (a) Proteins accumulating in lysogens infected with either λ gt11 or UM-2 were compared. Lysates were loaded on an 8% running, 5% stacking polyacrylamide gel, subjected to electrophoresis, and stained with Coomassie Brilliant Blue. Lane 1, mol. wt. markers; lane 2, E. coli Y1089; lane 3, E. coli Y1089 infected with λ gt11; lane 4, same as 3, but induced with IPTG; lane 5, E. coli Y1089 infected with UM-2, induced with IPTG; lane 6, same as 5 but without IPTG; lane 7, HCMV dense bodies. (b) Western blot of HCMV structural proteins and hybrid proteins with anti-gp58 serum. After transfer to nitrocellulose the membrane was incubated with anti-gp58, followed by Protein A coupled to horseradish peroxidase, then 4-chloro-1-naphthol. Lane 1, E. coli Y1089 infected with UM-2; lane 2, same as 1 but induced with IPTG; lane 3, HCMV virions; lane 5, E. coli Y1089 infected with λ gt11, induced with IPTG; lane 6, same as 5, without IPTG; lane 7, uninfected HFF.

Sepharose column (Maniatis et al., 1982). Double-stranded cDNA was synthesized (Gubler & Hoffman, 1983) starting with 10 μg of poly(A)+ RNA and oligo(dT)_{12-18} as the primer for the reverse transcription reaction. The cDNA was inserted into λ gt11 DNA without size fractionation. The λ gt11 vector (Young & Davis, 1983) was cleaved with EcoRI and treated with calf intestinal phosphatase to reduce intramolecular re-ligation. The cDNA was inserted between the phage arms using EcoRI linkers and packaged in vitro. A library was obtained from 100 ng of double-stranded cDNA which contained approximately 5 × 10^5 independent recombinants with 18% wild-type phage. The library was screened with antibody probes according to Young & Davis (1983) except that we used horseradish peroxidase coupled to Protein A, and 4-chloro-1-naphtholase as the detection system.
The antibodies against the 58K structural protein were raised in rabbits. Extracellular virions and dense bodies were purified by centrifugation through a glycerol-tartrate gradient (Talbot & Almeida, 1977). Approximately 1 mg total virion protein was subjected to preparative SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and the individual bands were visualized by Coomassie Brilliant Blue staining. The region corresponding to mol. wt. 56K to 58K was cut out, the protein extracted and used to immunize rabbits. Following the initial application, booster injections were given at 3 week intervals over a period of 7 months. Fig. 1(b) shows the specificity of the antiserum. In Western blots using purified virions, the serum recognized a single protein of 58K (lane 3). Depending on the virus preparation, weak reactions occasionally were observed with additional proteins of about 96K, 130K and 160K. The serum did not react with extracts from uninfected cells (Fig. 1b, lane 7). This antiserum was used to screen the cDNA library. A total of 150000 plaques were tested. Fifteen positive signals were obtained. One clone (UM-2) with an insert size of approximately 400 bases was chosen for further characterization. 

Escherichia coli strain Y1089 was infected with the recombinant phage at a m.o.i. of 5. Cells were grown at 32 °C with vigorous shaking to a density of 0.2 $A_{600}$. Isopropylthiogalactopyranoside (IPTG) was added to a final concentration of 5 mM, and the cells were quickly shifted to 42 °C for 15 min. After an additional 2 h at 37 °C cells were harvested by centrifugation and lysed in gel sample buffer (Laemmli, 1970) at 5% the original volume. Induction of the $\beta$-galactosidase gene in $\lambda gt11$-infected cells resulted in the synthesis of a 118K protein (Fig. 1a, lane 4). This protein was not synthesized in uninfected or infected non-induced cells (Fig. 1a, lanes 2, 3). Upon induction, cells infected with recombinant phage from clone $\lambda$ UM-2 synthesized a protein which was slightly bigger than $\beta$-galactosidase (Fig. 1a, lane 5). This protein was also made in neither uninfected nor infected non-induced cells. The corresponding Western blot is shown in Fig. 1(b). The anti-gp58 serum reacted with a single protein in induced $\lambda$ UM-2-infected cells (lane 2). It did not recognize a protein in $\lambda$ gt11-infected cells (lanes 5, 6). This indicated that recombinant clone $\lambda$ UM-2 was synthesizing a fusion protein with an HCMV glycoprotein part. The cDNA insert of 400 bp was hybridized with each of eight cosmid clones spanning the whole genome of HCMV (Fleckenstein et al., 1982). Cosmid pCM1029 (see Fig. 4) containing the fragments HindIII-D and -F hybridized with the cDNA (data not shown). A more detailed Southern blot analysis of this region confined the hybridizing HCMV DNA fragment to a 1.4 kb EcoRI-BamHI fragment located at the right end of HindIII F (Fig. 2, 4).

The translation of a 400 bp cDNA insert, in addition to the open reading frame for $\beta$-galactosidase, could give rise to a fusion protein of approximately 130K. The fusion protein synthesized in clone $\lambda$ UM-2 was clearly smaller. However, Western blot data and the hybridization with a fragment of HCMV DNA under stringent conditions make the possibility of this clone being unrelated to HCMV very unlikely. An obvious mechanism for the synthesis of such a small fusion protein is the location of a translational stop signal within the cDNA. Since cDNA is synthesized from the 3' end of mRNA and therefore might contain 3' untranslated sequences, this possibility appears to be most likely. In order to confirm this, we subcloned the cDNA insert in an orientation-dependent way. A 1.3 kb KpnI-EcoRI fragment was subcloned into M13 mp19, thus placing the EcoRI site that is located at the $\beta$-galactosidase-HCMV junction at the beginning of the sequence (Fig. 3). The sequence indeed showed a stop codon 27 bases downstream from the $\beta$-galactosidase-HCMV junction. It also determined the polarity of the coding strand within UM-2. It was therefore used to analyse the direction of transcription on the HCMV genome. The EcoRI-BamHI fragment of pCM5004 was subcloned into M13 mp10 and mp11 so that the ssDNA of these clones represented each strand of the fragment. The identity of the clones was verified by taking advantage of the asymmetrically located PstI site. Single-stranded DNA of UM-2 was hybridized to ssDNA of both EcoRI-BamHI clones and hybridization was analysed by gel retardation of the resulting hybrid (data not shown).

From these experiments we conclude that the mRNA coding for gp58 is transcribed from right to left in the U1 orientation of the HCMV AD169 genome as shown in Fig. 4. This places the 3' terminus of the open reading frame coding for gp58 into a 1.4 kb BamHI/EcoRI fragment.
Fig. 2. Hybridization of recombinant phage UM-2 with subfragments of the HindIII F region of HCMV. (a) Ethidium bromide staining of a 1% agarose gel of various subfragments of the HindIII F-D junction. Lanes 1, 2, λ HindIII (lane 1) and φX174 HaeIII (lane 2) size markers; lane 3, EcoRI + BamH1 digest of pCM5004; lanes 4 to 6, EcoRI digestions of pCM5002, pCM5004 and pRR3 respectively. The insert in pRR3 is illustrated in Fig. 4. All HCMV fragments were cloned in pACYC184 which results in the production of an EcoRI digestion 4 kb fragment. (b) Southern blot hybridization with a nick-repair-labelled probe of a gel with lanes as in (a).

Fig. 3. Arrangement of UM-2 cDNA sequences within λ gt11 and M13 mp19. The nucleotide and amino acid sequences surrounding the EcoRI site of β-galactosidase are shown. The UM-2 sequence was determined by the dideoxy method. It should be noted that the GAATTCC sequence in the UM-2 construct is derived from the EcoRI linkers used in the original cDNA cloning procedure.
which is located between map positions 0.344 and 0.350 of the viral genome. Since the unglycosylated precursor of group A glycoproteins has been shown to be a 95K polypeptide, this indicated that the final product, the dominant surface glycoprotein gp58, is derived by proteolytic cleavage and represents the carboxy-terminal part of the primary translation product. Northern blot analyses with nick translated labelled clone UM-2 revealed several size classes of late RNA molecules between 2.5 kb and 8 kb. The most abundant RNA type was a discrete 4 kb transcript.

In summary, this study provided evidence that the glycoprotein A complex of HCMV is encoded by a DNA sequence in the right end of the HindIII-F fragment of virion DNA. Sequence analysis of a cDNA clone determined the direction of transcription and translation to be from right to left. The coding sequence for the dominant virion glycoprotein gp58 is the 3' terminal part of the translational frame for the group A glycoprotein complex precursor of HCMV. With the exception of the mRNA coding for the unusual gp280 of EBV (Beisel et al., 1985), all major glycoprotein transcripts of other herpesviruses are not spliced. Thus, it seemed likely that the gA coding sequence is entirely contained within a single reading frame of about 2.5 kb, as indicated in Fig. 4. Recently we learned that, in the course of shotgun sequencing in this genomic HCMV DNA region, an open reading frame of about this size has been found. The potential polypeptide encoded by this reading frame shows homology to glycoprotein B (gB) of HSV (B. Barrell & T. Kouzarides, personal communication). The gene coding for the gA complex of HCMV is located immediately upstream from the viral DNA polymerase gene (R.
Short communication

Heilbronn, G. Jahn, A. Bfirkle, U.-K. Freese, B. Fleckenstein & H. zur Hausen, unpublished results). Both genes are transcribed in the same direction. This is remarkable, since the same sequence arrangement was found in EBV. This virus has also an open reading frame (BALF-4) of approximately 2-1 kb which is located just upstream from the DNA polymerase gene (Baer et al., 1984). The putative polypeptide encoded by BALF-4 shows significant homology to gB of HSV (Pellett et al., 1985). The gB of HSV is one of the most abundant virion glycoproteins; it is a major target of the host immune response and is involved in adsorption of the virion envelope to the cellular membrane. The gB has a mol. wt. of 120K and is derived from a 97K precursor, about the same size postulated for the gA precursor of HCMV (Pereira et al., 1984). The gB in turn shares antigenic epitopes with a 63K glycoprotein of varicella-zoster virus (Edson et al., 1985). This lends support to the unifying concept of an ancestral glycoprotein gene that is generally conserved in herpesviruses of all subgroups. Thus, it is even more surprising to what extent herpesviruses are heterogeneous in phenotype with regard to the immunologically dominant surface glycoproteins. HSV has, among other abundant glycoproteins, a high mol. wt. gB; the equivalent genes of varicella-zoster virus and HCMV code for a large precursor which is processed proteolytically to the dominant envelope glycoproteins. In EBV, however, the most abundant surface glycoprotein (gp350/280) is entirely unrelated.

It is a pleasure to acknowledge the excellent technical assistance of Barbara Schütze. We also thank Gerhard Jahn for help in preparing the antisera. This work was supported by Bundesministerium für Forschung und Technologie, Projekträger Biotechnologie.

REFERENCES


(Received 7 February 1986)