Sequence variation within neutralizing epitopes of the envelope glycoprotein B of human cytomegalovirus: comparison of isolates from renal transplant recipients and AIDS patients

Donna M. Roy,* Jane E. Grundy† and Vincent C. Emery

Division of Communicable Diseases, Royal Free Hospital School of Medicine, Rowland Hill Street, Hampstead, London NW3 2QG, U.K.

The envelope glycoprotein B of human cytomegalovirus (CMV) is a major target of the neutralizing antibody response against this virus, and hence has importance as a potential subunit vaccine. PCR was utilized to amplify DNA encoding the dominant antigenic determinant on this molecule, AD-1 (codons 552 to 635), and DNA sequencing was carried out in order to compare nucleotide variation in AD-1 between clinical isolates of CMV and the laboratory strain AD169. Wild-type CMV strains isolated from AIDS patients were not only more likely to possess nucleotide substitutions (19/24 compared to 5/25, \(P < 0.0001\)) than those from renal transplant recipients, but they also exhibited a greater degree of nucleotide sequence divergence (6.94 versus 0.82 substitutions/1000 bp, \(P < 0.0001\); 96.0 to 100% versus 99.4 to 100% similarity). Increased sequence variation in the AIDS patients did not correlate with absolute peripheral blood CD4+ T cell level (\(r = 0.33, P > 0.1\)). Only two strains from AIDS patients and one strain from the renal transplant recipients possessed nucleic acid substitutions that resulted in codon changes, indicating that AD-1 is relatively well conserved amongst clinical isolates of CMV. The demonstration of strains with codon changes within neutralizing epitopes, however, highlights the importance of taking into consideration the presence of these strains within the wild-type virus population when preparing subunit vaccines.

Cytomegalovirus (CMV) is an important pathogen in the immunocompromised individual; disease may result from primary infection, reinfection, or following reactivation of latent virus. In renal transplant recipients, CMV infection may result in significant disease, and has been implicated in graft failure (Glenn, 1981). In patients with AIDS, CMV is the most common cause of life-threatening viral opportunistic infection (Jacobson & Mills, 1988). As a consequence of underlying immune defects, it may not be possible to stimulate cell-mediated immune responses against CMV in these individuals. Hence, humoral immunity may have a particularly important role in modifying the clinical severity of progressive CMV infection, as demonstrated by the protective efficacy of prophylactic immunoglobulin preparations against severe CMV disease (Snydman et al., 1987; Snydman, 1990). The effective mechanism of antibody-mediated immunity may be in part neutralization of virions in the blood during systemic dissemination, thus limiting the spread of infection to multiple organs, since cell-free virus has been detected in the plasma (Spector et al., 1992) and serum (Ishigaki et al., 1991) of immunocompromised individuals. A live attenuated vaccine has been shown to reduce the severity of disease in renal transplant recipients (Plotkin et al., 1991), but concerns about the potential for latency, reactivation and oncogenicity have prompted a search for a candidate subunit vaccine immunogen.

At least three envelope glycoprotein complexes of CMV have been identified as targets for neutralizing murine monoclonal antibodies: UL55, gCl, gp55-116 or glycoprotein B (gB) (Britt, 1984; Gretch et al., 1988a); gClII or gp47-52 (Kari et al., 1986; Gretch et al., 1988b); and UL75, gClIII, gp86 or glycoprotein H (gH) (Rasmussen et al., 1984; Cranage et al., 1988). Purified (Gonczol et al., 1986) and recombinant (Marshall et al., 1990; Britt et al., 1988) preparations of gB have been shown to induce humoral immunity in experimental animals. In seronegative human volunteers, purified gB induced both neutralizing antibodies and lymphoproliferative responses, and in naturally immune individuals booster responses were generated (Gonczol et al., 1990). Vaccinia virus–gB recombinants have been utilized to assess humoral immunity against this target in convalescent human sera, and sera obtained from
individuals immunized with the live attenuated vaccine strain (Towne) of CMV. These studies have demonstrated that between 40 and 88% of total serum neutralizing activity in these individuals was directed against the envelope glycoprotein gB (Britt et al., 1990a; Gonczol et al., 1991). Furthermore, antibodies to a recombinant adenovirus–gB protein in naturally immune individuals correlated significantly with serum neutralizing activity (Marshall et al., 1992).

These data indicate that gB-specific antibody appears to be the major component of the neutralizing antibody response against CMV. Several discontinuous and linear neutralizing epitopes have been identified on the gB molecule. The human anti-gB antibody response predominantly recognizes a discrete region between amino acids 552 and 635 (Wagner et al., 1992), termed antigenic determinant-1 (AD-1). This restricted antibody response against human gB is also observed in mice (Britt et al., 1990b, 1991; Kneiss et al., 1991). The biased antibody response to the AD-1 region of gB may be important since a significant correlation was observed between neutralizing antibody titres in human sera and reactivity in ELISAs against a prokaryotically expressed polypeptide containing AD-1 (Kropff et al., 1993).

An important consideration in selecting possible candidates for a subunit vaccine of CMV is identifying the extent of sequence heterogeneity for that target protein within the wild-type virus population. To this end we have previously performed sequence analysis of a linear neutralizing epitope which is located within AD-1, corresponding to amino acids 609 to 627 (Utz et al., 1989), and showed that clinical strains of CMV isolated from immunocompromised individuals demonstrated sequence variation within this epitope, at both the nucleotide and amino acid levels (Darlington et al., 1991). To confirm and extend our previous work, we utilized PCR to amplify CMV DNA for direct sequencing, in order to assess sequence variation in the whole of AD-1 for two immunocompromised patient groups (see below) for which a future vaccine would be beneficial.

Clinical strains of CMV were collected over the period 1989 to 1991 from renal transplant recipients, and human immunodeficiency virus (HIV)-infected individuals classified as having AIDS according to CDC criteria. The median peripheral blood absolute CD4+ T cell level for the AIDS patient group was $37 \times 10^6/\text{l}$ (range $10^4$ to $208 \times 10^6/\text{l}$). Virus isolates, collected from sources including urine, saliva and bronchoalveolar lavage, were propagated in human embryo lung fibroblasts, and DNA was extracted for analysis within five passages of primary isolation by the Hirt method (Hirt, 1967). Laboratory CMV strains AD169, Towne and Davis were obtained from the ATCC, and propagated for two passages before DNA extraction.
cencing reactions were analysed by urea-PAGE by the protocols described in Sambrook.

PCR reactions to confirm the fidelity of the sequencing were primer A, 5' ACAACAAACCGATTGC-CGCGCAGT and primer B, 5' AGGGCCGTAGCCGGG-TCGACTACC which correspond to nucleotides 81876 to 81899 and 81374 to 81397 of the DNA sequence of AD169 (Chee et al., 1990). Reaction mixtures contained 1·5 mM-MgCl₂, 25 mM-Tris–HCl pH 8·4, 17 mM-(NH₄)₂SO₄, 10 mM-2-mercaptoethanol, 0·002 % gelatin, 200 mM-dNTPs, 0·01 units/μl Taq polymerase (Amplitaq, Perkin-Elmer Cetus), 1 ng/ml of each primer, and 1 μg of template DNA. Temperature cycling, performed in a programmable thermal cycler (Hybaid), consisted of 94 °C for 6 min, 60 °C for 1 min and 72 °C for 2 min for one cycle; 94 °C for 1 min 15 s, 60 °C for 1 min 15 s and 72 °C for 2 min for 32 cycles; 94 °C for 1 min 15 s, 60 °C for 1 min 15 s and 72 °C for 10 min for two cycles. The resulting 525 bp PCR products were analysed by electrophoresis in a 1·2 % agarose gel containing ethidium bromide. Template DNA for sequencing was prepared by purification of PCR products using a Gene Clean II kit (Bio 101), according to the manufacturer’s instructions.

Purified PCR amplification products were sequenced directly using the internal oligonucleotide primers C, 5' GAGGACACAACTCCTGTTGGGCA and primer D, 5' GTCGACGGTGGAGATACTGCTGAGG, corresponding to nucleotides 81683 to 81708 and 81558 to 81583 of the AD169 DNA sequence. Direct sequencing was achieved by adoption of the dideoxynucleotide chain termination procedure (Sanger et al., 1977), using DMSO (Winship, 1989) and the Sequenase 2.0 sequencing kit (United States Biochemicals). Completed sequencing reactions were analysed by urea–PAGE by the protocols described in Sambrook et al. (1989). Each virus strain was sequenced three times from three separate PCR reactions to confirm the fidelity of the Taq polymerase in the PCR procedure. Statistical analysis of the resulting sequence data was performed by the chi-square test.

The nucleotide sequences of 25 clinical isolates of CMV from 24 renal transplant recipients and 24 isolates from 23 AIDS patients, containing the region of gB encoding AD-1, were compared with the reference strain AD169. The frequency of base substitutions occurring at different positions within the region examined are shown for the two groups in Fig. 1, with results summarized in Table 1.

Virus isolates from AIDS patients were more likely to have base substitutions than renal transplant recipients. In addition, isolates from the AIDS patient group (similarity 96 to 100%) possessed a greater degree of nucleotide sequence divergence than those from renal transplant recipients (99·4 to 100% identity). Seventeen clinical isolates from AIDS patients had between one and six nucleotide changes. Two other strains from AIDS patients showed respectively 21 (A1) and 17 (A2) substitutions. Two of the 24 strains were isolated from different sites from the same patient at the same time, and possessed an identical single base change. In contrast, in the renal transplant recipient group, five virus isolates exhibited between one and three nucleotide changes. Two of the 25 strains were obtained from the urine of one patient at different times, and possessed the same single nucleotide change. The frequency of nucleotide substitution observed in the renal transplant recipient group (0·1 %) is comparable to that observed in our previous study (Darlington et al., 1991) for immunocompromised individuals (0·3%).

The nucleotide sequence of virus strains from the two clinical groups were also compared to the reference strain Towne, which is 97·5 % identical to AD169 in the gB gene. Interestingly, isolates from the AIDS patient group appeared to show more sequence similarity with the Towne strain of CMV (97·1 to 100 %) than with AD169. The differences observed between the two clinical groups of isolates, however, remained significant.

Two causes of artefactual sequence alterations are possible, firstly that Taq DNA polymerase introduced errors, and secondly that the selection of these variants was affected by cell culture passage. Sequence data obtained upon sequencing of a primary PCR reaction were always verified upon resequencing of DNA amplified in subsequent PCR reactions. No differences in the sequence data were noted for any sample, confirming the fidelity of the Taq DNA polymerase in the PCR reaction under the conditions used in this study. In addition, when the two virus isolates from AIDS patients that showed the highest sequence divergence between all the isolates (A1 and A2) were plaque-purified, the virus strains recovered from individual plaques for both patients demonstrated complete nucleotide sequence identity with the parent strains. The sequence obtained for the laboratory strain Towne is in agreement with that previously published for the gB sequence of Towne (Spaete et al., 1988). A second laboratory strain, Davis, is exactly the same as AD169 in this region. Although there is a possibility that propagation of virus isolates in

<table>
<thead>
<tr>
<th>Clinical group</th>
<th>Total no. of isolates in group</th>
<th>No. of isolates with substitutions (%)</th>
<th>Substitutions/1000 bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>24</td>
<td>19 (79·2)*</td>
<td>6·94†</td>
</tr>
<tr>
<td>Renal transplant</td>
<td>25</td>
<td>5 (20)*</td>
<td>0·82†</td>
</tr>
</tbody>
</table>

* P < 0·0001.
† P < 0·0001.

Table 1. Nucleotide sequence variation amongst clinical CMV isolates

The nucleotide sequences of 25 clinical isolates of CMV from 24 renal transplant recipients and 24 isolates from 23 AIDS patients, containing the region of gB encoding AD-1, were compared with the reference strain AD169. The frequency of base substitutions occurring at different positions within the region examined are shown for the two groups in Fig. 1, with results summarized in Table 1.

Virus isolates from AIDS patients were more likely to have base substitutions than renal transplant recipients. In addition, isolates from the AIDS patient group (similarity 96 to 100%) possessed a greater degree of nucleotide sequence divergence than those from renal transplant recipients (99·4 to 100% identity). Seventeen clinical isolates from AIDS patients had between one and six nucleotide changes. Two other strains from AIDS patients showed respectively 21 (A1) and 17 (A2) substitutions. Two of the 24 strains were isolated from different sites from the same patient at the same time, and possessed an identical single base change. In contrast, in the renal transplant recipient group, five virus isolates exhibited between one and three nucleotide changes. Two of the 25 strains were obtained from the urine of one patient at different times, and possessed the same single nucleotide change. The frequency of nucleotide substitution observed in the renal transplant recipient group (0·1 %) is comparable to that observed in our previous study (Darlington et al., 1991) for immunocompromised individuals (0·3%).

The nucleotide sequence of virus strains from the two clinical groups were also compared to the reference strain Towne, which is 97·5 % identical to AD169 in the gB gene. Interestingly, isolates from the AIDS patient group appeared to show more sequence similarity with the Towne strain of CMV (97·1 to 100 %) than with AD169. The differences observed between the two clinical groups of isolates, however, remained significant.

Two causes of artefactual sequence alterations are possible, firstly that Taq DNA polymerase introduced errors, and secondly that the selection of these variants was affected by cell culture passage. Sequence data obtained upon sequencing of a primary PCR reaction were always verified upon resequencing of DNA amplified in subsequent PCR reactions. No differences in the sequence data were noted for any sample, confirming the fidelity of the Taq DNA polymerase in the PCR reaction under the conditions used in this study. In addition, when the two virus isolates from AIDS patients that showed the highest sequence divergence between all the isolates (A1 and A2) were plaque-purified, the virus strains recovered from individual plaques for both patients demonstrated complete nucleotide sequence identity with the parent strains. The sequence obtained for the laboratory strain Towne is in agreement with that previously published for the gB sequence of Towne (Spaete et al., 1988). A second laboratory strain, Davis, is exactly the same as AD169 in this region. Although there is a possibility that propagation of virus isolates in
Chou, 1990; Brytting AD-1 region appears to remain stable compared with sequence data obtained for the same alignments of isolates from AIDS patients (A1, A2) and renal transplant recipient (R1) with amino acid sequence 532 to 706 of reference strain AD169. Shaded region indicates AD-1.

cell culture might select or produce variant strains, the available data suggest that sequences obtained for regions of the major immediate early region, DNA polymerase, gH and gB genes via PCR amplification of DNA direct from clinical material, are maintained when compared with sequence data obtained for the same strains from culture-grown virus (Darlington et al., 1991; Chou, 1990; Brytting et al., 1992). In the present study, passage of clinical isolates in culture was minimal (three to four passages), but the laboratory strain AD169 was extensively passaged (> 50 times) in our laboratories and still showed complete identity with the parent strain (data not shown), demonstrating that the sequence of the AD-1 region appears to remain stable in vitro.

Translation of the nucleotide sequences of variant clinical isolates demonstrated that the majority of base substitutions were silent. However, two strains from the AIDS patient group (8.3 %) and a single strain from the renal transplant recipient group (40 %) possessed nucleotide changes that led to alterations in the amino acid sequence of the region of gB examined (Fig. 2). Virus strain A1 (from an AIDS patient) exhibited three amino acid changes and, in addition to a mutation at position 642 (S → Y), two of these, at positions 584 (A → V) and 612 (L → F), were within AD-1. The virus isolate A2, from the second AIDS patient, possessed a cluster of six amino acid substitutions within the amino-terminal half of AD-1, between amino acids 560 and 590: position 565 (N → T), 566 (V → I), 568 (E → D), 570 (P → T), 582 (N → T) and 588 (Y → D). Finally, the virus strain R1 from a renal transplant recipient had one amino acid substitution within AD-1, at position 584 (A → V).

Examining sequence variation in the linear neutralizing epitope (amino acids 609 to 626) within AD-1, the virus isolate from an AIDS patient (A1) demonstrated a base substitution at position 612, from a leucine to a phenylalanine. Our previous study with isolates from a group of immunocompromised patients (Darlington et al., 1991) revealed three virus strains also with codon changes at this position, from leucine to histidine, valine and phenylalanine, respectively. It is interesting that other studies have also demonstrated codon changes at this same position in clinical isolates (Chou & Dennison, 1991; Lehner et al., 1991): the possibility exists that in the face of immunological pressure this position represents a mutational 'hotspot'. Further studies to assess the functional significance of the codon changes, and to explore the possibility that variant isolates may be antibody 'escape mutants', are now in progress.

At the nucleotide and amino acid levels the data presented here suggest that the AD-1 region, spanning amino acids 552 to 635, displayed a high degree of amino acid sequence conservation in all strains. These data are consistent with previous studies examining sequence variation within this region in immunocompromised patient groups (Darlington et al., 1991; Chou & Dennison, 1991; Lehner et al., 1991). It is possible that functional requirements constrain the amount of variation that can exist in the region under study. The herpes simplex virus (HSV) gB homologue has been shown to be essential for virus entry into cells (Cai et al., 1988). The gB molecule of CMV may fulfil a similar role, since recent data have suggested that the gB molecule may be involved in fusion events leading to virus penetration of cells (Tugizov et al., 1992). Immunoaffinity-purified gB from CMV was found to bind a molecule of Mr 31K (Aldish et al., 1990), which may correspond to a putative cellular receptor described for CMV (Taylor & Cooper, 1990; Rasmussen et al., 1991), and whose expression also correlates with infectability of cells (Nowlin et al., 1991). Furthermore it has been suggested that CMV, like HSV, may attach to cells through interaction of viral envelope glycoproteins, namely gC-II, with heparin sulphate on the cell membrane surface (Kari & Gehrz, 1992). Glycoprotein B has also been shown to bind to heparin to a lesser extent, and hence may be involved in viral attachment in addition to viral penetration (Kari & Gehrz, 1993).

Within the gB sequence examined in this study, spanning amino acids 532 to 706, heterogeneity was not randomly distributed. The majority of nucleotide and amino acid variation observed in both of the clinical groups was located in the first third of this region (amino acids 532 to 590). In addition to part of AD-1, this region is also known to contain other neutralizing epitopes (Qadri et al., 1992; Kneiss et al., 1991; Banks et al., 1989). It may be that parts of the gB molecule outside AD-1 are functionally less constrained, and immunological pressure on these regions therefore may generate variation that is not deleterious to virus growth.

In this study, virus isolates from AIDS patients were found to exhibit a greater degree of nucleotide and amino acid sequence divergence than those strains from renal transplant recipients. It is possible that in the face of the declining ability of AIDS patients to mount an
to a pathogenic strain (Weise et al., 1987). Host factors such as immune status, underlying both B and T cell immunity, probably play a role in the generation of diversity in this biologically significant region. Not only does the region under study contain both linear and discontinuous antibody epitopes, but at least one cytotoxic T cell epitope has been mapped between amino acids 619 and 628 (Utz et al., 1992).

Studies are now under way to determine whether the sequence heterogeneity seen in this region of gB in clinical isolates from AIDS patients is also reflected in other biologically important regions of gB and gH.

This work was supported by Action Research.

References


gB, and influenza virus M1, is determined by common structural features of the HLA-A2.1 molecule. *Journal of Immunology* **149**, 214–221.


(Received 8 April 1993; Accepted 16 July 1993)