Identification of the genes encoding two capsid proteins of herpes simplex virus type 1 by direct amino acid sequencing

Frazer J. Rixon, Matthew D. Davison and Andrew J. Davison

1MRC Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR and 2Department of Biochemistry, University of Leicester, Leicester LE1 7RH, U.K.

Amino-terminal amino acid sequencing was carried out on proteins purified from herpes simplex virus type 1 capsids. The sequences of two capsid proteins (VP19C and VP23) showed them to be encoded by genes UL38 and UL18, respectively. The product of UL38 has been shown to be essential for capsid assembly, but no role has previously been assigned to the product of UL18.

Viewed in the electron microscope, herpes virions have a characteristic appearance. An icosahedral capsid is embedded in an amorphous layer known as the tegument which is enclosed by a glycoprotein-containing lipid envelope. Gibson & Roizman (1972) identified six proteins as components of capsids isolated from the nuclei of cells infected with herpes simplex virus type 1 (HSV-1). Heilman et al. (1979) and Cohen et al. (1980) subsequently identified a seventh capsid protein of Mr 12000 (henceforth referred to as 12K). The properties of these proteins, and the nomenclature used by each group, are summarized in Table 1. Map locations have been proposed for the genes encoding three capsid proteins [VP5, VP19C and VP22a in the nomenclature of Gibson & Roizman (1972)]. VP5 is commonly referred to as the major capsid protein, and is thought to form the main component of hexavalent capsomers (Vernon et al., 1981; Schrag et al., 1989). Using a polyclonal antibody against VP5 to assay the in vitro translation products of mRNAs which had been selected by hybridization to cloned DNA sequences, Costa et al. (1984) demonstrated that VP5 is encoded by gene UL19 (Davison & Scott, 1986a; McGeoch et al., 1988). Mutants with temperature-sensitive lesions in this gene fail to assemble capsids at the non-permissive temperature (Weller et al., 1987). VP19C is a DNA-binding protein (Braun et al., 1984a), and has been proposed to be either a constituent of the pentavalent capsomers (Vernon et al., 1981) or an internal capsid component (Braun et al., 1984b). It has been mapped using recombinants between HSV-1 and HSV-2 to a region of the genome containing part of gene UL39, all of UL40 and UL41 and part of UL42 (Braun et al., 1984a). VP22a is involved in packaging viral DNA into capsids and in capsid maturation. Unlike the other capsid proteins, VP22a is highly processed, comprising many phosphorylated and non-phosphorylated forms which resolve into numerous spots when examined by two-dimensional gel electrophoresis (Braun et al., 1984b). Preston et al. (1983) showed that a mutant containing a temperature-sensitive lesion in gene UL26 is defective in the processing of VP22a, and proposed that UL26 encodes VP22a. The mutant accumulates empty capsids at the non-permissive temperature. Braun et al. (1984a) have suggested that VP21 (Table 1) may be a slowly migrating form of VP22a. Map locations have not been reported for the genes encoding the remaining capsid proteins (VP23, VP24 and 12K).

We have identified the genes encoding VP19C and VP23 by sequencing the amino termini of the purified capsid proteins. In order to obtain sufficient protein for sequencing, capsids were prepared from batches of 10 roller bottles of BHK C13 cells. The cells were infected at 0·002 p.f.u./cell and harvested after incubation for 3 days at 37 °C. Capsids were isolated using the procedure of Irmierre & Gibson (1985). Briefly, nuclei from infected cells were subjected to three cycles of freezing on solid CO₂ and thawing at 37 °C. Nuclear debris was then pelleted and the supernatant subjected to sucrose gradient centrifugation. Additional capsids could be extracted from the nuclear debris by repeating this procedure up to four times. After centrifugation through 5 to 40% (w/w) sucrose gradients at 40000 r.p.m. for 20 min, two capsid bands [A (coreless) and B (cored) (Gibson & Roizman, 1972)] were visible. These were collected and examined by electron microscopy (Fig. 1a).
Table 1. Comparison of HSV-1 capsid protein sizes and nomenclatures

| Gene* | $M_r \times 10^{-3}$ predicted | $M_r \times 10^{-3}$† | Name‡ | $M_r \times 10^{-3}$§ | Name¶ | $M_r \times 10^{-3}$|| |
|-------|-------------------------------|----------------------|-------|------------------------|-------|------------------------|
| UL19¶ | 149                           | 150                  | VP5   | 155                    | NC1   | 154                    |
| UL38  | 50–3                          | 54                   | VP19C | 53                     | NC2   | 50                     |
|       |                               | 42                   | VP21  | 44                     | NC3   | 40                     |
| UL26**| 62.5                          | 40                   | VP22a | 38.8                   | NC4   | 38                     |
| UL18  | 34.3                          | 34                   | VP23  | 33                     | NC5   | 33                     |
|       |                               | 24                   | VP24  | 25                     | NC6   | 26                     |
|       |                               | 12                   |       |                        | NC7   | 12                     |

* From McGeoch et al. (1988).
† From this paper.
‡ From Gibson and Roizman (1972). VP19 was renamed VP19C by Heine et al. (1974).
§ From Cohen et al. (1980).
¶ From Heilman et al. (1979).
¶ Mapped by Costa et al. (1984) and Davison & Scott (1986a).
** Mapped by Preston et al. (1983).

Subsequent analyses were performed using the more abundant B capsids.

Capsid proteins were separated by SDS–PAGE on a 5 to 15% (w/v) gradient gel (Fling & Gregerson, 1986) and electroblotted to polyvinylidene difluoride (PVDF) membranes (Millipore) as described by Matsudaira (1987). The transfer buffer was 50 mM-glycine, 50 mM-Tris–HCl pH 10, 10% (v/v) methanol. Aliquots containing approximately 5% of each capsid preparation were subjected to SDS–PAGE on the same gel as the samples that were electroblotted, but this portion of the gel was stained with Coomassie blue without electroblotting. An example is shown in Fig. 1(b). The profile closely resembles those obtained by other groups, and the estimated $M_r$ values of major bands correspond closely to those described previously (Table 1). Our assignments of the various other nomenclatures to the bands are given in Table 1. Inspection of Fig. 1(b) also reveals a spectrum of minor proteins in addition to the seven previously recognized capsid proteins. Minor proteins were detected in all the capsid preparations that were examined. However, we believe these to be contaminating cellular species since they were not present in the autoradiographic profiles of capsids labelled with [35S]methionine.

Samples from two capsid preparations were subjected to SDS–PAGE and electroblotted separately. Portions of PVDF membrane containing stained bands corresponding to those designated VP5, VP19C, VCP23 and VP24 (Fig. 1b) were excised from each of the two blots and sequenced separately on an Applied Biosystems 470A gas-phase sequencer. Duplicate samples of VP19C and VP23 were sequenced from each of the two blots. VP22a and 12K were excised and sequenced from only one blot. No attempt was made to sequence VP21 or the other minor bands. No amino acid residues could be assigned for VP5, VP22a, VP24 and 12K. It is possible that modification of the amino termini (e.g. acetylation) of these proteins in vitro makes them refractory to the sequencing chemistry. Sequencing of proteolytic fragments generated from internal regions of the proteins may be necessary in order to obtain sequence data from these proteins.

We successfully determined the amino acid sequences of the amino termini of VP19C and VP23. Unambiguous assignments were made for 10 and 11 residues of VP19C and VP23, respectively. In single-letter code the amino acid sequences obtained were MKTNPLPATP for VP19C and MLADGFETDIA for VP23. The initial yield for these samples was estimated to be about 5% (the amount of protein in each band was estimated by comparison with blots of known amounts of standard proteins). This value is consistent with the sequenced species being the major component of each band, since control experiments where known amounts of standard proteins were electroblotted and sequenced gave initial yields of between 5 and 10%.

When the sequences obtained for VP19C and VP23 were compared with the amino acid sequences of proteins predicted from the complete HSV-1 DNA sequence (McGeoch et al., 1985, 1986, 1988; Perry & McGeoch, 1988), they aligned uniquely and precisely with the amino termini of the proteins encoded by genes UL38 and UL18, respectively. We conclude that VP19C is encoded by UL38, and VP23 by UL18.

No protein or function has previously been mapped to UL18. This gene encodes a protein of predicted $M_r$ of 34268, a value very similar to the $M_r$ of 34000 for VP23 estimated from SDS–PAGE. A different location for the gene encoding VP23 was derived by Lemaster & Roizman (1980) from an analysis of recombinants between HSV-1 and HSV-2. The protein involved,
Fig. 1. Purification of capsid proteins. (a) Electron micrograph of purified B (cored) capsids. Two μl of sample was adsorbed to a parlodion-coated grid, blotted dry and stained with 3% sodium silicotungstate at pH 7. The bar represents 200 nm. (b) SDS-PAGE analysis of the capsid preparation shown in (a). The sample was subjected to SDS-PAGE and stained as described in the text. Major protein bands are indicated on the right using the nomenclature of Gibson & Roizman (1972), except for the 12K band which they did not describe. The positions and sizes (×10^{-3}) of Mr standards are indicated on the left.

However, was subsequently shown to be a component of the tegument and not a capsid protein (Braun et al., 1984b). The role of VP23 is obscure at present and, as no conditional lethal mutations have been mapped in UL18, it is not known whether VP23 is essential for capsid formation. VP23 is present in both cored and coreless capsids, and surface labelling studies indicate that it is at least partially exposed on the capsid surface (Braun et al., 1984b). The location and number of molecules of this protein per capsid has led to the suggestion that VP23 forms the trimeric structures that link adjacent hexameric capsomers (Schrag et al., 1989). We note that UL18 (VP23) and UL19 (VP5) are transcribed as a 3′ coterminal family of mRNAs (Costa et al., 1984). The significance of this is unknown. The amino acid sequences of VP23 and VP5 appear unrelated and there is no evidence that these genes have arisen by gene duplication.

With respect to UL38, some information on its function has been reported by Pertuiset et al. (1989). They described a temperature-sensitive mutant with a lesion in UL38, which made viral DNA and late proteins but failed to assemble capsids at the non-permissive temperature. Hence, they proposed that the UL38 product functions in capsid morphogenesis, either as an integral capsid component or as a non-structural protein required for capsid assembly. Our results now identify this protein as a constituent of the capsid, where it presumably forms an essential structural component. UL38 encodes a protein with a predicted Mr of 50260, a size which corresponds well with the Mr of 54000 for VP19C estimated from SDS-PAGE. For both UL18 (VP23) and UL38 (VP19C), the similarities between the molecular masses predicted from the DNA sequence and estimated by SDS-PAGE, and the presence of an unmodified amino terminus, suggest that these two capsid proteins do not undergo extensive post-translational processing. However, it has been shown by two-dimensional gel electrophoresis under sequential non-denaturing then denaturing conditions, that VP19C and VP5 are linked in the capsid by disulphide bonds (Zweig et al., 1979). Braun et al. (1984a) used recombinants between HSV-1 and HSV-2 to map VP19C to a region of the genome close to, but excluding, UL38. We are unable to reconcile their map location for the gene encoding VP19C with ours, but cannot dismiss the possibility that two different capsid proteins comigrate when subjected to SDS-PAGE, and have been mapped to their respective genes by different mapping techniques.

With the addition of UL18 (VP23) and UL38 (VP19C) to UL19 (VP5) and UL26 (VP22a), we now know the genes encoding four of the seven major capsid proteins. In addition to that of HSV-1 (an alphaherpesvirus), complete DNA sequences are known for varicella-zoster virus (an alphaherpesvirus; Davison & Scott, 1986b), human cytomegalovirus (a betaherpesvirus; B. Barrell, personal communication) and Epstein–Barr virus (a gammaherpesvirus; Baer et al., 1984). Amino acid sequence comparisons in all cases reveal counterparts for UL18, UL19, UL26 and UL38. Conservation of the proteins that constitute the capsid is expected since herpesviruses from the three subfamilies (alpha-, beta- and gammaherpesviruses) have been shown to possess morphologically similar capsids composed of a similar spectrum of proteins (reviewed by Dargan, 1986).

We wish to thank Professor J. H. Subak-Sharpe for critical reading of the manuscript and Mrs Clare Addison for excellent technical assistance. M. Davison was supported by Science and Engineering Research Council grant no. GR/E74779.
References


COSTA, R. H., COHEN, G., EISENBERG, R., LONG, D. & WAGNER, E. (1984). Direct demonstration that the abundant 6-kilobase herpes simplex virus type 1 mRNA mapping between 0.23 and 0.27 map units encodes the major capsid protein VP5. Journal of Virology 49, 287-292.


(Received 5 December 1989; Accepted 1 February 1990)