Volume 13 Number 3 1985 Nucleic Acids Research

Detailed analysis of the mRNAs mapping in the short unique region of herpes simplex virus type 1

Frazer J.Rixon and Duncan J.McGeoch

Medical Research Council Virology Unit, Institute of Virology, Church Street, Glasgow G11 5JR, UK

Received 5 December 1984; Accepted 7 January 1985

ABSTRACT

We have analysed the mRNAs which map within the short unique (U_S) region of the herpes simplex virus type 1 (HSV-1) genome. U_S has a total length of 12979 base pairs (1) and is extensively transcribed with approximately 94% of the total sequence present in cytoplasmic mRNAs and 79% of the total sequence considered to be protein coding. There are several examples of overlapping functions and multiple use of DNA sequence within this region.

U_S contains 12 genes (1) which are expressed as 13 mRNAs. Two of these mRNAs are thought to arise from the same gene since they differ only slightly in the positions of their 5' ends and probably specify the same polypeptide. 11 of the 13 mRNAs are arranged into four nested families with unique 5' ends and common 3' co-termini. The other two mRNAs have unique 5' and 3' ends.

INTRODUCTION

Analysis of the arrangement and expression of genes in herpesviruses is not as far advanced as that of the smaller DNA viruses. This is mainly attributable to the large genome size (greater than 80 megadaltons of duplex DNA) of herpesviruses. However, it is already apparent for herpes simplex virus (HSV), that certain aspects of gene arrangements and transcription patterns differ markedly from those of the smaller DNA viruses. Earlier studies have established that the HSV genome is extensively transcribed with many overlapping mRNAs, the most frequent arrangement of which is as nested families with unique 5' ends and common 3' ends (2-6). Most HSV mRNAs are unspliced, and in the few cases where splicing has been described it occurs in the untranslated portions of the mRNAs concerned and appears to have no effect on their coding potential (7-10). 3' co-terminal families of mRNAs have been described in

adenoviruses and papovaviruses but in these cases the families also have common 5' termini and individual mRNAs are produced by differential splicing (11-15); in many cases the generation of individual polypeptides from overlapping genes is also a result of differential splicing. In general the extent of gene overlap reported for HSV is less than that found in the smaller DNA viruses. Only one example is published for HSV of the use of overlapping reading frames (6); however, several cases have been described where the 5' ends and promoter sequences of particular mRNAs lie within the polypeptide coding sequences of overlapping mRNAs (2,6).

To date, HSV-1 transcription has largely been analysed in the absence of extensive DNA sequence data. In this paper we give the first description of the transcription pattern of an extensive region of the herpes simplex virus type 1 (HSV-1) genome for which the complete DNA sequence is known, namely the 12979 bp short unique region (U_S)(1).

MATERIALS AND METHODS

Cells and virus

Baby hamster kidney 21 (C13) cells were grown as monolayers in rotating 80 oz. bottles. All infections were carried out using HSV-1 (Glasgow strain 17) at 37°C. For the production of immediate-early (IE) RNA, cell monolayers were infected at a multiplicity of infection of 50 p.f.u./cell. The cell monolayers were pretreated and maintained in medium containing cycloheximide as previously described (16). For the production of early and late RNA, cell monolayers were infected at a multiplicity of infection of 10 p.f.u./cell and incubated for the appropriate time. Cytoplasmic RNA was prepared using the method of Kumar and Lindberg (17).

Cloning Procedures

Fragments of HSV-1 DNA, generated by using restriction endonucleases, were cloned into pAT 153 and grown in *E. coli* K12 HB101. The procedures used for cloning and isolation of cloned virus DNA were those described by Davison and Wilkie (18).

Northern blot analysis

Cytoplasmic polyadenylated RNA was electrophoresed on

denaturing formaldehyde agarose gels and transferred to nitrocellulose (19). Prehybridisation and hybridisation of the DNA probe to the nitrocellulose were performed in 50% formamide, 1 x Denhardt's solution (20)(0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 2 x SSC (1 x SSC = 0.15 M NaCl, 0.015 M sodium citrate) and 50 ug/ml calf thymus DNA. Prior to hybridisation the DNA probe was denatured in 100% formamide at 100°C for 10 min and rapidly chilled on ice. Hybridisation was performed at 45°C for 20 h. Blot strips were washed repeatedly in 2 x SSC at 60°C and exposed to Kodak X-Omat-S film at -70°C.

Structural analysis of mRNAs

The DNA/RNA hybridisation procedures and the nuclease S1 and exonuclease VII digestion procedures were carried out as described previously (8).

Denaturing polyacrylamide gels containing 9M urea were run in 90 mM Tris, 90 mM boric acid, pH 8.3, 1 mM EDTA essentially as described by Maxam and Gilbert (21). Samples were dissolved in deionised formamide and denatured at 90°C for 2 min before loading. Electrophoresis was carried out at room temperature for 3-6h at 40W. The radiolabelled bands were detected by autoradiography.

RESULTS

Fig. 1 shows restriction enzyme site maps for the short unique (U_S) region of the HSV-1 strain 17 genome and identifies the cloned DNA fragments used in the following analyses.

Northern Blotting Analysis

Initial examination of the mRNAs mapping in U_S was performed by probing blot strips containing oligo-dT selected RNA made under IE, early (3h PI) and late (6h PI) conditions (Fig 2). With IE mRNA a single band of about 2 kb was detected with pGX156, pGX35, pGX45 and pGX55 (data not shown). With 3h mRNA a single major additional band of about 2.8 kb was detected on blot strips probed with pGX156, pGX45 and pGX43 (Fig. 2A, Tracks 1,3 and 4). With 6h mRNA, pGX156 gave a complicated pattern of bands and hybridisation with a series of cloned DNA fragments from across U_S gave the following results (Fig. 2B).

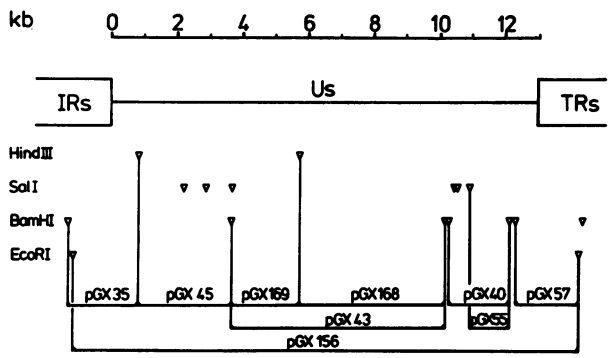

FIG. 1 Diagram of the short unique region (U_S) of HSV-1 and part of the inverted short repeat regions (IR_S and TR_S) showing the restriction enzyme sites and cloned DNA fragments used to analyse the mRNAs mapping in U_S. The scale in kilobases (kb) is numbered from the first base in U_S as described in the text.

pGX35 identified a single 2 kb band (Track 2). pGX45 identified bands of about 3.3, 2.8 to 2.6, 2.0 and 1.5 kb (Track 3). pGX43 identified a number of bands from around 3.3 to 2.6 kb, plus additional bands of around 1.6 and 1.0 kb (Track 4). pGX55 identified bands of around 2.7, 2.0, 1.6, 1.3 and 0.6 kb (Track 5).

Due to clustering of the bands around 2.7 and 2.0 kb and ignorance of their orientation, the interpretation of these Northern blots must be treated with caution. However, a number of points can be made. The 2.0 kb band detected with pGX35, pGX45 and pGX55 represents the two similar sized IEmRNAs, US1 and US12 (7,8), and the 1.6 and 1.3 kb bands detected with pGX55 represent US11 and US10 (6), which have previously been mapped to these fragments. The detection of a band of around 2.8 kb with pGX45 and pGX43 at 3h (Fig 2A, Tracks 3 and 4) indicates that an mRNA of this size is transcribed across the junction between these fragments. A number of bands are only detected with a single probe (other than pGX156) suggesting that they map entirely within these fragments. These include the 1.5 kb band in pGX45 (Track 3), the 1.0 kb band in pGX43 (Track 4) and the 0.6 kb band in pGX55 (Track 5).

The map locations of the mRNAs identified by Northern

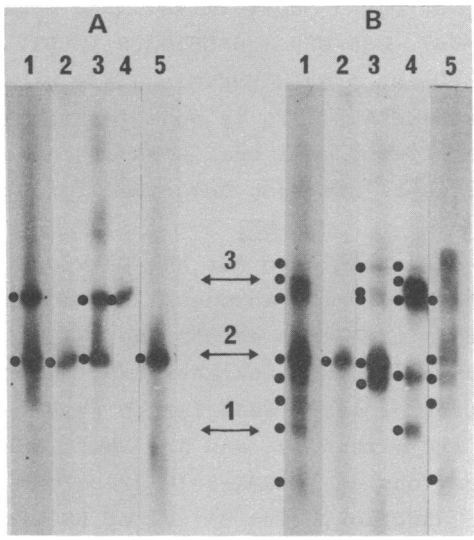

FIG. 2 Northern blotting analysis of mRNAs mapping in U_S. Panel A shows the pattern obtained with early (3h) mRNA and panel B the pattern obtained with late (6h) mRNA. The blot strips were probed with the following cloned DNA fragments: 1. pGX156, 2. pGX35, 3. pGX45, 4. pGX43, 5. pGX55. The arrows indicate the approximate positions of 3, 2 and 1 kb bands. Dots to the left of each track indicate the positions of the bands described in the text. The labelling present above the 3 kb position in panel B, track 5 is not thought to represent mRNA since it was not observed in equivalent mRNA preparations.

blotting were determined more precisely using nuclease digestion procedures.

Characterisation of Individual mRNAs

U_S contains all or part of twelve genes which are numbered US1 to US12 from the left to the right in accordance with the nomenclature of McGeoch *et al.* (1). This nomenclature is also used here for the mRNAs and polypeptides specified by these genes. The length of U_S in the clones sequenced by McGeoch *et al.* (1) and used in the analyses described here is taken as 12979 bp, and the numbering of the nucleotide sequence is from left to right beginning with the first base in U_S adjacent to the IR_S/U_S junction.

The mRNAs mapping within U_S fall into 6 groups based on shared 3' ends. US1 and US2 have unique 3' termini and the

remainder belong to four 3' co-terminal mRNA families comprising US3/US4, US5/US6/US7, US8/US9 and US10/US11/US12. In the following descriptions, single positions are given for the 5' and 3' ends of each mRNA. This is done for simplicity and more detailed analysis might reveal that some of these mRNAs have multiple start sites as has been described for US6 (22).

US1

US1 is the name used here for the previously described major IEmRNA (IEmRNA-4) which encodes a 68K polypeptide (IE68)(23,24). This mRNA maps across the IR_S/U_S junction and is spliced with a 248 bases long untranslated leader joined to an unspliced 3' portion of 1420 bases length (7,8). The leader and intron sequences lie within IR_S and are identical in sequence to the equivalent portions of IEmRNA-5 (US12 below) which maps across the TR_S/U_S junction. The 3' end of US1 is located at position 1356. The first AUG occurs within U_S, 40 bases from the junction with IR_S, and initiates an open reading frame which extends to position 1299.

US2

Northern blotting analysis had identified a 1.5 kb mRNA within pGX45. Examination of the DNA sequence suggested that a leftward transcribed mRNA was located downstream from the 3' end of US1. A fragment of DNA, cloned into the single stranded vector M13, which would hybridise specifically to such an mRNA, was used as a probe against the Northern blots and confirmed that the 1.5 kb mRNA originated from this region (data not shown).

To locate the 3' end of this mRNA, nuclease digestion analysis was performed using pGX45 which had been TaqI digested and 3' labelled. The 995 bp TaqI fragment, mapping between positions 919 and 1913, generated two nuclease S1 and exonuclease VII resistant bands of 438 and 495 bases length (data not shown). The 438 bases long band is formed by the 3' portion of US1 and the 495 bases long band is formed by the 3' portion of US2. To confirm this designation the analysis was repeated using a 447 bp HinfI/DdeI fragment (position 1302 to 1748) which was uniquely 3' labelled at the DdeI site. This generated a single 330 bases long nuclease S1 resistant band

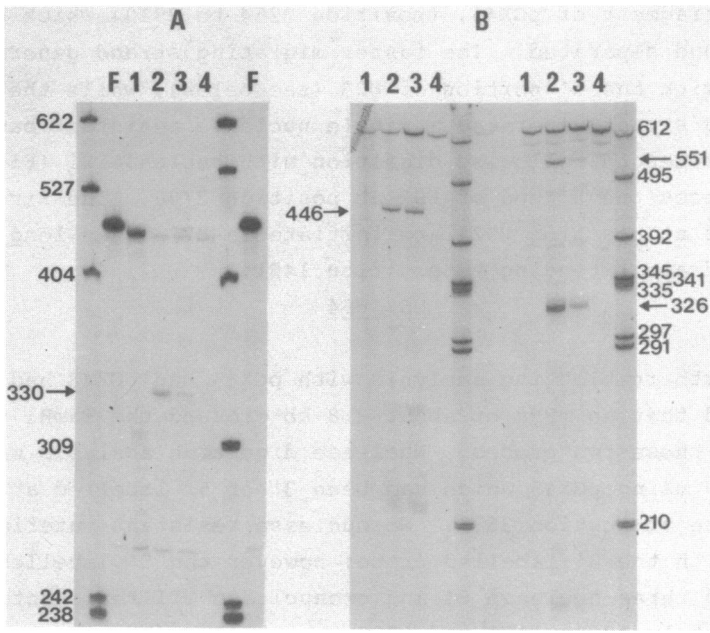

FIG. 3 Nuclease S1 analysis of US1, US2 and US3. Probes complementary to the 3' portion of US2 and 5' portions of US2 and US3, were hybridised to the following mRNA samples: 1. 15 ug of cytoplasmic mRNA made under IE conditions. 2. 15 ug of cytoplasmic mRNA prepared at 4h PI. 3. 15 ug of cytoplasmic mRNA prepared at 7h PI. 4. 20 ug of mock infected cytoplasmic mRNA. All samples were digested with nuclease S1. The DNA probes used were the 447 bp HinfI/DdeI subfragment of pGX45 which was 3' labelled at the DdeI site (panel A) and the slower (tracks on left) and faster (tracks on right) migrating strands of the 5' labelled, 657 bp SalI subfragment of pGX45 (panel B). The sizes of the protected fragments are given alongside the bands. Tracks marked with an F show unhybridised probe. 3' labelled HpaII digests of pAT153 and pBR322 DNA (panel A) and 5' labelled HincII digested øX 174 DNA (panel B) were used as size standards

(Fig. 3A). These analyses place the 3' end of US2 at position 1419, 63 bp from the 3' end of US1 at position 1356.

To locate the 5' end of US2, nuclease digestion analysis was performed using pGX45 which had been 5' labelled at the unique BstEII site at position 1459. A single resistant band approximately 1240 bases long was found following either nuclease S1 or exonuclease VII digestion (data not shown). The 5' end was located more precisely using a 5' labelled 657 bp

SalI subfragment of pGX45, (position 2254 to 2910) which had been strand separated. The faster migrating strand generated hybrids with the 5' portion of US3 (see below), while the slower migrating strand generated a single nuclease resistant band of 446 bases length following digestion with nuclease S1 (Fig 3B), which places the 5' end of US2 at position 2700. The first AUG in US2 is at position 2324 and initiates a 873 bases long open reading frame extending to position 1451.

US3/US4

US3

Northern blotting analysis with pGX45 and pGX43 had indicated that an mRNA of about 2.8 kb crossed the BamH1 site defining these two probes. Nuclease digestion analysis was performed using pGX45 which had been 3' or 5' labelled at the BamH1 site at postion 3689. No nuclease resistant material was formed with the 3' labelled probe; however the 5' labelled probe generated three nuclease S1 and exonuclease VII resistant bands approximately 1300, 1100 and 1020 bases long (data not shown).

To locate these possible 5' ends more precisely the same strand separated 657 bp SalI fragment (positions 2254 to 2910) described under US2 was used. The faster migrating strand generated two resistant bands of 551 and 326 bases length following digestion with nuclease S1 (Fig. 3B). These bands correspond to the 1300 and 1100 bases long bands formed with 5' labelled, BamH1 digested pGX45. A third band of around 240 bases which would correspond to the 1020 bases long band was not seen. We believe that this 1020 bases long band was a result of displacement of the DNA probe from the hybrid formed with the US3 mRNAs by the overlapping portion of the US2 mRNA. This would be less important at the lower temperatures used with single stranded DNA probes, when the greater stability of an RNA/RNA duplex would be less significant.

A summary of these data is given in Fig 4. US3 appears to possess two distinct 5' termini at positions 2360 (US3A) and 2585 (US3B). For each of these 5' ends the first downstream AUG is at position 2618 and this initiates a 1443 bases long open reading frame which extends to position 4061. Thus these two subcomponents of US3 have identical coding capacities. The two

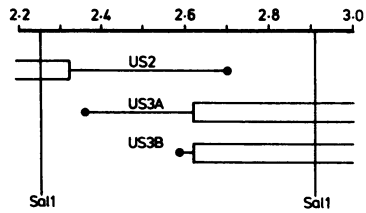

FIG. 4 Arrangement of the 5' portions of the US2 and US3 mRNAs. The scale at the top gives the distance in kb from the IR_S/U_S junction as described in the text. The vertical lines indicate the Sall restriction enzyme sites used in mapping these 5' ends. US2 and US3 are transcribed in opposite orientations. The untranslated portion of each mRNA is represented by a line and the open reading frame downstream from the first AUG by an open box. US3 has two components (US3A and US3B) with distinct 5' ends, both of which initiate translation at the same AUG. The region of overlap between US2 and US3 is largely untranslated but the 5' end of US2 lies within the predicted US3 coding sequences and upstream control sequences for US3A may extend into the US2 coding sequences.

5' termini of US3 both overlap the unique 5' end of US2 which maps at position 2700. However, neither extends to within the proposed polypeptide coding sequences of US2 (beyond position 2324). The 5' end of US2 at position 2700 does overlap the proposed polypeptide coding sequences of US3 beyond position 2618 over a length of 82 bp.

The 3' end of US3 was located using pGX169 which was 3' labelled at the BamHl site at position 3688. This generated a single nuclease resistant band of 1240 bases length (data not shown) placing the 3' end around position 4930 (see below).

US4

The 3' end of US3 maps approximately 850 bp downstream from the end of the open reading frame at position 4061. To identify any mRNA mapping within this 3' untranslated portion, a 5' labelled, 776 bp, BamHl/Xmal subfragment of pGX169 (positions 3686-4461) was used. In addition to the completely protected material generated by US3, a 337 bases long nuclease resistant band was found following nuclease Sl or exonuclease VII digestion (Fig. 5A). This indicated that the 5' end of another mRNA (US4) is located at position 4125. The position of the 5' end was confirmed using a 445 bp Hinfl fragment (positions 3867

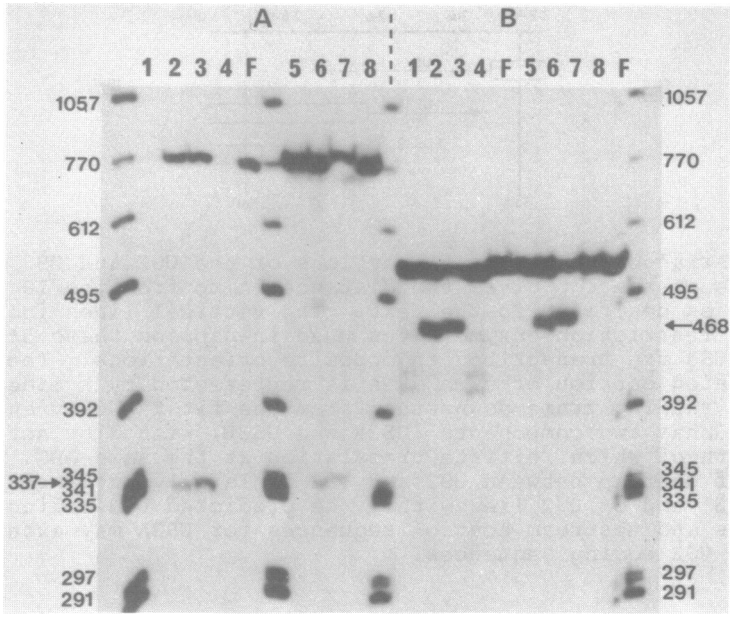

FIG. 5 Nuclease S1 and exonuclease VII analysis of US3 and US4. The hybridisations were performed as described for Fig. 3, ie. tracks 1 and 5. IEmRNA. Tracks 2 and 6. 4h mRNA. Tracks 3 and 7. 7h mRNA. Tracks 4 and 8. mock infected mRNA. Tracks marked with an F show unhybridised probe. Samples 1-4 were digested with nuclease S1 and samples 5-8 were digested with exonuclease VII. The DNA probes used were a 5' labelled, 776 bp BamH1/Xmal subfragment of pGX169 (panel A) and a 3' labelled, 545 bp Xmal subfragment of pGX169 (panel B). The sizes of the protected fragments are given alongside the bands. 5' labelled HinclI digested øX 174 DNA was used as a size standard.

to 4311), which generated a 187 bases long nuclease resistant product (data not shown). The first AUG downstream from the 5' end is at position 4140 and initiates a 714 bases long open reading frame which extends to position 4854.

The common 3' end of US3 and US4 was located precisely using the 545 bp Xmal subfragment of pGX169 (positions 4460 to 5004). This generated a 468 bases long nuclease resistant product (Fig. 5B) placing the 3' co-terminus of US3 and US4 at position 4927. This position was confirmed using a 544 bp DdeI fragment (positions 4483 to 5037), which generated a 445 bases long nuclease resistant band (data not shown).

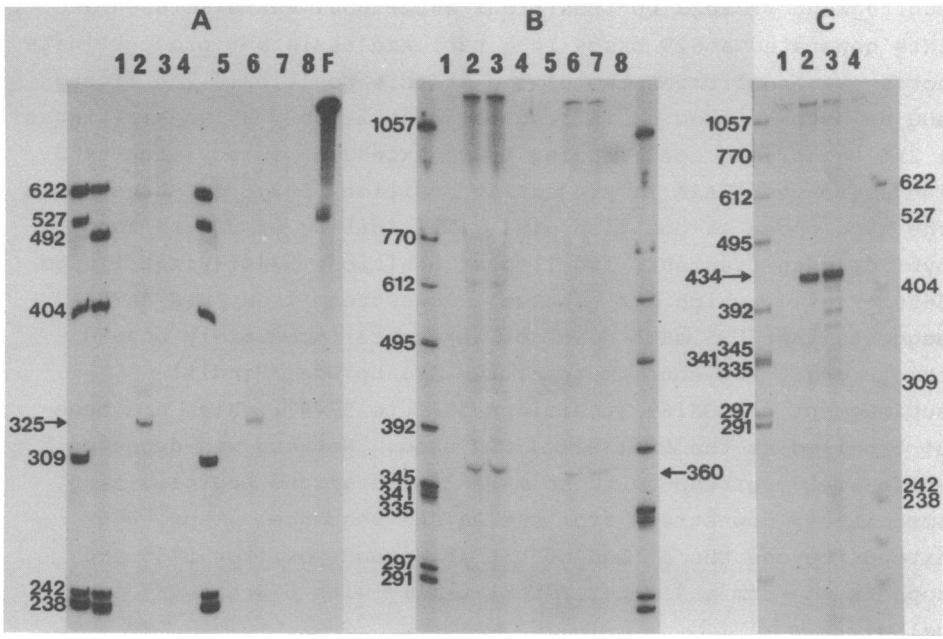

FIG. 6 Nuclease S1 and exonuclease VII analysis of US5, US6 and US7. The hybridisations were performed as described for Fig. 3, ie. tracks 1 and 5. IEmRNA. Tracks 2 and 6. 4h mRNA. Tracks 3 and 7. 7h mRNA. Tracks 4 and 8. mock infected mRNA. Track F in panel A shows unhybridised probe. Samples 1-4 were digested with nuclease S1 and samples 5-8 were digested with exonuclease VII. The DNA probes used were a 5' labelled, 1428 bp RsaI subfragment of pGX169 (panel A), a 5' labelled, 1725 bp HindIII/BstEII subfragment of pGX168 (panel B) and a 3' labelled, 1308 bp BstEII subfragment of pGX168. The sizes of the protected fragments are given alongside the bands. 3' labelled HpaII digests of pAT153 and pBR322 DNA were used as size standards in panel A and 5' labelled HincII digested øX 174 DNA was used as a size standard in panels B and C.

US5/US6/US7

US5

Nuclease digestion analysis with pGX169 which had been 5' labelled at the HindIII site (position 5744) generated a resistant band of around 780 bases length (data not shown). The 5' end of this mRNA (US5) was located precisely using a 5' labelled, 1428 bp, RsaI subfragment of pGX169 (positions 3924 to 5351). This gave a 325 bases long nuclease resistant product (Fig. 6A) placing the 5' end at position 5026. A HinfI

subfragment of 1029 bp (positions 4622-5650) was also used and this generated a 625 bases long nuclease resistant product (data not shown) confirming the 5' end at this position. The first AUG downstream from the 5' end is at position 5127 and initiates a 276 bases long open reading frame extending to position 5403. The sequence AATAAA is present at position 5589, downstream from the stop codon at position 5403. The ability of US5 to form hybrids with fragments labelled at positions 5744 (HindIII) and 5650 (HinfI), which are both well downstream from this AATAAA, suggests that US5 mRNA does not terminate immediately beyond this signal. To confirm this, the 673 bp DdeI/HindIII subfragment of pGX169 (positions 5071 to 5744), which had been 3' labelled at the DdeI site, was used. No band was detected which would correspond to an mRNA 3' end around position 5600, immediately downstream from the AATAAA sequence. Thus, US5 extends beyond the 5' end of US6 at around position 5735 and appears to form a 3' co-terminal family with US6 and US7 (see below).

US6

This mRNA, which encodes glycoprotein D, has been extensively described in a number of laboratories (4,10,22) and no attempt was made to analyse it further. The position of the 5' end and TATA box are taken from Everett (22).

US7

The 5' end of US7 was examined using a 5' labelled 1725 bp HindIII/BstEII subfragment of pGX168 (positions 5740 to 7464). This generated a completely protected band (due to hybridisation with US6 which is transcribed right across this fragment), and a band of 360 bases length formed with the 5' portion of US7 (Fig. 6B). This places the 5' end of US7 at position 7090; ie in a similar location to that described previously (4,10). The first AUG in US7 is at position 7181 and initiates a 1170 bases long open reading frame which extends to position 8351.

US6 and US7 form a 3' co-terminal family to which a further member, US5, can now be added. The 3' end of this group of mRNAs was located using pGX168 which had been 3' labelled at the HindIII site (position 5743). This generated a nuclease resistant band of around 2700 bases length (data not shown).

The 3' end was located precisely using a 3' labelled 1308 bp BstE11 subfragment of pGX168 (positions 7995 to 9302). This gave a 434 bases long nuclease resistant band (Fig. 6C) which places the 3' end of US5, US6 and US7 at position 8429. This agrees with the previous assignment for these mRNAs (4,10).

US8/US9

US8

This mRNA is thought to encode glycoprotein E (1). Nuclease analysis was performed using pGX168 which had been 5' labelled at the BamH1 site at position 10146. This generated a single nuclease S1 and exonuclease VII resistant band of 1580 bases length (data not shown). To locate the 5' end more precisely, hybridisation was performed with a 5' labelled 680 bp BstN1 subfragment of pGX168 (positions 8037 to 8717). This generated nuclease resistant bands of around 152 bases length (Fig. 7A) which place the 5' end of US8 at position 8566. The first AUG downstream from the 5' end of US8 is at position 8639 and initiates a 1650 bases long open reading frame which extends to position 10289.

The 3' portion of US8 was examined using pGX40 3' labelled at the BamH1 sites at positions 10236 and 12074. Two nuclease resistant products 853 and 561 bases long were detected (data not shown). The 561 bases long band corresponds to the previously analysed 3' ends of US10, US11 and US12 (6-8) at position 11514, and the 853 bases long band represents the 3' end of US8 (see below).

US9

Northern blotting had identified a 0.6 kb mRNA which was detected with pGX55. Examination of the nucleotide sequence for the region encoding US8 revealed that a 270 bases long open reading frame, extending from an AUG at position 10708 to position 10978, occupied the untranslated 3' portion of US8 downstream from the TAA stop codon at position 10289. To determine the possible presence of any mRNA mapping within this region, the 384 bp Sall subfragment of pGX40 (positions 10498-10881) was isolated and 5' labelled. This generated nuclease resistant bands of around 241 bases length which are formed by the 5' portion of US9 (Fig. 7B) and locate the 5' end

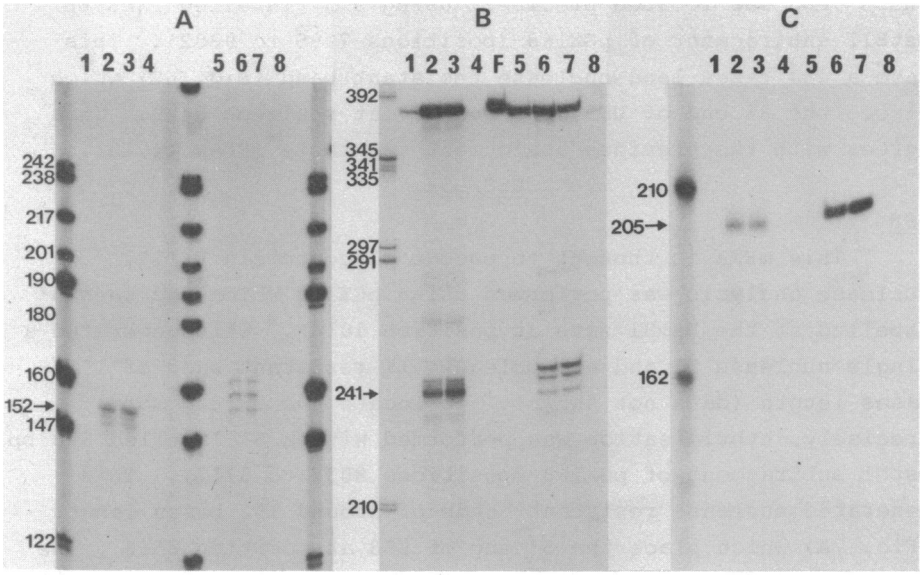

FIG. 7 Nuclease S1 and exonuclease VII analysis of US8 and US9. The hybridisations were performed as described for Fig. 3, ie. tracks 1 and 5. IEmRNA. Tracks 2 and 6. 4h mRNA. Tracks 3 and 7. 7h mRNA. Tracks 4 and 8. mock infected mRNA. Track F shows unhybridised probe. Samples 1-4 were digested with nuclease S1 and samples 5-8 were digested with exonuclease VII. The DNA probes used were a 5' labelled, 680 bp BstN1 subfragment of pGX168 (panel A), a 5' labelled, 384 bp Sall subfragment of pGX40 (panel B) and pGX55 3' labelled at the Sall site. The sizes of the protected fragments are given alongside the bands. 3' labelled Hpall digests of pAT153 and pBR322 DNA were used as size standards in panel A and 5' labelled Hincll digested øX 174 DNA was used as a size standard in panels B and C.

at position 10641.

To locate precisely the common 3' end of US8 and US9, pGX55 3' labelled at the Sall site at position 10881 was used. This resulted in a 205 bases long nuclease resistant product (Fig. 7C) which places the 3' end at position 11086.

US10/US11/US12

These three mRNAs which have been described elsewhere (6) form a 3' co-terminal family one member of which (US12) is IEmRNA-5. Their 3' co-terminus is at position 11514 and their unique 5' ends are at positions 12561 (US10), 12855 (US11) and in TR_S (US12). The portion of US12 which maps within TR_S is

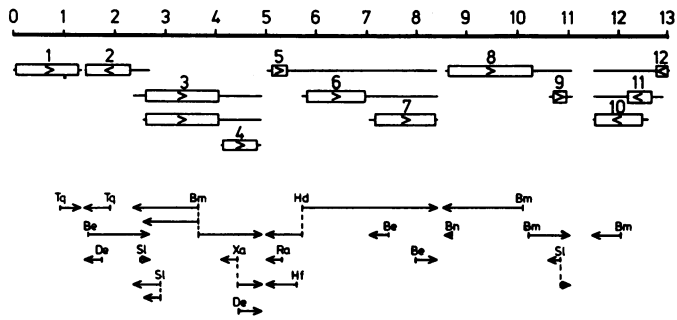

FIG. 8 Summary of the mRNAs mapping within U_S. The scale at the top is marked in kilobases. The upper part of the diagram shows the overall pattern of genes in U_S. The mRNA species are numbered 1 to 12 from left to right as described in this paper and in McGeoch et al. (1). The two forms of US3 which differ only in the locations of their 5' ends are illustrated. The untranslated portion of each mRNA is indicated by a line and the probable polypeptide coding portion by the closed boxes. The arrow heads within the boxes indicate the orientation of each mRNA. The lower part of the figure shows a summary of the nuclease digestion mapping data described in the text. The horizontal lines indicate the sizes of the protected DNA fragments from the restriction enzyme site used to map the mRNA to the appropriate 5' or 3' end. The nature of the restriction enzyme site used to generate each protected fragment is given by the following letters: Bm = BamHl, Be = BstEll, Bn = BstNl, De = Ddel, Hd = Hindlll, Hf = Hinfl, Ra = Rsal, Sl = Sall, Xa = Xmal. Where a particular restriction enzyme site was used in more than one analysis this is indicated by a vertical dashed line.

spliced and has an identical sequence to the equivalent portion of US1 (IEmRNA-4) which maps within IR_S (7,8,25,26).

Fig. 8 shows the transcription pattern for U_S and gives a summary of the mapping data presented above.

DISCUSSION

We believe that the mRNAs described here comprise all the major species expressed from the U_S region of the HSV-1 genome. Northern blotting detected a band of about 3.3 kb (Fig. 2B, track 3). None of the mRNAs detected during nuclease digestion analysis would give a band of this size, and in a Northern blotting study of this region by Watson et al. (4), using the HSV-1 Patton strain, no equivalent band was found. Further

analyses would be necessary to determine the exact nature and origin of this band.

The data presented here represent the first comprehensive description of the mRNAs mapping within the U_S region of HSV-1. However, a number of previous studies have examined particular genes in U_S. The two IE genes which extend into U_S from the flanking short repeat regions have been examined by several groups (7,8,23-29). Recently we showed that one of these (US12) is a member of a 3' co-terminal family (6). The present study indicates that the other IE mRNA (US1) does not overlap any of the other U_S mRNAs. The two polypeptides encoded by the mRNAs which overlap US12 (33K and 21K) had earlier been mapped to their approximate genome locations by Lee et al. (30) using in vitro translation of mRNA selected against cloned DNA fragments. They also mapped a number of other polypeptides within U_S, including glycoproteins D and E, a 42K polypeptide which mapped to the left of gD and a 55K polypeptide which co-mapped with gD. It seems likely that the 42K polypeptide is specified by US4 and the 55K polypeptide is specified by US7 (1). Watson et al. (4) have mapped the mRNA encoding gD (US6) and its downstream 3' co-terminal partner (US7) and determined the nucleotide sequence of the gD gene. Ikura et al. (10) have also mapped the gD mRNA and its downstream co-terminal partner. Certain polypeptides have been mapped within U_S by analysis of HSV-1/HSV-2 intertypic recombinants. These include the gD, gE, IE68, IE12, 55K and 21K polypeptides (28,31-36). In general all these mapping data agree with our mRNA analyses, with candidate mRNAs apparent for all the polypeptides mentioned above. In addition we have identified a further four genes (US2, US3, US5 and US9) for which there was little previous evidence. The polypeptides specified by these genes are described in more detail in McGeoch et al.(1). It seems likely that similarly detailed analysis of other regions of the HSV-1 genome will uncover many more previously unrecognised genes.

In all the mRNAs identified in U_S appropriate transcriptional control sequences are present around the 5' and 3' ends. Fig. 9A shows a comparison of the sequences around the 5' ends of US1-US12 giving the relative positions and sequences

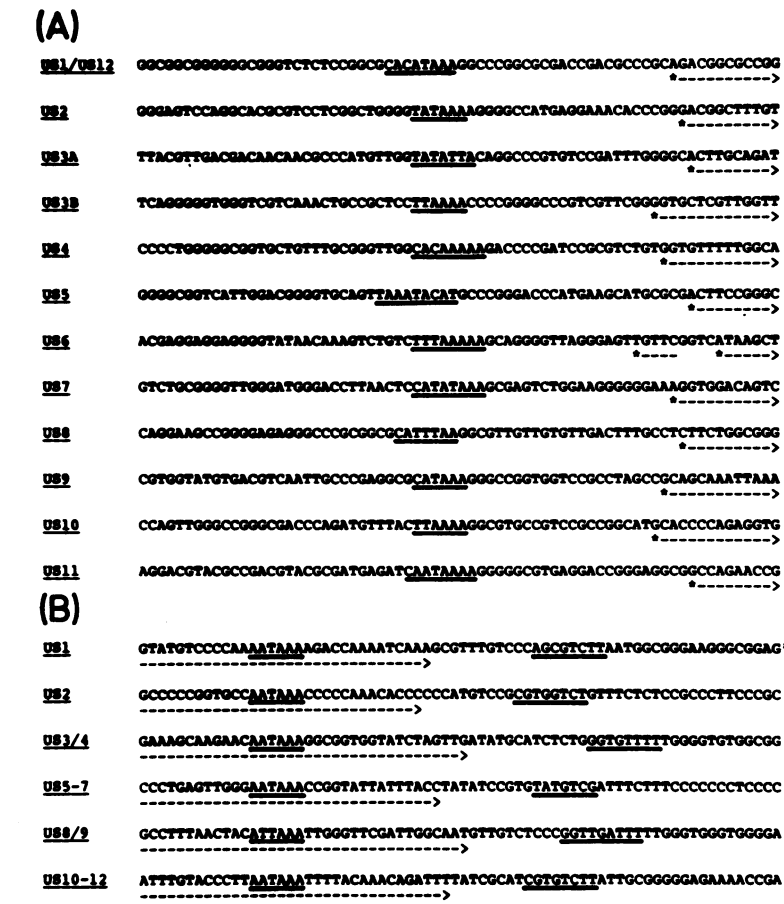

FIG. 9 Sequences around the 5' and 3' ends of the US mRNAs. All sequences are shown in the 5'-3' orientation. Sequence data are from Murchie and McGeoch (26) (5' terminal regions of US1 and US12 mRNAs) and McGeoch et al. (1).

Fig. 9A shows sequences around the 5' ends of the 13 US mRNAs which map completely or partly in U_S. A single line is shown for US1 and US12 since these two 5' ends map within IR_S and TR_S respectively and have identical sequences. Arrows below the sequence (*--->) define the position of each 5' end. The two positions given for US6 correspond to those described by Everett (22). The upstream 'TATA' signals (37) are underlined.

Fig. 9B shows sequences around the 3' ends of US1 and US2 and the four 3' co-terminal ends constituted by US3/US4, US5/US6/US7, US8/US9 and US10/US11/US12. Arrows below the sequence (--->) define the position of each 3' end. The 'AATAAA' and 'ATTAAA' polyadenylation signals (38) and the GT rich sequences downstream from the 3' ends, which are also thought to play a role in processing of mRNA 3' termini (41, McLauchlan et al. manuscript in preparation), are underlined.

of the upstream 'TATA' homologies (37). In five of the six examples shown in Fig. 9B an AATAAA polyadenylation signal (38) is present a short distance upstream of the 3' terminus. The sixth example (the US8/US9 3' co-terminus) has the sequence ATTAAA at this position. This sequence has previously been found to be a rare alternative to the normal AATAAA form (39,40). In all the US mRNAs other than US11 (6) the first ATG downstream from the 5' end initiates the longest available open reading frame and indicates the most probable site for the beginning of translation. The only uncertainty is in US5 where another in frame ATG (position 5019) is present 7 bp upstream from the position given for the 5' end. A slight error in mapping the 5' end of US5, which proved the most difficult U_S mRNA to analyse, would incorporate this ATG in the mRNA. Nevertheless we feel that the ATG at position 5127 remains the most probable start site for translation in US5.

Sequence analysis of U_S, using cloned DNA fragments, gave a length of 12979 bp (1). However, it is known that U_S contains several regions of variable length which are situated in both protein coding and non-coding regions (1). U_S encodes 12 polypeptides which are specified by open reading frames with a combined length of 10189 bp. Thus 79 % of the total length of U_S is polypeptide coding. The 2790 bp of intergenic sequence can be sub-divided into 2063 bp of untranslated 5' and 3' mRNA sequences and 727 bp of sequence which does not appear as cytoplasmic mRNA. Detailed analysis of the US6 promoter has shown that signals essential for control of transcription of this mRNA extend at least 80 bp upstream from this 5' end (22), and other studies suggest that sequences with a role in polyadenylation are present up to 30-40 bp downstream from the 3' end (41,42, McLauchlan et al. manuscript in preparation). Assuming that these values are approximately the same for the other mRNAs in U_S, it becomes clear that there is very little sequence in U_S which is not directly implicated in the structure and expression of genes. On the contrary there are several cases of dual usage of sequence. The most obvious example of this is the use of overlapping reading frames to encode genes US10 and US11 (6). Other examples include mRNA 5' ends which

map within the coding regions of overlapping mRNAs (US2, US10 and US11), overlap of mRNA 5' ends of mRNAs transcribed in opposite orientations from different DNA strands (US2 and US3), and cases where probable promoter sequences lie within protein coding sequences of overlapping mRNAs (promoters for US3A, US4, US10 and US11). Similar overlaps have been described from other regions of the HSV-1 genome (2,3). Within U_S, the only extensive region which has no obvious role in gene expression lies between positions 11086 and 11514. This region contains 10 copies of a tandemly reiterated 15 bp sequence between positions 11107 and 11259 (different copy numbers of this sequence are present in other cloned DNA fragments (1)). Such tandem reiterations are found in a number of other untranscribed regions of HSV-1 and are thought to comprise non-functional DNA produced by illegitimate recombination (43). Thus only around 250 bp of this region is unique sequence to which no function can yet be ascribed.

The degree of sequence utilisation found in U_S suggests that gene compression in HSV is intermediate between that found in the smaller DNA viruses and that in the eukaryotic genome where overlapping of functions is rare.

ACKNOWLEDGEMENTS

We thank Professor J.H. Subak-Sharpe for helpful discussion and comment on this manuscript and Mr. D. Joicey for expert technical assistance.

REFERENCES

1. McGeoch, D.J., Dolan, A., Donald, S. and Rixon, F.J. (1985) J. Molec. Biol. (in press).
2. McLauchlan, J. & Clements, J.B. (1983) J. Gen. Virol. 64, 997-1006.
3. Hall, L.M., Draper, K.G., Frink, R.J., Costa, R.H. & Wagner, E.K. (1982) J. Virol. 43, 594-607.
4. Watson, R.J., Colberg-Poley, A.M., Marcus-Sekura, C.J., Carter, B.J. & Enquist, L.W. (1983) Nucleic Acids Res. 11, 1507-1522.
5. Costa, R.H., Draper, K.G., Banks, L., Powell, K.L., Cohen, G., Eisenberg, R. & Wagner, E.K. (1983) J. Virol. 48, 591-603.
6. Rixon, F.J. and McGeoch, D.J. (1984) Nucleic Acids Res. 12, 2473-2487.
7. Watson, R.J., Sullivan, M. & Vande Woude, G.F. (1981) J. Virol. 37, 431-444.

8. Rixon, F.J. & Clements, J.B. (1982) Nucleic Acids Res. 10, 2241-2255.
9. Frink, R.J., Eisenberg, R., Cohen, G. and Wagner, E.K. (1983) J. Virol. 45, 634-647.
10. Ikura, K., Betz, J.L., Sadler, J.R. and Pizer, L.I. (1983) J. Virol. 48, 460-471.
11. Broker, T.R., Chow, L.T., Dunn, A.R., Gelinas, R.E., Hassell, J.A., Klessig, D.F., Lewis, J.B., Roberts, R.J. and Zain, B.S. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 531-553.
12. Berk, A.J. and Sharp, P.A. (1978) Cell 14, 695-711.
13. Chow, L.T., Broker, T.R. and Lewis, J.B. (1979) J. Mol. Biol. 134, 265-303.
14. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. and Ysebaert, M. (1978) Nature 273, 113-120.
15. Soeda, E., Arrand, J.R., Smolar, N., Walsh, J.E. and Griffin, B.E. (1980) Nature 283, 445-453.
16. Clements, J.B., Watson, R.J. & Wilkie, N.M. (1977) Cell 12, 275-285.
17. Kumar, A. & Lindberg, U. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 681-685.
18. Davison, A.J. & Wilkie, N.M. (1981) J. Gen. Virol. 55, 315-331.
19. Spandidos, D.A. and Paul, J. (1982) EMBO J. 1, 15-20.
20. Denhardt, D.T. (1966) Biochem. Biophys. Res. Commun. 23, 641-646.
21. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
22. Everett, R.D. (1983) Nucleic Acids Res. 11, 6647-6666.
23. Anderson, K.P., Costa, R.H., Holland, L.E. and Wagner, E.K. (1980) J. Virol. 34, 9-27.
24. Mackem, S. and Roizman, B. (1980) Proc. Natl. Acad. Sci. U.S.A. 76, 4117-4121.
25. Watson, R.J. and Vande Woude, G.F. (1982) Nucleic Acids Res. 10, 979-991.
26. Murchie, M.-J. and McGeoch, D.J. (1982) J. Gen. Virol. 62, 1-15.
27. Watson, R.J., Preston, C.M. & Clements, J.B. (1979) J. Virol. 31, 42-52.
28. Marsden, H.S., Lang, J., Davison, A.J., Hope, R.G. & MacDonald, D.M. (1982) J. Gen. Virol. 62, 17-27.
29. Clements, J.B., McLauchlan, J. & McGeoch, D.J. (1979) Nucleic Acids Res. 7, 77-91.
30. Lee, G.T.Y., Para, M.F. & Spear, P.G. (1982) J. Virol. 43, 41-49.
31. Marsden, H.S., Stow, N.D., Preston, V.G., Timbury, M.C. & Wilkie, N.M. (1978) J. Virol. 28, 624-642.
32. Preston, V.G., Davison, A.J., Marsden, H.S., Timbury, M.C., Subak-Sharpe, J.H. and Wilkie, N.M. (1978) J. Virol. 28, 499-517.
33. Morse, L.S., Pereira, L., Roizman, B. and Schaffer, P.A. (1978) J. Virol. 26, 389-410.
34. Ruyechan, W.T., Morse, L.S., Knipe, D.M. and Roizman, B. (1979) J. Virol. 29, 677-697.
35. Hope, R.G., Palfreyman, J., Suh, M. and Marsden, H.S. (1982) J. Gen. Virol. 58, 399-415.

36. Para, M.F., Goldstein, L. and Spear, P.G. (1982) J. Virol. 41, 137-144.
37. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science 209, 1406-1414.
38. Fitzgerald, M. & Shenk, T. (1981) Cell 24, 251-260.
39. Hagenbuchle, O., Bovey, R. and Young, R.A. (1980) Cell 21, 179-187.
40. Jung, A., Sippel, A.E., Grez, M. and Schutz, G. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5759-5763.
41. McLauchlan, J. and Clements, J.B. (1983) EMBO J. 2, 1953-1961.
42. McDevitt, M.A., Imperiale, M.J., Ali, H. and Nevins, J.R. (1984) Cell 37, 993-999.
43. Rixon, F.J., Campbell, M.E. and Clements, J.B. (1984) J. Virol. 52, 715-718.