Multiple Spliced and Unspliced Transcripts from Human Cytomegalovirus Immediate-Early Region 2 and Evidence for a Common Initiation Site within Immediate-Early Region 1

RICHARD M. STENBERG,[†] PAMELA R. WITTE, AND MARK F. STINSKI*

Department of Microbiology, School of Medicine, University of Iowa, Iowa City, Iowa 52242

Received 6 May 1985/Accepted 1 August 1985

Human cytomegalovirus immediate-early (IE) region 2 (0.732 to 0.740 map unit) begins 35 nucleotides downstream of IE region 1 (Stenberg et al., J. Virol. 49:190-199, 1984). A series of mRNAs that have different splicing patterns are transcribed from region 2. There is an unspliced 1,589-nucleotide exon present in minor amounts and two spliced exons (836 and 289 nucleotides) present in larger amounts. The IE region 2 exons were found to be spliced onto the first three exons of region 1. Therefore, under IE conditions the region 1 promoter-regulatory region can promote transcription of region 2. Promoter sequences (i.e., CAAT and TATA boxes) are found upstream of the 5' end of IE region 2 but presumably function poorly at IE times after infection. The transcriptional regulation of these IE genes and the possible functional roles of the proteins are discussed. We postulate that a series of unique but related proteins are made from the region 2 transcripts. Some of these proteins should contain the same 169 amino-terminal residues as the major IE 72-kilodalton protein encoded by IE region 1 (Stenberg et al., J. Virol. 49:190-199, 1984). Variations in the amino acid sequences of the region 2 proteins could occur at either the amino terminus, the carboxy terminus, or both termini.

Human cytomegalovirus (CMV), a member of the herpesvirus family, is a large double-stranded DNA virus with a genome 240 kilobases (kb) long (6, 8, 11, 20). Like the cycles of other herpesviruses, the replicative cycle of CMV can be broken down into the following three phases: immediate early (IE), early, and late (4, 7, 14, 26, 27). Mutants of herpes simplex virus in the IE genes have been shown to be unable to establish latency (9). Thus, the expression of herpesvirus IE genes and the manner in which viral gene products regulate subsequent gene expression play an important role in determining the outcome of infection.

Infection of permissive cells with CMV in the presence of inhibitors of protein synthesis leads to accumulation of viral IE RNA (4, 10, 14, 26, 27). This RNA is transcribed from restricted regions of the viral genome. Previously, we designated three transcription units within *Xba*I fragment E as IE regions 1, 2, and 3 (21). IE region 1 codes for an abundant 1.95-kilobase (kb) mRNA (18, 20). This RNA was translated in vitro, and an abundant amount of a protein with an apparent molecular weight of 72,000 was synthesized (21, 28). Previously we described the splicing pattern of the 1.95-kb mRNA, the sequence of the DNA, and the predicted amino acid sequence of the protein (18). The actual molecular weight of the major IE protein according to its amino acid sequence is 64,000.

IE region 2 lies adjacent to region 1 and codes for a series of at least four mRNAs ranging in size from 1.10 to 2.25 kb. The gene products of these viral mRNAs were determined by in vitro translation of hybrid-selected RNA. The predominant protein encoded by this region has a molecular weight of 56,000, as determined by denaturing polyacrylamide gel electrophoresis. Several other proteins ranging in molecular weight from 42,000 to 16,500 were also detected (21). The

presence of a diverse population of mRNAs and proteins originating from IE region 2 suggested that various RNA splicing events might occur in this relatively small region of the viral genome.

In this paper we describe the structure of the RNAs originating from IE region 2. Under IE conditions, the viral mRNAs are generated through a complex series of splicing events in which the exons in IE region 2 can be linked to IE region 1 exons. A series of IE region 2 proteins that differ either at their carboxy or amino termini could arise. The relative abundance of the IE region 2 mRNAs changes during the course of infection. Spliced mRNAs predominate at IE times after infection. In contrast, unspliced mRNAs appear at early times and are predominant at late times. The role that these viral proteins might play in regulation of CMV gene expression is discussed. The role of the 72-kilodalton (kDa) IE protein from IE region 1 in autoregulation and the relationship of this phenomenon to transcription in IE region 2 are discussed in the accompanying paper (17).

MATERIALS AND METHODS

Cell culture and virus propagation. The growth of human foreskin fibroblasts and the propagation of human CMV strain Towne have been described previously (19).

Preparation of RNA. IE polysome-associated RNA was prepared in the presence of 100 μ M anisomycin as described by Stinski et al. (21). Whole-cell RNA was prepared by the guanidinium hydrochloride lysis method essentially as described by Strohmman et al. (22).

Recombinant plasmid DNAs. The CMV recombinant plasmid DNAs used in this study and the methods used for their purification have been described previously (21, 25).

Restriction enzymology and radioactive DNA probes. Uniquely end-labeled DNA fragments were generated as previously described (18), except that calf intestine alkaline phosphatase (Boehringer Mannheim Biochemicals, Indian-

* Corresponding author.

[†] Present address: Department of Microbiology, West Virginia University School of Medicine, Morgantown, WV 26506.

FIG. 1. Northern blot analysis of human CMV IE region 2 mRNAs. Polysome-associated IE RNA was subjected to denaturing agarose gel electrophoresis and transferred to nitrocellulose as described in Materials and Methods. The map (C) is a detailed restriction enzyme map of IE region 2. Abbreviations: B, *Bam*HI; C, *Cla*I; N, *Nco*I; S, *Sst*I; Sa, *Sal*I; X, *Xho*I. The *Taq*I profile is shown above the linear map. (A) RNA hybridized to 32 P-labeled nick-translated probe DNA extending from the *Sal*I site to the *Sst*I site. (B) RNA hybridizing to 32 P-labeled nick-translated *Pst*I fragment D of *Xba*I fragment E (21). This probe hybridized only to the 121-nucleotide leader of the 1.95-kb mRNA of IE region 1 (18, 21). The sizes of the RNAs are expressed in kilobases. The fractional length of the genome is indicated in map units.

apolis, Ind.) was substituted for bacterial alkaline phosphatase.

RNA analysis. RNA was analyzed by the nuclease mapping technique of Berk and Sharp (1), as previously described (18). IE region 2 mRNA was purified by hybridizing 65 μ g of IE polysome-associated RNA to the *Sal*I-*Sst*I fragment of region 2 DNA bound to diazobenzoyloxymethyl-paper as previously described (21). Denaturing formaldehyde-1.5% agarose gels were prepared and subjected to electrophoresis as described by Lehrach et al. (12). After electrophoresis, the RNA was transferred to nitrocellulose by the method of Thomas (23).

DNA sequencing. The DNA sequence analysis was performed by the chemical degradation method of Maxam and Gilbert (13), as previously described (18).

Protein structure determination. Two-dimensional protein structures were determined by the methods of Chou and Fasman (2) and Cohen et al. (3).

RESULTS

Physical map of IE region 2. A detailed restriction map of the *Xba*I-E region of CMV strain Towne and the relative location of IE region 2 have been previously described (21). Figure 1 shows a more detailed restriction map of IE region 2. This map was generated by the partial digestion procedure of Smith and Birnstiel (16) in conjunction with double

restriction digestion of recombinant plasmid DNAs. The restriction sites were verified by the DNA sequence.

Northern analysis of IE region 2 RNAs. We previously examined the mRNAs from IE region 2 by Northern blot analysis and found five size classes of RNA originating from a DNA region approximately 2 kb long (21). However, the results of our previous studies suggested that the 1.95-kb mRNA species detected with the DNA described above might have been due to contamination of the probe with region 1 DNA (21). Therefore, IE region 2 DNA from the *Sal*I site to the *Sst*I site (Fig. 1) was purified by multiple electrophoresis steps, 32 P nick translated, and hybridized to a Northern blot of IE RNA. Figure 1A shows that four RNA size classes were detected (2.25, 1.70, 1.40, and 1.10 kb). However, there was no RNA species present at the position expected for the 1.95-kb size class, indicating that our previous probes had some DNA contamination from IE region 1 (21). The position of the IE region 1 1.95-kb mRNA was determined by hybridizing a recombinant DNA from IE region 1 (pXEP22 [21]) that contained only the 121-nucleotide leader sequence of the major IE mRNA (Fig. 1B). Only the 1.95-kb RNA was detected after a short exposure of the autoradiogram. However, after longer exposures we detected additional mRNA species (data not shown). These observations are discussed below.

Nuclease mapping of IE region 2 RNAs. Our preliminary data obtained by using uniformly labeled DNAs as probes in

FIG. 2. Nuclease mapping of CMV IE RNA from region 2. Polysome-associated IE RNA was hybridized to ^{32}P -end-labeled DNA probes, digested with either mung bean nuclease (MB) or exonuclease VII (Exo), and subjected to alkaline gel electrophoresis. The gels were dried and autoradiographed. (A) 5'-Labeled *XhoI*-to-*BglII* probe. (B) 3'-Labeled *NcoI*-to-*SalI* probe. (C) 5'-Labeled *TaqI*-to-*XhoI* probe. (D) 3'-Labeled *XhoI*-to-*TaqI* probe. The stars indicate the labeled ends. Abbreviations: B, *BamHI*; S, *SalI*; T, *TaqI*; N, *NcoI*; X, *XhoI*; Bg, *BglII*. Minus and plus signs indicate the absence and presence of virus-specific RNA in the hybridization mixture, respectively. The sizes of protected probes are indicated (in nucleotides).

nuclease mapping studies indicated that the RNAs from IE region 2 crossed the *XhoI* site. The *XhoI* site of the *XhoI*-to-*SalI* fragment (Fig. 2) was uniquely 5' end labeled and hybridized to IE polysome-associated RNA. Under these conditions, the RNA failed to protect this fragment from mung bean nuclease digestion (data not shown). To test for transcription in the opposite direction, we 5' end labeled the *XhoI*-to-*BglII* fragment at the *XhoI* site and used this probe in a mung bean nuclease analysis of IE RNA. A single 630-nucleotide protected band was detected after alkaline agarose gel electrophoresis (Fig. 2A). Therefore, a 5' end was located 630 nucleotides upstream from the *XhoI* site, and transcription of IE region 2 was from right to left on the prototype arrangement of the CMV strain Towne genome. To probe for the 3' ends of region 2 RNAs, the 3' end of the *NcoI* site was labeled, and the fragment from *NcoI* to *SalI* was used for hybridization studies with IE RNA. The hybrids were treated with mung bean nuclease or exonuclease VII and subjected to alkaline agarose gel electrophoresis. Mung bean nuclease digestion generated two bands, a major 220-nucleotide band and a minor 950-nucleotide band (Fig. 2B). Digestion with exonuclease VII caused a shift in the 220-nucleotide band to a major 950-nucleotide band and a minor 850-nucleotide band (Fig. 2B). However, the 220-nucleotide band also was present after exonuclease VII digestion. These data suggested the presence of spliced and unspliced RNAs. The major 3' end of these RNAs was at a position 950 nucleotides downstream from the *XhoI* site. The minor 850 nucleotide band present after exonuclease VII

digestion indicated the presence of a second 3' end. However, preliminary data indicated that this 3' end is a possible donor splice site which may link to exons located downstream (data not shown). The band at 220 nucleotides generated by exonuclease VII digestion may represent artifactual nicking of the intron loop, followed by digestion to the splice point or another 3' end.

To analyze the splice points of the two exons, the fragment from the *TaqI* site to the *XhoI* site was labeled either 5' at the *TaqI* site (Fig. 2C) or 3' at the *XhoI* site (Fig. 2D) and hybridized to IE RNA. Mung bean nuclease digestion indicated that the donor site of the first exon was approximately 220 nucleotides downstream from the *XhoI* site (Fig. 2D). This was in agreement with the data in Fig. 2B since the *NcoI* and *XhoI* sites were very close together. When the 5' end of the *TaqI* site was used, a 150-nucleotide band was generated (Fig. 2C). Since the *TaqI* site was located approximately 150 nucleotides upstream from the major 3' end, these data indicated that there was a 3' exon approximately 300 nucleotides long. The band at 830 nucleotides represented the full length of the *TaqI*-*XhoI* probe, either protected by unspliced RNA or as a reassociated probe. A sequence analysis of cDNAs confirmed the presence of both spliced and unspliced RNAs (data not shown). The structures of the exons originating from IE region 2 are summarized in Fig. 2.

Sequence analysis of IE region 2. To better understand the expression of IE region 2, the entire region was sequenced by the method of Maxam and Gilbert (13). In addition, the

-60 -40 -20 -1
 ATCCTACTCA ATAAACGTGT CACGCCTGTG AAACCGTACT AAGTCTCCCG TGTCTTCTTA TCACCATCAG
 +1
 GTGACATCCTCGCCCAGGCTGTCAATCATGCCGGTATCGATTCCAGTAGCACCGGCCCCACGCTGACAACCCAC
 aspilleleualaglinalavalasnhisalaglyileasperserserthrglyprothrleuthrthrhis
 75
 TCTTGCAGCGTTAGCAGCGCCCTCTTAACAAGCCGACCCACCAGCGTCGCGGTTACTAACACTCCTCTCCCC
 sercysservalargargalaproleuasnlysprothrprothrservalalavalthrasnthrproleupro
 150
 GGGGCATCCGCTACTCCCGAGCTCAGCCCGCGTAAGAAACCGCGCAAAACCACGCGTCCTTTCAAGGTGATTATT
 glyalaseralathrprogluleuserproarglylsproarglysthrthrargprophelysvalileile
 225 252
 AAACCGCCCGTGCCTCCCGCGCCTATCATGTGCCCCCTCATCAACAGGAAGACATCAAGCCCCGAGCCCGACTTT
 lysproprovalproproalaproleuemetleuproleuilelysglngluaspileysprogluproasp phe
 300
 ACCATCCAGTACCGCAACAAGATTATCGATACCGCCGGCTGTATCGTGATCTCTGATAGCGAGGAAGAACAGGGT
 thrileglntyrargasnlysileileaspthralaglycysilevalileseraspsergluglugluglngly
 375
 GAAGAAGTCGAAACCCGCGGTGCTACCGCGTCTTCCCTTCCACCGGCGAGCGGCACGCCGCGAGTGACCTCTCCC
 glugluvalgluthrargglyalathralasererproserthrglyserglythrproargvalthrserpro
 450
 ACGCACCCGCTCTCCAGATGAACCAACCCTCTCTTCCCGATCCCTTGGGCGGCGCCGATGAAGATAGTTCCTCT
 thrhisproleuserglmetasnhisproproleuproaspproleuglyargproaspgluaspserserser
 525
 TCGTCTTCTCCTGCAGTTCGGCTTCGGACTCGGAGAGTGAGTCCGAGGAGATGAAATGCAGCAGTGGCGGAGGA
 sersersersercysserseralaseraspsergluserglusergluglumetlyscysserserglyglygly
 600
 GCATCCGTGACCTCGAGCCACCATGGGCGCGCGGTTTTGGTGGCGCGGCCTCCTCCTCTCTGCTGAGCTGCGGC
 alaservalthrsererhishisglyargglyglyphglyglyalaalaserserserleuleusercysgly
 625
 CATCAGAGCAGCGGCGGGCGAGCACCGGACCCGCAAGAAGAAGAGCAAACGCATCTCCGAGTTGGACAACGAG
 hisglnserserglyglyalaserthrglyproarglylslysserlysargileserleuleuaspasnglu
 750
 AAGGTGCGCAATATCATGAAAGATAAGAACACCCCTTCTGCACACCCAACGTGCAGACTCGGCGGGGTGCGGTC
 lysvalargasnilemetlysplysasnthrprophecysthrproasnvalglnthrargargglyargval
 825 836
 AAGATTGACGAGGTGAGCCGATGTTCCGCAACACCAATCGCTCTCTTGAGTACAAGAACCTGCCCTTCACGATT
 lysileaspgluvalserargmetpheargasnthrasnargserleuglutyrlyspaspleuprophet hile
 900
 CCCAGTATGCACAGGTGTTAGATGAGGCCATCAAAGCCTGCAAAACCATGCAGGTGAACAACAAGGGCATCCAG
 prosermethisglvalleuaspglualeilelysalacyslysthrmetglvalasnasnlysglyilegl
 925
 ATTATCTACACCCGCAATCATGAGGTGAAGAGTGAGGTGGATGCGGTGCGGTGTCGCTGGGCACCATGTGCACC
 ileiletyrthrargasnhisgluvallysserglualaspalavalargcysargleuglythrmetycysthr
 1050 1092
 TGGCCCTCTCCACTCCCTTCTCATGGAGCACACCATGCCCGTGACACATCCACCCAAAGTGGCGCAGCGCACAC
 trpproserproleuprosersertrpserthrprocyspro
 1125
 CCGATGCTTGTAACGAAGGCGTCAAGGCCGCGTGGAGCCTCAAAGAATTGCACACCCACCAATTATGCCCCCGTT
 1200
 CCTCCGATTACCGCAACATGATCATCCACGCTGCCACCCCGTGGACCTGTTGGGCGCTCTCAACCTGTGCCTGC
 1275 1301
 CCCTGATGCAAAAGTTTCCCAAACAG GTCATGGTGCATCTTCTCCACCAACCAGGGTGGGTTTCATGTGCCT
 valmetvalargilepheserthrasnglnglyglyphemetleupro
 1349
 ATCTACGAGACGGCCACGAAGGCCCTACGCCGTGGGGCAGTTTGAGCAGCCACCGAGACCCCTCCCGAAGACCTG
 iletyrgluthralathrlysalatyravalglylnphegluglnprothrgluthrproprogluaspleu
 1424 1481
 GACACCCTGAGCCTGGCCATCGAGGCAGCCATCCAGGACCTGAGGAACAAGTCTCAGTAAGTAAAACTGGAAA
 aspthrleuserleualaileglualaalaileglnaspleuargasnlysser gln
 1499 1567
 GAGAGACATGGACTCTGTACATAGTGATTCCCCGTGACAGTATTAACGTGTGGTGAAGAATGCTGTTTAAATAAAA
 1574
 GTAGCTTTTTTTATAC

FIG. 4. Summary of the structure of IE region 2 exons. RNAs from IE region 2 are transcribed from right to left on the viral DNA. The designated sequences are read from right to left. The arrow labeled 1 indicates the 5' end of the RNA. The thick and thin lines represent mRNA exons and their introns, respectively. The medium-intensity lines represent open reading frames with their ATG initiation codons and either TGA or TAA termination codons. AATAAA indicates the location of the polyadenylation signal. Abbreviations: S, *Sst*I; B, *Bam*HI; N, *Nco*I; X, *Xho*I; C, *Cl*aI; Bam, *Bam*HI. The fractional length of the genome is indicated in map units.

locations of the 5' and 3' ends, as well as the intron-exon splice junctions, were determined by mung bean nuclease analysis and electrophoresis of the protected DNA adjacent to the DNA sequence. IE region 2 has typical CAAT and TATA sequences upstream from the 5' end at positions -62 to -59 and -34 to -29, respectively (Fig. 3). The gene extended 1,589 nucleotides to the 3' end. At a position 18 to 23 nucleotides upstream from the 3' end was a consensus polyadenylation addition sequence, AATAAA (15). The ATG sequence at nucleotide 252 initiated a single open reading frame which extended 280 amino acids until it reached a TGA stop codon at nucleotide 1092. Thus, a 30-kDa protein was generated. A smaller but related protein could be produced by splicing nucleotide 836 onto nucleotide 1301, with the open reading frame extending to a TAA stop codon at nucleotide 1481. This RNA had an open reading frame which extended for 255 amino acids and coded for a 27-kDa protein. A minor spliced RNA which seemed to terminate 110 nucleotides upstream from the designated 3' end (Fig. 2B) actually contained a donor splice site that eliminated the TAA stop codon at nucleotide 1481 and consequently allowed translation to continue. Two RNAs were believed to utilize this splice junction, join onto additional exons, and terminate at a 3' end approximately 2 kb further downstream. These RNAs were present in the infected cells at very low levels under IE conditions and require further investigation. The structures of the RNA exons originating from IE region 2 and their open reading frames are summarized in Fig. 4. The RNA that splices to a third exon further downstream was present in minor amounts.

Analysis of region 1-region 2 splicing. The structure and relative abundance of the RNAs determined from the mung bean nuclease analysis did not agree with the structure and relative abundance determined by Northern blot analysis (Fig. 1). The mung bean nuclease analysis indicated that the RNAs were smaller than 1.0 kb (Fig. 2), and, therefore, it

could not account for the abundant 1.70- and 2.25-kb RNAs. Since region 2 contains only 1,589 nucleotides, a 2.25-kb RNA could not be generated entirely from this region. Previous hybridization studies indicated that these abundant RNAs could not be detected downstream from region 2, suggesting that they did not splice to exons further downstream (21; unpublished data). However, it was possible that these RNAs originated from IE region 1. This hypothesis was based on several pieces of evidence, as follows. (i) Long-term exposures of IE RNA Northern blots hybridized to region 1 probes often reveal additional RNAs that are usually masked by the intense exposure of the 1.95-kb major IE RNA (unpublished data). (ii) The 5' noncoding region of IE region 2 has an open reading frame that is continuous with the exons of region 1 (Fig. 3); this is atypical of 5' noncoding regions in eucaryotic mRNAs. (iii) The 5' end of IE region 2 contains a consensus acceptor splice sequence with an AG dinucleotide beginning upstream at position -2 and the typical pyrimidine-rich stretch 5' to the splice site (Fig. 3).

To investigate the possibility that the more abundant mRNAs from IE region 2 might actually be linked to IE region 1 by differential RNA splicing, it was necessary to purify IE region 2 mRNAs of the abundant 1.95-kb major IE mRNA from region 1. The fragment of DNA extending from the *Sall* site to the *Sst*I site in region 2 was purified and bound to diazobenzylxymethyl-paper. IE region 2-specific polysome-associated RNA was hybrid selected and used in both mung bean nuclease and Northern blot analyses. For the mung bean nuclease analysis, a fragment extending from the *Xho*I site in region 2 to a *Hind*III site upstream from IE region 1 was labeled at the 5' end of the *Xho*I site. This probe was then hybridized to region 2 selected mRNA. The results demonstrated that the mung bean nuclease analysis generated a predominant band (band 4) which indicated a 5' end at the designated location in region 2 (Fig. 5A). Digestion with exonuclease VII caused the appearance of three higher-molecular-weight bands, which indicated that there were 5'

FIG. 3. DNA and amino acid sequences of IE region 2. The entire sequence of IE region 2 was determined by using the method of Maxam and Gilbert (13) as previously described (18). The nucleotide at position +1 represents the 5' end of IE region 2. The upstream region is numbered in the opposite direction. The positions of the CAAT and TATA sequences are indicated (boxes). Sequences of interest are underlined, numbered, and explained in the text.

FIG. 5. Analysis of IE region 2 selected RNA. IE polysome-associated RNA (65 μ g) was hybridized to the *Sall*-to-*SstI* DNA fragment bound to diazobenzyloxymethyl-paper as previously described (21). The IE region 2 RNA was eluted and analyzed. (A) Selected RNA was hybridized to a DNA probe extending from a 5'-labeled *XhoI* site in region 2 to a *HindIII* site upstream of IE region 1. For the methods used see the legend to Fig. 2. The bands are labeled with numbers; the sizes of these bands correspond to the locations of the 5' ends of the exons shown in map A. Also shown is the location of the probe relative to the exons. Minus and plus signs indicate the absence and presence, respectively, of virus-specific RNA in the hybridization mixture. MB, Mung bean nuclease; Exo, exonuclease VII; M, markers. (B) Selected RNA was subjected to denaturing agarose gel electrophoresis and analyzed as described in the legend to Fig. 1. The blot was hybridized to a 32 P-labeled nick-translated *TaqI* fragment which was located as shown on restriction enzyme map B. The sizes of the RNAs are indicated (in kilobases). (C) IE polysome-associated (IE) and region 2 selected (S) RNAs were hybridized to a 3'-labeled *BstNI*-to-*SstI* probe and digested with mung bean nuclease. The minus sign indicates that no virus-specific RNA was included. The *BstNI* site was located 10 nucleotides downstream from the 5' end of IE region 1. Map C shows the location of the probe and the approximate location of the TATA box. The thick line above the TATA box represents the location of exon 1 of IE region 1. The size of the protected band is indicated (in nucleotides). The fractional length of the genome is indicated in map units.

ends at each of the three small exons in region 1 (Fig. 5A). However, specific hybrids that extended into the major exon (exon 4) of IE region 1 were not detected. The location of the three small exons in IE region 1 has been described previously (18). Ideally, if all RNAs from region 2 originated from the region 1 transcription unit, then we should have seen a single mung bean nuclease band at position 4 shift to a single band at position 1 after exonuclease VII digestion (Fig. 5A). This was not reflected in the relative abundances of the bands. However, this type of analysis is dependent on the stability of the hybrid during exonuclease VII digestion. Previously, we demonstrated that the three small exons in region 1 were unstable during exonuclease VII digestion, with intermediates consistently being generated (18). The RNAs from region 2 contained an additional large intron approximately 1,500 nucleotides long that was likely to make

these hybrids even less stable. Therefore, this pattern is not so surprising.

To further demonstrate that small exons from IE region 1 were actually spliced to exons in IE region 2, RNA selected by region 2 DNA was subjected to electrophoresis on a denaturing agarose gel. The RNA was transferred to nitrocellulose and hybridized to a 32 P-labeled *TaqI* fragment which hybridized only to the three small exons of region 1. Figure 5B shows that all of the region 2 mRNAs hybridized to the DNA coding for the upstream small exons of IE region 1. In contrast, the 1.95-kb IE region 1 RNA was not detected. We also used a 3' *BstNI* probe (Fig. 5C) which hybridized only to 121-nucleotide exon 1 of the 1.95-kb major IE mRNA to test whether this region hybridized to IE region 2-selected RNA. We used both total IE RNA and IE region 2-selected RNA in the mung bean analysis. Figure 5C

A

B

FIG. 6. (A) Splicing of the sequences between IE region 1 and 2 exons and the corresponding amino acid sequence. The nucleotide at position +1 of IE region 1 is shown, as are the locations of the splice points between the various exons, as previously described (18). The splicing of IE region 1 exon 3 to the 5' end of IE region 2 is shown. (B) Structures of the RNAs and their locations on the genomic map. Abbreviations: S, *Sst*I; B, *Bam*HI; Bg, *Bgl*II. The thick lines represent the exons of the 1.95-kb IE region 1 mRNA. The medium-intensity lines represent sequences associated with IE region 2 exons. The thin lines represent introns. The locations of the ATG start codons and TGA or TAA stop codons are indicated. The sizes of the RNAs are indicated (in kilobases), and the sizes of the encoded proteins (as determined by the sequence) are also shown (in kilodaltons [kd]). The fractional length of the genome is indicated in map units.

shows that the selected RNA from IE region 2 protected the probe, generating bands identical to those generated by total IE RNA. The top band was the expected 111-nucleotide band. The location of a 110-nucleotide band was identified, and this band probably represented breeding at the 3' end of the hybrid as well as the bottom band. This pattern of multiple bands was consistent with our previous findings (18). Therefore, region 2 mRNAs were linked to region 1 exons.

Proposed structure of IE region 2 mRNAs. The sequence analysis of the first three exons of IE region 1 showed that these exons could splice onto region 2 such that the open reading frame from region 1 was continuous with region 2

(Fig. 6A). Beginning at the ATG sequence in exon 2 of region 1, an open reading frame extended through the region 1 small exons into region 2; depending on how the exons were spliced, different but related proteins could be generated. Figure 6B shows the proposed structures of the RNAs originating from region 2, as well as region 1. As Fig. 6B shows, we generated a series of mRNAs that could code for proteins that varied at the amino terminus, the carboxy terminus, or both termini. In addition, at least two region 2 proteins could be expected to have the same amino-terminal portion as the region 1 major IE protein. It is likely that this region codes for some important functional domain since it is highly conserved. The 1.4-kb RNA is difficult to understand.

FIG. 7. Time course of IE region 2 RNAs. Whole-cell RNA was isolated from uninfected cells (HF) or infected cells at 2, 5, 10, 24, 48, and 72 h after infection (indicated at the top) and hybridized to a 3' *Cla*I probe (shown at the bottom). The hybrids were digested with mung bean nuclease and subjected to alkaline agarose gel electrophoresis. The positions of the spliced (S) and unspliced (U) region 2 exons are indicated. M, Molecular weight size markers.

The hybridization data indicated that this RNA was linked to a IE region 1 exon. However, the size of the RNA molecule indicates that it cannot have all three region 1 exons. Therefore, the structure of this RNA (Fig. 6B) is hypothetical. The 1.7-kb unspliced RNA in region 2 is discussed below in more detail.

Appearance of region 2 mRNAs throughout the replicative cycle. To better understand how region 2 is expressed over time, we analyzed the steady-state accumulation of RNAs originating from IE region 2 at different intervals after infection. These analyses were performed by hybridizing whole-cell RNA from an equivalent number of infected cells to a 3'-labeled *Cla*I (Fig. 7) probe in region 2. The hybrids were digested with mung bean nuclease and subjected to electrophoresis under alkaline conditions. The spliced RNAs, which normally accumulated to a higher level than the unspliced RNAs under IE conditions, were approximately equal in abundance to the unspliced RNAs at 5 h after infection in the presence of protein synthesis (Fig. 7). The levels of unspliced and spliced RNAs diminished somewhat by 10 and 24 h after infection. However, the unspliced RNA level increased dramatically by 48 and 72 h after infection (Fig. 7). Therefore, region 2 was expressed at late times after infection and primarily as unspliced RNA.

DISCUSSION

In this paper we describe an analysis of the RNA originating from human CMV IE region 2. Our data demonstrate that IE region 2 codes for four species of RNA (2.25, 1.70, 1.40, and 1.10 kb). Detailed structures for the three larger RNAs are proposed. The structure of the minor 1.10-kb species is

still under investigation. We also identified a second 1.70-kb RNA species which is most abundant at late times after infection. Transcription of these RNAs is from right to left on the CMV genome, which is consistent with the findings of Wilkinson et al. (28). These RNAs differ at the level of RNA splicing and thereby generate mRNAs that can be translated to different, but related proteins. Transcription of region 2 originates in IE region 1, indicating that the strong region 1 promoter and its upstream control sequences influence region 2 expression at IE times after infection. We previously demonstrated that the IE region 1 promoter is very strong at IE times after infection (21, 24). An event occurring at the level of transcription termination, polyadenylation, or RNA splicing may control the relative abundance of region 1 and region 2 mRNAs in the cytoplasm of infected cells. Alternatively, the abundance of these RNAs may be influenced by RNA turnover.

The virus could code for a series of proteins, all of which contain the same open reading frame but which differ at the amino termini, the carboxy termini, or both due to differential splicing. Computer-derived two-dimensional structures of the 48- and 51-kDa region 1-region 2 proteins as predicted by the method of Chou and Fasman (2) are shown in Fig. 8A and B. The molecular weights of the proteins, as determined from their amino acid compositions, are slightly lower than those of proteins synthesized by RNA isolated by hybridization selection and analyzed by denaturing polyacrylamide gel electrophoresis (21). These hypothetical proteins are essentially identical except that they differ at the carboxy-terminal end beginning with amino acid 365 (Fig. 8). There are no remarkable structural differences between these two proteins within this region, but something as subtle as the

FIG. 8. Two-dimensional structure of IE region 2 proteins. The structures of the 51-kDa (A) and 48-kDa (B) proteins were predicted by using the method of Chou and Fasman (2), as described previously (3). The locations of the amino (NH₂) and carboxy (COOH) termini are indicated. The open octagons indicate hydrophilic regions, and the concentric octagons indicate hydrophobic regions. The sizes of the octagons indicate the degree of hydrophilicity or hydrophobicity. The amino acids are numbered every 20 residues. The arrows indicate the location of the splice junction between region 1 and region 2 (arrow 1), the beginning of the variable carboxy termini (arrow 2) and the beginning of the region 2 amino acid sequence starting at the ATG (Met) at nucleotide 252 (arrow 3) (see Fig. 3).

variable carboxy terminus of the 48- and 51-kDa proteins might play a role in their function.

With the exception of the amino terminus from amino acids 1 to 169 (Fig. 8), the 48- and 51-kDa proteins contain a C-terminal sequence identical to that of the smaller 27- and 30-kDa proteins (Fig. 6B). The presence or absence of the amino-terminal 169 amino acids in the 48- and 51-kDa proteins could affect the ability or lack of ability of the proteins to positively or negatively regulate CMV gene expression.

As mentioned above, the 27- and 30-kDa proteins are identical to the 48- and 51-kDa proteins except that they are missing the amino-terminal 169 amino acids derived from IE region 1. These proteins may differ in function in one of two ways. First, they may be altered by simply lacking an active site or binding site which is present in the amino termini of the larger proteins. Second, they may function by freeing an internal active site which is masked by the amino terminus in the larger proteins.

It was interesting that in the presence of de novo viral protein synthesis, the pattern of region 2 RNAs was different than the pattern in the absence of protein synthesis. This observation indicates that one function of the IE genes may be to negatively regulate the level of IE mRNA in the cytoplasm of infected cells. In the accompanying paper we demonstrate that the IE region 1 72-kDa protein autoregulates transcription of IE region 1. Therefore, it is likely that the 72-kDa protein also negatively regulates the expression of the RNAs from region 2 that are linked to region 1 exons. This hypothesis is consistent with the decrease in relative abundance of region 2 spliced transcripts at 10 h after infection in the presence of viral protein synthesis. What is most interesting about the expression of region 2 transcripts is the tremendous increase in the level of unspliced 1.70-kb RNA at late times after infection. The phenomenon of IE region 1 autoregulation described in the accompanying paper may explain why the mRNA for the 30-kDa protein accumulates to the highest levels at late times after infection. We propose that at early and late times after infection, the region 2 promoter functions on its own to produce primarily the unspliced RNA from region 2, while at IE times region 2 expression originates primarily from the region 1 promoter. After the IE 72-kDa protein negatively regulates the region 1 promoter (17), the region 2 promoter is preferentially selected by the transcription complex.

ACKNOWLEDGMENTS

We thank R. Eisenberg and G. Cohen for assistance in determining the computer-derived secondary structures of the viral proteins. We also thank W. Goins, C. M. Stoltzfus, and C. Grose for critical reviews of the manuscript.

This work was supported by Public Health Service grant AI13562 from the National Institute of Allergy and Infectious Diseases, by grant MV-193 from the American Cancer Society, and by grant 1-697 from the National Foundation of the March of Dimes. M.F.S. is the recipient of Public Health Service Research Career Development Award AI100373 from the National Institute of Allergy and Infectious Diseases. R.M.S. is the recipient of Public Health Service Postdoctoral Fellowship AI06829 from the National Institute of Allergy and Infectious Diseases.

LITERATURE CITED

1. Berk, A. J., and P. A. Sharp. 1977. Sizing and mapping of early adenovirus-mRNAs by gel electrophoresis of S_1 endonuclease digested hybrids. *Cell* 12:721-732.
2. Chou, P. Y., and G. D. Fasman. 1974. Prediction of protein

conformation. *Biochemistry* 13:222-245.

3. Cohen, G. H., B. Dietzschold, M. Ponce de Leon, E. Golub, A. Varrichio, L. Pereira, and R. J. Eisenberg. 1984. Localization and synthesis of an antigenic determinant of herpes simplex virus glycoprotein D that stimulates the production of neutralizing antibody. *J. Virol.* 49:102-108.
4. DeMarchi, J. M. 1981. Human cytomegalovirus DNA: restriction enzyme cleavage maps and map locations for immediate early, early and late RNAs. *Virology* 114:23-38.
5. DeMarchi, J. M. 1985. Post transcriptional control of human cytomegalovirus gene expression. *Virology* 124:390-402.
6. DeMarchi, J. M., M. L. Blankenship, G. D. Brown, and A. S. Kaplan. 1978. Size and complexity of human cytomegalovirus DNA. *Virology* 89:643-646.
7. DeMarchi, J. M., C. A. Schmidt, and A. S. Kaplan. 1980. Pattern of transcription of human cytomegalovirus in permissively infected cells. *J. Virol.* 35:277-286.
8. Geelen, J. L. M. C., G. Walig, P. Wertheim, and J. van Der Noordaa. 1978. Human cytomegalovirus DNA. I. Molecular weight and infectivity. *J. Virol.* 26:813-816.
9. Gerdes, J. C., H. S. Marsden, M. L. Cook, and J. G. Stevens. 1979. Acute infection of differentiated neuroblastoma cells by latency-positive and latency-negative herpes simplex virus ts mutants. *Virology* 94:430-441.
10. Jahn, G., E. Knust, H. Schmolla, T. Sarre, J. A. Nelson, J. K. McDougell, and M. Fleckenstein. 1984. Predominant immediate early transcripts of human cytomegalovirus AD169. *J. Virol.* 49:363-370.
11. Kilpatrick, B. A., and E.-S. Huang. 1977. Human cytomegalovirus genome: partial denaturation map and organization of genome sequences. *J. Virol.* 24:261-276.
12. Lehrach, D., D. Diamond, J. M. Wozney, and H. Boedtkes. 1977. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical re-examination. *Biochemistry* 16:4743-4751.
13. Maxam, A. M., and W. Gilbert. 1980. Sequencing with base specific chemical cleavages. *Methods Enzymol.* 65:499-560.
14. McDonough, S. H., and D. H. Spector. 1983. Transcription in human fibroblasts permissively infected by human cytomegalovirus strain AD169. *Virology* 125:31-46.
15. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' Non-coding region sequences in eucaryotic messenger RNA. *Nature (London)* 263:211-214.
16. Smith, H. O., and M. L. Birnstiel. 1976. A simple method for DNA restriction mapping. *Nucleic Acids Res.* 3:2387-2399.
17. Stenberg, R. M., and M. F. Stinski. 1985. Autoregulation of the human cytomegalovirus major immediate-early gene. *J. Virol.* 56:676-682.
18. Stenberg, R. M., D. R. Thomsen, and M. F. Stinski. 1984. Structural analysis of the major immediate early gene of human cytomegalovirus. *J. Virol.* 49:190-199.
19. Stinski, M. F. 1978. Sequence of protein synthesis in cells infected by human cytomegalovirus: early and late virus-induced polypeptides. *J. Virol.* 26:686-701.
20. Stinski, M. F., E. S. Mocarski, and D. R. Thomsen. 1979. DNA of human cytomegalovirus: size heterogeneity and defectiveness resulting from undiluted passage. *J. Virol.* 31:231-239.
21. Stinski, M. F., D. R. Thomsen, R. M. Stenberg, and L. C. Goldstein. 1983. Organization and expression of the immediate early genes of human cytomegalovirus. *J. Virol.* 46:1-14.
22. Strohman, R., P. Moss, J. Micou-Eastwood, D. Spector, A. Probyla, and B. Peterson. 1977. Messenger RNA for myosin polypeptides: isolation from single myogenic cell cultures. *Cell* 10:265-273.
23. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* 77:5201-5205.
24. Thomsen, D. R., R. M. Stenberg, W. F. Goins, and M. F. Stinski. 1984. Promoter regulatory region of the major immediate early gene of human cytomegalovirus. *Proc. Natl. Acad. Sci. USA* 81:659-663.
25. Thomsen, D. R., and M. F. Stinski. 1981. Cloning of the human cytomegalovirus genome as endonuclease Xba I fragments.

- Gene 16:207-216.
26. **Wathen, M. W., and M. F. Stinski.** 1982. Temporal patterns of human cytomegalovirus transcription: mapping the viral RNAs synthesized at immediate early, early, and late times after infection. *J. Virol.* **41**:462-477.
 27. **Wathen, M. W., D. R. Thomsen, and M. F. Stinski.** 1981. Temporal regulation of human cytomegalovirus transcription at immediate early and early times after infection. *J. Virol.* **38**:446-459.
 28. **Wilkinson, G. W. G., A. Akrigg, and P. J. Greenaway.** 1984. Transcription of the immediate early genes of human cytomegalovirus strain AD169. *Virus Res.* **1**:101-116.