Current Topics in Microbiology and Immunology 154

Cytomegaloviruses

Current Topics in Microbiology 154 and Immunology

JAMES K. MCDOUGALL PhD.
Research Professor of Pathology
University of Washington,
Member, Fred Hutchinson Cancer Research Center,
1124, Columbia Street, Seattle, WA 98104 , USA

ISBN 3-540-51514-3 Springer-Verlag Berlin Heidelberg NewYork
ISBN 0-387-51514-3 Springer-Verlag NewYork Berlin Heidelberg

Library of Congress Catalog Card Number 15-12910
Printed in Germany

Typesetting: Thomson Press (India) Ltd., New Delhi
Offsetprinting: Saladruck, Berlin; Bookbinding: B. Helm, Berlin
2123/3020-543210 Printed on acid-free paper.

Cytomegaloviruses

Edited by J. K. McDougall

With 58 Figures

Springer-Verlag
Berlin Heidelberg NewYork
London Paris Tokyo HongKong

Preface

Named for the enlarged, inclusion-bearing cells characteristic of infection by these viruses, cytomegaloviruses present a significant challenge to both microbiologist and immunologist. Although most primary infections in humans are subclinical, cytomegalovirus can be associated with a wide spectrum of disease, particularly when infection occurs in the immunocompromised individual or as a result of congenital or perinatal infection. Although reinfection with cytomegalovirus has been demonstrated, most recurrent and persistent infections result from the reactivation of latent virus. Cytomegaloviruses, like other members of the Herpesviridae family, have the capacity to establish latency after a primary infection but the mechanisms for establishing the nonreplicating but reactivatable state have not been defined. The factors responsible for the spectrum of manifestations of cytomegalovirus infection are largely undetermined but host immunological function, route of infection, and size of inoculum all contribute to the extent and severity of disease.

Cytomegaloviruses have the largest genomes in the herpesvirus family, approximately 240 kilobase pairs, providing a potential coding capacity for more than 200 proteins of which less than one-fourth have been mapped and described. There are many similarities to other herpesviruses in genome structure and gene expression; for example, three temporal classes of genes can be identified as α (immediate-early), β (early), and γ (late) products. The first five chapters of this volume review and describe recent developments in understanding the transcription and regulation of these gene classes. Analysis of the molecular biology of cytomegaloviruses has progressed more slowly than that of herpes simplex virus or Epstein-Barr virus but has now received a significant boost from the progress made by Barrell's group in sequencing the complete genome of human cytomegalovirus strain AD 169. Their analysis of the open reading frames in the AD 169 sequence is in the sixth chapter. A knowledge of proteins involved in the sequential

steps of virus replication and assembly can provide the basis for antiviral strategies as well as for an understanding of the host's immune response, a subject dealt with in the last section of this book. Efforts to establish a basis upon which effective prevention or modulation of cytomegalovirus infection can be developed are making good progress.

We hope that the reviews and the new results presented herein will provide encouragement to researchers in this field and to those dealing with often insidious consequences of infection by cytomegalovirus.

James K. McDougall
Seattle, Spring 1989

List of Contributors

(Their addresses can be found at the beginning of the respective chapters)

ABENES, G. B.
BARRELL, B. G.
BANKIER, A. T.
BECK, S.
BOHNI, R.
BROWN, C. M.
CERNY, R.
CHEE, M. S.
CHERRINGTON, J. M.
DEL VAL, M.
FLECKENSTEIN, B.
GHAZAL, P.
GNANN, Jr., J. W.
GÖNCZÖL, E.
HORSNELL, T.
HUTCHISON, III, C. A.
ISOM, H. C.
JAHN, G.
KLUCHER, K. M.
KOSZINOWSKI, U. H.
KOVZARIDES, T.
MACH, M.
MANNING, W. C.
MARTIGNETTI, J. A.
MOCARSKI, E.
NELSON, J. A.
PLOTKIN, S.
PREDDIE, E.
RABERT, D. K.
RASMUSSEN, L.
REDDEHASE, M. J.
SAMBUCETTI, L. C.
SATCHWELL, S. C.
SPECTOR, D. H.
STAMMINGER, T.
TOMLINSON, P.
WESTON, K. M.
WRIGHT, D. A.
YIN, C. Y.

List of Contents

Gene Expression

Immediate-Early Transcription Regulation of Human Cytomegalovirus

T. STAMMINGER and B. FLECKENSTEIN

1 Genomic Localization of Immediate-Early Genes

Herpesvirus genomes are expressed in three temporally regulated phases during productive infection. The first period of transcription, commonly termed "immediate early" (IE), follows entry of the virus into the host cell. It is independent of de novo synthesis of viral proteins. In general, a second ("early") phase follows, when a number of regulatory proteins and viral enzymes are synthesized. During the third phase ("late"), which begins with onset of virion DNA replication, viral structural proteins are synthesized (HONESS and ROIZMAN 1974, 1975). Immediate-early proteins are thought to exert important regulatory functions in the switch from the IE to the early phase. If permissive cells are infected in the presence of cycloheximide or anisomycin, both potent inhibitors of protein synthesis, viral IE RNA is accumulated. This approach was applied to investigate IE transcription in various strains of human cytomegalovirus (HCMV) (WATHEN and STINSKI 1982; MCDONOUGH and SPECTOR 1983; WILKINSON et al. 1984; DEMARCHI 1981). The IE RNA of HCMV arises from a few distinct regions of the viral genome. The region of highest transcriptional activity is localized between map units 0.66 and 0.77 in the genomes of all HCMV strains that have been investigated (Fig. 1). This region is part of the large unique compartment of the viral genome, for instance corresponding to the *Hin*dIII-E fragment of HCMV strain AD169 and the *Xba*I-E and -N fragments of strain Towne. It comprises approximately 20 kb.

The four transcription units with most abundant IE expression have been termed IE-1 to IE-4 (JAHN et al. 1984; STINSKI et al. 1983) (Fig. 1). The major transcript from the IE-1 gene was found to be a polyadenylated mRNA of 1.9 kb (JAHN et al. 1984; STINSKI et al. 1983). The IE-2 gene is expressed into mRNA of

Institut für Klinische und Molekulare Virologie der Universität Erlangen-Nürnberg, Loschgestraße 7, D-8520 Erlangen, FRG

Current Topics in Microbiology and Immunology, Vol. 154

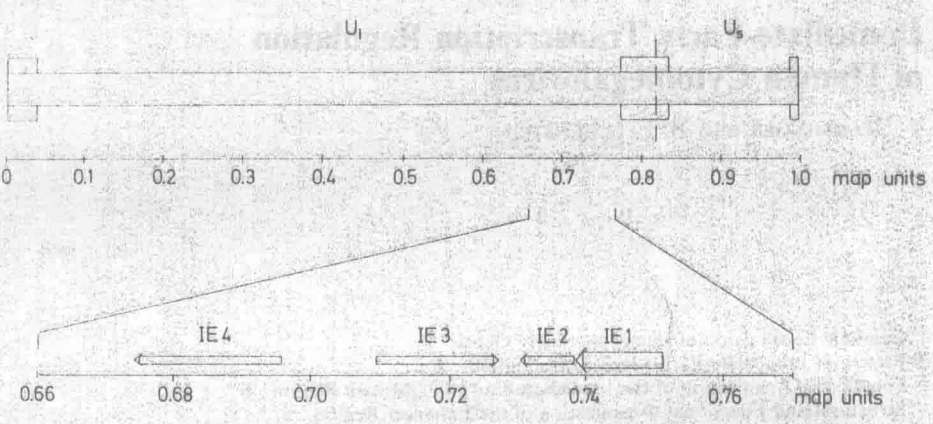

Fig. 1. Genomic localization of the four major IE transcription units within the HCMV genome. The four major IE transcription units (IE-1 to IE-4) are located between map units 0.66 and 0.77 of the HCMV genome. The entire genome has about 230 kb. U_l, long unique segment; U_s, short unique segment of the viral genome

approximately 2 kb (JAHN et al. 1984; STINSKI et al. 1983; WILKINSON et al. 1984). Some controversy remained about the IE-3 region (map units 0.709–0.728) which codes for a 2.2-kb transcript in strain AD169 and for a 1.95-kb message in the strain Towne of HCMV. While JAHN et al. (1984), WILKINSON et al. (1984), and STINSKI et al. (1983) classified this transcript as immediate early, STAPRANS and SPECTOR (1986) reported early transcription from this region of the viral genome. The IE-4 gene appeared highly unusual in structure and expression pattern. It is transcribed into a 5-kb RNA. While part of this RNA is polyadenylated, the majority of the transcripts were found in nonpolyadenylated fractions. In contrast to other IE genes of HCMV, the DNA coding for the 5-kb RNA is also expressed during late virus replication in high quantities. No compelling evidence could be found for a protein to be encoded (PLACHTER et al. 1988) within the entire nucleotide sequence of the IE-4 gene. This genomic region is devoid of translational reading frames greater than 200 nucleotides in the direction of transcription; part of it was shown to be capable of transforming primary embryonic rat fibroblasts and NIH3T3 cells in vitro (NELSON et al. 1982, 1984).

Low-abundance IE transcription was also reported from a number of other genomic regions (DEMARCHI 1981; MCDONOUGH and SPECTOR 1983; WATHEN and STINSKI 1982). At least four additional IE-mRNAs that are differentially spliced could be assigned to the unique short HCMV genome region (WESTON 1988).

2 Pattern of Immediate Early Gene Transcription

The IE-1 region, also referred to as the major IE gene, codes for the most abundant type of IE RNA with a size of about 1.9 kb under the experimental conditions of cycloheximide block (STINSKI et al. 1983; WILKINSON et al. 1984; JAHN et al. 1984). The predominant RNA is transcribed from right to left in the prototype

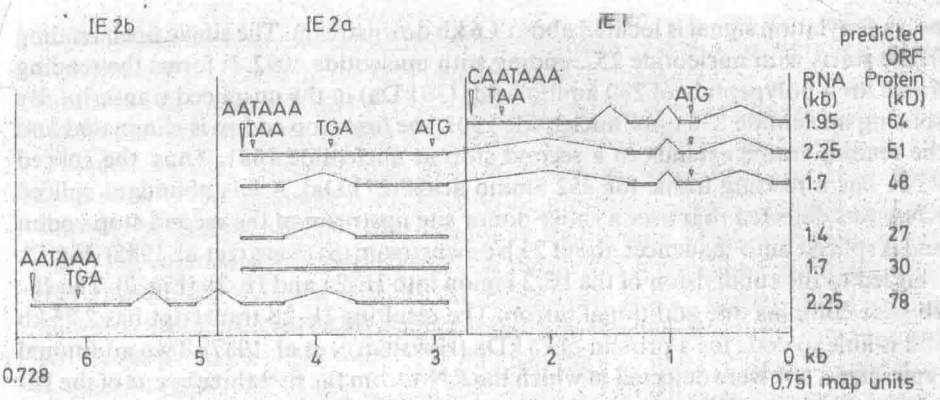

Fig. 2. Splicing pattern of IE-1 and IE-2-specific transcripts. Direction of transcription is *from right to left*. The exons of the six known mRNAs originating from this region are given in *double lines*. Start and stop codons for reading frames and polyadenylation sites are indicated by *triangles*. (Data taken from HERMISTON et al. 1987). After submission of this manuscript STENBERG et al. (1989) reported that HCMV strain Towne does not have a stop codon at nt. 1092 within the IE-2 sequence. Thus the unspliced 1.7 kb message derived from IE-2 is capable of coding for a protein of 40 kd whereas the spliced messages of 1.7 and 2.25 kb code for proteins of 55 and 86 kd, respectively

arrangement of the HCMV genome (0.739–0.751 map units) (Fig. 2). The gene has been determined in nucleotide sequence, and the transcripts were analyzed by nuclease mapping (STENBERG et al. 1984; AKRIGG et al. 1985). The major 1.9-kb IE transcript is made up of three small (121, 88, and 185 nt) and one large exon (1342 nt). The three introns (827, 114, and 70 nt) are located near the 5′ end of the gene. The single open reading frame starts in the second exon and has the coding capacity for a polypeptide of 491 amino acids; the molecular weight was calculated to be 63.8 kDa (AKRIGG et al. 1985). If the mRNA was selected by hybridization to immobilized DNA of the IE-1 region and translated in vitro, it could be shown that it encodes the predominant protein found in infected cells within the first hour after infection (STENBERG et al. 1984). The protein is quickly accumulated in the nucleus of infected cells (MICHELSON-FISKE et al. 1977). The molecular weight of this IE protein was estimated by polyacrylamide gel electrophoresis to be approximately 75 kDa (BLANTON and TEVETHIA 1981; CAMERON and PRESTON 1981; GIBSON 1981; WILKINSON et al. 1984). A proline-rich region near the N terminus of the predominant IE protein might explain anomalous mobility in SDS polyacrylamide gels due to disruption of the helical protein structure (AKRIGG et al. 1985). Since it could be shown that the major IE protein is phosphorylated (GIBSON 1981), it was suggested that posttranslational modification may contribute to the unexpected migration pattern (STINSKI et al. 1983).

The IE-2 region lies immediately downstream to region 1 (Fig. 2). Northern blot analyses detected at least four classes of mRNA in the size range from 1.1 to 2.25 kb (STENBERG et al. 1985; AKRIGG et al. 1985). This suggested that the various mRNAs are formed through differential splicing from a rather small part of the genome. The IE-2 gene is transcribed in the same direction as the major IE-1 gene. A functional promoter with the CAAT and TATA sequence elements was found located between the 3′ terminus of the IE-1 gene and the transcription initiation site of IE-2. A

polyadenylation signal is located about 1.6 kb downstream. The single open reading frame starts with nucleotide 252, ending with nucleotide 1092. It forms the reading frame for a polypeptide of 280 amino acids (30 kDa) in the unspliced transcript. By splicing nucleotide 836 onto nucleotide 1301, the first stop codon is eliminated and the reading frame extends to a second stop at nucleotide 1481. Thus, the spliced RNA has a reading frame for 252 amino acids (27 kDa). A less abundant spliced RNA was detected that uses a splice donor site upstream of the second stop codon and is spliced onto sequences about 2 kb downstream (STENBERG et al. 1985) (Fig. 2). This led to the subdivision of the IE-2 region into IE-2a and IE-2b (Fig. 2). The IE-2b gene contains one additional intron. The resulting IE-2b transcript has 2.25 kb and is able to code for a protein of 78 kDa (HERMISTON et al. 1987). Two additional types of mRNA were detected in which the RNA from the first three exons of the IE-1 gene are differentially spliced to region 2 transcripts, extending the number of IE-2 proteins (Fig. 2). Remarkably, the relative abundancy of the various IE-2 transcripts varies with the time course of HCMV infection; the spliced IE-2 transcripts were preferentially found during the initial IE phase, while, later on in the replication cycle, unspliced transcripts from this genomic region dominated (STENBERG et al. 1985). The predominant IE-2 protein had a molecular weight of 56 kDa, whereas several other proteins in the molecular weight range from 16.5 to 42 kDa could also be detected (STINSKI et al. 1983).

Immediate-early proteins are assumed to exert regulatory functions by acting in *trans* on *cis* elements such as promoters or enhancers. EVERETT and DUNLOP (1984) could show that cotransfection of the herpes simplex virus type 1 glycoprotein D or the rabbit β-globin gene with the entire major IE gene region (*Xba*I-E fragment of strain Towne, *Hin*dIII-E fragment of strain AD169) resulted in activation of the heterologous promoters by the HCMV IE gene products. HERMISTON et al. (1987) found that a plasmid expressing the IE-1 and IE-2 genes activated the adenovirus E-2 promoter. In this experiment, the IE-1 protein did not function independently as gene activator, while the IE-2 proteins did. Transactivation by IE-2 proteins was augmented in the presence of IE-1 polypeptide. It suggested that IE-2 gene products might exert maximal stimulatory effects only if combined with the IE-1 protein. Since transactivator and inhibitor effects in various cell lines depend on numerous factors (such as concentration of transactivating proteins, their relative abundancy, differentiation of cells), far more work will be required to define the regulatory effect of each single IE-1/2 protein at various stages of replication and persistence in different cell types.

3 Functional Organization of the Immediate Early-1-Upstream Region

The IE-1 gene is efficiently expressed immediately following entry of virions into permissive cells (MICHELSON-FISKE et al. 1977). This could be taken as a hint that strong *cis*-regulatory elements may govern the IE-1 promoter. TATA and CAAT

sequence elements are found within 65 bp upstream of the transcription initiation site. These sequences form the IE-1 promoter and are required for a low level of gene expression in intact cells and in cell-free transcription systems (STINSKI and ROEHR 1985; GHAZAL et al. 1987, 1988b). A strong enhancer was found upstream of the IE-1 promoter by the "enhancer trap" CI (WEBER et al. 1984; BOSHART et al. 1985). Linearized SV40 genomes with deletion of the two major functional enhancer domains were cotransfected with randomly fragmented DNA of the HCMV IE region. Spontaneous intracellular ligation resulted in replication-competent SV40-type recombinant viruses with HCMV DNA sequences replacing the genuine enhancer. The respective HCMV DNA fragments, representing the upstream region between nucleotides −528 and −118, fulfilled the essential criteria of a eukaryotic enhancer element (BOSHART et al. 1985). It strongly activates transcription in a wide variety of cells, including *Xenopus laevis* and *Drosophila melanogaster* cells (BOSHART et al. 1985; FÖCKING and HOFSTETTER 1986; SINCLAIR 1987). It is able to stimulate transcription from its cognate promoter and from heterologous promoters (STINSKI and ROEHR 1985; BOSHART et al. 1985). The transcription-enhancing effect could also be demonstrated in a cell-free transcription system (THOMSEN et al. 1984; GHAZAL et al. 1987). The nucleotide sequences between −120 and −65 were also able to stimulate transcription at least by a factor of 10 (GHAZAL et al. 1988b), suggesting that the transcription-enhancing domain extends beyond the enhancer sequence described by BOSHART et al. (1985). Thus, HCMV enhancer activity is localized at least between nucleotides −520 and −65. Enhancer sequences of comparable strength were also found upstream of the IE genes of the mouse cytomegalovirus (MCMV) (DORSCH-HÄSLER et al. 1985) and the simian cytomegalovirus (SCMV) (JEANG et al. 1987). No comparable enhancer could be found within all other herpesvirus genomes investigated so far. Thus, strong constitutive enhancers seem to be characteristic of cytomegalotype herpesviruses.

Although the functional relevance of the enhancer in human cytomegalovirus is obvious, the mechanisms by which these sequences mediate the enhancing effect remain to be elucidated. While most proteins binding the HCMV enhancer have not been identified so far, four high-affinity binding sites for nuclear factor 1 (NF-1) have been mapped between nucleotides −780 and −610 in the upstream region of the IE-1 enhancer/promoter (HENNIGHAUSEN and FLECKENSTEIN 1986; JEANG et al. 1987) (Fig. 3). One additional strong NF-1-binding position appeared in the first intron of the IE-1 gene (HENNIGHAUSEN and FLECKENSTEIN 1986). The NF-1-binding sites of the HCMV IE upstream sequences fall into regions that have been shown to be

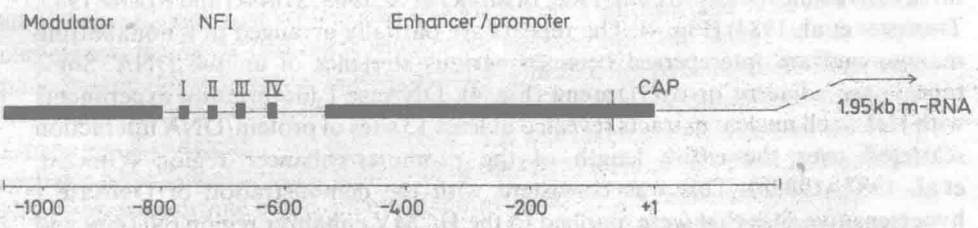

Fig. 3. The four functional domains in the upstream sequence of the major IE gene

sensitive to DNAase I in the active gene, but are not sensitive in the silent gene (NELSON and GROUDINE 1987). A cluster of at least 20 NF-1-binding sites has been observed upstream of the IE regulatory region of SCMV (JEANG et al. 1987). As the dominant IE gene of SCMV is expressed in a broad spectrum of cell types (LAFEMINA and HAYWARD 1988), it has been speculated that NF-1 may influence cell-type-specific expression of the linked IE-1 gene (JEANG et al. 1987; HENNIGHAUSEN and FLECKENSTEIN 1986).

Additional DNAase I hypersensitive sites were mapped in the region further upstream of the HCMV IE-1 promoter between nucleotides −1185 and −750, suggesting that this region also acts as target for *trans*-regulating factors (NELSON and GROUDINE 1987). To investigate the role of these elements, the corresponding fragments were placed in front of homologous and heterologous promoters linked to the CAT reporter gene. Transient expression assays showed that the element was able to modulate gene expression in nonpermissive cells in a negative manner, whereas it positively influenced expression in permissive cells (NELSON et al. 1987). This was confirmed by transcription in vitro; nuclear extracts of various cell lines were tested for the ability to mediate transcription in the presence or absence of the IE-1 upstream region between nucleotides −1145 and −524 (LUBON et al. 1989). This region, termed modulator region, mediates transcriptional repression in certain cell lines, whereas it augments transcription in other cells. It suggests a cell-specific regulatory mechanism of IE-1 gene expression.

In summary, the complex IE-1 upstream region of HCMV can be divided into at least four domains. The promoter (nucleotides −65 to +3) appears sufficient for a low level of transcription. The enhancer (−520 to −65) stimulates downstream gene expression from cognate and foreign promoters to a high level in a wide variety of cell types. The modulator segment (nucleotides −1145 to −750) which acts on gene expression in a differentiation-dependent fashion is located upstream of the cluster of NF-1-binding sites (−780 to −610) (Fig. 3).

4 Structural and Functional Organization of the Enhancer Region

The nucleotide sequence of the IE-1 enhancer of HCMV revealed four characteristic types of repeat elements with 17, 18, 19, and 21 bp, respectively, each represented three to five times (AKRIGG et al. 1985; BOSHART et al. 1985; STINSKI and ROEHR 1985; THOMSEN et al. 1984) (Fig. 4). The repeats are partially arranged in a nonabutting manner and are interspersed between various stretches of unique DNA. Some repeats are adjacent or overlapping (Fig. 4). DNAase I footprinting experiments with HeLa cell nuclear extracts revealed at least 13 sites of protein/DNA interaction scattered over the entire length of the promoter/enhancer region (GHAZAL et al. 1987, 1988b). This was consistent with the demonstration of DNAase I hypersensitive sites that were mapped to the HCMV enhancer region (NELSON and GROUDINE 1986). Sites of DNA-protein interaction correlated well with repeat