INTRODUCTION

Human cytomegalovirus (HCMV), a member of the herpesvirus group, is a common agent of worldwide occurrence. About 50 to 80% of the adult population in Central Europe or North America are infected with HCMV (Onorato et al., 1985). Primary infections of children or adults usually remain without overt disease. Subsequent to initial infection, HCMV persists in the organism; the virus may be reactivated under conditions of immunosuppression, sometimes resulting in severe illness. HCMV can be transmitted by close personal contact and by blood transfusion (Adler, 1983) or organ transplantation (Pass et al., 1983). HCMV is frequently transmitted from mother to child, either in utero or during the perinatal phase (Dworsky et al., 1983). Prenatal transmission often results in severe disturbance of development and disease that may become manifest at birth, such as thrombocytopenia, hepatitis, splenomegaly, and microcephaly (Stagno & Witek, 1985). Frequent late sequelae are hearing loss and mental retardation (Pass et al., 1980). Disease following post-natal HCMV infection is gaining importance, as conditions of severe immunosuppression or paralysis of the immune system are more often seen with the expansion of transplantation medicine and with the spread of AIDS.

Most medical problems associated with HCMV cannot, at present, be adequately addressed. Repeated standard virus isolation attempts have sometimes been unsuccessful in severe active HCMV disease. Serodiagnostics are not always reliable in determining active HCMV disease, because antibody titres frequently do not correlate with the severity of an illness. Although successful trials for chemotherapy have been accomplished with ganciclovir, life-threatening HCMV infections are not yet amenable to chemotherapy (DHPG Collaborative Study Group, 1986). Prevention methods against prenatal HCMV infections have not been introduced, and there seems to be little prospect of general vaccination programmes. Although the severity of the clinical problems has been recognized for a long time, understanding of natural history, virus structure and pathogenesis at the molecular level has been gained only in recent years. This review discusses some recent developments of the molecular biology of HCMV and their implications for diagnosis.

GENOME STRUCTURE

The genome of HCMV is, as with other herpesviruses, linear dsDNA of high complexity. The first attempts to determine the $M_t$ of HCMV DNA led to the remarkable observation that the genomic DNA exceeded other herpesvirus genomes in size to a significant degree (DeMarchi et al., 1978). Intact virion DNA, defined by its infectivity in cell culture, appeared to be approximately 230 to 240 kb in length (Geelen et al., 1978; Lakeman & Osborn, 1979). HCMV DNA purified from virus particles had a buoyant density in CsCl gradients of 1.716 g/ml, equivalent to an average $G + C$ content of 56% (Huang et al., 1973; Ebeling et al., 1983b). The two criteria, genome size and average $G + C$ content, appear to be characteristic for the beta-herpesviruses. This subgroup is defined not only by molecular characteristics; the members also share a number of pathogenic properties. Other beta-herpesviruses are cytomegalovirus strains from chimpanzee (Wroblewska et al., 1979) and various Old World primates.
M. MACH, T. STAMMINGER AND G. JAHN

Fig. 1. Schematic representation of the isomeric forms of HCMV DNA. The restriction maps for the enzymes HindIII and EcoRI are given in the upper section. L, long segment; S, short segment; U, unique segment; I, internal repeat; T, terminal repeat.

(Huang & Roche, 1978), viruses from mice (Ebeling et al., 1983a; Mercer et al., 1983), rats (Meijer et al., 1984), and guinea-pigs.

The first restriction maps of HCMV DNA were mainly constructed by cross blot experiments using purified virion DNA (Weststrate et al., 1980; DeMarchi, 1981). This was made much easier when virion DNA was cloned in plasmids (Thomsen & Stinski, 1981; Oram et al., 1982; Spector et al., 1982) and cosmids (Fleckenstein et al., 1982), from which detailed physical gene maps could be derived (Greenaway et al., 1982) (Fig. 1). Restriction maps, as well as electron microscopy of alkali-denatured HCMV DNA, indicated that the viral genome consists of two regions, designated as the S (short) and L (long) segments. Each segment is flanked by a pair of inverted repetitions (LaFemina & Hayward, 1980; Ebeling et al., 1983b). Both segments can be orientated in either direction, resulting in four isomeric forms of virion DNA of approximately equal proportions.

The physical maps were always found to be collinear. Restriction enzyme cleavage patterns of various laboratory and wild-type strains revealed distinct differences that usually allow the identification of individual strains and a molecular epidemiology to be established (Huang et al., 1976, 1980b; Grillner & Blomberg, 1984; Chandler & McDougall, 1986). The differences most commonly found were the presence or absence of restriction endonuclease cleavage sites, and some heterogeneity was found at the junction of the long and short components (Weststrate et al., 1983). Southern blot hybridizations with cloned DNA probes showed that all genomic regions from 21 fresh clinical isolates were conserved to the extent that they hybridized identically, using a cloned library of the laboratory strain AD169 (Chou et al., 1984). In summary, this indicates that HCMV strains are relatively stable in nucleotide (nt) sequence and are well conserved, most heterogeneity being represented by a limited restriction enzyme cleavage polymorphism due to sequence divergence. Weston & Barrell (1986) have determined a 46 kb DNA sequence containing the short unique region (U_s), the short repeat (R_s) and part of the long repeat (R_L). This is the largest piece of continuous HCMV DNA sequence reported so
far. Analysis of the sequence has revealed that IRs has a much higher G + C content than Us; the two regions are 74.15% and 56.37% respectively, against an HCMV genome average of 57% (Ebeling et al., 1983b). There is a significant rise in percentage G + C content at the Us-IRs junction, a slight decline across IRs, and then a further drop in IRl, to a mean value of 53.5%. The low G + C content in IRl contrasts with results obtained for varicella-zoster virus (VZV; Davison & Scott, 1986) and herpes simplex virus type 1 (HSV-1; Murchie & McGeoch, 1982) where both the long and short repeats have a G + C content consistently higher than the unique sequence. The current attempts to determine the nucleotide sequence of the entire HCMV genome will contribute considerably to an understanding of the functional genome organization and its evolutionary position within the herpesvirus group.

The DNA of HCMV has some nucleotide sequences in common with mammalian cellular DNA (Peden et al., 1982). Rüger et al. (1984a) and Shaw et al. (1985) found five regions with cell-related nucleotide sequences in the genome of HCMV strain AD169. It has been determined that the nucleotide sequence homologies of HCMV DNA are different from those in simian cytomegalovirus (SCMV) strains, where tandem repeats of CA dinucleotides appeared to hybridize with highly repetitive sequence elements in cellular DNA (Jeang & Hayward, 1983). The DNA sequences which were found to hybridize with the avian retrovirus oncogene v-myc (Spector & Vacquier, 1983; Gelmann et al., 1983) are very high in G + C content (Rasmussen et al., 1985b), probably causing non-specific hybridization signals. There was no indication that viral DNA shared homologies with the cellular proto-oncogene c-myc (Rüger et al., 1984a; Rasmussen et al., 1985b). The highly homologous sequence of the EcoRI b fragment in HCMV strain AD169 DNA contains a promoter for polymerase III transcripts that are synthesized late in virus replication (Marshalek et al., 1989).

Infection of permissive fibroblast cultures gives rise to joining of the termini of viral DNA resulting in either circular or concatemeric replicative forms of HCMV DNA. These forms are abundant in the first 48 h after infection of permissive cells (LaFemina & Hayward, 1983). HCMV DNA usually contains a short sequence, designated the ‘a’ sequence, that is present at both viral termini of virion DNA and is found in inverted orientation at the junction between the short and long components of the genome (Tamashiro et al., 1984). The a sequence is sometimes amplified to tandem repeats and flanked by additional short repeat elements (Mocarski et al., 1987; Tamashiro & Spector, 1986). The a sequence of HCMV can function as a cleavage/packaging signal for defective genomes of HSV (Spaete & Mocarski, 1985). This complementation in trans suggests identical functions of the a sequences of HCMV and HSV, providing a cleavage signal for the monomerization of virion DNA from concatemeric precursors during virus replication.

**IMMEDIATE EARLY GENE EXPRESSION**

Expression of herpesvirus genomes occurs in a temporally regulated manner (Hones & Roizman, 1974, 1975). It can be categorized in three main phases designated as immediate early (IE), early and late. IE genes are the first to be transcribed following entry of the virus into the host cell and their transcription is independent of de novo synthesis of viral proteins. After the IE proteins have been synthesized a switch from restricted to extensive transcription of the genome takes place. The IE proteins are thought to mediate this switch and play a crucial role in regulating viral gene expression.

Infection of permissive cells with HCMV in the presence of cycloheximide or anisomycin, both potent inhibitors of protein synthesis, leads to the accumulation of viral IE RNA. This approach has been used to investigate IE transcription of different HCMV strains (Wathen & Stinsky, 1982; McDonough & Spector, 1983; Wilkinson et al., 1984; DeMarchi, 1981). It was shown that viral IE RNA arises from only a few distinct regions of the genome. Although differences in the pattern of lower abundance transcription have been reported (Wilkinson et al., 1984) the region of highest transcriptional activity maps to a common position between map units (m.u.) 0·66 and 0·77 within the genomes of all HCMV strains investigated so far. This region of approximately 20 kb, located within the large unique compartment of the genome, corresponds to the HindIII E fragment of strain AD169 and the XbaI E and N fragments in
strain Towne. It is of interest that this transcription pattern contrasts with the pattern seen in HSV-infected cells in which the major IE transcripts originate from the repeat sequences (Clements et al., 1977; Easton & Clements, 1980).

Structural organization of the major IE transcriptional units of HCMV

Four major transcription units designated IE1 to IE4 are localized within the region of highest IE transcriptional activity (Jahn et al., 1984; Stinski et al., 1983) (Fig. 2). mRNAs arising from IE1 to IE3 could be found in the polyadenylated fraction of RNA (Jahn et al., 1984; Stinski et al., 1983). Conflicting information exists regarding the IE3 region (0.709 to 0.728 m.u.) which in strain AD169 encodes a 2.2 kb mRNA, and in strain Towne encodes a 1.95 kb mRNA (Jahn et al., 1984; Wilkinson et al., 1984; Stinski et al., 1983) (Fig. 2). Whereas the latter groups classified IE3 to be 'immediate early', Staprans & Spector (1986) reported an early transcript from this region of the genome. A 5 kb transcript, corresponding to the IE4 transcription unit, seems to be unusual; part of this RNA appeared polyadenylated, but the majority of the transcript was found in the non-polyadenylated fraction. Unlike the other IE genes, the DNA encoding the 5.0 kb IE RNA is also transcribed in high quantities during the late phase of virus gene expression. When searching for an open reading frame (ORF) within IE4, firm evidence could not be found for a protein-encoding sequence in this region (Plachter et al., 1988). A DNA fragment previously reported to be able to transform rodent cells in vitro is totally contained within the IE4 coding region (Nelson et al., 1982, 1984).

IE region 1, which is also referred to as the major IE gene, encodes the most abundant species of IE RNA, with a size of 1.95 kb (Stinski et al., 1983; Wilkinson et al., 1984; Jahn et al., 1984). This predominant RNA is transcribed from right to left on the prototype arrangement of the viral genome (0.739 to 0.751 m.u.). Transcription of this gene is controlled by a strong cis-acting element, located immediately upstream of the 5' end of the RNA. The structure of the gene has been investigated in detail by nucleotide sequence analysis and nuclease mapping of the corresponding transcript (Stenberg et al., 1984; Akrigg et al., 1985). It encodes a spliced transcript of about 1.9 kb, made up of three small exons (121, 88, 185 nt) and one large exon (1342 nt). Three introns (827, 114 and 170 nt) are located near the 5' end of the gene. A single ORF starts in the second exon of the gene and has a coding capacity for 491 amino acids. The $M_r$ of the encoded protein was calculated to be 63,8K (Akrigg et al., 1985). In vitro translation of mRNA selected by hybridization to IE region 1 demonstrated that this region encodes the predominant IE protein found in infected cells within 1 h after infection (Stenberg et al., 1984). The protein accumulates primarily in the nucleus of the cell (Michelson-Fiske et al., 1977). The $M_r$ of the predominant IE protein, which varied slightly between different strains of HCMV, was estimated to be 75K by SDS–PAGE (Blanton & Tevethia, 1981; Cameron & Preston, 1981; Gibson, 1981b; Wilkinson et al., 1984). This value differs considerably from that calculated from the sequence. A proline-rich region near the amino terminus might lead to anomalous mobility in SDS–PAGE and thus explain this discrepancy (Akrigg et al., 1985). Stinski et al. (1983) suggested that post-translational modifications might also contribute to heterogeneous migration in SDS–PAGE. It could be shown that the major IE protein is phosphorylated (Gibson, 1981b).

IE region 2 lies adjacent to region 1. Northern blot analysis detected at least four classes of mRNA, according to size, ranging from 1.1 to 2.25 kb, originating from a DNA region of approximately 2 kb (Stenberg et al., 1985; Akrigg et al., 1985). This suggests that various splicing events occur in this relatively small part of the genome. Nuclease mapping experiments revealed the direction of transcription, which is orientated from right to left (as for the major IE1 transcript), as well as the presence of both spliced and unspliced transcripts. Typical CAAT and TATA boxes were found in the nucleotide sequence, suggesting that a functional promoter is located immediately upstream of the 5' end of the mRNA start site. The gene extends about 1.6 kb downstream with a polyadenylation signal about 20 nt upstream of the 3' end. A start codon at nt 252 initiates a single ORF which extends to the first stop codon at nt 1092. This ORF can encode 280 amino acids, which corresponds to a protein of 30K. By splicing nt 836 to nt 1301, the first stop codon can be eliminated and the ORF extended to a second stop codon at nt
Fig. 2. Organization of HCMV IE genes. The upper part of the figure shows the genomic localization of the four major IE transcription units of HCMV. The mRNA structures, sizes and predicted ORFs for IE genes IE1 and IE2 are illustrated in the lower section [data taken from Hermiston et al. (1987)]. Open boxes represent mRNA exons and thin lines introns. Start and stop codons of the respective ORFs, and polyadenylation signals are indicated.
This RNA has an ORF of 255 amino acids (predicted Mr of 27K for the protein). A minor spliced RNA has been detected, which uses a splice site just upstream of the second stop codon and can splice to sequences approximately 2 kb further downstream (Stenberg et al., 1985). This has led to the subdivision of this IE region into IE2a, containing the sequence to the first polyadenylation signal, and IE2b, which contains the part localized downstream. Further investigation of IE2b revealed the existence of two exons. Thus the minor RNA which originates in IE2a and is spliced onto IE2b has a length of 2.25 kb and encodes a protein of 78K (Hermiston et al., 1987). The detection of different species of RNA, in which the first three exons of region 1 are differentially spliced onto region 2 exons, further complicates the splicing pattern observed for IE2 RNAs, and extends the range of encoded proteins (Fig. 1). It is also remarkable that the pattern of region 2 RNAs changes during the time course of HCMV infection: spliced transcripts were detected preferentially under IE conditions, whereas later in the replication cycle, unspliced transcripts from this region dominate (Stenberg et al., 1985). The gene products of viral IE2 mRNAs were determined by in vitro translation of hybrid-selected RNA. The predominant protein encoded by this region has an Mr of 56K, whereas several other proteins ranging in Mr from 16.5K to 42K were also detected (Stinski et al., 1983).

Functional aspects of IE gene products

IE gene products are thought to function as regulatory proteins by acting in trans on cis-regulatory elements such as promoters or enhancers. Transfection assays with recombinant plasmids containing a reporter gene linked to a eukaryotic control element were used (in most cases) to investigate these trans-acting effects. Everett & Dunlop (1984) were able to show that transcription from two heterologous promoters, from the genes for HSV-1 glycoprotein D and rabbit β-globin, could be stimulated upon cotransfection with the XbaI E fragment of HCMV strain Towne or the HindIII E fragment of strain AD169. The results, which located a trans-acting function within the major IE transcription unit, were strengthened by complementation studies; the same DNA fragment of HCMV could complement an adenovirus 5' IE mutant, which was defective in its Ela gene region (Spector & Tevethia, 1986). Since IE1 is the most abundantly transcribed gene and the corresponding protein is the predominant protein at times of IE gene expression in the infected cell (Blanton & Tevethia, 1981; Stinski, 1978; Stinski et al., 1983), this gene was thought to be the most probable candidate for exerting a trans-acting function. Moreover, its expression is regulated very strictly during the time course of infection; it reaches a maximum at about 4 to 5 h after infection whereas barely detectable levels of RNA are produced at late times (Jahn et al., 1984; Stenberg et al., 1984). Stenberg & Stinski (1985) tried to investigate the role of this gene by expressing the IE1 region in COS cells, which resulted in detectable levels of IE1 RNA and protein. By using a deletion mutant, which expressed a protein lacking 145 amino acids at the carboxy terminus, they could show that the intact IE1 gene product is necessary to down-regulate its own expression in this system. Therefore they proposed a negative autoregulatory mechanism of IE1 gene expression in analogy to HSV IE genes (DeLuca et al., 1984; Preston, 1979, 1981). This mechanism was assumed to be responsible for the dramatic decrease of IE1 transcription at late times. As COS cells contain the large T antigen of simian virus (SV40), possible interactions between these two regulatory proteins cannot be ruled out in this system. Similar results were reported by Davis et al. (1987), who found that cotransfection of an IE1 expression plasmid together with a chloramphenicol acetyltransferase (CAT) hybrid gene, driven by the HCMV IE1 regulatory region, down-regulates gene expression by a factor of eight in HeLa cells. However they could also report a stimulatory effect of HCMV IE gene products on the long terminal repeat of human immunodeficiency virus type 1, which was fused to the CAT gene. This stimulatory trans-acting effect was not due to the IE1 gene product but could be produced by an expression plasmid containing IE2 sequences. Hermiston et al. (1987), who investigated the effect of IE1 and IE2 proteins on CAT gene expression driven by the adenovirus E2 promoter (E2CAT) obtained similar results. In this system the IE1 gene product could not function independently in activating the E2 promoter whereas the IE2 gene products could. Sequences from region IE2a, with ORFs which could encode two small proteins of 30K and 27K (Fig. 2), were able to stimulate the expression from E2CAT to a high level independently of IE1. A plasmid with the IE1 regulatory regions and
sequences coding for IE2 proteins (27K, 30K, 48K and 51K) also stimulated E2CAT expression, but to a lower level (Fig. 2). However expression of this plasmid was augmented by IE1 sequences, suggesting that IE2 might exert its maximum stimulatory effect only in combination with IE1. These results were confirmed by investigating the ability of different IE genes of HCMV to complement mutants of adenovirus E1a and to stimulate expression from the early E2 promoter of adenovirus (Tevethia et al., 1987). It was observed that the stimulatory effect of both IE region 1 and 2 together is about 10-fold higher than that of IE2 alone. In contrast to Tevethia et al. (1987) and Hermiston et al. (1987), Pizzorno et al. (1988) reported the absence of a cooperative effect from IE1 and IE2. Furthermore they attributed both the stimulatory and the negative autoregulatory effect of the IE genes to a protein encoded by IE2 sequences. Investigation of HSV trans-acting proteins revealed that the type of cell line used can influence the observed trans-acting function to a significant degree (O'Hare & Hayward, 1985). The conflicting results may be explained by the fact that Pizzorno et al. (1988) used Vero cells for their experiments, whereas Hermiston et al. (1987) worked with permissive human fibroblasts.

**Regulatory elements in the IE1 upstream region**

IE1 is the region transcribed in highest abundance at immediate early times of infection. IE gene expression occurs independently of de novo protein synthesis and the IE1 gene product can be detected within 1 to 2 h after infection. The activation of IE1 gene expression is one of the first steps of transcriptional interaction between virus and host cell and therefore plays a key role in regulating viral gene expression. A remarkable feature of the IE1 upstream region (also referred to as the IE1 promoter/regulatory region; Thomsen et al., 1984), is its complex sequence composition (Fig. 3). A CAAAT sequence and a TATA box are found within 65 bp upstream of

---

*Fig. 3. Organization of the IE1 upstream region. The IE1 upstream region can be subdivided into three main domains, which comprise the modulator, the NF1 binding sites, and the enhancer/promoter region. The nucleotide sequence of the enhancer/promoter is shown in the lower section of the figure. The four repeat elements of 17, 18, 19 and 21 bp are illustrated by different underlining [(.-.-)], 17 bp repeat; (-----), 18 bp repeat; (- - - -), 19 bp repeat; (-----), 21 bp repeat). Nucleotide sequence elements which are protected in DNase I footprinting experiments are indicated by large brackets. Bold brackets and thin brackets correspond to protected regions of each DNA strand. Small brackets show GC boxes, which represent potential Sp1 binding sites.*
the mRNA 5' terminus. These sequences form the IE1 promoter element and are required for a minimal level of gene expression in both in vivo and in vitro transcription assays (Stinski & Roehr, 1985; Ghazal et al., 1988b). The upstream region of the promoter contains a series of different repeat sequences, distributed up to position −509 from the mRNA cap site (Thomsen et al., 1984; Akrigg et al., 1985; Stinski & Roehr, 1985; Boshart et al., 1985) which enhance expression of the HCMV major IE gene (Thomsen et al., 1984; Stinski & Roehr, 1985; Boshart et al., 1985). By cotransfection with enhancer-negative SV40 DNA (the 'enhancer trap'; Weber et al., 1984), a DNA fragment of HCMV was identified which can substitute for the SV40 enhancer. This sequence, which is located upstream of the transcription initiation site of IE1 between nt −118 and −524, fulfils the essential criteria for a eukaryotic enhancer element (Boshart et al., 1985). It strongly activates transcription in a wide variety of cells, including Xenopus laevis kidney cells (Boshart et al., 1985; Foecking & Hofstetter, 1986), and is able to stimulate transcription from its cognate promoter and from heterologous promoters (Stinski & Roehr, 1985; Boshart et al., 1985). The enhancing effect of this sequence was reproduced in an in vitro system (Thomsen et al., 1984; Ghazal et al., 1987). Investigation of the role of sequences between positions −120 and −65 revealed that they are able to stimulate transcription by at least 10-fold (Ghazal et al., 1988b), suggesting that the transcription-enhancing domain extends beyond the enhancer sequence described by Boshart et al. (1985). Very strong enhancer sequences are also located upstream of IE genes of the mouse cytomegalovirus (MCMV) (Dorsch-Häsler et al., 1985) and SCMV (Jeang et al., 1987). No comparable enhancers could be identified within any other herpesvirus genome investigated so far. Thus, these strong constitutive enhancers seem to be characteristic for cytomegalovirus herpesviruses.

Four types of repeat elements with 17, 18, 19 and 21 bp, respectively, each represented three to five times within the IE1 upstream region, contribute to the complexity of the enhancer/promoter sequence (Boshart et al., 1985; Akrigg et al., 1985; Stinski & Roehr, 1985; Thomsen et al., 1984). The repeats are arranged in a non-abutting manner and are interspersed between various unique DNA sequences (Fig. 2). Footprinting experiments with HeLa cell nuclear extracts revealed at least 13 sites of protein–DNA interaction distributed over the promoter/enhancer region (Ghazal et al., 1987; 1988b). This confirmed earlier results, in which DNase I hypersensitive sites, serving as indicators of DNA–protein interaction, were mapped to the HCMV enhancer region (Nelson & Groudine, 1986). The location of DNA–protein interactions correlated mainly with the repeat sequences, indicating that each of the repeat motifs acts as a target sequence for specific transcription factors (Ghazal et al., 1987, 1988a). The pattern of protein–DNA interaction is even more complex because unique sequences are also protected. The extent of protection for a repetitive sequence varies within members of a class of repeat elements as well as between different classes (Ghazal et al., 1988b). These features are consistent with the model of a large nucleoprotein complex forming on the enhancer sequence and allowing for extensive protein–protein interactions. To investigate the functional role of different repeat elements, competition assays were performed both in vivo (Fickenscher et al., 1989) and in vitro (Ghazal et al., 1988a), in which synthetic oligonucleotides corresponding to repeat elements competed for factors of the transcriptional machinery. It was shown that competition with the 18 and 19 bp element efficiently diminished the stimulatory effect of the HCMV enhancer on transcription. The 17 bp repeat element does not appear necessary for efficient enhancer function (Stinski & Roehr, 1985; Ghazal et al., 1988a). A 12 nt core sequence of the palindromic 19 bp repeat element is conserved in the IE upstream sequences of cytomegalovirus herpesviruses, but not within the known sequences of other herpesviruses. The inner palindromic core of the 19 bp motif corresponds to a transcription regulatory element in various hormone genes (Silver et al., 1987; Montminy et al., 1986), which is able to act as a 3',5' cyclic AMP response element. The 18 bp repeat motif is also contained within the IE upstream region of MCMV but could not be detected within SCMV regulatory sequences (Dorsch-Häsler et al., 1985; Jeang et al., 1987). A similar sequence also forms part of a different, more recently described enhancer of HCMV, located in the short unique region of the genome (Weston, 1988).

An octameric DNA sequence, present in three of the 18 bp repeats, resembles the cellular heat shock treatment core consensus. It was suggested that this element might be involved in
transcriptional activation of IE1 gene expression following heat shock in a cell line harbouring several stable integrated copies of the IE1 gene (Boom et al., 1986; Geelen et al., 1987). The 18 bp repeat matches over 10 bp with the NF-kappa-B binding site of the kappa Ig enhancer (Picard & Schaffner, 1984; Sen & Baltimore, 1986). No homologies could be found for the 21 bp repeat motif.

Although the functional relevance of different sequence elements of the enhancer region has been confirmed, further investigations will be necessary to elucidate the mechanisms by which these elements mediate the enhancer effect. Although the enhancer binding proteins have not been identified so far, four high affinity binding sites for nuclear factor 1 (NF1) have been mapped between positions −780 and −610 in the upstream region of the IE1 enhancer/promoter (Hennighausen & Fleckenstein, 1986; Jeang et al., 1987). The NF1 binding sites of the HCMV IE upstream region coincide with regions shown to be sensitive to DNase I in the active gene, but not sensitive in the silent gene (Nelson & Groudine, 1986). A cluster of at least 20 NF1 binding sites has been observed upstream of the IE promoter/regulatory region of SCMV (Jeang et al., 1987). As SCMV is able to express IE genes in a broader spectrum of cell types (LaFemina & Hayward, 1988), it has been speculated that NF1 might have an influence on cell type-specific expression of IE genes (Jeang et al., 1987; Hennighausen & Fleckenstein, 1986). Additional DNase I hypersensitive sites were found in the region further upstream between positions −750 and −1145, suggesting that this region acts also as a target for trans-acting factors (Nelson & Groudine, 1986). To investigate the role of this upstream element, it was placed in front of homologous and heterologous promoters linked to the CAT reporter gene. In transient expression assays this element was able to modulate gene expression in non-permissive cells in a negative manner, whereas it positively influences expression in permissive cells (Nelson et al., 1987). This was confirmed by in vitro transcription studies in which nuclear extracts of various cell lines were tested for their ability to mediate transcription in the presence or absence of the IE1 upstream region between positions −1145 and −524 (Lubon et al., 1989). This region, also referred to as the modulator region, participates in certain cell lines in transcriptional repression, whereas it augments transcription in other cells. This suggests a cell-type specific regulatory mechanism of IE1 gene expression (Lubon et al., 1989).

In summary the complex IE1 upstream region can be divided into three main domains. The promoter between positions +3 and −65 is required for a minimal level of transcription. The enhancer domain between positions −520 and −65 is able to stimulate downstream gene expression to a high level in a wide variety of cell types. The domain upstream of the enhancer/promoter appears to modulate IE gene expression in a cell type-specific and differentiation-dependent manner.

**IN VITRO TRANSFORMATION AND ONCOGENESIS**

Numerous types of human malignant tumours, such as neuroblastoma, Wilm's tumour (Wertheim & Voute, 1976), prostate cancer (Sanford et al., 1978), and testicular cancer (Mueller et al., 1988) have been linked to HCMV by seroepidemiological evidence. The virus was isolated from biopsies of cervical cancer (Melnick et al., 1978, 1979) and adenocarcinoma of the colon (Hashiro et al., 1979). In vitro hybridizations using radioactively labelled purified virion DNA seemed to indicate that biopsy materials from colon carcinomas can contain HCMV-specific DNA sequences (Huang & Roche, 1978). This was confirmed by reassociation kinetics in the liquid phase (Roche et al., 1981). Further studies, however, applying cloned probes of HCMV DNA could not substantiate the idea that viral genome fragments are consistently present in adenocarcinomas (Bričkaček et al., 1980) or other forms of colorectal tumours (Rüger & Fleckenstein, 1985). These studies, using cloned HCMV DNA from the entire viral genome, did not provide any indication that certain DNA fragments are consistently present in tumour tissues.

Kaposi's sarcoma was among the first tumours suspected to be associated with HCMV infections. This was primarily based on seroepidemiology; elevated antibody titres were measured in the sera of European and American patients with Kaposi's sarcoma in its classical form (Giraldo et al., 1972a, 1978), and furthermore, virus particles, virion proteins and viral
DNA were found in tumour biopsies (Boldogh et al., 1981; Giraldo et al., 1972b, 1975, 1980; Drew et al., 1982). Kaposi's sarcoma is associated with depressed functions of the cellular immune system; elevated antibody titres can be the consequence of endogenous virus reactivation, but serology is not applicable for proving a causal relationship between virus infection and tumour initiation or development. Thus the assumption that Kaposi's sarcoma may be related to HCMV has rested mainly on the demonstration of virus-specific nucleic acids in tumour specimens by reassociation kinetics with purified, radioactively labelled HCMV DNA probes. However this was challenged when cloned DNA probes became available for Southern blot analysis. First, it was possible to identify viral DNA sequences with homologies to cellular DNA and to avoid those parts of the genome in the search for virus-specific DNA sequences in tumour biopsies. Cosmid- and plasmid-cloned probes representing about 70% of the HCMV genome were used for Southern blot hybridizations, excluding all sequences of virus DNA with homology to cellular DNA. Applying the selected probes, no HCMV DNA fragments were found in European Kaposi's sarcoma or tumours of the African type or AIDS-related Kaposi's sarcomas (Rüger et al., 1984b,c). Also, in situ hybridizations with cloned probes gave no evidence for the presence of HCMV nucleic acids in specimens of African Kaposi's sarcoma (Ambinder et al., 1987). Therefore the more recent studies with cloned DNA did not confirm that HCMV infection is consistently associated with or causally related to Kaposi's sarcoma, although some features of pathogenesis and epidemiology have suggested a viral cause for this proliferative disease. In summary, there is no conclusive evidence for a role of HCMV in the aetiology of human neoplastic diseases.

In parallel with the seroepidemiology and with the search for virus DNA in tumours, numerous experiments have been performed to determine the oncogenic potential of HCMV by transformation studies in cell culture (reviewed by Macnab, 1987). The virus resembles oncogenic DNA viruses in its ability to stimulate host cell DNA synthesis (St. Jeor et al., 1974), transcription and cellular enzyme activities (Isom, 1979). Virus particles inactivated by u.v. radiation were shown to transform hamster embryo cells to an oncogenic phenotype when injected into newborn hamsters (Albrecht & Rapp, 1973). In vitro transforming capacity was correlated with a short segment of IE gene 4 (Nelson et al., 1982) that appeared not to code for protein (Nelson et al., 1984; Plachter et al., 1988). In addition, a second region with possible transforming capacity was described (El-Beik et al., 1986). However stable transformants containing virus-specific DNA sequences could not be isolated, casting doubt on the active role of HCMV in the maintenance of a transformed phenotype (Galloway & McDougall, 1983). Thus in vitro transformation assays with cloned HCMV DNA could not support the notion that HCMV is an oncogenic virus expressing defined genes in transformed culture cells or in neoplastic tissues.

**EARLY GENES AND PROTEINS**

The second (early, E) phase of herpesvirus gene expression is dependent on IE gene products. These genes have been defined experimentally by their transcription in the presence of inhibitors of viral DNA synthesis such as phosphonoacetic acid and phosphonoformic acid. This definition is most likely to be an oversimplification because some transcripts which can be detected during the early phase are also found at late times.

The early transcripts of HCMV originate from most parts of the genome (DeMarchi, 1981; Wathen et al., 1981; Wathen & Stinski, 1982; McDonough & Spector, 1983; Chang et al., 1989). Regions from which stable early RNA could not be detected include several areas on the L segment, as well as sequences encompassing the L-S junction and the righthand end of the genome (DeMarchi, 1984).

Several groups have described the organization and expression of major early genes derived from the long inverted repeats of the HCMV genome (Hutchinson et al., 1986; Greenaway & Wilkinson, 1987; McDonough et al., 1985). Three poly(A)° RNAs of 2.7, 2.0 and 1.2 kb were detected from this genomic region during the early and late phases. The major RNA of 2.7 kb mapped within the EcoRI O and EcoRI WN fragments in the long repeat. According to S1 nuclease and exonuclease VII protection analyses there was no indication of splicing. A
potential ORF encoding a polypeptide of 170 amino acids could be identified close to the 5' terminus within the sequence of the gene encoding the 2.7 kb RNA. The transcript contains a long untranslated 3' trailer sequence with numerous stop codons (McDonough et al., 1985). Attempts to identify a gene product using hybrid-selected in vitro translation and various prokaryotic expression systems have been unsuccessful (Greenaway & Wilkinson, 1987). A number of additional early transcripts were mapped to the internal portion of the long repeat (McDonough et al., 1985). One of these RNAs is an unspliced 1.2 kb transcript, found in the early phase which accumulates at late times. It could be translated in vitro after hybrid selection of specific DNA fragments for the mRNA, resulting in a protein of approximately 37K (Hutchinson & Tocci, 1986). Staprans & Spector (1986) described an early transcript of 2.2 kb, located at the right end of the long unique segment. However, it should be mentioned that this transcript was classified as IE by other groups (see previous section). This transcript, encoded by EcoRI fragments R and d of the HCMV strain AD169, showed a complex spliced structure with 5' and 3' coterminial ends. Four phosphorylated proteins of 84K, 50K, 43K and 34K were immunoprecipitated with an antisera directed against a synthetic oligopeptide corresponding to the predicted amino terminus of the protein. The 43K protein was the most abundant of these polypeptides, and similar to the others, was associated with the nuclei of the infected cells (Wright et al., 1988).

The HCMV DNA polymerase has been described in detail. This enzyme was distinguished from cellular polymerases by chromatographic behaviour, template primer specificity, and by the requirement of KCl or ammonium sulphate for maximum activity (Huang, 1975; Nishiyama et al., 1983). The purified virus-induced polymerase consists of two proteins of about 140K and 58K. Various nucleotide triphosphate analogues were tested as substrates for purified HCMV DNA polymerase and the cellular DNA polymerases. Each derivative had a more inhibitory effect on the virus-specific DNA polymerase than on cellular DNA polymerases (Mar et al., 1985a, b). The genomic localization, gene sequence and transcription analysis of the HCMV polymerase have been recently published (Heilbronn et al., 1987; Kouzarides et al., 1987). The gene is located within the long unique region of the genome (HindIII F fragment) of strain AD169, and is transcribed into a 5-4 kb RNA which is present at early and late times after infection (Heilbronn et al., 1987). The sequence of the polymerase gene shows homology to the predicted Epstein–Barr virus (EBV) and HSV polymerases both on the DNA and the amino acid level. More than 24% of their sequences is precisely conserved, and deletions and insertions that appear to have occurred are all in non-conserved areas of the protein (Kouzarides et al., 1987a, b). Viral enzymes usually play an important role in the mechanism of action of anti-herpetic drugs, so that the characterization of this particular enzyme and its coding sequence was of interest. Nucleoside analogues such as acyclovir have been proven to be selective and effective by inhibition of the HSV-encoded DNA polymerase after activation by the virus-encoded thymidine kinase (Elion et al., 1977). The potency of this anti-herpetic nucleoside analogue has been correlated with its selective phosphorylation by the virus-encoded thymidine kinase. However no evidence has been found for the presence of an HCMV-specific thymidine kinase activity in infected cells (Estes & Huang, 1977), and this might explain why acyclovir is not sufficiently effective against HCMV infection. It has been demonstrated that other nucleoside analogues, such as ganciclovir [9-(1,3 dihydroxy-2-propoxymethyl)-guanine; DHPG], were effective inhibitors of HCMV multiplication in vitro and in vivo. Ganciclovir, however, seems to be phosphorylated to DHPG monophosphate by a cellular deoxyguanosine kinase and then further phosphorylated to the triphosphate by other cellular enzymes. This appears to happen in HCMV-infected cells with an efficiency 10-fold higher than in non-infected cells. Patients with -CMV retinitis or gastrointestinal disease stabilized or improved under DHPG therapy (DHPG Treatment Study Group, 1986).

Some other early proteins with functions such as DNA binding have been described recently and their coding sequences determined (Mocarski et al., 1985; Anders & Gibson, 1988; Kemble et al., 1987). An ssDNA-binding protein is present in HCMV-infected cells which has properties analogous to those of HSV strain ICP8 (Anders et al., 1987). The gene has been mapped in the centre of the long unique segment of the HCMV genome (BamHI K fragment of strain Towne
and *BamH1* M fragment of strain AD169) by screening a λgt11 library of HCMV DNA fragments with a monoclonal antibody (MAb) (Kemble et al., 1987). In addition, sequence conservation allowed the identification of the HCMV gene and the SCMV (Colburn) major DNA-binding protein gene by Southern blot hybridization with a probe specific to HSV strain ICP8 (Anders & Gibson, 1988). Early HCMV proteins and viral enzymes have also been reviewed recently (Landini & Michelson, 1988).

**STRUCTURAL PROTEINS AND THEIR CODING SEQUENCES**

There are numerous reports describing the structural proteins of HCMV. Recently, Landini & Michelson (1988) summarized more than one-hundred publications on HCMV-encoded proteins and tried to classify these proteins on the basis of *M*, and type of modification. Infectious virus particles probably contain about 35 to 40 structural proteins. All or most of those proteins are also found in non-infectious enveloped particles (NIEPs), a form of extracellular particle which does not contain DNA (Gibson & Irmiere, 1984). NIEPs also contain a protein of 36K which has been termed the "assembly protein". The third type of particles produced in tissue culture are designated 'dense bodies'. They do not contain a capsid and consequently are devoid of the respective structural proteins. Some 80 to 90% of the protein mass in dense bodies of laboratory strains are accounted for by a phosphoprotein of approximately 65K (pp65) (Gibson & Irmiere, 1984). Extensive characterization of HCMV proteins is hampered by the low abundance of most polypeptides in virions. On a structural basis, analyses have been carried out for capsid proteins (Irmiere & Gibson, 1985; Gibson, 1981a) and some phospho- and glycoproteins (Roby & Gibson, 1986; Stinski, 1977). A virus-associated kinase has been identified as a structural component and has been characterized (Roby & Gibson, 1986; Brit & Auger, 1986b). Structural proteins are important targets for the humoral and cellular immune response. Envelope proteins are the essential targets for neutralizing antibodies. Identification and characterization of these proteins are important for several reasons, the most obvious being the development of a subunit vaccine or the determination of antigenic variability in different strains of HCMV. In this section we focus on those structural proteins whose coding regions have been identified.

It seems that the nomenclature of HCMV proteins today is in a state of individualism rather than rationalism. Table 1 lists reported descriptions for the structural proteins discussed in the section. For the purpose of this review, a nomenclature has been chosen which gives the *M* (* × 10^-3*) of the protein preceded by the prefix p for protein, pp for phosphoprotein and gp for glycoprotein. Until final agreement has been reached on the nomenclature of HCMV proteins, these assignments must be considered tentative.

**Phosphoproteins**

Extracellular virions of various strains of HCMV contain three predominant proteins (pp150, pp71 and pp65) and at least nine minor phosphorylated species ranging in *M* from 122K down to 24K (Roby & Gibson, 1986). As in HSV, the phosphoproteins of HCMV appear to be located in the matrix region of the particle (Gibson, 1981a, 1983; Gibson & Irmiere, 1984). NIEPs contain the same phosphorylated proteins and one extra major species, the 36K assembly protein. In dense bodies the pp65 is the predominant phosphorylated protein, whereas the other two species (pp150 and pp71) are absent or present at less than 1% of their relative amounts in the other particles (Roby & Gibson, 1986). Immunoblot analyses have shown that the phosphoproteins are among the species most consistently recognized by human sera. In particular, pp150, pp65 and pp28 are highly reactive and may be of diagnostic value (Landini et al., 1985).

So far, the coding regions for five phosphoproteins (pp150, pp71, pp67, pp65 and pp28) have been identified. Using a cDNA expression library constructed in the vector λgt11 and monospecific rabbit antiserum, the gene encoding pp150 was mapped to the long unique segment of the viral genome within *HindIII* fragments J and N (Jahn et al., 1987b). This region contains two ORFs of 3144 and 2014 nt respectively, neither of which is large enough to encode a polypeptide of 150K. The longer frame encodes a polypeptide of 1048 amino acids which...
Table 1. Reported M, values and designation of HCMV structural proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pp150</td>
<td>149K, basic, phosphoprotein (BPP)</td>
<td>Roby &amp; Gibson (1986)</td>
</tr>
<tr>
<td>pp150</td>
<td>149K upper matrix proteins (UM)</td>
<td>Roby &amp; Gibson (1986)</td>
</tr>
<tr>
<td>pp150</td>
<td>149K phosphorylated viral protein</td>
<td>Nowak et al. (1984a, b)</td>
</tr>
<tr>
<td>pp65</td>
<td>69K lower matrix protein (LM)</td>
<td>Roby &amp; Gibson (1986)</td>
</tr>
<tr>
<td>pp65</td>
<td>69K matrix-like protein</td>
<td>Nowak et al. (1984a, b)</td>
</tr>
<tr>
<td>pp65</td>
<td>69K glycoprotein</td>
<td>Irmiere &amp; Gibson (1983)</td>
</tr>
<tr>
<td>pp28</td>
<td>24K phosphoprotein</td>
<td>Meyer et al. (1988)</td>
</tr>
<tr>
<td>pp28</td>
<td>24K, phosphoprotein</td>
<td>Re et al. (1985)</td>
</tr>
<tr>
<td>gp58</td>
<td>gp58-55</td>
<td>Gönczöl et al. (1986)</td>
</tr>
<tr>
<td>gp61</td>
<td>gp58-55</td>
<td>Landini et al. (1987a)</td>
</tr>
<tr>
<td>gB</td>
<td>gp58-55</td>
<td>Cranage et al. (1986)</td>
</tr>
<tr>
<td>gA</td>
<td>gp58-55</td>
<td>Pereira et al. (1984)</td>
</tr>
<tr>
<td>gp62</td>
<td>gp58-55</td>
<td>Benko &amp; Gibson (1986)</td>
</tr>
<tr>
<td>gp50-52</td>
<td>gp58-55</td>
<td>Law et al. (1985)</td>
</tr>
<tr>
<td>gp66</td>
<td>gp58-55</td>
<td>Rasmussen et al. (1984)</td>
</tr>
<tr>
<td>gp66</td>
<td>gp58-55</td>
<td>Cranage et al. (1988)</td>
</tr>
</tbody>
</table>

constitutes a 113K protein. This frame also contains the DNA sequence homologous to the pp150-specific cDNA. The size discrepancy between the sequence-derived \( M_r \) value and the \( M_r \) estimated from PAGE may be partly due to particular characteristics of the phosphoprotein i.e. high content of proline residues, a basic overall nature and a high degree of phosphorylation. However it is still possible that splicing events do occur. Computer analysis of the primary translation product indicated a remarkable accumulation of hydrophilic regions which in most cases coincide with \( \beta \)-pleated sheets. This could explain the high antigenicity of the protein. A 450bp fragment of the pp150 gene was expressed as a fusion protein with a truncated \( \beta \)-galactosidase. It was sufficient for recognition by human sera and for induction of antibodies in rabbits which allows \textit{in situ} detection of HCMV antigen in infected cells (Scholl \textit{et al.}, 1988).

The most abundant phosphoprotein in laboratory strains of HCMV is the lower matrix protein (Gibson & Irmiere, 1984). From the location of the gene and from the DNA sequence it is clear that the protein originally described as glycoprotein 64 (gp64) of the Towne strain (Clark \textit{et al.}, 1984; Pande \textit{et al.}, 1984) and the phosphoprotein 65 (pp65) of strain AD169 (Nowak \textit{et al.}, 1984a; Rüger \textit{et al.}, 1987) are the same. Both polypeptides map to the corresponding \textit{HindIII} fragments \( H \) and \( N \) of strain Towne and fragments \( L \), \( b \), and \( c \) of strain AD169. In addition, the partial amino acid sequence reported (44 amino acids) for gp64 (Pande \textit{et al.}, 1984) matches perfectly to the hypothetical sequence derived from the DNA sequence (positions 1813 to 1954) of the pp65 gene (Rüger \textit{et al.}, 1987). The gene appears to be unusual in structure. It does not contain a typical \textit{TATA} consensus sequence as found in most eukaryotic promoter sequences. Another unusual feature of its mRNA is that it is a potential bicistronic messenger. At its 3' end it contains the reading frame encoding another structural phosphoprotein of 71K. It is not clear, however, whether this reading frame is actually translated into a protein from pp65 mRNA or whether pp71 is encoded entirely by a rare 1-9 kb mRNA which is 3'-coterminal with the 4 kb mRNA encoding pp65 (Rüger \textit{et al.}, 1987). Again, the sequence-derived \( M_r \) values for pp65 and pp71 differ significantly from those estimated by PAGE (61.5K and 62.9K respectively). Using MAbs combined with immune electron microscopy, the pp65 has been localized in infected cells
and HCMV particles. The protein was detected in the nuclei and in the cytoplasm, exclusively within the dense body matrix (Landini et al., 1987b). Using SV40-transformed monkey kidney cells transfected with HCMV DNA and a MAAb, Davis et al. (1984) mapped an additional phosphorylated viral protein in the size range of 65 to 67K to the EcoRI G fragment of strain Towne. Although this polypeptide is encoded by a 1.9 kb mRNA and is located near the reported map position for pp65, the sequence comparison shows that the two polypeptides are not related (Davis & Huang, 1985). Protein kinase activity has been associated with an HCMV-encoded protein of approximately 68K (Mar et al., 1981; Britt & Auger, 1986b; Michelson et al., 1984, 1985; Roby & Gibson, 1986). It is not clear whether pp65 or pp67 contain the enzymic activity.

Recently the gene encoding another phosphoprotein, pp28, was identified. It is located within the HindIII R region of strain AD169 (Meyer et al., 1988). It is transcribed into an abundant 1.3 kb RNA, present exclusively at late times in infection. Fusion proteins containing a large portion of pp28 in addition to β-galactosidase have been shown to be able to substitute for authentic viral antigen in an immunoblot analysis with HCMV-positive human sera (Meyer et al., 1988).

In summary, the genes encoding the major phosphoproteins have been identified. None of the characterized phosphoproteins show homology to reported sequences from the other human herpesviruses. Since some of these proteins can induce high titres of antibodies, they can be used to develop more defined and perhaps more sensitive and specific methods for serodiagnosis of previous exposure to HCMV. The recognition by human sera of HCMV polypeptides expressed in a prokaryotic system points the way for the development of such diagnostic tools.

**Glycoproteins**

The identification of virion glycoproteins and their coding regions is of particular interest for the investigation of the immune response against HCMV. Subunit vaccines, based on the antigenicity of surface glycoproteins, should be an attractive alternative to the development of conventional HCMV vaccines. A number of authors have described numerous glycoproteins using metabolic label procedures and/or immunoprecipitation with human sera or MAbs. Investigation of HCMV-specific glycoproteins is hampered by the low abundance of these polypeptides in virions or infected cells and the formation of complexes of various sizes. Several disulphide-linked complexes have been identified in the envelope of HCMV virions or in dense bodies (Britt, 1984; Farrar & Greenaway, 1986; Gönczöl et al., 1986; Kari et al., 1986; Law et al., 1985; Rasmussen et al., 1985a). Recently Gretch et al. (1988a) analysed glycoprotein complexes via fractionation in sucrose density gradients followed by immunoprecipitation with MAbs. Three distinct families, designated gcI, gcII and gcIII were observed. The number of the complexes and their sizes as well as their constituents are summarized in Table 2. The complexes can be separated into their individual components by reducing agents, indicating that they are held together by disulphide bonds. The three glycoprotein families are likely to be important targets of the human immune response to HCMV infection. MAbs as well as monospecific sera specific for the gcI family are capable of neutralizing HCMV in the presence or absence of complement (Britt, 1984; Cranage et al., 1986; Kari et al., 1986; Rasmussen et al., 1985a, 1988; Britt et al., 1988). MAbs that recognize gp86 do not require complement for neutralization (Kari et al., 1986; Rasmussen et al., 1984). Both healthy HCMV-seropositive adults as well as congenitally infected children are able to mount a B cell response to proteins of the gcI and gcII complex (Liu et al., 1988). However when the T and B cell responses to the envelope glycoproteins were compared no correlation was found between antibody production and lymphocyte proliferation, suggesting that HCMV proteins other than gcI and gcII are important in stimulating T cell proliferative responses. One antigenic protein eliciting a proliferative response has been identified as the 64K matrix protein (pp65) (Forman et al., 1985).

The gene encoding gp58 was mapped to the right end of the HindIII F fragment of HCMV AD169 (Mach et al., 1986). The nucleotide sequence revealed that the ORF has the capacity to encode a polypeptide of 102K (Cranage et al., 1986). The gp58 is derived from a glycosylated precursor molecule of approximately 160K via proteolytic cleavage (Britt & Auger, 1986a) and
Table 2. Glycoprotein complexes: three families, gcI, gcII and gcIII

<table>
<thead>
<tr>
<th>Complex*</th>
<th>$M_r$ (× 10^-3)</th>
<th>Designated name in this review</th>
</tr>
</thead>
<tbody>
<tr>
<td>gcI</td>
<td>250–300, 190, 160</td>
<td>Reduced gp58</td>
</tr>
<tr>
<td>gcII</td>
<td>93–300</td>
<td>Reduced gp86</td>
</tr>
<tr>
<td>gcIII</td>
<td>240</td>
<td>Reduced gp58</td>
</tr>
</tbody>
</table>

*Gretch et al. (1988a).

represents the carboxy-terminal part of the precursor polypeptide (Mach et al., 1986). No information is available regarding the fate of the amino-terminal part of the protein or the site of processing. However a potential cleavage site can be deduced from sequence comparison with homologous proteins. The predicted amino acid sequence of the gp58 ORF shows homology with that of glycoprotein B of HSV, the BALF-4 product of Epstein–Barr virus (EBV), gpII of VZV (Cranage et al., 1986) and gII of pseudorabies virus (PRV) (Robbins et al., 1987). So far proteolytic processing has been shown to occur for the respective polypeptides of VZV, PRV and HCMV. From the amino acid sequence of the amino-terminal part of gpII, Keller et al. (1986) identified the cleavage site to be RSRR/Saa433. The counterparts in PRV and HCMV contain a homologous motif near the calculated cleavage sites (RARR/S503 in gII, RTRR/S460 in gp58). Recently the sequence TKR/S460 has been identified as the processing site in HCMV strain Towne (Spaete et al., 1988). No such sequence is present in glycoprotein B of HSV. With the information available at the present time, it cannot be excluded that the BALF-4 gene product of EBV is also cleaved. A 125K protein has been identified which originates from this reading frame (Emini et al., 1987). However, a potential cleavage site (RRRR/D431), is present in the molecule and moreover, a polypeptide of 56K has been identified in the envelope of the virus (Edson & Thorley-Lawson, 1981).

The gene encoding gp86 is localized within the HindIII L fragment of strain AD169 (Cranage et al., 1988). Again, this protein has homologous counterparts in the other human herpesviruses, namely glycoprotein H of HSV, gpIII of VZV and the BXLF2 gene product of EBV (Cranage et al., 1988). However the homology is very weak and restricted to the carboxy-terminal parts of the molecules. Both gp58 and gp86 were expressed in vaccinia virus (Cranage et al., 1986, 1988; Britt et al., 1988). The gp58 gene products could be identified in recombinant vaccinia virus-infected cells by immunoprecipitation with murine MAbs against gp58. The properties of HCMV gp58 expressed via recombinant vaccinia virus were indistinguishable from gp58 synthesized in HCMV-infected fibroblasts; in particular the proteolytic processing steps were observed. Laboratory animals immunized with the vaccinia virus recombinant expressing gp58 developed neutralizing antibodies. Expression of gp86 in vaccinia virus resulted in a protein which is different in some aspects from gp86 in HCMV-infected cells. The electrophoretic mobility of the vaccinia virus product was slightly greater than the HCMV-associated gp86 and, more importantly, the protein is localized at the nuclear membrane rather than on the surface of the cells (Cranage et al., 1988).

It has been shown that HCMV from urine specimens binds $\beta_2$ microglobulin (McKeating et al., 1987; Grundy et al., 1987a). The binding enhances the infectivity of the virus and enables class I HLA molecules to be used as a virus receptor (Grundy et al., 1987b). The virus itself encodes a potential glycoprotein which shows homology to several class I antigens (Beck & Barrell, 1988). The gene is located within the HindIII O fragment and its structure resembles a pseudogene rather than the architecture of cellular HLA coding sequences, which are composed of eight exons. The HCMV class I homologue is 1104 bp in length and does not contain any suitable GT/AG splicing signals. This is comparable with the average length of about 1100 bp for known class I exons. The similarities between the hypothetical HCMV class I protein and various cellular class I molecules are in the 20% range. These findings could provide an explanation for the binding of $\beta_2$ microglobulin and would support the findings of Grundy and colleagues. However the mechanism of HCMV binding to $\beta_2$ microglobulin remains speculative, unless it can be proved that the HCMV class I gene product is located in the envelope of infectious virus and is binding to $\beta_2$ microglobulin.
Proteins from the US region

Within the short unique region, Weston & Barrell (1986) have identified at least 38 regions that may code for proteins. The number and distribution of polyadenylation signals in relation to the ORFs indicate that many of the mRNAs generated will have 3' coterminal ends, a phenomenon common to all herpesviruses (McGeoch et al., 1985). Despite the similarities in structural organization, the HCMV short unique region is not related at the protein level to the US components of other human herpesviruses. Homology studies of the 38 reading frames in US and TRs/IRs indicated that 24 fall into five families, a phenomenon not encountered in the other human herpesviruses. Hydrophilicity plots of the predicted protein products of US indicate that about half have hydrophobic profiles similar to those of known membrane proteins. These proteins are of two types. Eight proteins have hydrophobic regions at the amino and carboxy termini, which could be involved in membrane translocation and anchorage and also have potential sites for N-linked glycosylation. Therefore these proteins may represent membrane-anchored glycoproteins. The Mr of the primary translation products is in the range of 21K to 27K. Glycoproteins in a similar size range have been identified in infected cell extracts by a number of groups (Pereira et al., 1984; Stinski, 1977). The second type of proteins are polypeptides with several (usually seven) highly hydrophobic putative transmembrane regions. These polypeptides, ranging in Mr from 26K to 41K, may therefore represent integrated membrane proteins. The fact that most proteins encoded by US resemble structural proteins is consistent with the classification of transcripts from this area as late (McDonough & Spector, 1983; McDonough et al., 1985).

Of particular interest is the HXLF (HindIII X left reading frame) gene family, a group of five genes which share regions of homology and are arranged in tandem. Gretch et al. (1988b) showed that at least two of these genes encode a glycoprotein(s) of Mr 47 to 52K (the gclII family). A characteristic feature of this glycoprotein is the diffuse migration in SDS–PAGE. With the methods used (SP6 polymerase-directed synthesis of RNA, translation in rabbit reticulocyte lysates and immunoprecipitation with MAbs) only products of the reading frames HXLF1 and HXLF2 could be detected. Whether the remaining genes are transcribed and translated in vivo remains open to question. Britt & Auger (1985) also described a virion envelope protein, which migrates highly diffusely in polyacrylamide gels, of estimated Mr 52K to 65K. It is possible that the gp47–52 and the 52K to 65K envelope protein are the same species, especially as MAbs against both proteins are capable of neutralizing virus in tissue cultures (Britt & Auger, 1985; Kari et al., 1986).

DIVERSITY OF GENE PRODUCTS FROM HUMAN HERPESVIRUSES

On the basis of DNA hybridization studies, it seems that the human herpesviruses have diverged significantly at the DNA level. With the exception of HSV types 1 and 2, which are closely related, the DNAs of the herpesviruses show only 5% homology to each other determined by conventional hybridization assays (Huang & Pagano, 1974). However DNA sequence data of EBV, VZV and HSV have revealed that the different herpesviruses possess a number of gene products with amino acid homology and in some areas have a similar gene layout. The discovery of similarities in gene arrangement and gene products suggests that proteins performing similar functions have been conserved during evolution of the three subfamilies of alpha-, beta- and gammaherpesviruses. Unfortunately the characterization of the HCMV genome on the DNA sequence level has lagged behind that of HSV (McGeoch et al., 1988), VZV (Davison & Scott, 1986) and EBV (Baer et al., 1984) whose complete DNA sequences have been established. HCMV sequence data reported so far comprise the short unique region (Weston & Barrell, 1986) and the 20 kb HindIII F fragment (Kouzarides et al., 1987a). Whereas no homology has been found between the US regions of HCMV and those of the other herpesviruses (see above) some interesting relationships in terms of conserved gene products and gene organization were discovered within HindIII F (Kouzarides et al., 1987b). The HindIII F fragment contains nine potential coding regions. Four of these show homology to regions encoding characterized proteins of the other human herpesviruses: namely, DNA polymerase, the gp58 equivalent, a protein which has been termed ICP18.5 in HSV-1 and which is believed to affect the transport
of the viral glycoproteins (Pellet et al., 1986) and the major DNA binding protein. The genes are arranged in a collinear fashion within a block that is perfectly conserved in EBV (see Fig. 4). With the exception of the DNA polymerase gene, this gene arrangement is conserved in the other herpesviruses. However, the DNA polymerase genes of HSV and VZV are inverted in their relative position and are located four genes away from the collinear position. Directly downstream of the HCMV polymerase is a block of five genes that show homology to reading frames in EBV. Once again, the five HCMV frames have the same relative organization as their EBV counterparts, but according to the prototype orientation of the EBV genome, the EBV frames are inverted relative to those of HCMV. In addition, the two blocks are separated by 92 kb on the EBV genome. The same relative arrangement of reading frames occurs in VZV (Davison & Taylor, 1987). The events leading to this long range rearrangement are unclear at present. On the other hand, there are a number of gene products from the U + regions of HCMV that do not show a detectable homology to other herpesviruses; for instance, all the HCMV-specific phosphoproteins identified so far. However, it is still possible that these genes are located within conserved gene blocks and that functional homologies are conserved which are not recognized in the amino acid sequences. Davison & Taylor (1987) have proposed such a homology for counterparts of EBV and VZV proteins that do not show amino acid homology but a conserved gene location and similar predicted secondary structure.

The relationships between the human herpesviruses on the genomic level provide evidence that these viruses diverged from an ancestral herpesvirus. It is clear that HCMV, HSV, EBV and VZV share a number of conserved genes that were probably part of the progenitor’s genetic make-up.

**Diagnostic Procedures**

An active HCMV infection resulting from a primary or reactivated latent infection cannot be diagnosed on clinical symptoms alone because the symptoms may range from the lack of an overt illness to severe organ damage. The infection in the adult may be asymptomatic; or various syndromes, including mononucleosis, hepatitis or pneumonitis, may occur. In the immunocompromised patient the clinical consequences of an HCMV infection are often life-threatening.

For the diagnosis of an active infection, viral isolation is usually required. Before cell culture was established HCMV diagnosis was restricted to histopathological demonstration of characteristic large cells with intranuclear and cytoplasmic inclusions. However, standard histology may be negative in a viruric patient. For isolation of HCMV in cell cultures, human
fibroblasts are the best supporters for growth of the virus. These cultures include those prepared from human embryonic tissue, foreskins and serially passaged diploid human foetal lung cells. Cells inoculated with patient material containing virus show a c.p.e. between a few days and several weeks, probably depending on the concentration of infectious virus in the specimen. Initially, foci of enlarged, rounded cells appear on the fibroblast monolayer. The c.p.e. of HCMV is sufficiently characteristic for the identification of the virus without further specific
confirmation. Viral antigens may be demonstrated from the infected cell culture by protein gel electrophoresis, immunofluorescence or electron microscopy. Detection of HCMV in cell culture is time-consuming and can be improved by modifying the culture procedures. Methods have been described for more rapid and specific detection of HCMV antigens in cell culture. These procedures allow the detection of viral antigens by immunofluorescence or immunoperoxidase-staining within 16 to 48 h (Espy & Smith, 1987; Sutherland et al., 1987; Swenson & Kaplan, 1985). Usually MAbs against IE or early proteins are used which allow the detection of viral antigens in infected cells before c.p.e. occurs. In addition, centrifugation of monolayers during inoculation might enhance viral infectivity (Thiele et al., 1987). These assays in combination with MAbs offer a rapid and specific method for diagnostic laboratories. Each new method must be calibrated against established assays and should be tested for applicability and reproducibility to a variety of HCMV strains. For example, the enhancement of infectivity by centrifugation, can be influenced by the growth state of the fibroblasts and the viral strains used and their adaptation to in vitro conditions. MAbs have the potential pitfall of not recognizing the respective antigen in a wide variety of HCMV strains. For example, two of 21 CMV strains were not recognized by a MAb against the major IE viral protein (Chou & Scott, 1988). Despite this potential drawback, MAbs could serve as powerful tools for the rapid diagnosis of active HCMV infection (Emanuel et al., 1987). Fig. 5(a) shows an immunofluorescence of human foreskin fibroblasts (HFF) 24 h after inoculation with HCMV strain AD169, using a commercially available MAb against an early protein (DuPont). Detection of viral antigen in the cytoplasm of infected HFF 72 h after inoculation is shown in Fig. 5(b). The antibody used in this example was a monospecific rabbit antiserum against a recombinant protein expressing an immunogenic portion of the large viral phosphorylated matrix protein pp150 (Jahn et al., 1987b;
Oligopeptides, or monospecific antisera against recombinant proteins were used for these transmission. This is a powerful tool for determining the molecular epidemiology of HCMV. Endonuclease digestion of viral DNA offers a method of identifying patterns of HCMV hybridization techniques provide a useful method for rapid viral diagnosis and for combination with in situ hybridization of viral nucleic acid. With this method, it was possible to describe more specifically the HCMV encephalitis and the spread of HCMV in the infected brain. Although encephalitis due to HCMV had been reported before, it was not possible to isolate the virus from the cerebrospinal fluid or brain tissue. The detection of viral antigens in tissues by immunohistochemistry provides a very helpful tool for analysing the spread of the virus in various organs and may provide better insight into the pathogenesis of HCMV. Evidence for active or silent HCMV infection can also be obtained by in situ hybridization. A number of authors described this technique for the detection of viral DNA or RNA in tissue specimens. With this method, it was possible to detect viral nucleic acid in cytomegalic cells and in cells which did not show morphological changes. The identification of occult HCMV infection may be important to improve our understanding of tissue specificity and pathogenesis of HCMV.

The dot blot hybridization has been demonstrated to be a reliable method for testing clinical specimens for the presence of HCMV. Southern blot hybridizations are more sensitive and allow more detailed studies on HCMV infection. It was possible to show that women attending a clinic for sexually transmitted diseases as well as patients with AIDS might have infection with more than one HCMV strain. In general, the hybridization techniques provide a useful method for rapid viral diagnosis and for quantification of viral shedding. The comparison of HCMV isolates via restriction endonuclease digestion of viral DNA offers a method of identifying patterns of HCMV transmission. This is a powerful tool for determining the molecular epidemiology of HCMV infection. Epidemiologically related strains such as those from congenitally infected infants and from their mothers can be analysed. By this method it could be shown that endogenous HCMV appears to be the most frequent source of recurrent infection and also that exogenous reinfection occurs, but less commonly.

Specific HCMV antibodies can be demonstrated by several test systems. The complement fixation test (CFT) has been commonly used for determining the titre of HCMV antibody. In comparison to other tests, the CFT is less sensitive and the CF titre takes longer to give a positive result than corresponding titres in other test systems. However HCMV viruria may occur without the presence or increase of the CF titre. It is recommended, that HCMV serology should include a more sensitive assay, such as the ELISA. These assays may be adapted to quantify specific antibodies of each immunoglobulin class. However, IgM assays may give false positive results and the absence of detectable HCMV IgM antibodies during HCMV infection in patients with AIDS or in newborns is a well described phenomenon. Other serological tests, not commonly used, include the indirect immunofluorescence, the radioimmunoassay, the indirect haemagglutination assay, the immune adherence haemagglutination assay and the neutralization test. The neutralization test may be applicable for showing strain specificity. However these latter tests are tedious and not routinely used. The virus strain from the patient is preferable, but even if the homologous virus is on hand, cultivation for weeks may be required before sufficient quantities of cell-free infectious virus can be produced. Culture-adapted HCMV strains are usually prepared for the determination of neutralizing antibody titres.

HCMV antigens for use as reagents in antibody test systems are usually prepared from a
Review: Molecular biology of HCMV

laboratory-adapted strain such as AD169. These cell extracts do not represent well-defined antigen preparations and could include antigens which show cross-reactivity with other human herpesviruses. Improved HCMV serological test systems may be based on highly specific immunogenic proteins without homologous counterparts in other herpesviruses. One approach is the expression of immunogenic epitopes and the establishment of recombinant ELISAs for the detection of HCMV antibodies.

FUTURE PERSPECTIVES

Investigation of the structure of the IE genes of HCMV and characterization of factors involved in the regulation of these genetic elements have considerably improved our understanding of how transcription of herpesvirus genes might be regulated. The participation of cellular factors in these processes makes this viral system an attractive model to study gene regulation in general.

The identification and characterization of immunologically important structural phospho- and glycoproteins and their coding regions will provide tools to define antigenic domains of the virus such as epitopes involved in induction of neutralizing antibodies or regions of proteins which could be used as antigens in serological tests. Considering the size of the genome and the little information on structural proteins available so far this is certainly an area where a lot of work needs to be done.

Aspects of cellular immunity were not covered in this article; however this is also an area of research where interesting data have been published recently, especially in the mouse system. The completion of the HCMV DNA sequence will certainly be a milestone in HCMV research and will stimulate many further studies.

Work from this laboratory was supported by the Deutsche Forschungsgemeinschaft and the Wilhelm Sander-Stiftung.

REFERENCES


Review: Molecular biology of HCMV


