Human cytomegalovirus origin of DNA replication (oriLyt) resides within a highly complex repetitive region

(herpesvirus/electroporation/DNA repeats/betaherpesvirus)

MARIE J. O. MASSE[†], SAMUEL KARLIN[‡], GABRIEL A. SCHACHTEL^{‡¶}, AND EDWARD S. MOCARSKI^{†§}

[†]Department of Microbiology and Immunology, and [‡]Department of Mathematics, Stanford University, Stanford, CA 94305-5402

Contributed by Samuel Karlin, March 4, 1992

ABSTRACT A global analysis of the 230-kilobase-pair (kbp) human cytomegalovirus genome revealed three regions that were very rich in repeated sequences. The region with the highest content of inverted and direct repeats lies between 92,100 and 93,500 bp, upstream of the gene encoding the single-stranded DNA binding protein. Cloned restriction fragments containing this region were able to replicate when trans-acting factors were provided by virus infection in a transient replication assay. With this assay, the region between 92,210 and 93,715 bp on the viral genome was defined as the minimal replication origin, oriLyt. The sequence composition and repeats within oriLyt were used to divide the region into two domains that may be important in origin function. Sequences flanking either the left or right side of the minimal oriLyt contributed to efficient replication; however, these sequences were not essential for origin function. Thus, the region of the viral genome with the most striking concentration of direct and inverted repeats corresponds to the oriLyt of human cytomegalovirus.

Herpesvirus genomes carry two distinct classes of DNA replication origins—one for productive viral growth (1) and one that maintains the viral genome during latency (2). Lytic replication origins (ori_S and ori_L) were first characterized in herpes simplex virus (HSV; refs. 3–5). In this and other alphaherpesviruses, lytic replication origins are characterized by two conserved elements: a binding site for the viral origin-specific DNA binding protein and an adjacent A+T-rich region (6–8). Although this type of origin is relatively simple, sequences flanking the core elements apparently contribute to the efficiency of replication (9). Replication origins, including those of adenoviruses, papovaviruses, and Epstein–Barr virus (EBV; refs. 10–15), commonly have binding sites for critical viral proteins and an A+T-rich sequence where initial unwinding is either known or suspected to occur. Structurally, these origins consist of arrays of direct repeat (DR) and inverted repeat (IR) elements that may either be essential core elements or play a stimulatory role as auxiliary components in origin function. Some of these elements are sites of interaction with sequence-specific DNA binding proteins and may allow for redundancy and cooperativity in protein–DNA interactions. Such arrays of repeat elements are often components of transcriptional enhancers that stimulate replication in cis (10).

The gammaherpesvirus EBV (15) and the betaherpesvirus simian cytomegalovirus (CMV) (16) carry lytic replication origin sequences that are among the most structurally complex known. For example, the EBV lytic replication origin (oriLyt) is bipartite, stretching over 1 kilobase pair (kbp). One part contains a series of IR and palindromic elements positioned adjacent to an A+T-rich sequence and the second part con-

tains arrays of IR and DR elements. These arrays function in a manner such that it may be replaced in part by a heterologous transcriptional enhancer (15). In simian CMV, an oriLyt has been mapped upstream of the single-stranded DNA binding protein (ssDBP) gene on the viral genome, spanning >1.3 kbp and composed of three domains replete with IR and DR elements and at least one A+T-rich sequence (16). Preliminary evidence using a transient replication assay (16) and ganciclovir-induced chain termination (17) has suggested that human CMV oriLyt maps to a region of the viral genome positionally analogous to oriLyt of simian CMV and ori_L of HSV.

Based on the studies with EBV and simian CMV oriLyt, we predicted that the human CMV oriLyt would be structurally complex, and we have performed a global analysis of repeat frequency of the entire CMV genome (Fig. 1). Using the statistical methods described by Karlin (18), we found the region between 92,100 and 93,500 bp to be remarkably rich in closely spaced DR and IR elements. After establishing a highly sensitive transient replication assay, this structurally complex region was found to contain the cis signals required for recognition as an origin for lytic replication.

MATERIALS AND METHODS

Cells, Virus, and Replication Assay. Human foreskin fibroblast cells and CMV (AD169) were propagated as described (19). Electroporation methods (20) were modified to achieve 50% efficiency as follows: 5×10^6 cells suspended in Opti-Medium (GIBCO/BRL) were electroporated (24°C, 220 V; 996 μ F) with test (10–40 μ g) and control (pKS or pMT11; 1–2 μ g) DNA in a final vol of 400 μ l. After 24–36 hr, half of the cell cultures were infected with CMV (AD169) at a multiplicity of infection of 2–5 plaque-forming units per cell and were incubated at 37°C until 80% cytopathic effect was observed (5–7 days). DNA was extracted from pelleted cells as described (21) and digested overnight with *Hind*III in the presence of RNase A and in the presence or absence of *Dpn* I (New England Biolabs; ref. 22). The digested DNAs were electrophoretically separated, transferred to nylon (Hybond N+, Amersham), hybridized at 65°C (23, 24) with ³²P-labeled pBR322 (25), and autoradiographed on Kodak XAR film. All plasmids were tested at least three times.

Plasmid Constructs. All plasmid clones were constructed in pGEM-3Zf(+) (Promega) as shown in Fig. 2A. pON2400 was derived from CMV (strain AD169) plasmid clone pCM1029 (27) by *Cla* I digestion and ligation. Plasmids pON2401,

Abbreviations: CMV, cytomegalovirus; HSV, herpes simplex virus; EBV, Epstein–Barr virus; oriLyt, lytic origin of replication; DR, direct repeat; IR, inverted repeat; ssDBP, single-stranded DNA binding protein; ORF, open reading frame.

[¶]Present address: Mathematics Division, University of Texas, San Antonio, TX 78285.

[§]To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

FIG. 1. Global analysis of inverted repeat content of the human CMV genome. Top line shows the CMV genome with inverted repeats of the L (*b* and *b'*) and S (*c* and *c'*) components as solid boxes. The *a* sequences are shown as small open boxes at the genome termini and the L-S junction. *U_L* and *U_S* unique sequences are denoted by a thin line. Middle line shows a *Hind*III restriction map of human CMV (AD169). Graph at bottom represents frequency of inverted repeats (≥ 8 bp long, separated by a distance of ≤ 150 bp, with no mismatch allowed) scored by using a sliding window of 1 kbp. The number of IRs summed over 1-kbp segments is plotted versus position on the genome.

pON2402, and pON2403 were made from *Kpn* I, *Sac* I/*Kpn* I, and *Sac* I fragments, respectively, isolated from pON2400. pON2404 was derived by *Dra* I digestion of pON2401. pON2405 was derived by *Dra* I digestion and pON2406 was derived by *Sal* I digestion of pON2403. pON2407, pON2408,

pON2409, pON2410, and pON2411 were made from pON2404 using *Pvu* I, *Bss*HII, *Mlu* I, *Bam*HI, and *Eco*RI, respectively. Constructs were characterized by dideoxynucleotide sequencing (28) with Sequenase (United States Biochemical).

FIG. 2. Mapping oriLyt. (A) Restriction map of fragment *Hind*III/*Cla* I from within *Hind*IIIID based on the sequence of CMV (AD169) (26) with the *Hind*III (F/D) and *Eco*RI (V/I) sites indicated. Region defined in this study as being rich in IR (Figs. 1 and 3) is shown by a hatched box ≈ 1.6 kbp upstream of the ssDBP open reading frame (ORF) (open arrow). Below the restriction map, a series of plasmid clones used to define oriLyt are depicted by brackets. Right column summarizes the results obtained with these clones in the transient replication assay. Replication is indicated as follows: ++, readily detected replication; +, fair replication; +/-, poor replication; -, no replication. Minimal oriLyt defined in this study is shown by stippled box. (B) Autoradiograms resulted from transient replication assays. Positions of *Hind*III linearized test plasmids are indicated by solid arrows. *Hind*III-linearized control plasmids pKS and pMT11, included to ensure complete *Dpn* I digestion, are indicated by the hatched arrows. Products resulting from *Dpn* I digestion of methylated plasmid migrated in the bracketed region. Although a complete analysis was performed for all tested plasmids, it is shown only for pON2404 (lanes 7-11) and included the following: linearized plasmid DNA as a size marker (lane 7), DNA extracted from transfected/uninfected cells (lanes 8 and 11), restricted (lane 11) or not (lane 8), by *Dpn* I, and transfected/CMV infected cells (lanes 9 and 10). For pON2400, pON2402, pON2403, pON2405, pON2406, pON2407, and pON2408 (lanes 1-6 and 12-19), only DNAs isolated from transfected/CMV-infected cells are shown linearized with *Hind*III in the presence or absence of *Dpn* I. The amount of transfected test plasmid was 10 μ g (lanes 1-4, 18, and 19), 20 μ g (lanes 5, 6, 16, and 17), 30 μ g (lanes 7-11, 14, and 15), and 40 μ g (lanes 12 and 13). Size markers are shown on the right and apply only to lanes 5-19. pON2400 and pON2402 are 13 and 8.8 kbp, and *Dpn* I digestion controls (hatched arrow) pMT11 and pKS are 2.3 and 2.9 kbp.

RESULTS

Repeat Frequency Analysis. Three rigorous statistical methods were used to assess frequency and complexity of repeats across the human CMV genome sequence: (i) all occurrences of perfect IRs 8 bp or longer within 150-bp segments using a 1-kb sliding window were identified; (ii) all long global and local occurrences of IR and DR were compared with repeats generated in random sequences of the same base composition as a control for statistical significance (29); and (iii) all arrangements of possible exact palindromes (10 bp or longer) were compared to determine the statistical chance of the observed ordering of palindromes. The three methods independently identified the same region between 92,100 and 93,500 bp as having the greatest density and complexity of repeats. In addition, two previously identified transcriptional enhancers [at 174,000–175,000 (30–32) and 195,000–196,000 bp (33)] also contained significant numbers of repeats. According to method *i*, the highest concentration of IRs was within the region between 92,000 and 94,000 bp (Fig. 1). Furthermore, method *ii* revealed that this region included different 29- or 30-bp DRs flanked by overlapping IRs plus four 32-bp closely spaced DRs with internal IRs (Table 1), some of which have been previously noted (17). Finally, analysis *iii* showed this region to have the longest highly significant run of an exact palindrome 10 bp or longer (without other intervening 10-bp palindromes), a 10-bp palindrome GCGCATGCGC at bp 92,570, 92,643, 92,709, 92,747, 92,783, and 92,859. Except for the two enhancer regions (TTGACGTCAG occurs at bp 173,863, 174,049, 174,132, and 174,185; TGTCGCGACA appears at bp 195,112, 195,151, 195,221, and 195,262), no other such runs of exact palindromes were found. As a statistical

control, it should be noted that a random rearrangement of the observed 10-bp palindromes would only occur as a run of four or more in a genome the size of CMV with probability $\leq 10^{-5}$ and, for comparison, a similar analysis for significant runs of palindromes in the EBV genome (34) reveals only two regions: oriP and oriLyt (18).

Definition of the Minimal Human CMV oriLyt. When compared to the CMV genome map (26), the region richest in repeats was found to lie upstream of the ORF encoding the ssDBP (UL57) and was previously predicted to include two ORFs, UL60 and UL61. To test whether the region between bp 92,000 and 94,000 contained or was part of oriLyt, we derived the series of plasmid clones shown in Fig. 2A and tested their ability to replicate in a transient replication assay. Every experiment included a control plasmid that lacked oriLyt as a monitor for complete digestion of input methylated DNA by *Dpn* I. The resulting autoradiograms are shown in Fig. 2B. All replication of the test plasmids was dependent on CMV infection such that test plasmids were fully digested by *Dpn* I when introduced into cells that were left uninfected (Fig. 2B, compare lanes 8 and 11).

We first confirmed the observation of Anders and Puntieri (16) and showed that CMV (AD169) sequences in cosmid pCM1029 (27), which carries *Hind*III and -F, replicated when introduced into cells that were subsequently infected with CMV (data not shown). As shown in Fig. 2, the test plasmid pON2400 (the leftmost 12-kbp *Hind*III/*Cla* I fragment from this *Hind*III) and the series of plasmid clones derived from this *Hind*III/*Cla* I fragment, including pON2402 (*Sac* I/*Kpn* I; 5.5 kbp) and pON2403 (*Sac* I/*Sac* I; 4.4 kbp), replicated as indicated by test plasmid resistance to *Dpn* I digestion. Plasmid pON2403 replicated less efficiently than

Table 1. Repetitive sequences in minimal oriLyt

Repeat	Sequence	Length [†]	Coord [‡]	Dist [§]
Domain I				
IR1	ACCCCTGACAAgCC	14(1)	92,145	437
DR1	TGACGTAGCGAGCGtAGCGAGCTACGTCA	29(1)	92,227	38
IR2*	CATACGgTGACGTAGC	16(1)	92,220	11
IR2	CTGATGACGTAGC	13(0)	92,290	11
IR3	CAGTCCGTTTTaCGTA	16(1)	92,324	95
IR4	GAtGCTAGGCgaCGCCCC	19(3)	92,344	3
IR5	TAGGGGcGGA-GCcT	14(3)	92,363	0
DR2	CCGTGCGCATGCGCaG	16(1)	92,566	54
	gaGTGCGCATGCGCCG	16(2)	92,639	47
	CgGgGCGCATGCGCgG	16(3)	92,705	22
DR2*	CCcTGCGCATGCGCCGGTAAAAATTCCTACTGGG	32(1)	92,746	4
	tCGTGCGCATGCGCCaGTAAtttTTCCTACTaGa	32(7)	92,782	6
	CaGTGCGCATGCTCCGGTAAAAATTCCTACTaGa	32(4)	92,820	6
	CCGTGCGCATGCGCCGGTAtttTTCCTACTGGG	32(3)	92,858	
IR6	Underlined in DR2 series			
DR3	GAAAACCTATAtA	13(1)	92,668	294
Domain II				
IR7	CCCCGGCCGAGGcCCGG	17(1)	93,097	80
IR8	GCCCCtCCCACcGCCtGGa	20(3)	93,131	6
DR4	CACCCtCCcCGGCCGaGGcCCGGcg-CCGG	30(8)	93,091	32
IR9	CCCCCGCTCGa--CCCCC	19(3)	93,223	62
IR10	GGCC-CGGcCGGgct-GGGACCCC	25(6)	93,248	6
IR11	CCGGGGTCCGCG-CCcttCCCCtCC	24(4)	93,427	6

IR and DR sequences within domains I and II of minimal oriLyt as depicted in Fig. 3 and discussed in the text are shown. For each set of repeats, the nucleotide sequence of only the first copy of a DR or left arm of an IR is presented. Mismatches are indicated by lowercase letters and gaps are indicated by hyphens. For DR2 and DR2*, the sequence of each copy is presented and lowercase letters indicate mismatch to consensus. Underlining is used to emphasize shared sequences between different sets of repeats.

[†]Length (bp) of repeat with number of mismatched nucleotides (or gaps) indicated in parentheses. For DR2 and DR2*, the mismatch is relative to consensus.

[‡]Map coordinate of first nucleotide of first copy of each repeat on the CMV (AD169) genome. For DR2 and DR2*, the position of the first nucleotide of each copy is given.

[§]Distance (bp) between last nucleotide of first repeat and first nucleotide of second repeat. Terms used by Hamzeh *et al.* (17) in describing repeats were as follows: DR1, 28-bp repeat; IR4, 38-bp palindrome; IR5, 28-bp palindrome; DR2, 19-bp repeat; IR6, *Fsp* I/*Sph* I repeat; DR3, 13-bp repeat.

pON2402 and pON2405 (*Dra* I/*Sac* I; 3.6 kbp) replicated less efficiently than pON2404 (*Dra* I/*Kpn* I; 2.6 kbp), establishing that test plasmids terminating at the *Sac* I site (bp 93,715) replicated ≈ 10 times less efficiently than those terminating at the *Kpn* I site at bp 94,861. Consistent with this, pON2406 (*Sal* I/*Sac* I; 2.4 kbp) replicated with the same low efficiency as pON2405. Plasmids that carried sequences extending to the *Kpn* I site at 94,861 bp, such as pON2407 (3.3 kbp), replicated efficiently; however, deletion of an additional 255 bp to the *Bss*HII site reduced replication efficiency at least 10-fold. Further deletion of sequences to the *Mlu* I site at 92,018 bp (pON2409) or the *Eco*RI site at 92,210 bp (pON2411) resulted in even lower levels of replication even when 30 μ g of plasmid DNA was transfected (data not shown). Further analysis determined the right limit of oriLyt using pON2410, a plasmid that carried a 354-bp deletion relative to pON2405. pON2410 was never observed to replicate in multiple attempts, even when up to 65 μ g of this plasmid was introduced into cells (data not shown). Thus, minimal oriLyt mapped between 92,210 and 93,715 bp (the *Eco*RI and *Sac* I sites) on the viral genome. This region did not replicate well in the absence of at least 700 bp of rightward or 1100 bp of leftward flanking sequences. The coordinates of minimal oriLyt overlapped substantially with the IR-rich region, which when analyzed in detail lies between 92,100 and 93,500 bp on the viral genome (Fig. 3).

Two lines of evidence showed that the test plasmids replicated autonomously and did not integrate into viral DNA. First, when linearized by *Hind*III, all tested clones displayed a single band with the same size as the input test plasmid, as illustrated for pON2404 (Fig. 2B, lanes 7–10). Second, the DNA extracted from cells transfected with pON2404 and digested with *Cla* I, which did not cut pON2404, migrated as a high molecular weight species that was resistant to *Dpn* I (data not shown), consistent with a rolling circle mechanism of DNA replication that generates concatemeric DNA as its product. Thus, all test plasmids replicated autonomously and not by integration into the viral genome.

oriLyt Structure. Structurally, the minimal oriLyt could be divided into two domains described in Table 1 and shown in Fig. 3. Domain I, from 92,100 to 93,000 bp, was found to contain two prominent features. The first centered on two 29-bp repeats (DR1), noted previously (17), but which we found to be part of a more complex structure. The sequence DR1 may be depicted as *ABA'*, with *A* (9 bp) starting at 92,227 bp (underlined in Table 1) followed by *B* (12 bp) and then *A'*, the inverted complement of *A*, all within DR1 and overlapping IR2*. This structure was found again within the second copy of DR1 at 92,294 bp after a gap of 38 bp separating the two structures. The second prominent feature of domain I was the occurrence of seven copies of an exact 10-bp

palindrome GCGCATGCGC (IR6) embedded in seven copies of a DR (DR2/DR2*; Table 1 and Fig. 3). With an error occurring in only a single copy, IR6 forms an exceptional and highly significant run of palindromes, between 92,570 and 92,872 bp with gaps between copies of 73, 55, 28, 26, 28, and 28 bp. Four of the DRs, previously referred to as 19-bp repeats (17), were 32 bp long (DR2*) and were regularly spaced with short intervening sequences. In addition, three copies 16 bp long (DR2) were similar to the righthand side of DR2* (Table 1).

This domain was also found to contain 14 arrays of at least five continuous A-T bp, interspersed with 25 arrays of at least five G-C bp. The sequence between bp 92,404 and 92,442, at 72% A+T, was found to have the highest A+T content of any region within oriLyt, but, unlike origins in other systems, this region did not include a large continuous stretch free of any C-G bp.

In addition to an extremely rich G+C content throughout (80–90%), as noted previously (26, 35), domain II (93,000–93,600 bp) had a highly complex repeat structure that had not been previously recognized. The configuration centered on the two copies of DR4 with embedded IRs that could be portrayed as a 59-bp tandem DR. The region (93,090–93,210 bp) is shown in Fig. 3 and contained IR7 and IR8, two nested symmetrically positioned dyads, and DR4, which overlaps these dyads. Together these IRs can form a highly stable (G+C-rich) stem of 54 bp including 13 mismatches and a loop of 6 bp. The organization of this large repeat was as follows: (bp 93,097) 17 bp (IR7) . . . 17-bp gap . . . 20 bp (IR8) . . . 6-bp loop . . . 20 bp (IR8) . . . 17-bp gap . . . 17 bp (IR7) (bp 93,194). The two copies of this 59-bp repeat (bp 93,091–93,149 and 93,152–93,210) included portions of DR4, IR7, and IR8. This repeat was composed of a more conserved region, followed by a divergent sequence of 13 bp and a moderately conserved 17-bp sequence (11/17 matches). This 59-bp repeat was considered to be significant because its components were evenly spaced and the major portions of all stems fell within the conserved sequences of other repeat elements.

The last significant repeat in domain II was IR11. Given our observation that removal of the region including IR11 destroyed the ability of test plasmid pON2410 to replicate, it would appear that this motif was important in oriLyt function. In contrast to the complex structure of this core region, the auxiliary sequences flanking oriLyt contained significantly fewer repeats (see Fig. 1) and did not exhibit remarkable sequence structure.

DISCUSSION

Human CMV oriLyt is located within a structurally rich region upstream of the ssDBP gene positionally analogous to ori_L of HSV-1 and to oriLyt of simian CMV. Unlike these other

FIG. 3. Structural features within minimal oriLyt (92,100–93,500 bp). Minimal oriLyt was divided into two domains as depicted and percentage A+T (using a 50-bp sliding window) is shown above the sequence features. The following repeats in this region, whose sequences are given in Table 1, are diagrammed: DR1, DR2, DR3, and DR4; IR1–IR4 and IR7–IR11; and palindromes IR5 and IR6. The 59-bp repeat discussed in the text is bracketed.

origins, and in contrast to previous predictions (17), human CMV oriLyt lies more distant, upstream of that gene. As with EBV (15) and simian CMV (16), the human CMV oriLyt exhibits an elaborate sequence structure. In each of these systems, minimal oriLyt is composed of repeats; however, the precise role of individual repeats or of the overall structure is not yet clear. They may play roles as binding sites for either cellular or viral proteins important in promoting the initiation of DNA replication. The most extensive sequence identity to another herpesvirus DNA replication origin is a 53-bp stretch (bp 92,471–92,524) of human CMV oriLyt that exhibited a 72% level of identity to a large IR within the L-ori of pseudorabies virus (36), a sequence that bears little similarity to the conserved *oris/ori_L* of HSV and other alphaherpesviruses.

Simian and human CMV oriLyt do not share strong sequence identity, although they seem related in structural organization and density of repetitive motifs. The overall weak sequence similarity between these two CMVs may account for the fact that their origins cannot function interchangeably (16). The human CMV IR6 present in seven copies in human CMV is not found within simian CMV oriLyt. Two regions of sequence or structural similarity stand out. The first is an A+T-rich region located in an analogous position to the left of the region carrying the conserved motif IR6. The most A+T-rich sequence in human CMV oriLyt extends over 39 bp, contains 72% A+T, and is next to IR5 in domain I. This sequence may be analogous to a region of simian CMV that is 50 bp long. The second is the 59-bp repeat element (containing DR4, IR7, and IR8) that exhibits structural similarity to DR3 of simian CMV, having 66% overall sequence identity in this region. The underlying sequence and structural characteristics of the DR4–IR7–IR8 complex are analogous to a motif present in three copies in simian CMV, and called 47-bp DR3 (16). This simian oriLyt repeat can be extended to 65 bp (1873–1938 and 1941–1994) and the highest level of sequence similarity between the two viruses is

CTCCGtCCCTaCCCCCACCCTCCCC

within the rightmost region of their respective oriLyt.

One previously reported conserved motif, called R1 (16) or 10-bp repeats (17), does not appear to be important in human CMV oriLyt function. The twenty-five 10-bp repeat elements previously mapped to the region (17) were found in our study to be located between the ssDBP gene and oriLyt and not within the minimal oriLyt. Simian CMV oriLyt contains 16 copies of R1, and these have been implicated as important components for oriLyt function (16). The position of these repeats in human CMV between the ssDBP and the beginning of oriLyt suggests that they may play a regulatory role in ssDBP gene transcription and may possibly play a role in enhancing oriLyt function, as observed for other DNA viruses (9, 10, 15).

The efficiency of replication is enhanced by sequences flanking either side of the minimal origin, either from the *Pvu* I site (bp 91,495) to the *Eco*RI site on the left, or from the *Sac* I site to the *Kpn* I site (bp 94,861) on the right. Both flanking regions are required for maximal replication efficiency, although the sequences in these flanking regions are not homologous to one another or to sequences within minimal oriLyt. Stimulation of replication by flanking auxiliary sequences has been observed for other viral origins of replication and may be mediated through the presence of transcription signals or some other processes. Thus, we propose that the minimum oriLyt domain extends from 92,200 to 93,700 bp, while full efficiency of the oriLyt region requires sequences from 91,150 to 94,800 bp. More extensive mutagenesis is necessary; our limited deletion analyses suggest that IR11 in domain II may be critical for human CMV origin function. The role of the individual elements as sites for

interaction with sequence-specific DNA binding proteins expressed by the virus or host cell remains to be investigated.

M.J.M. was supported as a Chargée de Recherche by the Centre National de la Recherche Scientifique (France) and also in part by a grant from the Philippe Foundation. This work was supported by Grants AI20211 (to E.S.M.) and GM10452 and GM39907 (to S.K.) from the U.S. Public Health Service and by the Deutsche Forschungsgemeinschaft, Grant DFG-Scha-441/3-1 (to G.A.S.).

- Hammerschmidt, W. & Sugden, B. (1990) *Mol. Biol. Med.* **7**, 45–57.
- Middleton, T., Gahn, T. A., Martin, J. M. & Sugden, B. (1991) *Adv. Virus Res.* **40**, 19–55.
- Stow, N. D. & McMonagle, E. C. (1983) *Virology* **130**, 427–438.
- Weller, S. K., Spadaro, A., Schaffer, J. E., Murray, A. W., Maxam, A. M. & Schaffer, P. A. (1985) *Mol. Cell. Biol.* **5**, 930–942.
- Deb, S. & Doelberg, M. (1988) *J. Virol.* **62**, 2516–2519.
- Stow, N. D. & Davison, A. J. (1986) *J. Gen. Virol.* **67**, 1613–1623.
- Baumann, R. P., Yalamanchili, V. R. & O'Callaghan, D. J. (1989) *J. Virol.* **63**, 1275–1283.
- Camp, H. S., Coussens, P. M. & Silva, R. F. (1991) *J. Virol.* **65**, 6320–6324.
- Wong, S. W. & Schaffer, P. A. (1991) *J. Virol.* **65**, 2601–2611.
- DePamphilis, M. L. (1988) *Cell* **52**, 635–638.
- Villarreal, L. P. (1991) *Microbiol. Rev.* **55**, 512–542.
- Kelly, T. J. (1989) *Harvey Lect.* **85**, 173–188.
- Yang, L., Li, R., Mohr, I. J., Clark, R. & Botchan, M. R. (1991) *Nature (London)* **353**, 628–632.
- Reisman, D., Yates, J. & Sugden, B. (1985) *Mol. Cell. Biol.* **5**, 1822–1832.
- Hammerschmidt, W. & Sugden, B. (1988) *Cell* **55**, 427–433.
- Anders, D. G. & Punturieri, S. M. (1991) *J. Virol.* **65**, 931–937.
- Hamzeh, F. M., Lietman, P. S., Gibson, W. & Hayward, G. S. (1990) *J. Virol.* **64**, 6184–6195.
- Karlin, S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6915–6919.
- Spaete, R. R. & Mocarski, E. S. (1985) *J. Virol.* **56**, 135–143.
- Chu, G., Hayakawa, H. & Berg, P. (1987) *Nucleic Acids Res.* **15**, 1311–1326.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
- Peden, K. W. C., Pipas, J. M., Pearson-White, S. & Nathans, D. (1980) *Science* **209**, 1392–1396.
- Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
- Wahl, G. M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3683–3687.
- Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
- Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A. I., Kouzarides, T., Martignetti, J. A., Preddie, E., Satchwell, S. C., Tomlinson, P., Weston, K. M. & Barrell, B. G. (1990) *Curr. Top. Microbiol. Immunol.* **154**, 125–170.
- Fleckenstein, B., Muller, I. & Collins, J. (1982) *Gene* **18**, 39–46.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Karlin, S., Bucher, P., Brendel, V. & Altschul, S. F. (1991) *Annu. Rev. Biophys. Biophys. Chem.* **20**, 175–203.
- Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B. & Schaffner, W. (1985) *Cell* **41**, 521–530.
- Akrigg, A., Wilkinson, G. W. & Oram, J. D. (1985) *Virus Res.* **2**, 107–121.
- Thomsen, D. R., Stenberg, R. M., Goins, W. F. & Stinski, M. F. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 659–663.
- Weston, K. (1988) *Virology* **162**, 406–416.
- Baer, R., Bankier, A. T., Biggin, M. D., Deininger, P. L., Farrell, P. J., Gibson, T. J., Hatfull, G., Hudson, G. S., Satchwell, S. C., Seguin, C., Tuffnell, P. S. & Barrell, B. G. (1984) *Nature (London)* **310**, 207–211.
- Honess, R. W., Gompels, U. A., Barrell, B. G., Craxton, M., Cameron, K. R., Staden, R., Chang, Y. N. & Hayward, G. S. (1989) *J. Gen. Virol.* **70**, 837–855.
- Kupersmidt, S., DeMarchi, J. M., Lu, Z. Q. & Ben, P. T. (1991) *J. Virol.* **65**, 6283–6291.