Human Cytomegalovirus UL34 Binds to Multiple Sites within the Viral Genome

Ziqi Liu,^{a†} Bonita J. Biegelke^{a,b}

Molecular and Cellular Biology Program^a and Department of Biomedical Sciences,^b Ohio University Heritage College of Osteopathic Medicine, Athens, Ohio, USA

The human cytomegalovirus UL34 gene encodes a sequence-specific DNA binding protein that downregulates expression of the viral immune evasion gene US3. Analysis of the viral genome identified 14 potential UL34 binding sites. Using mobility shift experiments, UL34 bound to all predicted sites that were assayed (7 of 14). Furthermore, the UL34 binding site present within the regulatory region of the US9 gene downregulates expression in a manner similar to that seen for the US3 gene.

Although human cytomegalovirus (HCMV) infects a large percentage of adults worldwide, it is most commonly an opportunistic pathogen, causing clinically significant disease in immunosuppressed or immunocompromised patients (1). HCMV has a large double-stranded DNA genome (~235 kb) that is divided by repeat elements into unique long (UL) and short (US) regions. In cell culture, gene expression occurs in a precisely regulated pattern during the 5-day replication cycle. Intriguingly, of the ~220 genes carried by the virus, only ~25% of the viral genome is required for replication in the laboratory (2, 3). Included in the genes essential for viral replication is the UL34 gene.

UL34 is expressed from early times of infection throughout the viral replication cycle (4). Two UL34 proteins, an early and a late protein, are synthesized from the UL34 open reading frame. The two proteins differ by 21 amino-terminal amino acids that are present in the early protein but absent from the late protein. As yet, the only known function for the UL34 proteins is the downregulation of one of the viral immune evasion genes, the US3 gene. The UL34 proteins are sequence-specific DNA binding proteins that bind to an element (the UL34 binding site) in the US3 gene to repress US3 transcription (Fig. 1A) (5).

The US3 gene is located in a cluster of genes, US2 through US11 (6), that encode short glycoproteins, many of which interfere with immune recognition of virally infected cells. The US3

gene delays the processing of major histocompatibility (MHC) class I molecules through the endoplasmic reticulum, thus delaying the presentation of viral antigen and the host recognition of infected cells (see reference 7 for a review). The US3 gene is an immediate-early gene, transcribed at high levels immediately after infection (8). By 5 h postinfection, the interaction of UL34 proteins with the US3 gene results in a marked decline in the level of US3 transcripts.

Identification of multiple UL34 binding sites within the HCMV genome. The sequence-specific DNA binding activity of the UL34 proteins led us to examine the HCMV genome for other potential UL34 binding sites. Previous studies utilizing site-directed mutagenesis of the UL34 binding site in the US3 gene identified several critical nucleotides within the binding site (9). Based

Received 30 November 2012 Accepted 27 December 2012

Published ahead of print 2 January 2013

Address correspondence to Bonita J. Biegelke, Biegelke@ohio.edu.

† Deceased.

Copyright © 2013, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.03309-12

FIG 1 (A) Diagram of the US3 gene, with the sequence of the UL34 binding site shown. The positions of the nucleotides relative to the transcription start site are indicated; underlined nucleotides indicate positions where substitutions negatively impacted the function of the UL34 binding site (9). T indicates the position of the TATA box, the rectangle represents the US3 open reading frame, and the bent arrow indicates the transcription start site. (B) Diagram of the positions of the predicted UL34 binding sites within the HCMV strain AD169 genome. The gray rectangles represent repeated regions. The open rectangles indicate open reading frames or the location of the noncoding transcript (RNA 4.9 [18]) and are labeled with the name of the corresponding gene. The region identified as the origin (ori) is indicated by a black rectangle. The position of ori-lyt is indicated; the positions of US3, US9, and US11 are also indicated, with the locations of the UL34 binding sites depicted in Fig. 2 and 3. The black filled ovals represent the positions of the UL34 binding sites relative to each of the diagrammed elements.

TABLE 1 Predicted UL34 binding sites within the HCMV genome^a

Core element	Location	DNA sequence	Strand
		AAACACCGT $\frac{T}{G}$	
8,048	Between RL9 and RL10	CAAAACACCGTGACA	complementary
8,193	Between RL9 and RL10	TAAACACCGTGTTTC	complementary
18,966	Located within UL11	GGAAACACCGTTACA	
20,025	Located within UL13	CGAAACACCGTTGGT	
39,703	Located within UL31	GCAAACACCGTGTGG	complementary
41,014	Located within UL32	CAAAACACCGTGTCC	complementary
50,818	Located within UL37	CGAAACACCGTGTCC	
79,482	Located within UL54	TCAAACACCGTGGCG	
91,796	Located near <i>ori-lyt</i>	AAAAACACCGTGTTTC	complementary
92,2213	Located near <i>ori-lyt</i>	TGAAACACCGTGATG	
95,509	Located near <i>ori-lyt</i>	GTAAACACCGTTTGA	
197,217	5' of US3 ORF	CAAAACACCGTTCAG	complementary
202,929	5' of US9 ORF	CGAAACACCGTGCCT	
211,199	5' of US11 ORF	TAAACACCGTGTTTC	complementary

^a Numbers refer to the sequence of human cytomegalovirus AD169 (GenBank accession number [FJ527563.1](#)). The asterisks indicate binding sites assayed in electrophoretic mobility shift assays (Fig. 2).

on these studies, a 10-nucleotide core element that is critical for UL34-DNA interactions was identified. Using a motif search for the 10-nucleotide UL34 binding site, the HCMV genome was analyzed for other potential UL34 binding sites. Randomly, a 10-nucleotide element would appear only once within the HCMV genome. Our motif search identified an additional 13 occurrences of the UL34 binding site within the HCMV genome of strain AD169 (Fig. 1B and Table 1). Many of the predicted UL34 binding sites occur within a predicted protein-coding open reading frame; others are located outside open reading frames. Three predicted UL34 binding sites are located near the region comprising the essential components of the origin for lytic replication (Fig. 1B and Table 1) (10–12).

A search for predicted UL34 binding sites in two additional HCMV strains, Toledo (13) and VR1814 (14), identified 15 predicted binding sites. The UL34 binding sites present in the Toledo strain matched those present in strain AD169, with the following exceptions. The Toledo strain lacked a binding site in UL37, and additional binding sites were present, one in the *ori-lyt* region and one 5' of the UL89 open reading frame. The clinical isolate VR1814 also contained 15 predicted UL34 binding sites; again, the binding site locations were conserved with the AD169 strain, with the following exceptions. In strain VR1814, the binding site within UL54 was absent and two binding sites were present, one in the *ori-lyt* region and an additional binding site located between UL139 and UL140. The conservation of the predicted UL34 binding sites among different HCMV isolates further supports an important role for UL34-DNA interactions during the viral replication cycle.

We examined the ability of UL34 proteins to bind to seven of the predicted binding sites, utilizing electrophoretic mobility shift assays. Because of the known transcriptional regulatory role of UL34, binding sites located within regulatory regions of HCMV genes were chosen for analyses, along with the predicted binding sites within the *ori-lyt* region. UL34 and the control protein, luciferase, were synthesized using *in vitro* transcription/translation reactions; protein products were analyzed by SDS-gel electrophoresis to confirm the synthesis of the proteins (data not shown). The *in vitro*-synthesized proteins were incubated with radiolabeled double-stranded oligonucleotides containing predicted

UL34 binding sites (Fig. 2A). Oligonucleotide-protein mixtures were analyzed by gel electrophoresis. As illustrated in Fig. 2B, UL34 was able to bind to and alter the mobility of the oligonucleotides which contained a predicted UL34 binding site. The control protein, luciferase, had no effect on the migration of the radiolabeled DNA fragment. Titration of the amount of UL34 protein present in the DNA binding reactions altered the amount of the oligonucleotide shifted, as expected for a specific DNA-protein interaction. UL34 was unable to bind to DNA containing a mutated binding site (data not shown) (5). These data demonstrated

A.

Gene	Sequence		Oligos
US3	5'	TCAA*AAACACCGTT* CAGTCCACA 3'	115/116
US9	5'	GGAGCCG*AAACACCGTG* CCTC 3'	272/273
US11	5'	CCTAA*AAACACCGTG* TTCTCC 3'	274/275
RL10	5'	TAACA*AAACACCGTG* CACT 3'	402/403
<i>ori-lyt</i> 1	5'	ATTTAA*AAACACCGTG* TCG 3'	396/397
<i>ori-lyt</i> 2	5'	GCCATG*AAACACCGTG* ATGG 3'	400/401
<i>ori-lyt</i> 3	5'	TGTAA*AAACACCGTT* TGACCG 3'	404/405

B.

FIG 2 (A) Sequences of oligonucleotides used in electrophoretic mobility shift assays assessing pUL34 binding. The gene wherein the sequence is located is listed, as are the identifying numbers of the oligonucleotide pairs. The asterisk indicates that either a T or a G is functional in this position. (B) Electrophoretic mobility shift assays. *In vitro*-transcribed and -translated proteins, UL34 and luciferase (Luc.), were incubated with radiolabeled double-stranded oligonucleotides as listed in panel A. Numbers indicate the amounts of the *in vitro* translation reaction used in the assays; -, absence of any protein. Arrows indicate the shifted oligonucleotide-protein complexes.

FIG 3 (A) Diagram of the US9-US11 region (shown in reverse orientation relative to the conventional genomic position). Rectangles indicate open reading frames, T indicates the predicted TATA boxes, the bent arrow indicates the transcription initiation sites, and gray ovals indicate UL34 binding sites. (B) Sequences of the US9 and US11 wild-type and mutated UL34 binding sites. Mutated nucleotides are indicated by bold and lower-case type. (C) Reporter gene analyses of the US3, US9, and US11 regulatory regions containing either a wild-type (wt) or a mutant (mut) UL34 binding site. Levels of gene expression were measured in the presence of the major immediate-early genes and in the presence of a UL34 expression plasmid (black bars) or in the presence of a control plasmid in which the UL34 open reading frame is transcribed in the reverse orientation (gray bars). (D) Reporter gene analyses of the US3, US9, and US11 regulatory regions containing either a wild-type (wt) or a mutant (mut) UL34 binding site in the presence (dark-gray bars) or absence (light-gray bars) of viral infection. β -Galactosidase activity was determined by measuring the fluorescence of the cleavage product of 4-methylumbelliferyl- β -D-galactoside. Values are averages of two different transfections; error bars represent one standard deviation. Background levels of fluorescence obtained with a promoterless reporter gene plasmid were subtracted from the experimental values shown.

that UL34 binds specifically to sequences containing the 10-nucleotide core UL34 binding site. Although not all 14 predicted UL34 binding sites were examined for interactions with UL34, these data suggested that UL34 is likely to interact with all of the sites.

The UL34 binding site in the US9 gene contributes to down-regulation. The frequency of predicted pUL34 binding sites within the HCMV genome suggested that they have a functional role in viral gene expression and replication. In addition to US3, two other genes in the US2-US11 gene cluster, the early genes US9 and US11, also have UL34 binding sites within their regulatory

regions. US11 causes the dislocation of MHC class I molecules from the endoplasmic reticulum to the cytosol, where they are degraded (15). US9 localizes to the endoplasmic reticulum, but the function of the protein is still unclear (16). The locations of the UL34 binding sites within the US9 and US11 genes suggested that the interaction of pUL34 with these sequences might play a role in regulating gene expression (Fig. 3A), potentially providing a mechanism for coordinately regulating expression of this gene cluster.

To assess the contribution of the UL34 binding sites to the

regulation of US9 and US11 expression, reporter plasmids, in which expression of the reporter gene *lacZ* was regulated either by the US9 regulatory region (nucleotides –576 to –9 relative to the transcription start site) or by the US11 regulatory region (nucleotides –642 to +10), were constructed. In addition, reporter plasmids that contained mutated UL34 binding sites, substituting nucleotides to create a nonfunctional pUL34 binding site, were constructed (Fig. 3B). The resulting reporter plasmids were transfected into primary human diploid fibroblasts in the presence of the major immediate-early gene and with or without a UL34 expression plasmid. The levels of expression of the reporter gene *lacZ* were determined by assaying the amount of fluorescent cleavage product produced following incubation with medium containing methylumbelliferyl- β -D-galactoside (MUG). As controls, the US3 reporter plasmids, which either contained or lacked a pUL34 binding site, were assayed for levels of *lacZ* expression in the same experiments. Experiments were repeated multiple times; the data presented in Fig. 3C depict the average values obtained in one experiment.

The presence of UL34 markedly repressed the expression of the US3 promoter as previously shown (5); this repression was abrogated by mutation of the UL34 binding site (Fig. 3C). Expression directed by the US9 regulatory region was slightly decreased by the expression of UL34 in the presence of either a wild-type or a mutant UL34 binding site. This mirrors the slight repressive effect of UL34 seen on the US3 reporter construct that contains a mutant UL34 binding site. The presence or absence of UL34 had no effect on expression from the US11 regulatory region, either in the presence or absence of a functional UL34 binding site (Fig. 3C). Thus, although UL34 binds to sites within the US9 and US11 genes, no regulatory effect of UL34 was detected in transient-expression assays in the presence of the major immediate-early proteins.

We also assayed levels of expression from the US9 and US11 reporter plasmids in the presence or absence of viral infection. Reporter gene constructs were transfected into primary human diploid fibroblasts; cells were infected ~36 h after transfection. Reporter gene levels were measured 24 h postinfection as described above. Expression from the US3 promoter region resulted in low levels of reporter gene activity following infection; mutation of the UL34 binding site in the US3 regulatory region resulted in a 10-fold increase in expression (Fig. 3D). Similarly, low levels of gene expression were detected from the wild-type US9 construct while mutation of the UL34 binding site resulted in an ~6-fold increase in levels of expression. In contrast, the wild-type and mutant US11 promoter regions directed similar levels of *lacZ* expression.

The UL34 binding site in the US11 regulatory region is located upstream of the TATA box (nucleotides –100 to –90) (Fig. 3A), just 5' of the promoter region defined as critical for regulated gene expression (17). Although UL34 binds to sequences within the US11 regulatory region, UL34 or mutagenesis of the UL34 binding site had no measurable effect on US11 expression (Fig. 3C and D). The significance of the UL34-DNA interaction within the US11 regulatory region remains to be elucidated.

The UL34 binding site within the US9 regulatory region is located between the TATA box and the transcription start site (Fig. 3A), similar to the position of the UL34 binding site present in the US3 gene. Although UL34 had no effect on expression from the US9 regulatory region in transient-transfection assays, viral infection markedly increased the levels of expression from the US9

promoter region in the presence of a mutated UL34 binding site. These data suggested that UL34-DNA interactions contribute to the regulation of the US9 gene in the context of viral infection. These data further suggest that additional viral proteins, proteins other than the major immediate-early proteins and UL34, contribute to the activation of US9 expression. Downregulation of the US3 gene is important for efficient viral replication (Z. Liu and B. J. Biegalka, unpublished results); the results presented here suggest that downregulation of US9 expression may also be important for efficient viral replication.

In summary, we have identified multiple UL34 binding sites within the HCMV genome; all seven binding sites tested interact with UL34 in electrophoretic mobility shift assays. These data suggest that UL34 will bind to all of the predicted binding sites, including the sites located within protein-coding regions. Although two of the UL34 binding sites, the US3 and US9 binding sites, contribute to transcriptional regulation, the frequency of the binding sites suggests that UL34 will play additional roles in viral replication. The functional significance of the other UL34 binding sites, in particular, those within open reading frames and those in the region containing *ori-lyt*, remains to be clarified.

ACKNOWLEDGMENTS

This work was supported by the American Heart Association and the Ohio University Heritage College of Osteopathic Medicine.

We thank Erin Lester and Janet Hammer for expert technical assistance.

REFERENCES

1. Ho M. 2008. The history of cytomegalovirus and its diseases. *Med. Microbiol. Immunol.* 197:65–73.
2. Dunn W, Chou C, Li H, Hai R, Patterson D, Stolc V, Zhu H, Liu F. 2003. Functional profiling of a human cytomegalovirus genome. *Proc. Natl. Acad. Sci. U. S. A.* 100:14223–14228.
3. Yu D, Silva MC, Shenk T. 2003. Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proc. Natl. Acad. Sci. U. S. A.* 100:12396–12401.
4. Biegalka BJ, Lester E, Rana R. 2004. Characterization of the human cytomegalovirus UL34 gene. *J. Virol.* 78:9579–9583.
5. LaPierre L, Biegalka BJ. 2001. Identification of a novel transcriptional repressor encoded by human cytomegalovirus. *J. Virol.* 75:6062–6069.
6. Jones TR, Hanson LK, Sum L, Slater JS, Stenberg RM, Campbell AE. 1995. Multiple independent loci within the human cytomegalovirus unique short region down-regulate expression of major histocompatibility complex class I heavy chains. *J. Virol.* 69:4830–4841.
7. Liu Z, Winkler M, Biegalka BJ. 2009. Human cytomegalovirus: host immune modulation by the viral US3 gene. *Int. J. Biochem. Cell Biol.* 41:503–506.
8. Liu W, Zhao Y, Biegalka BJ. 2002. Analysis of human cytomegalovirus US3 gene products. *Virology* 301:32–42.
9. Biegalka BJ. 1998. Characterization of the transcriptional repressive element of the human cytomegalovirus immediate-early US3 gene. *J. Virol.* 72:5457–5463.
10. Anders DG, Kacica MA, Pari GS, Punturieri SM. 1992. Boundaries and structure of human cytomegalovirus *oriLyt*, a complex origin for lytic-phase DNA replication. *J. Virol.* 66:3373–3384.
11. Hamzeh RM, Lietman PS, Gibson W, Hayward G. 1990. Identification of the lytic origin of DNA replication in human cytomegalovirus by a novel approach utilizing ganciclovir-induced chain termination. *J. Virol.* 64:6184–6195.
12. Masse MJO, Karlin S, Schachtel GA, Mocarski ES. 1992. Human cytomegalovirus origin of DNA replication (*oriLyt*) resides within a highly complex repetitive region. *Proc. Natl. Acad. Sci. U. S. A.* 89:5246.
13. Cha TA, Tom E, Kemble GW, Duke GM, Mocarski ES, Spaete RR. 1996. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. *J. Virol.* 70:78–83.
14. Dargan DJ, Douglas E, Cunningham C, Jamieson F, Stanton RJ, Bal-

- uchova K, McSharry BP, Tomasec P, Emery VC, Percivalle E, Sarasini A, Gerna G, Wilkinson GWG, Davison AJ. 2010. Sequential mutations associated with adaptation of human cytomegalovirus to growth in cell culture. *J. Gen. Virol.* **91**:1535–1546.
15. Wiertz EJHJ, Jones TR, Sun L, Bogoy M, Geuze HJ, Ploegh HL. 1996. The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. *Cell* **84**:769–779.
 16. Mandic L, Miller MS, Coulter C, Munshaw B, Hertel L. 2009. Human cytomegalovirus US9 protein contains an N-terminal signal sequence and a C-terminal mitochondrial localization domain, and does not alter cellular sensitivity to apoptosis. *J. Gen. Virol.* **90**:1172–1182.
 17. Chau NH, Vanson CD, Kerry JA. 1999. Transcriptional regulation of the human cytomegalovirus US11 early gene. *J. Virol.* **73**:863–870.
 18. Gatherer D, Seirafian S, Cunningham C, Holton M, Dargan DJ, Baluchova K, Hector RD, Galbraith J, Herzyk P, Wilkinson GWG, Davison AJ. 2011. High-resolution human cytomegalovirus transcriptome. *Proc. Natl. Acad. Sci. U. S. A.* **108**:19755–19760.