Nontraditional Localization and Retention Signals Localize Human Cytomegalovirus pUL34 to the Nucleus

Bonita J. Biegalka

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

The human cytomegalovirus (HCMV) UL34 proteins localize to the nucleus. We identified a 204-amino-acid region, amino acids 40 to 244, as required for nuclear retention. A highly conserved 12-amino-acid section, amino acids 198 to 210, was required for nuclear localization. Although the nuclear localization and retention signals of pUL34 overlap the domain required for DNA-binding activity, the two regions are separable by point mutations. Our results presented here identify the first example of an HCMV protein with separable nuclear localization and retention signals.

Human cytomegalovirus (HCMV) is a common opportunistic pathogen, causing a variety of diseases in immunoincompetent individuals, including transplant recipients, people with HIV, and neonates (1). Despite a high degree of sequence conservation among clinical HCMV isolates, the majority of the viral genome is dispensable for replication in the laboratory, suggesting that many of the viral genes play an important role during infection and latency in the human host (2, 3). The UL34 gene is one of the ~50 viral genes (out of ~176 total genes) that are essential for replication in the laboratory.

The UL34 gene encodes two highly related proteins, with expression beginning 3 h postinfection and continuing throughout the replication cycle (4). Early and late forms of pUL34 are synthesized during infection; the early and late UL34 proteins are translated from the same open reading frame, with the early UL34 protein comprised of amino acids 1 to 407 and the late UL34 protein comprised of amino acids 22 to 407. The only known function of the UL34 gene is the downregulation of two nonessential viral genes, the US3 and US9 genes (5, 6). The UL34 proteins bind to a regulatory element in these genes to repress their expression (7). The region of pUL34 comprised of amino acids 22 to 244 is required for downregulation of US3 expression and for DNA-binding activity. Visualization of the UL34 proteins using either indirect immunofluorescent antibody staining or an enhanced green fluorescent protein (EGFP) tag identifies the nucleus as the predominant intracellular location (4).

Importation of proteins into the nucleus is mediated by a nuclear localization signal (NLS) in the target protein. The interaction of the target protein or cargo with the nuclear importation machinery allows the complex to be transferred through the nuclear pore (8). Shuttle proteins, or karyopherins, either alone or in the presence of adaptor proteins, such as importin α , bind to and transport the target protein in an energy-dependent process. Once inside the nucleus, Ran GTPase facilitates the separation of the cargo and shuttle proteins.

Proteins smaller than 40 kDa can readily diffuse into and out of the nucleus, proteins of 40 to 60 kDa generally utilize a nuclear localization signal to enter the nucleus, while proteins larger than 60 kDa in size require the presence of an NLS to enter the nucleus (9). NLSs have been characterized in many proteins; a classical NLS has a characteristic pattern of basic residues in loosely matching consensus sequences, the monopartite sequence, K(K/R)X(K/R), or the bipartite sequence, KRX₁₀₋₁₂KRXK (10). Not all nuclear

proteins contain or utilize classical NLSs for transport into the nucleus, and thus other amino acid sequences or the structure of the protein can play an important role in nuclear localization.

Many HCMV proteins localize to nuclei of infected cells, where they function in multiple ways, including regulation of transcription, alteration of the nuclear structure, and packaging of newly replicated viral genomes. Studies examining the NLSs in HCMV proteins have identified both classical and nonclassical NLSs. Other HCMV proteins also have a nuclear location; however, the regions of the proteins required for nuclear importation have not yet been identified. The studies described here were undertaken to identify the NLS(s) in the UL34 proteins.

Conservation of the UL34 open reading frame. The UL34 open reading frame encodes a 407-amino-acid protein comprised of a basic amino-terminal end and an acidic carboxy-terminal half. Amino acids 22 to 244 are required for DNA-binding activity (7). BLAST analyses with proteins in the NCBI protein database demonstrated that the UL34 open reading frame is conserved within other cytomegaloviruses but not in more distantly related herpesviruses. Conservation among the UL34 proteins varied, with identical amino acids ranging from 64% for chimpanzee CMV to 26% for elephantid herpesvirus 1 and similarities ranging from 82% to 48%. The conservation among the UL34 homologs was seen only in the amino-terminal 260 amino acids, where there are clusters of highly conserved amino acids and conserved cysteine residues (Fig. 1). The amino-terminal domain of the protein is predicted to form a helix-loop-helix conformation.

Location of predicted NLSs. UL34 proteins localize to nuclei, both following transfection and in infected cells (4). Computer analysis of the amino acid content of the UL34 protein using the computer algorithm PSORTII predicted the nuclear localization of the UL34 proteins (pUL34) and identified potential NLSs (see Fig. 4A) (11). A potential bipartite signal was predicted beginning at amino acid position 60 (RRNSNLFVFCTERDYRK; underlined in Fig. 1); this region is fairly well conserved among the cytomega-

Received 19 July 2013 Accepted 16 August 2013

Published ahead of print 21 August 2013

Address correspondence to Bonita J. Biegalka, biegalka@ohio.edu.

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

doi:10.1128/JVI.02025-13

FIG 1 Comparison of the UL34 amino acid sequences for human cytomegalovirus (row 1), panine herpesvirus 2 (row 2), macacine herpesvirus 3 (row 3), cercopithecine herpesvirus 5 (row 4), aotine herpesvirus 1 (row 5), saimiriine herpesvirus 3 (row 6), tupaiid herpesvirus 1 (row 7), murid herpesvirus 2 (row 8), caviid herpesvirus 2 (row 9), bat betaherpesvirus B7D8 (row 10), and elephantid herpesvirus 1 (row 11). Amino acid sequences were obtained from the protein database at the National Center for Biotechnology Information. The numbers correspond to the amino acids in the HCMV UL34 open reading frame. Cysteine residues are indicated by black shading; identical amino acids are indicated by dark gray shading; and amino acids with similar charges are indicated by light gray shading. The predicted helical regions are indicated by H. The asterisks indicate the methionine residues that correspond to the start of the early and late proteins. The predicted NLS is underlined.

loviruses. Additional monopartite NLSs were predicted beginning at amino acid 300 (KKPR) and at amino acid 306 (RRPR). Another program for predicting NLSs, NLStradamus (12), also predicted a potential NLS at amino acids 300 to 311; this region of the open reading frame shares little similarity overall among the cytomegaloviruses (data not shown).

Identification of UL34 nuclear localization and retention signals. We undertook a series of experiments to identify the NLS(s) in UL34. Plasmids that express portions of the UL34 coding region fused in frame to enhanced green fluorescent protein (EGFP) were constructed. Portions of the UL34 open reading frame were inserted in frame with the EGFP using plasmid pEGFP-C2 (Clontech) and either PCR or restriction enzyme digestion to generate the appropriate fragments for cloning. The UL34 sequences were obtained from plasmid pBJ374 (5), which contains the entire UL34 open reading frame. All plasmids were sequenced to confirm that the UL34 sequences were in frame with EGFP. The resulting plasmids were transfected into primary human diploid fibroblasts (HDF) using Effectene (Qiagen). Cells were fixed using 4% paraformaldehyde, permeabilized, and counterstained with diamidino-2-phenylindole (DAPI) to identify nuclei; following fixation, cells were observed using fluorescence microscopy. To

map the boundaries of the UL34 region necessary for nuclear localization, a series of plasmids which express amino or carboxy-terminal truncations of the UL34 open reading frame was constructed.

UL34 carboxy-terminal truncations were initially examined for their ability to localize to nuclei of transfected cells. Similar to the full-length protein (amino acids 1 to 407), proteins containing UL34 amino acids 1 to 374, 1 to 311, 1 to 287, and 1 to 244 all localized to the nuclei of transfected cells (Fig. 2A). The nuclear localization of the protein composed of amino acids 1 to 244 (predicted molecular size of 57.2 kDa) demonstrated that the predicted NLS located at amino acids 300 to 311 was not required for nuclear localization. Deletion of an additional 13 amino acids and expression of the EGFP fusion protein containing UL34 amino acids 1 to 231 (predicted molecular size of 53.3 kDa) resulted in an altered pattern of intracellular localization, where the fluorescent protein was detected in both the nucleus and the cytoplasm (Fig. 2), suggesting that this protein enters but is not retained in the nucleus.

In similar experiments, the contribution of the amino-terminal end of the protein to nuclear localization was also examined. The late UL34 protein is comprised of amino acids 22 to 407 and

FIG 2 Identification of the UL34 nuclear localization and retention signal. Plasmids expressing the indicated UL34 amino acids as EGFP fusion proteins were constructed and transfected into primary human fibroblasts. Proteins were visualized using fluorescence microscopy. (A) Carboxy-terminal truncations of the UL34 open reading frame; (B) amino-terminal truncations of the UL34 open reading frame; (C) intracellular location of EGFP; (D) amino- and carboxy-terminal truncations. The white rectangles below the images represent the UL34 open reading frame; the gray shaded areas represent the predicted NLSs. The numbers indicate the UL34 amino acids present in the fusion proteins. The arrows and arrowheads indicate the nuclei of the transfected cells.

FIG 3 Effects of deletions and mutations on the intracellular location of EGFP-UL34 fusion proteins. (A and B) Effects of deleting amino acids 189 to 230 or 198 to 210 from the UL34 open reading frame; (C) effects of mutating amino acids 42 and 43 or 45 and 46 on the intracellular location. Rectangles below the images represent the UL34 open reading frame; numbers indicate the UL34 amino acids present in the EGFP-UL34 fusion proteins. The gaps in the diagrammed proteins in panel B represent the deletion of amino acids 198 to 210.

localizes to the nucleus (Fig. 2B) (4). EGFP fusion constructs expressing UL34 amino acids 40 to 407, 45 to 407, 60 to 407, or 171 to 407 were analyzed as described above. The fusion protein expressing UL34 amino acids 40 to 407 localized to the nucleus, while deletion of an additional five amino acids (expression of a fusion protein containing amino acids 45 to 407 with a predicted molecular size of 69.8 kDa) was present in both the nucleus and the cytoplasm (Fig. 2B). Deletion of an additional 15 amino acids

(protein comprised of amino acids 60 to 407) or deletion of the amino-terminal 170 amino acids (expressing amino acids 171 to 407 in the fusion protein) also resulted in detection of the fluorescent protein throughout the cell (Fig. 2B). For a summary of these results, please see Fig. 4B.

The data in Fig. 2A and B suggest that UL34 amino acids 40 to 244 contain the NLS and, furthermore, that this region of pUL34 is required for nuclear retention. To confirm these results, EGFP

FIG 4 (A) Diagram of the UL34 open reading frame. The NLSs predicted by PSORTII are indicated by thick black lines; the basic (white rectangle) and acidic (gray rectangle) regions are depicted, with the predicted pI shown below the rectangles. The DNA-binding region of UL34 is delineated by a vertical line and a down arrow. (B) Summary diagram of the plasmids used in the experiments presented above. The UL34 amino acid positions in each EGFP fusion protein are given; the white rectangles represent proteins that are localized to the nucleus, the light gray rectangles represent proteins that are found throughout the cell, and the dark gray rectangles represent proteins that are found only in the cytoplasm. The predicted molecular sizes of the EGFP-UL34 fusion proteins are given in kilodaltons.

fusion proteins expressing UL34 amino acids 40 to 244, 40 to 231, or 60 to 244 were analyzed for their intracellular location (Fig. 2D). The EGFP-UL34 fusion protein comprised of amino acids 40 to 244 localized to and was retained in the nucleus. In contrast, the proteins containing UL34 amino acids 40 to 231 or 60 to 244 were detected throughout the cell (Fig. 2D).

Intriguingly, the EGFP fusion protein containing amino acids 45 to 407, although containing both predicted NLSs, was detected throughout the cell (Fig. 2A). This protein is too large to enter the nucleus through diffusion; the intracellular location of this fusion protein suggests that amino acids 40 to 45 play a critical role in retaining pUL34 in the nucleus. Similarly, the intracellular distribution pattern of the fusion protein expressing UL34 amino acids 1 to 231 (Fig. 2B) implicates amino acids 232 to 244 as important for nuclear retention.

The detection of fusion proteins containing UL34 amino acids 1 to 231 or amino acids 171 to 407 in the nucleus and cytoplasm of transfected cells, despite their predicted relative sizes of 53.2 kDa or 53.0 kDa, respectively, suggested that these proteins share a signal that mediates their nuclear import. The region shared between these two constructs does not include either predicted NLS but does include a very highly conserved region of the protein (Fig. 1). We postulated that this region might contribute to the

nuclear localization of the protein, despite lacking any defined NLS, and focused our deletion analyses on this region.

Initially, the conserved region comprised of amino acids 188 to 231 was deleted from the EGFP-UL34 expression plasmids. The late protein comprised of amino acids 22 to 407 localizes to the nucleus (Fig. 3A); however, deletion of amino acids 188 to 231 resulted in a loss of nuclear localization (Fig. 3A). Further analyses demonstrated that deletion of the most highly conserved amino acids from the UL34 protein, amino acids 198 to 210, was sufficient to result in a loss of nuclear localization (Fig. 3A). We also analyzed the contribution of the highly conserved region to the intracellular localization pattern of the fusion protein containing UL34 amino acids 45 to 407. Similar to the results obtained with the late UL34 protein, deletion of amino acids 188 to 231 or amino acids 198 to 210 also resulted in an inability of the protein to localize to the nucleus (Fig. 3A).

Additional plasmids expressing truncated versions of the EGFP-UL34 fusion proteins were constructed and analyzed for their patterns of localization following deletion of amino acids 198 to 210. Proteins synthesized from EGFP fusion plasmids expressing UL34 amino acids 1 to 231 or 40 to 231 were present in both the nucleus and the cytoplasm, while amino acids 40 to 244 resulted in only nuclear localization (Fig. 2D and 3B). Deletion of

amino acids 198 to 210 resulted in a loss of nuclear fluorescence for all three of the EGFP-UL34 fusion proteins (Fig. 3B). These data demonstrate the significance of the 12-amino-acid region, amino acids 198 to 210, to nuclear localization.

Separation of DNA-binding and nuclear retention signals.

The region comprising amino acids 40 to 244 that mediates nuclear localization and retention overlaps significantly with the region required for DNA-binding activity. To determine if nuclear retention is a function of DNA-binding activity, we assessed the intracellular location of two proteins with point mutations that have lost DNA-binding activity (7). The EGFP-UL34 fusion proteins expressing the UL34 open reading frame with point mutations for amino acids 42 and 43 or 45 and 46 (substituting the amino acids VA or NA for G⁴²K⁴³ or K⁴⁵L⁴⁶, respectively) localized to and were retained in the nucleus (Fig. 3C). These data demonstrate that the DNA-binding function of pUL34 is separable from nuclear localization and retention.

For successful replication, HCMV infection requires the importation of many viral proteins into the nuclei of infected cells, including the essential UL34 proteins. Despite containing amino acid sequences that resemble consensus NLSs, UL34 localizes to and is retained in the nucleus by using a nonconventional 204-amino-acid signal comprised of amino acids 40 to 244. The loss of nuclear localization following deletion of amino acids 198 to 210 demonstrates that this region plays a critical role in localizing pUL34 to the nucleus and, furthermore, that the predicted NLSs in the UL34 open reading frame do not contribute to the nuclear localization of the protein. In summary, import and retention of the UL34 protein in the nucleus are mediated by different portions of the protein. Amino acids 40 to 244 are required for retention of the protein in the nucleus, while a small 12-amino-acid region (amino acids 198 to 210) is essential for entry into the nucleus (Fig. 4). Despite overlapping extensively with the DNA-binding domain of UL34 (7), the intracellular localization signal is identified by point mutations as distinct from the DNA-binding domain.

Retention of proteins in the nucleus is much less well understood than nuclear import. Relatively few examples of nuclear retention signals have been identified. One is the nuclear retention signal in hnRNP C proteins, which is comprised of a 78-amino-

acid region (13). Because of the paucity of characterized nuclear retention signals, no consensus amino acid sequence has been defined. Our results presented here identify the first example of an HCMV protein with separable nuclear localization and retention signals.

ACKNOWLEDGMENTS

We thank Janet Hammer for technical assistance and Mark Berryman and Don Holzschu for helpful discussions. The following Ohio University undergraduate students contributed to these studies by constructing the indicated plasmids: Velma Lopez, a McNair scholar, constructed pBJ724, and Jessica Kunkel, a PACE student, constructed pBJ737.

This work was supported by the Ohio University College of Osteopathic Medicine and by the Baker Fund at Ohio University.

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. Liu Z, Biegalka BJ. 2013. Human cytomegalovirus UL34 binds to multiple sites within the viral genome. *J. Virol.* 87:3587–3591.
7. Lester E, Rana R, Liu Z, Biegalka BJ. 2006. Identification of the functional domains of the essential human cytomegalovirus UL34 proteins. *Virology* 353:27–34.
8. Pemberton LF, Paschal BM. 2005. Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* 6:187–198.
9. Dingwall C, Laskey RA. 1986. Protein import into the cell nucleus. *Annu. Rev. Cell Biol.* 2:367–390.
10. Dingwall C, Laskey RA. 1991. Nuclear targeting sequences—a consensus? *Trends Biochem. Sci.* 16:478–481.
11. Nakai K, Kanehisa M. 1992. A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* 14:897–911.
12. Nguyen Ba N, Pogoutse A, Provart N, Moses AM. 2009. NLStradamus: a simple hidden Markov model for nuclear localization signal prediction. *BMC Bioinformatics* 10:202. doi:10.1186/1471-2105-10-202.
13. Nakielnny S, Dreyfuss G. 1996. The hnRNP C proteins contain a nuclear retention sequence that can override nuclear export signals. *J. Cell Biol.* 134:1365–1373.