Virology **259,** 43–59 (1999)
Article ID viro.1999.9700, available online at http://www.idealibrary.com on IDEAL®

Sequence Requirements for the Nuclear Localization of the Murine Cytomegalovirus M44 Gene Product pp50

Lambert C. Loh,*[,1] Vicki D. Keeler,* and John D. Shanley†

*Department of Microbiology, University of Saskatchewan, 107 Wiggins Road, Saskatoon, Saskatchewan S7N 5E5 Canada; and †Department of Medicine, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06030-3212

Received October 23, 1998; returned to author for revision January 21, 1999; accepted March 10, 1999

The murine cytomegalovirus (MCMV) M44 gene product pp50 is normally present in the nuclei of virus-infected cells. During transient expression of pp50 in COS-1 cells, the phosphoprotein was readily detectable in the nuclei, indicating that it possesses a nuclear localization signal (NLS). Studies on the subcellular locations of N- and C-terminal deletion mutants of pp50 suggested that alterations in both the C terminus and the highly conserved N-terminal domains of pp50 affect nuclear localization. In particular, the C-terminal 11 amino acids of pp50, which includes a "KKQK" motif, were able to mediate the import of a β-galactosidase fusion protein into the nucleus. The pair of lysine residues in this motif constitutes an essential element of the C-terminal NLS as mutation of this motif to AAQK directly affected the nuclear localization of either pp50 or β-galactosidase fusion proteins containing the C-terminal portion of pp50. Furthermore our results indicated that the functionality of the C-terminal NLS is dependent on the structural integrity of the highly conserved N-terminal portion of the molecule, as deletion of amino acids 157–201 alone adversely affected nuclear localization. In the absence of a functional C-terminal NLS, the subcellular localization of pp50 is sensitive to potential conformational changes induced by mutations within the N-terminal half of the molecule. Under those circumstances, mutation of the YK residues at position 22–23 or deletion of amino acids 267–283 was sufficient to produce a protein that was impaired in nuclear import or retention.

INTRODUCTION

The transport of functional proteins to appropriate compartments within the cell is an important regulatory mechanism of cellular processes in eucaryotic organisms. For example, surface glycoproteins are sorted through the endoplasmic reticulum–Golgi pathway on their way to the cell surface, whereas transcription factors and DNA replication enzymes are translocated into the nucleus. In recent years, significant progress has been made in the identification of factors involved in nuclear import (Görlich, 1997). Within a cell, a nuclear envelope forms a barrier to passive diffusion between the nucleoplasm and the cytoplasm, and cytosolic proteins larger than ~60,000 can only be actively transported into the nucleus if they possess a nuclear localization signal (Dingwall and Laskey, 1986). Thus cellular proteins like the tumor suppressor p53 (Shaulsky et al., 1991) or the glucocorticoid receptor (Picard and Yamamoto, 1987), which are functionally dependent on their translocation into the nucleus, possess nuclear localization signals within their coding sequence to ensure targeting to the proper cellular compartment. The recent development of an *in vitro* nuclear import assay (Adam *et al.,* 1990) has facilitated the identification of several cellular proteins essential for mediating nuclear transport (Adam and Gerace, 1991; Görlich *et al.,* 1994), leading to a better understanding of the molecular mechanisms underlying this process.

Two distinct pathways for nuclear import have been identified. The more recently characterized M9 domain of 38 amino acids (Siomi and Dreyfuss, 1995), present in a subset of hnRNP (heterogeneous nuclear RNA-binding proteins), mediates nuclear import via transportin, a M9-specific receptor (Pollard *et al.,* 1996). In contrast, the "classical" nuclear localization signal (NLS), which mediates nuclear import of the vast majority of nuclear proteins via the importin-dependent pathway (Görlich and Mattaj, 1996; Görlich, 1997), has no clear consensus sequence. Instead it is characterized by the presence of clusters of basic amino acids that can be either continuous, like the SV40 T-antigen (Kalderon *et al.,* 1984a,b), or bipartite, as exemplified by nucleoplasmin (Dingwall *et al.,* 1988; Robbins *et al.,* 1991). Cytoplasmic proteins fused with functional NLSs can be actively transported into the nucleus (Gao and Knipe, 1992; Tillman *et al.,* 1994). However, residues surrounding a NLS or even distal sequences can affect its function, perhaps because of the need for such sequences to be exposed on

[1] To whom reprint requests should be addressed. Fax: (306) 966-4311. E-mail: loh@sask.usask.ca.

0042-6822/99 $30.00

the surface of a protein (Roberts *et al.,* 1987; Gao and Knipe, 1992; Yao *et al.,* 1995). It has also been demonstrated that the rate of import could be influenced by phosphorylation of nearby casein kinase II sites (Rihs *et al.,* 1991). Alternatively, proteins lacking a NLS can associate with one possessing a NLS, thereby gaining entry into the nucleus (Shaulsky *et al.,* 1990).

The murine cytomegalovirus phosphoprotein pp50, encoded by the M44 open reading frame (ORF), consists of 411 amino acids and is the homolog of the human cytomegalovirus (HCMV) polymerase accessory protein ppUL44 (Loh *et al.,* 1991, 1994; Pande *et al.,* 1991; Rawlinson *et al.,* 1996). The M44 ORF shares extensive sequence homology with its HCMV and human herpesvirus type 6 (HHV-6) homologs, with the greatest sequence conservation found within the N-terminal two-thirds of the molecule. The four regions with the highest degree of conservation between the three viruses are delineated by blocks in Fig. 1, with Regions 1 and 2 located within the N-terminal two-thirds of pp50, and the smaller Regions 3 and 4 located at the C-terminal end of the molecule. Although pp50 and the herpes simplex virus type 1 (HSV-1) UL42 gene product share almost no amino acid sequence homology, the two proteins are functionally analogous. Recent studies have shown that HCMV ppUL44 is a dsDNA-binding protein (Weiland *et al.,* 1994) essential for viral DNA replication (Pari and Anders, 1993; Pari *et al.,* 1993; Ripalti *et al.,* 1995). It was localized to the nuclei of virus-infected cells (Plachter *et al.,* 1992) and stimulated the activity of the viral DNA polymerase (Ertl and Powell, 1992; Weiland *et al.,* 1994).

As a homolog of HCMV ppUL44, pp50 is predicted to be required for viral DNA replication, which normally occurs in the nucleus. Thus the ability of pp50 to localize to the nucleus of infected cells is essential for its role in the virus replication cycle. Our laboratory has previously demonstrated that pp50 is a DNA-binding protein that localizes to the nuclei of MCMV-infected cells (Loh *et al.,* 1994). Results from size fractionation experiments suggested that pp50 is present predominantly as homodimers or in high molecular weight complexes with other viral or cellular proteins (Loh *et al.,* 1994) that are larger than the 60,000 limit for passive diffusion into the nucleus, thereby necessitating the presence of a NLS for nuclear localization. During transient expression, pp50 was found to be present exclusively in the nuclei of transfected cells (Loh *et al.,* 1994), suggesting that the protein possesses a functional nuclear localization signal. Analysis of the M44 ORF revealed a highly conserved NLS-like sequence at the C terminus of the protein. The putative NLS contains a stretch of basic amino acids "KKQK" (Fig. 7A, first line) and a potential casein kinase II phosphorylation site at the serine residue 15 amino acids upstream. Two additional clusters of basic amino acids are present between residues 5 and 8 ("RKVR," Fig. 7A, first line) and residues 169–174 ("KRNSKK," Fig. 7B) either of which could be a potential candidate for the NLS of pp50.

In this report, we describe experiments designed to identify pp50 sequences involved in nuclear localization. To this end, pp50 mutants were transiently expressed in COS-1 cells by transfection of cells with plasmids encoding mutant forms of the protein. The subcellular locations of pp50 mutants were determined by an indirect immunofluorescence assay using a panel of monoclonal antibodies recognizing epitopes located in different segments of the molecule. This approach allowed us to examine the subcellular location of pp50 independent of viral infection, where association with other viral proteins may influence its eventual subcellular localization. We employed deletion analysis to identify pp50 sequences required for nuclear localization of the protein. Furthermore chimeric proteins consisting of β-galactosidase fused in frame with putative NLSs were constructed to evaluate whether specific viral sequences were capable of directing nuclear import of the fusion proteins. Finally site-directed mutagenesis was used to identify amino acids that may affect the nuclear localization of pp50. Our results indicate that the pair of lysine residues within the "KKQK" motif at the extreme C terminus of the protein constitutes an essential element of a functional NLS that could mediate the active transport of β-galactosidase into the nucleus. In addition, the functionality of this NLS is likely to be dependent on the proper folding of the conserved N-terminal domains of pp50.

RESULTS

A panel of monoclonal antibodies recognizes epitopes in distinct domains of pp50

To utilize our panel of pp50-specific MAbs to detect mutant pp50 in immunofluorescence or Western-blotting experiments, we attempted to map epitopes recognized by these MAbs to specific segments of the protein. To this end, plasmid DNA encoding deletion mutants of pp50 were transfected into COS-1 cells which were harvested 48 h later. The solubilized proteins were separated by SDS–PAGE, immobilized on nitrocellulose filters, and probed with our panel of monoclonal antibodies by Western blotting. Reactivities of MAbs that did not recognize pp50 determinants by Western blotting were assessed by immunoprecipitation and indirect immunofluorescence. Results from experiments on selected MAbs are summarized in Table 1.

The six MAbs listed in Table 1 were selected for their reactivities against distinct segments of pp50. Five of these six MAbs (with the exception of 9A5) are reactive in Western blotting, indicating that they recognize linear epitopes. Briefly, MAb 31E11 is reactive with an epitope located within Region 1 (between amino acid residues

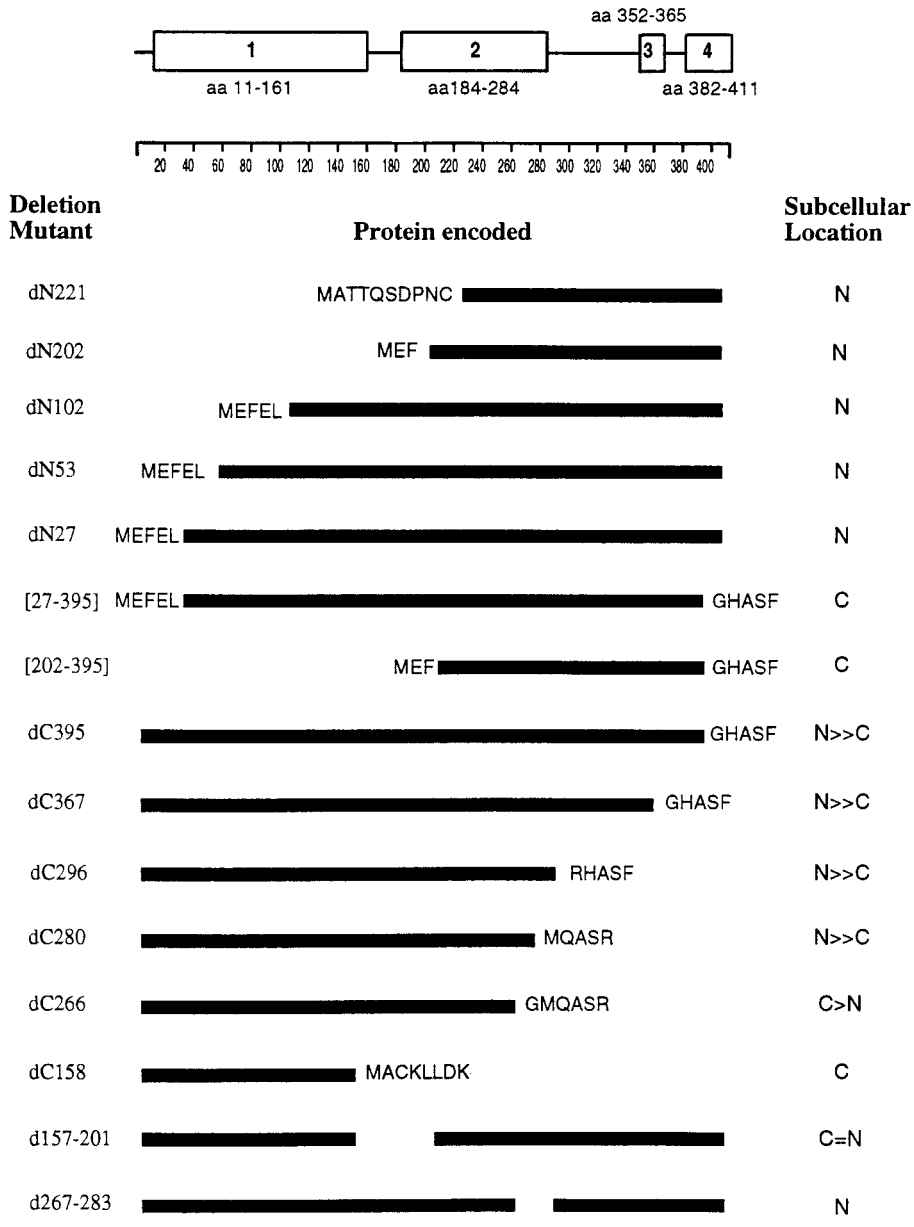

FIG. 1. Summary of pp50 deletion mutants. The first line represents the open reading frame of pp50 and delineates the locations of the four highly conserved blocks of amino acid residues (between MCMV, HCMV and HHV-6) with the appropriate residue numbers either below or above the numbered blocks. These conserved blocks were referred to as Regions 1–4 in the text. Directly underneath is a sequence ruler with tick marks every 20 amino acids along the entire length of the pp50 ORF. The solid lines underneath depict truncated pp50 sequences encoded by our panel of deletion mutants. The letters in front of or behind each line represent amino acid residues in single letter codes added to pp50 sequences as a result of providing translation start codons or universal translation stop codons to the truncated ORFs. Deletion mutants are labeled as described under Materials and Methods. Subcellular locations of the mutant proteins were determined by indirect immunofluorescence 48 h after transfecting COS-1 cells with the appropriate plasmids. The abbreviations used for describing the subcellular locations of deletion mutants are: N, nuclear; C, cytoplasmic.

102 and 158), whereas MAb 33C7 recognizes an epitope within a region (between amino acid residues 102 and 195) spanning the C-terminal half of Region 1, the N-terminal portion of Region 2, and the "linker" between Regions 1 and 2. Monoclonal antibody 25G11 is reactive with an epitope located in a region (between amino acid residues 202 and 296) encompassing most of Region 2 and sequences beyond its C-terminal end. Monoclonal antibody 9A5 is unique among the six MAbs in that it is reactive with pp50 in immunoprecipitation and indirect immunofluorescence assays but not in Western blotting. In addition, reactivity was lost if any deletions were introduced into conserved Region 2. Thus it is likely that 9A5 recognizes a conformation-dependent epitope and can be used to monitor proper protein folding within conserved Region 2. Last, although MAbs 5H10 and 3B9

TABLE 1

Reactivity of Monoclonal Antibodies with pp50 Deletion Mutants

MAbs	5H10	3B9	25G11	9A5	33C7	31E11
C-terminal deletions						
dC158	−	−	−	−	−	+
dC195	−	−	−	−	+	+
dC236	−	−	−	−	+	+
dC266	−	−	−	−	+	+
dC280	−	−	−	+	+	+
dC296	−	−	+	+	+	+
dC319	−	−	+	+	+	+
dC350	−	−	+	+	+	+
dC367	+	+	+	+	+	+
dC378	+	+	+	+	+	+
dC395	+	+	+	+	+	+
N-terminal deletions						
dN27	+	+	+	+	+	+
dN53	+	+	+	+	+	+
dN83	+	+	+	+	+	+
dN102	+	+	+	+	+	+
dN145	+	+	+	+	−	−
dN171	+	+	+	−	−	−
dN202	+	+	+	−	−	−
Internal deletions						
d157-201	+	+	+	−	−	+
d267-283	+	+	+	−	−	+

appeared to react with antigenic determinants located within the same region near the C terminus of pp50, MAb 5H10 but not 3B9 recognized an epitope conserved between MCMV pp50 and HCMV ppUL44 (Loh *et al.,* 1994), thereby demonstrating their unique specificities. Therefore with appropriate choices of antibodies, we were able to confirm the predicted sizes and expression of all the pp50 mutants described in this study by Western blotting (data not shown) and visualize their subcellular locations by immunofluorescence microscopy.

Subcellular location of pp50 N-terminal deletion mutants

We reasoned that if the "KKQK" sequence at the C terminus rather than the two N-terminal clusters of basic amino acids (residues 5–8 and residues 169–174) is the major determinant for nuclear localization, progressive N-terminal deletions into the pp50 ORF removing the latter would produce truncated proteins that are still predominantly localized to the nucleus. Thus our investigation began with the identification of the subcellular locations of pp50 N-terminal deletion mutants by indirect immunofluorescence using the panel of MAbs described in Table 1. The structures and observed subcellular locations of these mutants are summarized in Fig. 1, and staining patterns of selected mutants are shown in Fig. 2.

In MCMV-infected Balb/3T3 fibroblasts, pp50 was primarily localized to specific compartments of the nucleus (Fig. 2A). However, during transient expression of pp50 in COS-1 cells, the protein is distributed either in a diffuse manner throughout the nucleus or in a granular pattern within the nucleus (Fig. 2B). The subcellular staining pattern of two of these mutants (dN27 and dN202), something that is fairly representative of all the N-terminal deletion mutants examined, is shown in Figs. 2C and 2D. As illustrated in Fig. 1, Region 1 is progressively deleted in dN27, dN53, and dN102 until it is completely removed in dN202. In dN221, Region 1 and part of Region 2 are deleted. However, similar granular staining patterns in the nucleus were observed with all these mutants. Thus progressive deletion of Regions 1 and 2 along with the two aforementioned clusters of basic amino acids from pp50 does not appear to affect the subcellular localization of pp50 mutants, suggesting that a nuclear localization signal capable of directing the truncated protein to the nucleus is likely located within the C-terminal half of pp50. This was confirmed by the observation that removal of the last 16 amino acids of pp50 from either dN27 or dN202 completely abolished nuclear localization (mutants pp50[27–395] and pp50[202–395] in Figs. 2E and 2F).

Sequences at the extreme C terminus of pp50 can function as a NLS

One of the characteristics of a nuclear localization signal is its ability to direct the import of a normally

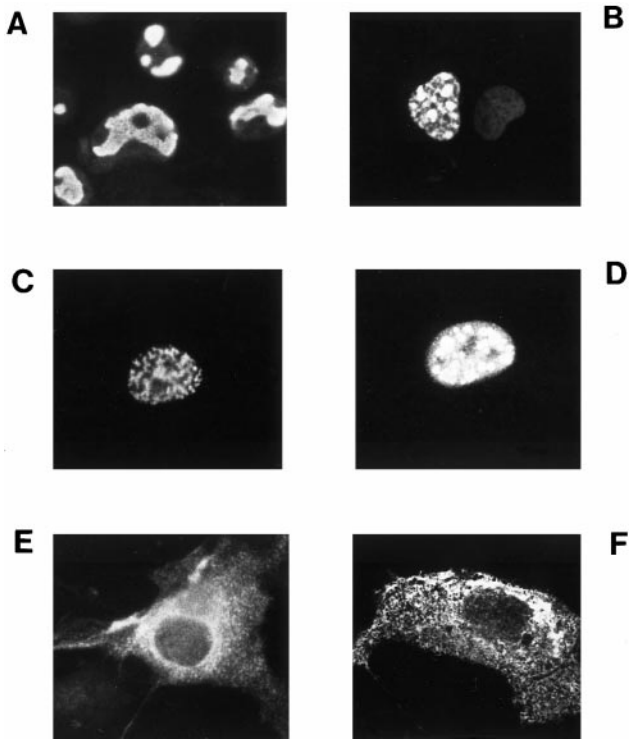

FIG. 2. Subcellular location of pp50 and N-terminal deletion mutants. Balb/3T3 fibroblasts infected with MCMV (A) were processed for immunofluorescence 20 h p.i. as described under Materials and Methods. Similarly, immunofluorescence microscopy was carried out with COS-1 cells grown on glass coverslips and transfected with plasmids pSVpp50 (B), pSVpp50dN27 (C), pSVpp50dN202 (D), pSVpp50[27–395] (E), and pSVpp50[202–395] (F). The subcellular locations of wild-type pp50 as well as N-terminal deletion mutants were identified by staining with MAb 3B9 48 h after transfection.

cytoplasm-resident protein into the nucleus. To test whether such a signal sequence exists within the C-terminal half of pp50, chimeric proteins consisting of β-galactosidase fused in frame with pp50 sequences were constructed as described under Materials and Methods and illustrated in Fig. 3. The subcellular locations of the chimeric proteins were determined in immunofluorescence experiments, and similar staining patterns were observed using either β-galactosidase specific or pp50-specific monoclonal antibodies. The predicted sizes of the fusion proteins were verified by Western blotting using the same antibodies (data not shown). The results of these experiments are summarized in Fig. 3, right.

As expected, β-galactosidase was predominantly present in the cytoplasm of COS-1 cells transfected with the β-galactosidase expressing plasmid pSVβgal (Fig. 4A). In contrast, the chimeric protein encoded by pSVβgal-pp50dN202 was exclusively localized to the nucleus (Fig. 4B). However, deletion of pp50 sequences beyond residue 395 from the chimeric protein effectively abolished nuclear localization (βgal-pp50[202–395] in Fig. 4C), suggesting that the C-terminal 16 amino acids of pp50 may be part of a nuclear localization signal that could mediate the nuclear import of the β-galactosidase fusion protein. This hypothesis is further strengthened by results showing that the last 11 amino acids of pp50, which include the highly conserved cluster of basic amino acids "KKQK" (Fig. 7A, first line), can direct nuclear import when fused in frame to the C terminus of β-galactosidase (βgal-pp50dN401 in Fig. 4D) and constitute a likely candidate for the NLS of pp50.

Subcellular location of pp50 C-terminal deletion mutants

We reasoned that if the C-terminal NLS described above is solely responsible for the nuclear localization of pp50, progressive deletion from that end of the molecule should result in a protein that fails to localize to the nucleus. The results of experiments using these mutants are summarized in Fig. 1, and staining patterns of selected C-terminal deletion mutants are shown in Fig. 5.

Our data indicate that stepwise deletions of the hydrophilic C-terminal tail of pp50 did not significantly affect nuclear localization (dC395 in Fig. 5A and dC296 in Fig. 5B). As illustrated in Fig. 1, sequences including the KKQK motif described in the previous section were deleted in dC395. The conserved Region 4 was deleted in dC367. Further deletions removing virtually the entire C-terminal tail, including conserved Regions 3 and 4, produced dC296. In each instance, the subcellular location as detected by immunofluorescence was predominantly nuclear with faint cytoplasmic staining being observed only occasionally. Even in dC280 where four amino acids at the C-terminal end of conserved Region 2 were deleted, a similar staining pattern was observed (data not shown). However, further deletions into Region 2 severely impaired nuclear localization, and a bright, granular, cytoplasmic fluorescence pattern together with faint, diffuse nuclear staining was observed when transfected cells were stained with MAb 31E11 (mutant dC266 in Fig. 5C). This suggested that sequences between residues 267 and 280, which do not contain clusters of basic amino acids, probably include amino acids involved in protein folding. Removal of these residues could result in conformational changes and affect nuclear localization indirectly. When Region 2 was removed by further deletions, the resulting truncated protein was found to be entirely resident in the cytoplasm with bright, granular cytoplasmic staining (mutant dC158 in Fig. 5D). Thus it would appear that the putative C-terminal NLS of pp50 may not be the only factor affecting nuclear localization of this protein. Additional determinants within the N-terminal half of the molecule, and Region 2 in particular, may influence the subcellular localization of pp50 mutants.

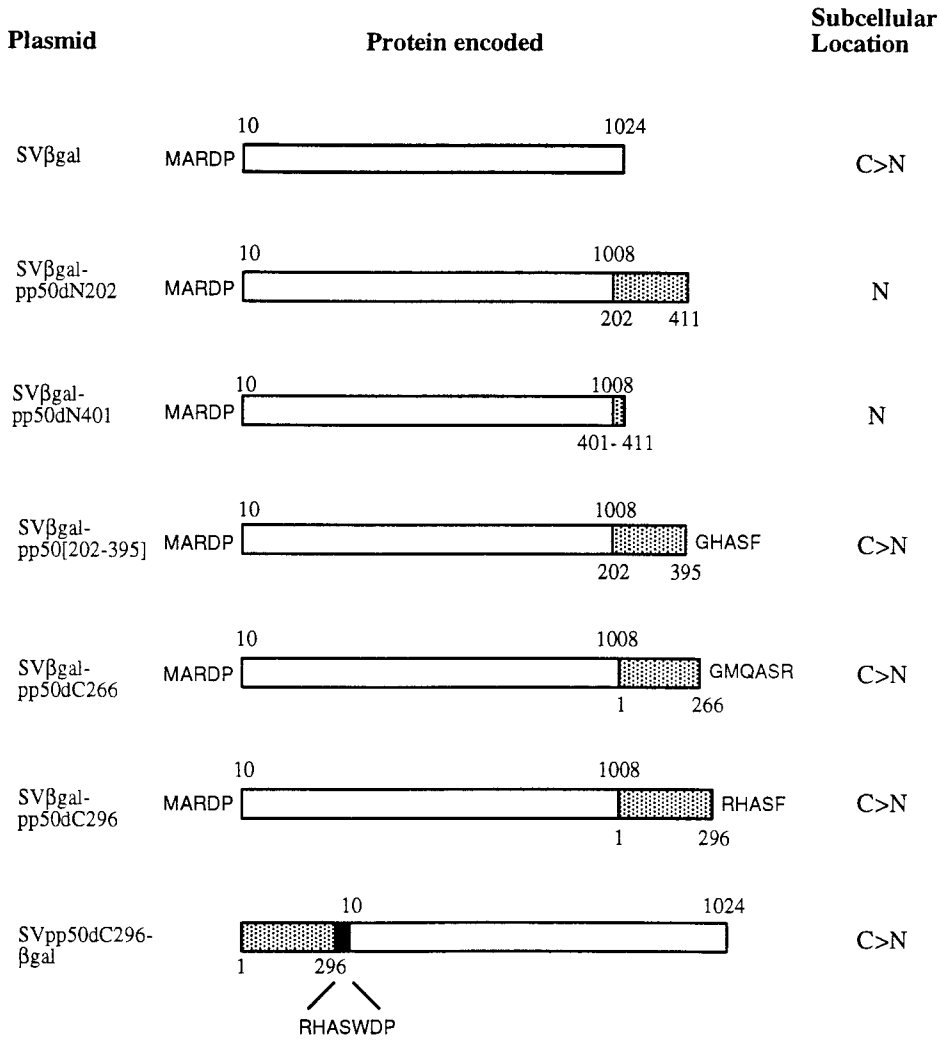

FIG. 3. Construction of β-galactosidase-pp50 chimeric proteins. Open bars represent β-galactosidase sequences, whereas shaded bars represent pp50 sequences. The numbers above each bar are β-galactosidase residue numbers, whereas those below each bar are pp50 residue numbers. The first 10 amino acids for β-galactosidase have been replaced by the sequence MARDP, providing unique restriction sites (*Nco*I, *Bam*HI, *Sma*I) for cloning purposes. The residues found at the end of the C-terminal truncated pp50 sequences are the result of universal stop codons present in the expression vector. The extraneous residues RHASWDP found between the pp50 and β-galactosidase sequences in SVpp50dC296-βgal are the result of inserting viral DNA between the *Nco*I and *Sma*I sites of the β-galactosidase gene. Subcellular locations of the fusion proteins were determined by indirect immunofluorescence 48 h after transfecting COS-1 cells with the appropriate plasmids. Abbreviations used for describing subcellular locations are the same as those used in Fig. 1.

The conserved N-terminal portion of pp50 cannot direct nuclear import of a β-galactosidase fusion protein

To determine whether the N-terminal sequences of pp50 contain a functional NLS, chimeras containing N-terminal sequences of pp50 fused in frame to the N-terminal end (pp50dC296-βgal) or C-terminal end (βgal-pp50dC296 and βgal-pp50dC266) of β-galactosidase sequences were constructed. The deletion mutant dC296 was chosen because it contains intact Regions 1 and 2 and is clearly localized to the nucleus in the absence of a functional C-terminal NLS (Fig. 5B). Fusion to the N-terminal end of β-galactosidase (pp50dC296-βgal) could be viewed as replacing the C-terminal tail of pp50 with foreign sequences and might be less disruptive to protein folding with respect to pp50. Fusion to the C-terminal end of β-galactosidase (βgal-pp50dC296) created a chimera that is more comparable to constructs used in characterizing the C-terminal NLS (Fig. 3). Parallel experiments using β-galactosidase fusions with the mutant dC266, which is primarily localized to the cytoplasm (Fig. 5C), would be used for comparison. The structures and subcellular locations of these fusion proteins are summarized in Fig. 3, bottom three lines.

As seen in Fig. 6, fusion proteins consisting of β-galactosidase sequences fused with N-terminal sequences

of pp50 were predominantly resident in the cytoplasm. In particular, the β-galactosidase-pp50 chimeras were seen in globular structures accumulating outside the nucleus. Thus our results indicated that the N-terminal portion of pp50 probably does not contain a functional NLS. Alternatively, structural constraints of the fusion proteins may prevent the surface exposure of a putative N-terminal NLS.

Lysine residues at the extreme C terminus constitute an essential element of the nuclear localization signal

One of the shortcomings of using N- or C-terminal deletion mutants is that amino acids unrelated to the wild-type gene product are almost inevitably added to the mutant as a result of providing translation start codons or universal termination codons to these truncated coding sequences. Although in our case, the additional amino acid sequences did not include stretches of basic amino acids that could affect the subcellular localization of the deletion mutants (Fig. 1), the possibility existed that these extraneous sequences may influence our results in a more subtle manner. For example, charged residues or amino acids with aromatic side chains could participate in protein folding and alter the conformation of the molecule. Therefore to define more precisely amino acid residues that affect pp50 nuclear localization, we carried out site-directed mutagenesis of selected pairs of amino acids and determined the subcellular locations of these site-specific mutants by immunofluorescence microscopy. The results of these studies and the structures of the mutants used in these experiments are summarized in Fig. 7A.

Having already demonstrated that the C-terminal 11 amino acids of pp50 constitute a functional NLS, we began by substituting the two lysine residues in the C-terminal KKQK motif with two alanine residues, creating the KK405AA mutant. Not surprisingly, this mutant is predominantly located in the nucleus (Fig. 8B) because we have already demonstrated that deletion

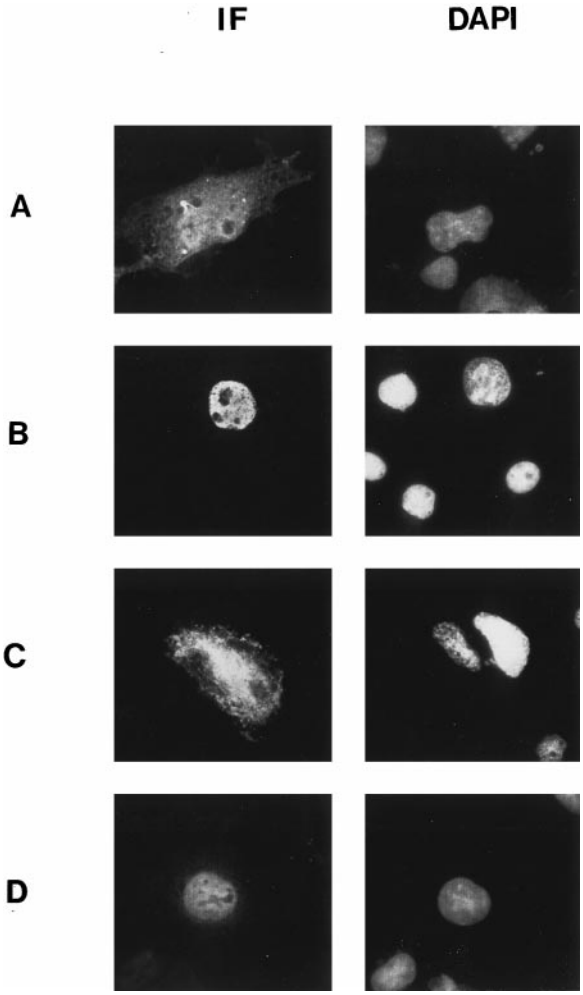

FIG. 4. The C terminus of pp50 contains a nuclear localization signal. C-terminal sequences of pp50 were fused in frame with the β-galactosidase ORF and placed behind the SV40 promoter in the plasmid vector pM2. The ability of pp50 sequences to direct the fusion protein to the nuclei of COS-1 cells transfected with these plasmids was evaluated by indirect immunofluorescence with a commercially available MAb specific for β-galactosidase. The plasmids used were pSVβgal (A), pSVβgal-pp50dN202 (B), pSVβgal-pp50[202–395] (C), and pSVβgal-pp50dN401 (D). DAPI staining was used to visualize the nuclei of transfected cells.

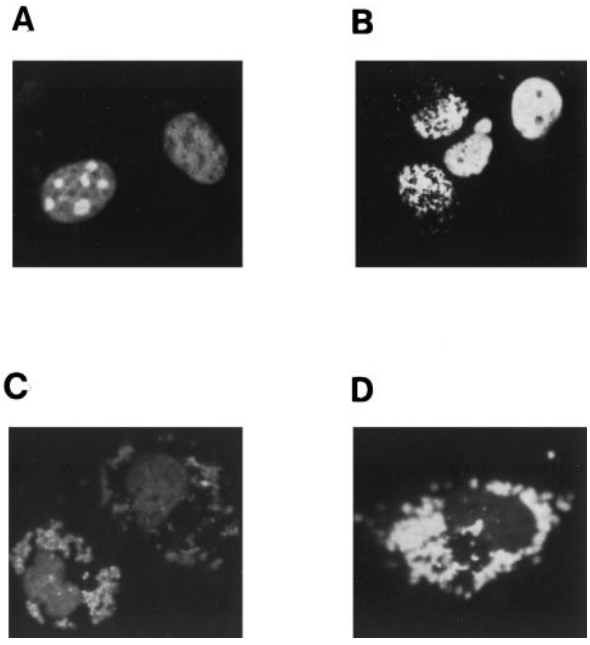

FIG. 5. Subcellular location of pp50 C-terminal deletion mutants. COS-1 cells grown on glass coverslips were transfected with plasmids pSVpp50dC395 (A), pSVpp50dC296 (B), pSVpp50dC266 (C), and pSVpp50dC158 (D) and processed for immunofluorescence 48 h later. The subcellular locations of the pp50 C-terminal deletion mutants were visualized by staining with MAbs 3B9 (A and B) and 31E11 (C and D).

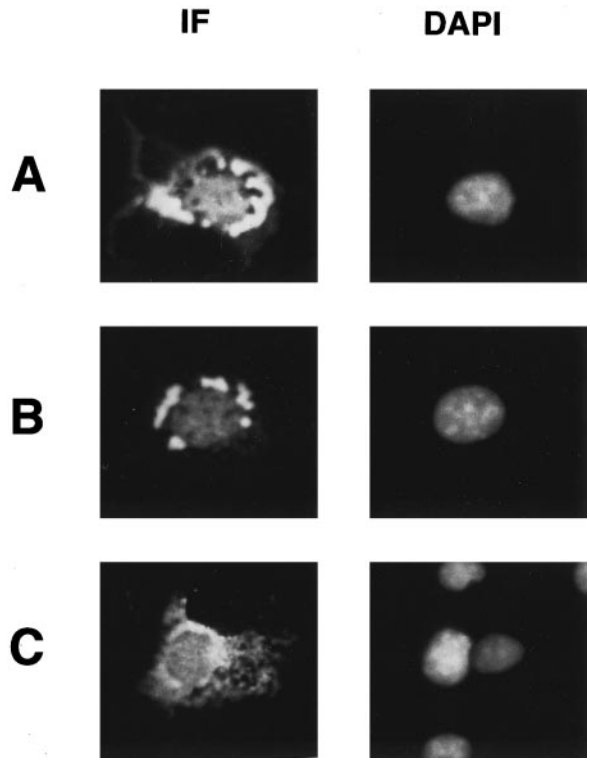

FIG. 6. The conserved N-terminal portion of pp50 cannot direct nuclear import of a β-galactosidase fusion protein. N-terminal sequences of pp50 were fused in frame with the β-galactosidase ORF and placed behind the SV40 promoter in the plasmid vector pM2. The ability of pp50 sequences to direct the fusion protein to the nuclei of COS-1 cells transfected with these plasmids was evaluated by indirect immunofluorescence with a commercially available MAb specific for β-galactosidase. The plasmids used were pSVβgal-pp50dC296 (A), pSVpp50dC296-βgal (B), and pSVβgal-pp50dC266 (C). DAPI staining was used to visualize the nuclei of transfected cells.

of the last 16 amino acids of pp50 did not abolish nuclear localization (dC395 in Fig. 5A). In contrast, the dN202-KK405AA mutant, in which the same lysine residues in the mutant dN202 were mutated, was detectable only in the cytoplasm (Fig. 8A), implying that these two basic residues are indeed an essential element of a NLS at the C terminus of pp50. A comparison between the mutants KK405AA and dN202-KK405AA suggests that additional determinants within the first 201 amino acids of pp50, representing the N-terminal amino acid sequences present in the KK405AA mutant but not in dN202-KK405AA, may influence the nuclear localization of pp50. By introducing the same KK to AA substitutions into N-terminal deletion mutants, we found that dN27-KK405AA was the mutant with the smallest N-terminal deletion that still failed to localize to the nucleus (Fig. 8C), suggesting that determinants present within the first 26 amino acids of the protein could affect nuclear localization in the absence of a functional C-terminal NLS.

Effect of mutations within the first 26 amino acids at the N terminus on nuclear localization

We considered two possibilities when residues within the first 26 amino acids of pp50 were selected for mutagenesis. First, the presence of a minor NLS could direct the import of a C-terminal truncated pp50 to the nucleus in the absence of the KKQK motif at the C terminus. In this scenario, the RKVR sequence starting at residue 5 constitutes the longest stretch of basic amino acids within the first 26 amino acids as well as conserved Region 1 and could potentially be a key element in a N-terminal NLS. This hypothesis could be tested by mutagenesis of the RK residues. Second, specific amino acid residues within that region could be essential for the proper folding of the protein and indirectly influence nuclear localization. We have selected the PP and YK residues starting at positions 12 and 22, respectively, for mutagenesis because prolines, charged residues, and amino acids with aromatic side chains such as tyrosines are often important determinants of protein structure. Furthermore, all three pairs of our targeted amino acids (RK, PP, and YK) are conserved between MCMV pp50 and HCMV ppUL44 (Fig. 7A). Thus double mutants in which RK, PP, or YK residues of the KK405AA mutant were substituted with SS, GA, or TG residues, respectively, were constructed. The structures of these mutants are shown in Fig. 7A.

As shown in Fig. 8D, the double mutant RK5SS-KK405AA was detectable exclusively in nuclei of transfected cells, implying that the R and K residues at positions 5 and 6 are not likely to be part of a NLS. The subcellular location of the double mutant PP12GA-KK405AA was predominantly nuclear although faint cytoplasmic staining was observed in a significant percentage of positive cells (Fig. 8E). Finally, the double mutant YK22TG-KK405AA appeared to be evenly distributed between the nucleus and cytoplasm of transfected cells (Fig. 8F). Thus the Y or K residues, or both residues together, could conceivably influence nuclear localization by affecting proper protein folding.

Effect of internal deletions within the conserved N-terminal domain on nuclear localization

The pair of YK residues at position 22–23 and the amino acids around them do not have the characteristics of a "classical" NLS. Therefore it is probable that the observed changes in subcellular localization are the result of conformational changes of the protein induced by their substitution with TG residues. We reasoned that internal deletions within the highly conserved N-terminal half of the protein with the potential of perturbing the stable conformation of pp50 would also influence its subcellular localization in a similar manner. For example, if amino acids interacting with the YK residues at posi-

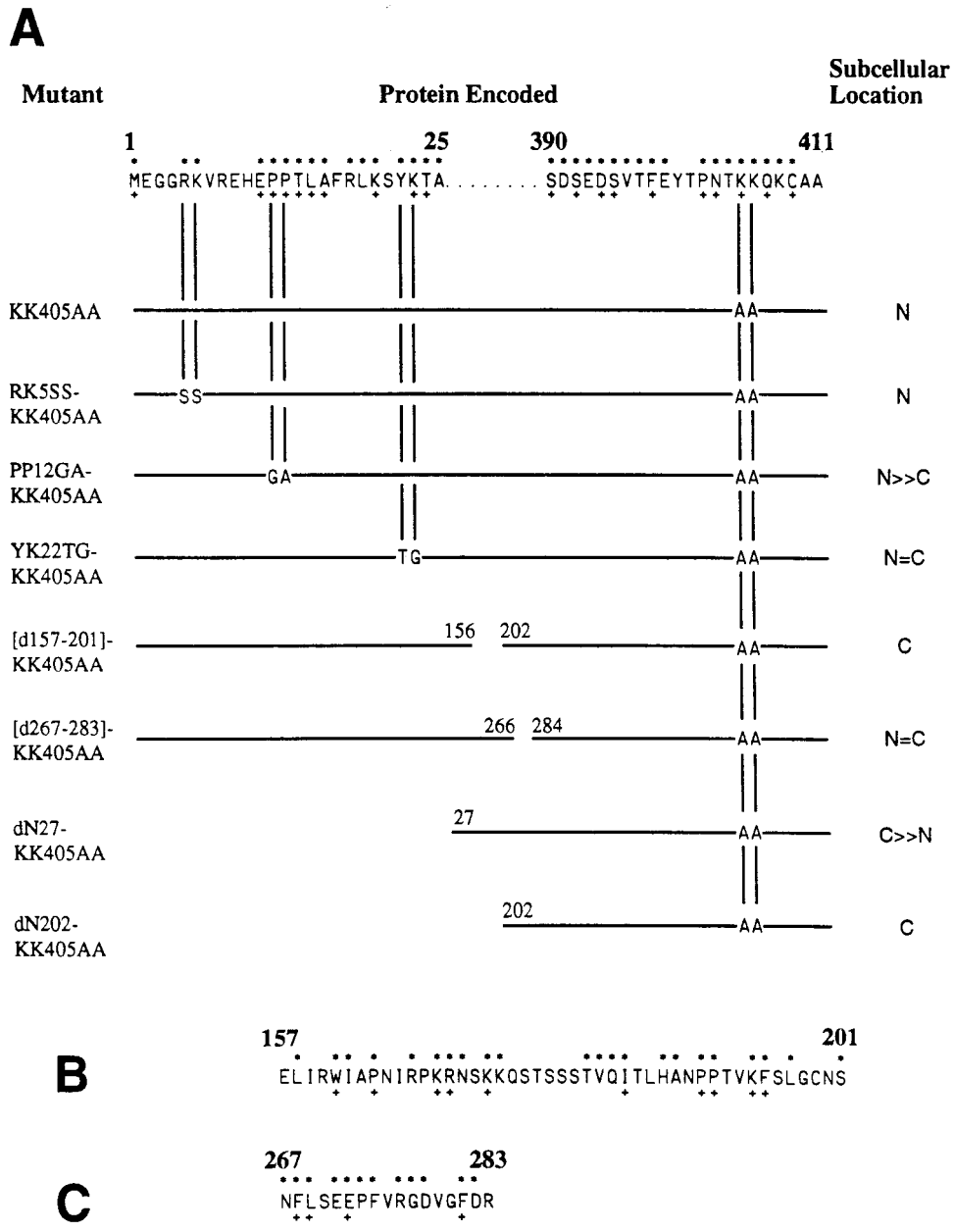

FIG. 7. Construction of site-specific pp50 mutants. (A) The first 25 and last 22 amino acids of pp50 are shown on the first line. Specific pairs of amino acid substitutions are shown for each mutant used in this study. Mutants are labeled according to the scheme described under Materials and Methods section. For deletion mutants (last four lines), numbers above the lines represent pp50 residue numbers. Subcellular locations of mutant proteins were determined by indirect immunofluorescence 48 h after transfecting COS-1 cells with the appropriate plasmids. Abbreviations used for describing subcellular locations are the same as those in Fig. 1. (B) Amino acids 157–201 (sequence deleted in d157–201). (C) Amino acids 267–283 (sequence deleted in d267–183). In all three panels, amino acids that are conserved among MCMV, HCMV, and HHV-6 are marked (+) underneath, whereas amino acids conserved only between MCMV and HCMV are marked (*) above. Numbers above each sequence represent the positions of the first and last amino acids shown.

tion 22–23 were deleted, the mutant would likely be impaired in nuclear import. To this end, mutants d157–201 and d267–283 were constructed with or without the KK405AA mutations, and their subcellular locations in transfected cells were determined by immunofluorescence microscopy. The structures of the four mutants are illustrated in Figs. 1 and 7A, and the internal sequences deleted from these mutants are shown in Figs. 7B and 7C.

These particular deletions were selected for initial analysis based on the following observations. First, residues 157–201 represent a relatively hydrophilic stretch of amino acids linking the more hydrophobic regions 1 and 2 (Loh *et al.,* 1994). Therefore it could function as a

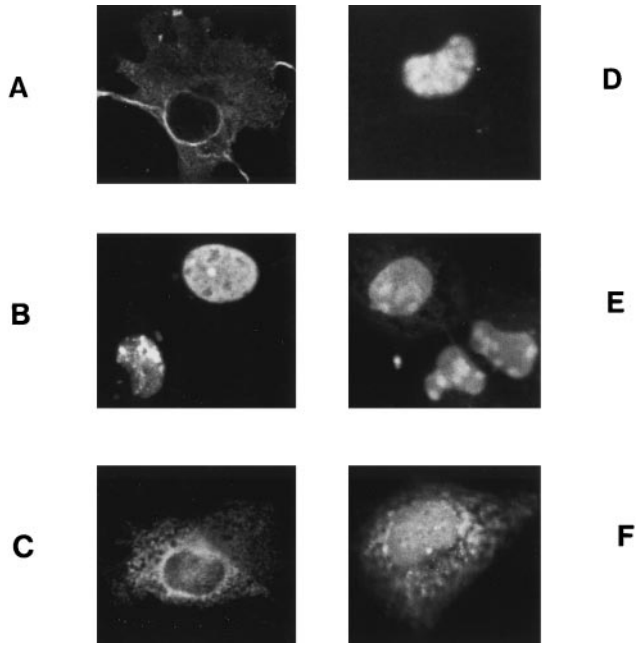

FIG. 8. Identification of amino acid residues affecting nuclear localization. COS-1 cells transfected with plasmids encoding pp50 carrying site-specific mutations were processed for immunofluorescence experiments 48 h after transfection. The plasmids used were pSVpp50dN202-KK405AA (A), pSVpp50-KK405AA (B), pSVpp50dN27-KK405AA (C), pSVpp50-RK5SS-KK405AA (D), pSVpp50-PP12GA-KK405AA (E), and pSVpp50-YK22TG-KK405AA (F). The subcellular locations of pp50 mutants were visualized by staining with MAb 3B9.

"linker" separating two independently folding and probably interacting domains. In the HSV-1 UL42 protein, insertion mutagenesis of a similar linker sequence abolished DNA polymerase binding and processivity (Digard *et al.,* 1993). Similarly, deletion of this stretch of amino acids may disrupt potential interactions between conserved Regions 1 and 2 by altering the distance separating the two domains. Second, our studies on C-terminal deletion mutants have already demonstrated that the loss of sequences between residues 267 and 280 severely impaired nuclear localization, possibly as a result of changes in protein conformation (compare dC266 and dC280 in Fig. 1).

The subcellular location of mutant d157–201 as determined by immunofluorescence microscopy was characterized by bright granular cytoplasmic staining as well as weak, diffuse nuclear staining and bright nuclear spots (Fig. 9A). Incorporation of the KK405AA mutations completely abolished nuclear localization (Fig. 9B). In contrast, the mutant d267–283 was resident entirely in the nuclei of transfected cells (Fig. 9C). Incorporation of the KK405AA mutations significantly impaired nuclear localization, showing a granular cytoplasmic staining pattern plus diffuse nuclear staining (Fig. 9D).

Thus the deletion of amino acids 157–201 alone is sufficient to impair nuclear localization. It is conceivable that the loss of these residues may introduce global conformational changes such that surface exposure of distal sequences like the C-terminal NLS is affected. In contrast, the deletion of amino acids 267–283 alone has little observable effect on nuclear localization (Fig. 9C). Perhaps the loss of these residues introduces conformational changes only within the N-terminal portion of pp50, resulting in the "masking" of a putative N-terminal NLS or the loss of association with cellular proteins that could facilitate the nuclear import of pp50. In that sense, the mutant d267–283 is similar to dN27 (Figs. 1 and 2C) in that the effect of the deletion is only obvious when the C-terminal NLS is nonfunctional.

Subcellular locations of pp50 mutants in a vaccinia/T7 expression system

One of the pathways for nuclear localization we have not considered thus far is passive diffusion into and retention in the nucleus through interactions with nuclear components. Because pp50 is a DNA-binding protein, it is conceivable that mutations affecting the dimerization but not DNA-binding properties of the protein could allow the monomeric form of pp50 to diffuse into the nucleus and be retained there by binding to cellular DNA. This hypothesis may account for the previously observed nuclear localization of mutants dC395, dC296, and

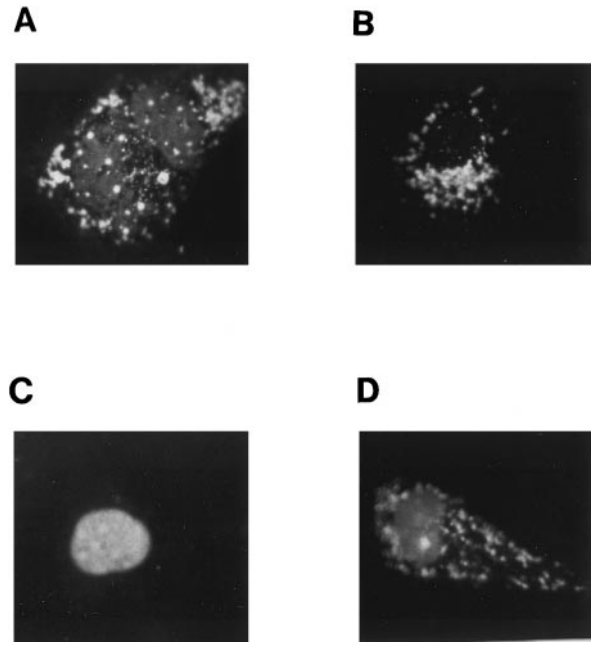

FIG. 9. Effect of internal deletions on the nuclear localization of pp50.COS-1 cells transfected with plasmids encoding pp50 carrying internal deletions within the highly conserved N-terminal domains were processed for immunofluorescence with MAb 3B9 48 h after transfection. The plasmids used in these experiments were pSVpp50d157–201 (A), pSVpp50[d157–201]-KK405AA (B), pSVpp50d267–283 (C), and pSVpp50[d267–283]-KK405AA (D).

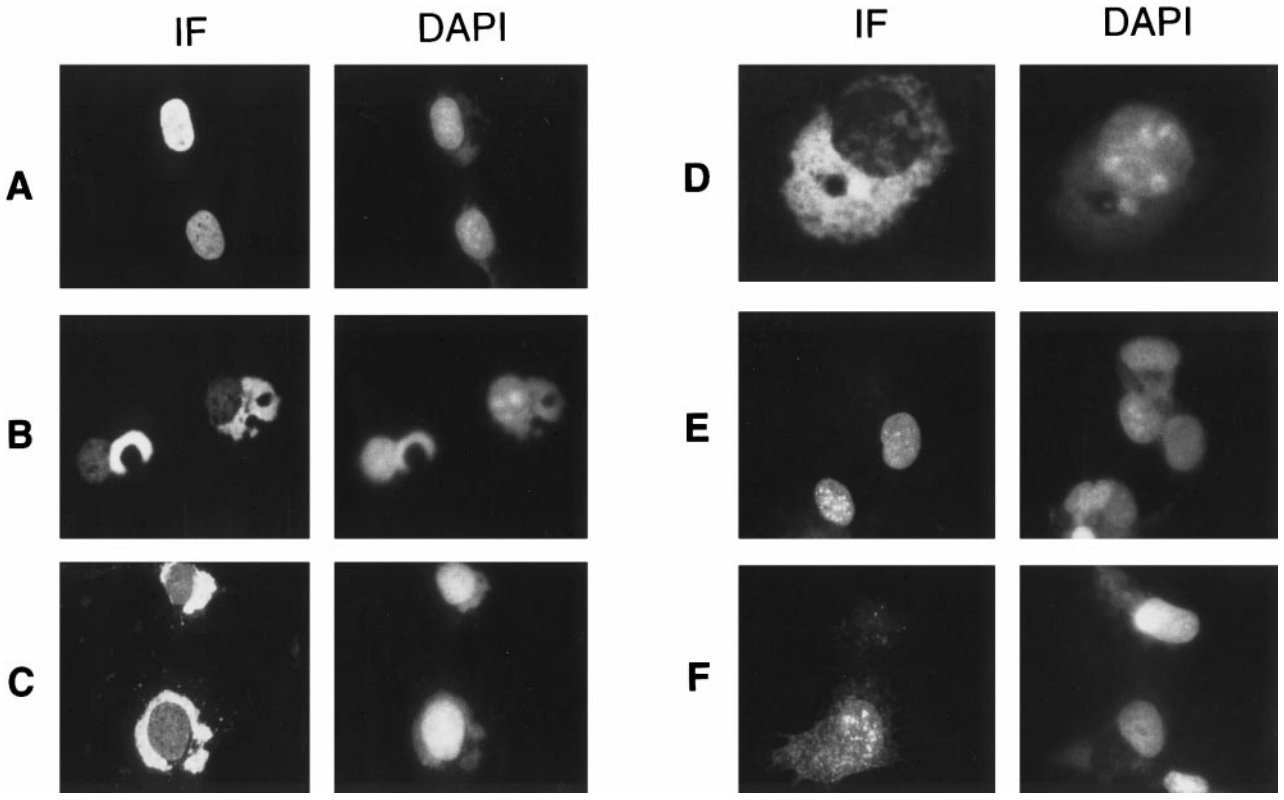

FIG. 10. Subcellular locations of pp50 mutants in a vaccinia/T7 expression system. CV-1 cells transfected with plasmids encoding wild-type (wt) and mutant pp50 ORFs under the control of the bacteriophage T7 promoter were infected with VV-T7 4 h after transfection and processed for indirect immunofluorescence (IF) with MAb 3B9 (A, B, and D–F) or 9A5 (C). The ORFs expressed in the different panels are wt pp50 (A), pp50dC395 (B), pp50dC296 (C), pp50KK405AA (D), pp50d267–283 (E), and pp50d157–201 (F).

pp50KK405AA (Figs. 5A, 5B, and 8B), which have clearly lost their C-terminal NLS. We can test this possibility indirectly by making use of the fact that vaccinia virus DNA replication takes place in the cytoplasm (Harford *et al.,* 1966), thereby allowing the cytoplasmic accumulation of viral DNA. Because passive diffusion into the nucleus is likely an inefficient process compared with active transport, mutant pp50 molecules without a functional NLS but <60,000 would be expected to at least partially localize to cytoplasmic sites of vaccinia DNA replication if their DNA-binding domain is intact. In contrast, pp50 molecules with a functional NLS would continue to be actively transported into the nucleus, showing the same subcellular distribution in the presence or absence of vaccinia virus infection. To this end, CV-1 cells were transfected with plasmids encoding various pp50 mutants under the control of the T7 promoter and subsequently infected with VV-T7 as described under Materials and Methods. The cells were then processed for indirect immunofluorescence with pp50-specific MAbs as described in the legend for Fig. 10.

As shown in Fig. 10, the expected presence of viral DNA in the cytoplasm was confirmed by DAPI staining of vaccinia-infected cells, where, in addition to the usual intense staining in the nucleus, fainter cytoplasmic staining was observed. Indirect immunofluorescence with pp50-specific MAbs revealed that wt pp50 was resident exclusively in the nuclei of vaccinia-infected cells. However, when the C-terminal NLS was rendered nonfunctional either by deletion (mutants dC395 and dC296 in Figs. 10B and 10C) or mutation of critical lysine residues (mutant pp50KK405AA in Fig. 10D), the mutant proteins were almost entirely resident in the cytoplasm. Examination of the subcellular distribution of these pp50 mutants revealed that their staining pattern roughly overlapped with the cytoplasmic DAPI staining, suggesting their colocalization with vaccinia DNA. Because these pp50 mutants possess intact N-terminal domains required for DNA-binding (Loh *et al.,* manuscript in preparation), these observations lend support to our previously stated hypothesis that the DNA-binding properties of pp50 mutants could influence their subcellular distribution. In contrast, pp50 mutants that have lost their DNA-binding ability as a result of deletions within the conserved N-terminal domains (pp50d267–283 and pp50d157–201 in Figs. 10E and 10F) showed similar subcellular distributions in the absence (Figs. 9A and 9C) or presence of vaccinia virus infection. In particular, the cytoplasmic distribution of pp50d157–201 did not overlap with the cytoplasmic DAPI staining. It should be noted

that both these mutants contain intact C-terminal NLSs and are expected to be actively transported into the nucleus if the C-terminal NLS is functional.

In summary, our data on the subcellular distribution of pp50 mutants in the vaccinia/T7 expression system and in conventional transfection experiments lend support to our hypothesis that the C-terminal NLS is the major determinant for the nuclear localization of pp50. However, in the absence of the C-terminal NLS, the DNA-binding and dimerization properties of mutant pp50 molecules may influence their subcellular localization by mediating retention in the cytoplasm or the nucleus depending on the presence of sufficient DNA in either location.

DISCUSSION

In this report, we have investigated the sequence requirements for the nuclear localization of MCMV pp50, gene product of the M44 ORF. Computer-aided analysis of the ORF revealed three isolated stretches of basic amino acids that are likely to be exposed on the surface of the protein: (1) "RKVR," starting at residue 5 (Fig. 7A, first line); (2) "KRNSKK," starting at residue 169 within a hydrophilic stretch of amino acids between Regions 1 and 2 (Fig. 7B); and (3) "KKQK," starting at residue 405 within Region 4 at the extreme C terminus of pp50 (Fig. 7A, first line). Based on this information, we hypothesized that the KKQK motif would constitute the most likely candidate for a NLS especially because a highly conserved casein kinase II phosphorylation site is nearby, further enhancing its similarity to the prototypical SV40 T-antigen NLS.

Subcellular localization experiments with our series of N- and C-terminal deletion mutants clearly supported this hypothesis (Fig. 1). Removal of the first basic cluster RKVR in dN27 and further deletion of the second basic cluster in dN202 produced gene products that were exclusively resident in the nuclei of transfected cells (Figs. 2C and 2D). Nuclear localization of dN27 or dN202 was abolished either by deletion of the last 16 amino acids of the ORF (mutants pp50[27–395] and pp50[202–395], Figs. 2E, 2F, and 3) or, more significantly, by the substitution of a pair of lysine residues at position 405–406 with alanines (mutants dN27-KK405AA and dN202-KK405AA, Figs. 7A, 8A, and 8C). Our hypothesis was further strengthened by results from our experiments with β-galactosidase-pp50 fusion proteins where deletion of the last 16 amino acids of pp50 abolished nuclear localization of the chimeric protein (Figs. 3, 4B, and 4C), and the presence of the last 11 amino acids alone was sufficient to mediate nuclear import of the fusion protein βgal-pp50dN401 (Figs. 3 and 4D). Thus we concluded that the last 11 amino acids of pp50, which contain a KKQK motif, constitute a functional NLS, and the pair of lysines within that motif is an important element of the C-terminal NLS. Furthermore we have demonstrated the absolute requirement of a functional C-terminal NLS for nuclear localization even if the truncated pp50 is small enough to passively diffuse into the nucleus (compare mutants dN202, [202–395] and dN202–KK405AA in Figs. 2D, 2F, and 8A).

Our hypothesis of a single C-terminal NLS was complicated by the observation that neither the substitution of the lysines at position 405 with alanines nor the deletion of the last 16 amino acids significantly affected nuclear localization if the highly conserved N-terminal two-thirds of the molecule remained intact (mutant KK405AA in Fig. 8B and mutant dC395 in Fig. 5A). Therefore it is likely that the putative C-terminal NLS functions together with additional determinants within the N-terminal domains of pp50 to mediate active transport of the protein into the nucleus. By introducing the KK405AA mutations into N-terminal deletion mutants as well as internal deletion mutants to inactivate the C-terminal NLS, we were able to identify three sets of N-terminal sequences that might play a role in nuclear localization in the absence of a functional C-terminal NLS. These sequences are: (1) the first 26 amino acids of pp50 (mutant dN27-KK405AA in Fig. 8C); (2) amino acids 157–201 located between Regions 1 and 2 (mutant [d157–201]-KK405AA in Fig. 9B); and (3) amino acids 267–283 at the C-terminal end of Region 2 (mutant [d267–283]-KK405AA in Fig. 9D).

Examination of these N-terminal sequences led us to speculate on their possible influences on nuclear localization. First, the presence of additional nuclear localization signals within these sequences could mediate nuclear import in the absence of the C-terminal NLS. For example, the tumor suppressor p53 is known to possess multiple NLSs within its coding sequences (Shaulsky *et al.,* 1990). Examination of the three sets of N-terminal sequences of pp50 described in the previous paragraph led to the following conclusions. There are no significant stretches of basic amino acids between residues 267 and 283 of the pp50 ORF. It is also unlikely that the RKVR sequence (positions 5–8) within the first 26 amino acids would function as a NLS because substitution of the RK residues with serines did not affect nuclear localization in the absence of a functional C-terminal NLS (mutant RK5SS-KK405AA in Fig. 8D). On the other hand, deletion of residues 157–201, which contains a stretch of basic amino acids KRNSKK, significantly impaired nuclear localization in the presence of an intact C-terminal NLS (mutant d157–201 in Figs. 9A and 10F) and completely abolished nuclear localization in the absence of a functional C-terminal NLS (mutant [d157–201]-KK405AA in Fig. 9B). Thus our data did not exclude the existence of a second NLS located between residues 157 and 201.

However, it should be noted that the intact N-terminal domain of pp50 cannot direct the nuclear import of a β-galactosidase-pp50 fusion protein (Fig. 6), making it quite unlikely that a NLS exists within this portion of the molecule. Nevertheless, it is possible that our failure to observe nuclear localization of the chimeric proteins may be the result of structural constraints introduced when the N-terminal sequences of pp50 are fused with a much larger protein like β-galactosidase, especially in view of the demonstrated sensitivity of the putative N-terminal NLS to small deletions at the N-terminal end (amino acids 1–26) or C-terminal end (amino acids 267–283) of this highly conserved section of the molecule. Therefore, to resolve the question of whether a N-terminal NLS exists, future experiments should be directed toward investigating the effect of mutating the basic residues within the KRNSKK sequence on nuclear localization.

A second possibility that may be applicable to all three sets of N-terminal sequences is that specific amino acids within these sequences may be essential for maintaining a proper and stable conformation of the molecule such that the C-terminal NLS is available for interaction with the importin heterodimer. For example, it has been hypothesized that deletion of sequences within the N terminus of HSV-1 ICP8 might alter the conformation of the protein and affect nuclear localization even though the C-terminal NLS was intact (Gao and Knipes, 1992). In another example, disruption of an alpha helix at the N terminus of HIV Vpr was sufficient to impair nuclear localization (Yao *et al.,* 1995). In our site-directed mutagenesis experiments designed to examine the effect of amino acid substitutions within the first 26 amino acids of pp50, the YK-to-TG mutations at residues 22 and 23 had the most significant effect on nuclear localization (mutant YK22TG-KK405AA in Fig.e 8F), followed by the more subtle effect seen with the PP-to-GA substitutions at residues 12 and 13 (mutant PP12GA-KK405AA in Fig. 8E). It is conceivable that the PP to GA substitutions may alter the rigid turns imposed by two consecutive prolines to a more flexible and perhaps less stable structure. In contrast, the Y-to-T substitution preserves the hydroxyl group but removes the aromatic ring structure, which could play a role in stabilizing protein folding via interactions with other aromatic and/or hydrophobic residues. Furthermore the K-to-G substitution could disrupt interactions with other negatively charged residues (e.g., carboxyl groups) and/or increase local flexibility. Thus both pairs of amino acid substitutions could potentially induce conformational changes in the protein and hence alter its subcellular location by affecting the surface exposure of putative nuclear localization signals. Similarly, the other two deleted sequences (residues 157–201 and 267–283 in Figs. 7B and 7C) also contain highly conserved prolines (residues 164, 190, and 191) and amino acids with aromatic side chains (tryptophan at residue 161, and phenylalanines at residues 195, 268, and 281) that could be important determinants in ensuring correct protein folding. As a result of the loss of one or more of these amino acids, the putative N-terminal NLS (probably the KRNSKK sequence at positions 169–174) may no longer be accessible for binding to the importin heterodimer. The effect of deleting amino acids 157–201 is particularly severe in that the mutant d157–201 is impaired in nuclear localization (Fig. 9A) even with an intact C-terminal NLS. It is conceivable that proper folding of the pp50 molecule requires specific interactions between the two highly conserved Regions 1 and 2, and removal of the linking sequence (residues 157–201) could alter the critical spacing between the two domains, resulting in a grossly misfolded protein and the masking of the C-terminal NLS.

Third, conformational changes induced in the N-terminal domains as a result of mutations described above could lead to the disruption of other biochemical properties of pp50 that might influence nuclear localization. For example, results from other experiments conducted in this laboratory suggested that both dimerization and the DNA-binding properties of pp50 may be affected by deletions into or within the highly conserved N-terminal domains (Loh *et al.,* manuscript in preparation). In the absence of a C-terminal NLS, monomeric pp50 mutants resulting from defects in dimerization or some of the more heavily truncated pp50 mutants under the 60,000 size limit could passively diffuse into and be retained in the nucleus by binding to cellular DNA if the N-terminal DNA-binding domain is intact (e.g., mutants dC395, dC296, and pp50KK405AA in Figs. 5A, 5B, and 8B). This conjecture is reinforced by results from subcellular localization experiments in the presence of vaccinia virus infection where nuclear import of these mutants was blocked in the presence of vaccinia virus infection. The observed cytoplasmic location of these mutants under such circumstances could be explained if the newly synthesized mutant proteins were retained at sites of viral DNA replication in the cytoplasm before they could diffuse into the nucleus as a result of their intact DNA-binding domains (Figs. 10B–10D). On the other hand, regardless of whether their sizes are above or below the passive diffusion limit, mutants that are defective in DNA-binding can only localize to the nucleus if they possess a functional C-terminal NLS (mutants dN27, dN202, and d267–283 with and without functional C-terminal NLSs, Figs. 2, 8, and 9). Otherwise, they remain resident in the cytoplasm.

Last, the possibility exists that improper folding could impair specific interactions with other cellular proteins that possess NLSs and prevent nuclear import of pp50 as part of a protein complex. Such protein–protein interactions might explain why pp50 mutants lacking the

C-terminal NLS could enter the nucleus even if their sizes are >60,000. We have not explored this aspect in depth because the identity of cellular proteins that can associate with pp50 is unknown.

In summary, our results suggested that the nuclear localization of pp50 depends on three factors: (1) the presence of the C-terminal NLS centered around the KKQK motif and (2) the structural integrity of the highly conserved N-terminal half of the protein encompassing Regions 1 and 2, (3) the unlikely but possible existence of a minor N-terminal NLS. The potential influence of biochemical properties such as DNA-binding and dimerization on nuclear localization could be important considerations in the construction of pp50 mutants for evaluating its role in the virus replication cycle.

MATERIALS AND METHODS

Cells and DNA transfections

COS-1 and CV-1 cells used in the transfection experiments were maintained in Dulbecco's minimum essential media (DMEM) supplemented with 10% fetal bovine serum. Transfection of plasmid DNA into COS-1 cells was carried out by electroporation as previously described (Loh *et al.,* 1994). Cells were harvested 48 h posttranfection for indirect immunofluorescence experiments or solubilized in immunoprecipitation buffer (Tris-buffered saline pH 8, supplemented with 1% NP-40, 1% deoxycholate, and 1 mM phenylmethylsulfonyl fluoride), and clarified by centrifugation in a microfuge for 20 min at 15,000 *g* in preparation for Western-blotting experiments.

The vaccinia/T7 expression system described by Fuerst *et al.,* (1986) was also used as a transient expression system for determining the subcellular location of wt or mutant pp50 proteins. Briefly, CV-1 or COS-1 cells were transfected with plasmid DNA containing wt or mutant pp50 ORFs under the control of the bacteriophage T7 promoter. Transfected cells were allowed to attach to tissue culture plates for 4 h before infection with VV-T7 (recombinant vaccinia virus stably expressing the T7 RNA polymerase, a gift from Dr. William Britt, University of Alabama) at a m.o.i. of 5. Infected cells were harvested 1 day later for immunofluorescence or Western-blotting experiments.

Monoclonal antibodies

Monoclonal antibodies (MAbs) were prepared as previously described (Chang and Balachandran, 1991; Loh *et al.,* 1991; Pande *et al.,* 1991). In particular, MAb 9A5 was raised against HHV-6 p41 but was cross-reactive with determinants present on MCMV pp50 (Loh *et al.,* 1994). Monoclonal antibodies 3B9, 5H10, 25G11, 33C7, and 31E11 were specific for MCMV pp50 (Loh *et al.,* 1991; Pande *et al.,* 1991; unpublished results from J. Shanley's laboratory). A monoclonal antibody against β-galactosidase was purchased from Gibco/BRL.

Indirect immunofluorescence

COS-1 cells transfected with DNA were grown on eight-well chamber slides (Gibco/BRL, Canada) or glass coverslips for 48 h before they were fixed in methanol and acetone at −20°C. Alternatively, cells on coverslips were fixed in 3.75% formaldehyde for 5 min and permeabilized by a 5-min incubation with phosphate-buffered saline (PBS) supplemented with 1% Triton X-100. The subcellular location of mutant pp50 was determined by staining with appropriate MAbs, and the nuclei of transfected cells were visualized by incubation with the fluorescent DNA-binding dye 4′,6-diamidino-2-phenylindole (DAPI, 1 μg/ml in a solution consisting of 50% glycerol and *p*-phenylene diamine at 1 mg/ml). In typical transfection experiments, 30–50% of the cells would be positive for expression of the protein encoded by the plasmid and representative photographs were taken with Tmax 400 film (Kodak).

Gel electrophoresis and Western blotting

SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and detection of mutant forms of pp50 by Western blotting were carried out as described in Loh *et al.,* (1991) using monoclonal antibodies described in a previous section.

Nomenclature for deletion and site-specific mutants

We have adopted a labelling scheme similar to that used by Digard *et al.* (1993). N-terminal deletion mutants and C-terminal deletion mutants were labeled dNX or dCX, respectively, where X refers to the first pp50-specific amino acid present after the deletion. Internal deletion mutants are labeled dX-Y when an amino acid sequence beginning with residue X and ending with residue Y is deleted. Site-specific mutants involving substitution of two consecutive amino acid residues are described as KKXAA if amino acids KK starting at residue X are substituted with amino acids AA. Expression plasmids in which these mutant ORFs are placed behind the SV40 promoter would be labeled as pSVpp50[ ] where [ ] represents the name of the ORF. For example, an expression plasmid encoding the mutant dN395 would be named pSVpp50dN395.

Site-directed mutagenesis

Site-directed mutagenesis was used to introduce useful restriction sites or specific mutations into the coding sequence of pp50. Both the Muta-Gene Phagemid *in vitro* mutagenesis kit (Bio-Rad) and the QuikChange Site-

directed mutagenesis kit (Stratagene) were used successfully for these experiments. In most cases, the mutagenesis primers were designed such that the mutated plasmids could be screened by the presence of new restriction sites. Subsequently the site-specific mutations introduced were verified by nucleotide sequencing, and the size and reactivity of the expressed mutant protein with MAbs were confirmed by Western blotting. Nucleotide sequencing of the mutant ORFs has not revealed any adventitious mutations arising elsewhere. The plasmid pTB16, which contained the intact ORF for pp50 (Loh *et al.,* 1994), was used as the initial template for site-directed mutagenesis. A unique *Nco*I site was introduced into the translation start codon of pp50 using the mutagenesis primer CATCTCCGCGGCCATGGAGGGTGGT to produce the plasmid pTB16N. This plasmid was used as the template for the construction of most of the site-specific mutants used in this study. The presence of unique *Nco*I and *Hin*dIII sites flanking the pp50 ORF facilitated the transfer of mutant ORFs into expression vectors.

To allow the excision of a 0.31-kb DNA fragment from pTB16N encoding the last 11 amino acids of pp50, site-directed mutagenesis was used to introduce a unique *Eco*RI site into the pp50 ORF using the mutagenesis primer GTGACCTTCGAATTCACTCCCAACACTA. This permitted the subsequent construction of plasmid pSVβgal-pp50dN401, which encodes a chimeric protein consisting of the C-terminal 11 amino acids of pp50 fused to the C terminus of β-galactosidase.

Construction of site-specific mutants

Mutations introduced into pairs of amino acids within the pp50 ORF were constructed as follows. The mutagenesis primer CACTCCCAACACTGCAGCCCAAAAGTGCGCGGCCTG was used to construct the mutant pp50-KK405AA where lysine residues at positions 405 and 406 were substituted with alanine residues as indicated by the underlined nucleotides. Similarly, the primers GGCCATGGAGGGTGGGAGCTCAGTTCGCGAGCACG, CGAGCACGAAGGCGCCACCCTGGCCTTTC, and CCTCAAGTCGACCGGGACCGCCATCCAGCAG were used to construct the mutants pp50-RK5SS, pp50-PP12GA, and pp50-YK22TG, respectively. Double mutants were constructed either by site-directed mutagenesis of single mutants, or whenever possible, by making use of unique *Nar*I or *Sph*I sites within the pp50 ORF to join together N- and C-terminal halves of pp50 coding sequences carrying the appropriate mutations. A diagram illustrating the mutations introduced into the pp50 ORF is shown in Fig. 7A.

Expression of site-specific mutants in COS-1 cells was accomplished by inserting a 1.6-kb *Bam*HI fragment containing the entire ORF of the pp50 mutant between the *Bgl*II-*Bam*HI cloning sites of the plasmid vector pM2, thereby placing the pp50 ORF behind the SV40 early promoter.

Construction of pp50 deletion mutants

The plasmid pTB16 was used as the template for the construction of nested sets of N- and C-terminal deletions as previously described (Henikoff, 1987). Appropriate choices of restriction enzymes ensured the preferential deletion into the coding sequences of the pp50 ORF while leaving the unique flanking *Eco*RI and *Hin*dIII sites intact so that the deletion constructs could be easily excised. Briefly, in constructing C-terminal deletion mutants, the plasmid was cleaved with *Pst*I and *Sma*I before digestion with Exonuclease III. Similarly, in the construction of N-terminal deletion mutants, cleavage with *Kpn*I and *Bam*HI before Exonuclease III digestion ensured the preferential deletion of coding sequences at the N terminus of pp50. The exact extent of the deletion was determined by nucleotide sequencing. For the expression of C-terminal deletion mutants, the truncated coding sequence was excised by *Bam*HI-*Hin*dIII digestion and inserted into the unique *Bgl*II-*Hin*dIII sites of the plasmid vector pM2 (Sadowski *et al.,* 1992), thereby positioning the SV40 early promoter in front of the translation start site and providing a universal translation termination sequence behind the truncated coding sequence. However, as shown in Fig. 1, extraneous amino acids were inevitably introduced into the C terminus of the mutants. In the case of N-terminal deletions, the truncated sequence was excised by *Eco*RI-*Hin*dIII digestion and inserted in-frame behind the six histidine residues (his-tag) of pRSET vectors (Invitrogen). The unique *Nco*I site of the vector then provided a translation start site immediately upstream of the truncated pp50 coding sequence. As a result, an ORF encoding a N-terminal truncated pp50 protein without the his-tag could be excised as a *Bam*HI-*Hin*dIII fragment and inserted into the pM2 vector as described above. Again, extra amino acids were introduced at the N terminus of the mutants as illustrated in Fig. 1.

Mutant pp50 carrying internal deletions were constructed as follows. Excision of a 135-bp *Sst*I fragment within the coding sequence of pp50 from pTB16N and religation produced the plasmid pTB16NdSst. The shortened ORF was placed behind the SV40 promoter in pM2 to create the expression plasmid pSVpp50d157–201. Similarly, the plasmid pSVpp50d267–283 was constructed by inserting a 0.6-kb *Sph*I-*Hin*dIII fragment representing the coding sequence for the hydrophilic C-terminal tail of pp50 in frame behind the C-terminal-truncated ORF found in the plasmid pSVpp50dC266.

The predicted sizes of all deletion mutants were confirmed by Western blotting. The names of deletion mutants used in epitope-mapping experiments are shown in

the first column of Table 1, whereas the structures of deletion mutants used in subcellular localization experiments are illustrated in Fig. 1.

Construction of β-galactosidase fusion proteins

The β-galactosidase gene was modified to facilitate the construction of fusion proteins. Sequences containing unique restriction sites (*Nco*I, *Bam*HI, and *Sma*I) replaced those encoding the first nine amino acids of the gene. As a result, the first five amino acids of the modified β-galactosidase in single letter codes are MARDP (Fig. 3), being followed immediately by the 10th amino acid of authentic β-galactosidase. The chimera pp50dC296-βgal was constructed by inserting the coding sequence for the mutant dC296 between the *Nco*I and *Sma*I sites at the 5′ end of the β-galactosidase gene. Alternately, a unique *Eco*RI site located at the extreme 3′ end of the β-galactosidase gene was used for the insertion of DNA fragments to generate chimeric proteins consisting of the first 1004 amino acids of β-galactosidase fused in frame with defined segments of pp50. The entire ORF of the chimeric protein could then be excised as a *Nco*I-*Hin*dIII fragment and placed behind the SV40 promoter in the plasmid vector pM2 for expression in COS-1 cells. The structures of chimeric proteins used in this study are illustrated in Fig. 3.

ACKNOWLEDGMENTS

We thank Dr. Bala Chandran for his gift of MAb 9A5. This work was supported by Grant MT-8196 from the Medical Research Council of Canada (MRC) and Saskatchewan Health/MRC bridge funding to L. Loh.

REFERENCES

Adam, S.A ., and Gerace, L. (1991). Cytosolic proteins that specifically bind nuclear localization signal are receptors for nuclear import. *Cell* **66,** 837–847.

Adam, S. A., Sterne-Marr, R. E., and Gerace, L. (1990). Nuclear protein import in permeabilized mammalian cells requires soluble cytoplasmic factors. *J. Cell Biol.* **111,** 807–816.

Chang, C. K., and Balachandran, N. (1991). Identification, characterization and sequence analysis of a cDNA encoding a phosphoprotein of human herpesvirus 6. *J. Virol.* **65,** 2884–2894; correction **65,** 7085.

Digard, P., Chow, C. S., Pirrit, L., and Coen, D. M. (1993). Functional analysis of the herpes simplex virus UL42 protein. *J. Virol.* **67,** 1159–1168.

Dingwall, C., Robbins, J., Dilworth, S. M., Roberts, B., and Richardson, W. D. (1988). The nucleoplasmin nuclear location sequence is larger and more complex than that of SV40 large T antigen. *J. Cell Biol.* **107,** 841–849.

Dingwall, C., and Laskey, R. A. (1986). Protein import into the cell nucleus. *Annu. Rev. Cell Biol.* **2,** 367–390.

Ertl, P. F., and Powell, K. L. (1992). Physical and functional interaction of human cytomegalovirus DNA polymerase and its accessory protein (ICP36) expressed in insect cells. *J. Virol.* **66,** 4126–4133.

Fuerst, T. R., Niles, E. G., Studier, F. W., and Moss, B. (1986). Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA* **83,** 8122–8126.

Gao, M., and Knipe, D. M. (1992). Distal protein sequences can affect the function of a nuclear localization signal. *Mol. Cell. Biol.* **12,** 1330–1339.

Görlich, D. (1997). Nuclear protein import. *Curr. Opin. Cell Biol.* **9,** 412–419.

Görlich, D., and Mattaj, I. W. (1996). Nucleocytoplasmic transport. *Science* **271,** 1513–1518.

Görlich, D., Prehn S., Laskey, R., and Hartmann, E. (1994). Isolation of a protein that is essential for the first step of nuclear protein import. *Cell* **79,** 767–778.

Harford, C., Hamlin, A., and Riders, E. (1966). Electron microscopic autoradiography of DNA synthesis in cell infected with caccinia virus. *Exp. Cell Res.* **42,** 50–57.

Henikoff, S. (1987). Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods Enzymol.* **155,** 156–165.

Kalderon, D., Richardson, W. D., Markham, A. F., and Smith, A. E. (1984a). Sequence requirements for nuclear localization of SV40 large T antigen. *Nature* **311,** 33–38.

Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E. (1984b). A short amino acid sequence able to specify nuclear location. *Cell* **39,** 499–509.

Loh, L. C., Balachandran, N., and Britt, W. J. (1991). Characterization of a membrane-associated phosphoprotein of murine cytomegalovirus (pp50) and its immunological cross-reactivity with a human cytomegalovirus protein. *Virology* **183,** 181–194.

Loh, L. C., Britt, W. J., Raggo, C., and Laferté, S. (1994). Sequence analysis and expression of the murine cytomegalovirus phosphoprotein pp50, a homolog of the human cytomegalovirus UL44 gene product. *Virology* **200,** 413–427.

Pande, H., Campo, K., Shanley, J. D., Creeger, E. S., Artishevsky, A., Gallez-Hawkins, G., and Zaia, J. A. (1991). Characterization of a 52K protein of murine cytomegalovirus and its immunological cross-reactivity with the DNA-binding protein ICP36 of human cytomegalovirus. *J. Gen. Virol.* **72,** 1421–1427.

Pari, G. S., and Anders, D. G. (1993). Eleven loci encoding trans-acting factors are required for transient complementation of human cytomegalovirus oriLyt-dependent DNA replication. *J. Virol.* **67,** 6979–6988.

Pari, G. S., Kacica, M. A., and Anders, D. G. (1993). Open reading frames UL44, IRS1/TRS1, and UL36–38 are required for transient complementation of human cytomegalovirus oriLyt-dependent DNA synthesis. *J. Virol.* **67,** 2575–2582.

Picard, D., and Yamamoto, K. R. (1987). Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *EMBO J.* **6,** 3333–3340.

Plachter, B., Nordin, M., Wirgart, B. Z., Mach, M., Stein, H., Grillner, L., and Jahn, G. (1992). The DNA-binding protein P52 of human cytomegalovirus reacts with monoclonal antibody CCH2 and associates with the nuclear membrane at late times after infection. *Virus Res.* **24,** 265–276.

Pollard, V. W., Michael, W. M., Nakielny, S., Siomi, M. C., Wang, F., and Dreyfuss, G. (1996). A novel receptor-mediated nuclear protein import pathway. *Cell* **86,** 985–994.

Rawlinson, W. D., Farrell, H. E., and Barrell, B. G. (1996). Analysis of the complete DNA sequence of murine cytomegalovirus. *J. Virol.* **70,** 8833–8849.

Rihs, H., Jans, D. A., Fan, H., and Peters, R. (1991). The rate of nuclear cytoplasmic protein transport is determined by the casein kinase II site flanking nuclear localization sequence of the SV40 T-antigen. *EMBO J.* **10,** 633–639.

Ripalti, A., Boccuni, M. C., Campanini, F., and Landini, M. P. (1995). Cytomegalovirus-mediated induction of antisense mRNA expression

to UL44 inhibits virus replication in an astrocytoma cell line: Identification of an essential gene. *J. Virol.* **69,** 2047–2057.

Robbins, J., Dilworth, S. M., Laskey, R. A., and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequences. *Cell* **64,** 615–623.

Roberts, B. L., Richardson, W. S., and Smith, A. E. (1987). The effect of protein context on nuclear location signal function. *Cell* **50,** 465–475.

Sadowski, I., Bell, B., Broad, P., and Hollis, M. (1992). *GAL4* fusion vectors for expression in yeast or mammalian cells. *Gene* **118,** 137–141.

Shaulsky, G., Goldfinger, N., Ben-Zeev, A., and Rotter, V. (1990). Nuclear accumulation of p53 protein is mediated by several nuclear localization signals and plays a role in tumorigenesis. *Mol. Cell. Biol.* **10,** 6565–6577.

Shaulsky, G., Goldfinger, N., Tosky, M. S., Levine, A. J., and Rotter, V. (1991). Nuclear localization is essential for the activity of p53 protein. *Oncogene* **6,** 2055–2065.

Siomi, H., and Dreyfuss, G. (1995). A nuclear localization domain in the hnRNP A1 protein. *J. Cell Biol.* **129,** 551–560.

Tillman, U., Wagner, J., Boerboom, D., Westphal, H., and Tremblay, M. L. (1994). Nuclear localization and cell cycle regulation of a murine protein tyrosine phosphatase. *Mol. Cell. Biol.* **14,** 3030–3040.

Weiland, K. L., Oien, N. L., Homa, F., and Wathen, M. W. (1994). Functional analysis of human cytomegalovirus polymerase accessory protein. *Virus Res.* **34,** 191–206.

Yao, X. J., Subbramanian, R. A., Rougeau, N., Boisvert, F., Bergeron, D., and Cohen, E. A. (1995). Mutagenic analysis of human immunodeficiency virus type 1 Vpr: Role of a predicted N-terminal alpha-helical structure in Vpr nuclear localization and virion incorporation. *J. Virol.* **69,** 7032–7044.