Coat proteins of *Rice tungro bacilliform virus* and *Mungbean yellow mosaic virus* contain multiple nuclear-localization signals and interact with importin α

O. Guerra-Peraza,† D. Kirk,† V. Seltzer,‡ K. Veluthambi,§ A. C. Schmit,§ T. Hohn,†§ and E. Herzog†§

1Friedrich Miescher Institute, Maulbeerstrasse 66, 4058 Basel, Switzerland
2Institut de Biologie Moléculaire des Plantes, UPR-CNRS 2357, Université Louis Pasteur, 12 rue du Général Zimmer, 67084 Strasbourg Cedex, France
3Department of Plant Biotechnology, School of Biotechnology, Madurai Kamaraj University, Madurai 625021, India
4University of Basel, Botanical Institute, Plant Health Unit, Schoenbeinstrasse 6, 4056 Basel, Switzerland

Correspondence
E. Herzog
etienne.herzog@ibmp-ulp.u-strasbg.fr

Received 25 January 2005
Accepted 7 March 2005

Transport of the viral genome into the nucleus is an obligatory step in the replication cycle of plant pararetro- and geminiviruses. In both these virus types, the multifunctional coat protein (CP) is thought to be involved in this process. Here, a green fluorescent protein tagging approach was used to demonstrate nuclear import of the CPs of *Rice tungro bacilliform virus* (RTBV) and *Mungbean yellow mosaic virus - Vigna* (MYMV) in *Nicotiana plumbaginifolia* protoplasts. In both cases, at least two nuclear localization signals (NLSs) were identified and characterized. The NLSs of RTBV CP are located within both N- and C-terminal regions (residues 479KRPK/497KRK and 744KRK/758RRK), and those of MYMV CP within the N-terminal part (residues 3KR and 41KRRR). The MYMV and RTBV CP NLSs resemble classic mono- and bipartite NLSs, respectively. However, the N-terminal MYMV CP NLS and both RTBV CP NLSs show peculiarities in the number and position of basic residues. *In vitro* pull-down assays revealed interaction of RTBV and MYMV CPs with the nuclear import factor importin α, suggesting that both CPs are imported into the nucleus via an importin α-dependent pathway. The possibility that this pathway could serve for docking of virions to the nucleus is discussed.

INTRODUCTION

Viruses with nuclear replication phases usually exploit pre-existing cellular nucleo-cytoplasmic transport machinery to target the nuclear pore, where the viral genome is imported as either a nucleic acid macromolecule or a nucleoprotein complex. Nuclear targeting of virions is of particular importance in the case of animal and plant DNA viruses (Whittaker & Helenius, 1998; Greber & Fassati, 2003; Whittaker, 2003; Smith & Helenius, 2004). *Caulimoviridae* (plant pararetroviruses; Rothnie et al., 1994) and *Geminiviridae* are the two biggest families of plant DNA viruses (van Regenmortel et al., 2000). The double-stranded DNA genome of pararetroviruses accumulates in the nucleus, where the viral RNAs are produced and then exported to the cytoplasm for translation, encapsidation and reverse transcription (Rothnie et al., 1994). Pararetroviruses have therefore evolved means of transporting their infecting genome through the cytoplasm towards, and into, the nucleus. In the hepatitisvirus *Hepatitis B virus* (HBV – an animal pararetrovirus; Rothnie et al., 1994), the precore and core proteins contain a nuclear localization signal (NLS) located within the arginine-rich carboxyl-terminal sequence (Yeh et al., 1990; Eckhardt et al., 1991). Mature HBV capsids can be imported in intact form into the nuclear basket of the nuclear pore (Rabe et al., 2003). Import is dependent on phosphorylation of the HBV coat protein (CP) (Kann et al., 1999). In *Cauliflower mosaic virus* (CaMV), an NLS in the N-terminal region of the CP is necessary for viral infectivity (Leclerc et al., 1999). CaMV particles are thought not to enter the nucleus but to release their DNA into the nucleus after docking at the nuclear pore (Karsies et al., 2002).
In geminiviruses, the viral genome (circular single-stranded DNA) enters the nucleus, where virus replication, transcription and encapsidation occurs (Gafni & Epel, 2002). Whether intact viral particles or completely or partially uncoated DNA is imported is not yet clear. For movement to adjacent cells, the viral genome has to be transported to the cell periphery. Therefore, viral DNA must be shuttled in and out of the nucleus during the infection cycle. In geminiviruses with bipartite genomes, the DNA B gene product BV1 is involved in this shuttling process (Pascal et al., 1994; Sanderfoot et al., 1996), whereas the movement protein (BC1) redirects viral DNA to the plasmodesmata for cell-to-cell movement (Sanderfoot et al., 1996; Ward & Lazarowitz, 1999). In monopartite geminiviruses, these functions are mediated by the CP and the movement protein (in mastreviruses) (Liu et al., 1997, 1999; Kotlizky et al., 2000) or the CP, V1 and C4 proteins (in monopartite begomoviruses) (Rojas et al., 2001). The CP of bipartite geminiviruses is dispensable for systemic infection of natural hosts using laboratory procedures (viral DNA bombardment or agroinfection) (Pooma et al., 1996). However, it seems likely that under natural conditions the CP could play an important role in the nuclear targeting of both monopartite and bipartite geminivirus viirions during the initial steps of infection following insect transmission. Nuclear localization of geminivirus CPs has been described in <i>Tomato yellow leaf curl virus</i> (TYLCV) (Kunik et al., 1998; Rojas et al., 2001), <i>Maize streak virus</i> (MSV) (Liu et al., 1999) and <i>African cassava mosaic virus</i> (ACMV) (Unseld et al., 2001).

The process of viral nuclear import is facilitated by the presence of NLSs on specific viral proteins. Classical NLSs belong to one of two groups: (i) monopartite, consisting of a short stretch of basic amino acid residues (Kalderon et al., 1984) and (ii) bipartite, corresponding to two basic regions separated by a spacer of variable length (Robbins et al., 1991). Other types of NLS have also been described in several cellular and viral proteins (Nakielny & Dreyfuss, 1999; Cokol et al., 2000).

In the current model for nuclear import of classical NLS-containing proteins, the transport factors importin α and β (also called α and β karyopherins) form a heterodimer that binds to the NLS-protein (cargo) in the cytoplasm (Görlich et al., 1995a; Imamoto et al., 1995). Importin α provides the NLS-binding site (Adam & Gerace, 1991; Görlich et al., 1995b; Weis et al., 1996) whereas importin β is responsible for the docking of the trimeric cargo-importin-α/β complex and its subsequent translocation through the nuclear pore complex (NPC) (Moroianu et al., 1995). This process is terminated at the nucleoplasmic side of the nuclear pore by disassembly of the cargo-importin-α/β complex, after direct binding of Ran-GTP to importin-β (Rexach & Blobel, 1995; Moroianu et al., 1996; Görlich & Mattaj, 1996). Some viral proteins containing classical NLSs follow this standard pathway, interacting first with importin α, while others can interact directly with importin β (Truant & Cullen, 1999; Greber & Fassati, 2003). Protein import by unconventional mechanisms, occurring independently of members of the α and β factor family, has also recently been reported (Nakielny & Dreyfuss, 1999; Depienne et al., 2001).

NLSs have been identified in several plant DNA virus CPs. Nevertheless, there is little information regarding the transport mechanism(s) involved in CP nuclear import. First indications have been provided by the findings that TYLCV (Kunik et al., 1999) and CaMV (Karsies et al., 2002) CPs interact with importin α in the yeast two-hybrid system.

In this study, we investigated the nuclear import strategy of <i>Rice tungro bacilliform virus</i> (RTBV) (Hay et al., 1991) and <i>Mungbean yellow mosaic virus - Vigna</i> (MYMV) (Karthikeyan et al., 2004), members of the families <i>Geminiviridae</i> and <i>Geminiviridae</i>, respectively. We found that the CPs of RTBV and MYMV each contain at least two distinct NLSs. To investigate the pathway(s) by which these CPs are imported into the nucleus, we tested their interaction with importin α. We found that the CPs of both viruses are able to interact in vitro with importin α.

**METHODS**

**Construction of recombinant plasmids**

**Plasmids used for protoplast transfections.** pCK-EGFP was prepared by replacing the green fluorescent protein (GFP) sequence of pCK-GFP565C (Miller & Lindow, 1997) by the EGFP coding sequence from pEGFP (BD Biosciences Clontech) using the NcoI and XbaI cloning sites on both plasmids. All other pCK-EGFP-derived constructs were obtained by introducing DNA fragments flanked by appropriate restriction sites generated by PCR. Oligonucleotides used for PCR amplifications of the complete RTBV and MYMV CP sequences are listed in Table 1. A double enhancer 35S CaMV promoter and the 35S CaMV terminator control expression in these vectors.

pCK-EGFPBMX was derived from pCK-EGFP. The stop codon of the EGFP open reading frame (ORF) was removed, and BamHI, MluI and XbaI restriction sites were introduced downstream of the EGFP sequence.

pCK-EGFPChS was prepared by inserting the parsley chalcone synthase (ChS) coding sequence (without stop codon) (Haasen et al., 1999) between the BamHI and MluI restriction sites of pCK-EGFPBMX. pCK-EGFPChS encodes a fusion protein (EGFPChS) of 74 kDa, with a size above the exclusion limit (~60 kDa) of the nuclear pore (Mattaj & Engelmeier, 1998). A sequence corresponding to a glycine-rich peptide (a flexible peptide that should improve the structural separation of the EGFP and ChS parts) was introduced by PCR amplification at the 5′ end of the ChS ORF (see Table 1). At the 3′ end, the plasmid sequence encodes 28 additional C-terminal residues until the first stop codon is encountered.

pRTCPG1 and pMYCPG1 were obtained by inserting the complete CP coding sequence of RTBV and MYMV, respectively, into the NcoI site of pCK-EGFP. A sequence encoding a glycine-rich peptide was introduced at the 3′ end of CP ORFs during PCR amplification.

pGChSRTCP1, -2 and -6 and pGChSMYCP1, -2, -3, -8 and -12 were prepared by introducing complete or partial RTBV and MYMV CP sequences into the MluI and XbaI sites of pCK-EGFPChS.
Table 1. Oligonucleotides used for the PCR amplification of complete EGFP, ChS, RTBV and MYMV CP ORFs

<table>
<thead>
<tr>
<th>ORF</th>
<th>Oligonucleotides*</th>
<th>Plasmids†</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFP</td>
<td>GCCACCATGGTGAAGCAAGGCGGAG (+)</td>
<td>pCK-EGFPBMX</td>
</tr>
<tr>
<td></td>
<td>CTAGTCATAGCTACGCGTTCCGGATATCCGGTACGGTATCG (-)</td>
<td>pCK-EGFPCChS</td>
</tr>
<tr>
<td>ChS</td>
<td>CCGGATCCCTAAGCAGCCGGTGAGCGAAATCATCATA (+)</td>
<td>pRTCPG1</td>
</tr>
<tr>
<td></td>
<td>GGAACGGGTGGATGAAAGTACGGCCTTCC (-)</td>
<td>pRTC1</td>
</tr>
<tr>
<td>RTBV CP</td>
<td>TCAGCCATTTGAGCCCACTAAAAGCACTCAAAG - 2440 (+)</td>
<td>pMYCP1</td>
</tr>
<tr>
<td></td>
<td>CATTGGTTCCCCACCTCTGTATATTCATATTTCTACAGCTATTTAAGATC (-)</td>
<td>pMYCP1</td>
</tr>
<tr>
<td>RTBV CP</td>
<td>TCAGCCATTTGAGCCCACTAAAAGCACTCAAAG - 2440 (+)</td>
<td>pGChS</td>
</tr>
<tr>
<td>MYMV CP</td>
<td>GAACCATGGGAAACGGAATACG (+)</td>
<td>pGChS</td>
</tr>
<tr>
<td></td>
<td>CATGGCATTGCTGCTACCCTCTGCATATTCCATTTGAATCGAATCATAA - AAATAGATGCCG - 1042 (-)</td>
<td>pGChS</td>
</tr>
<tr>
<td>MYMV CP</td>
<td>GGAACCGGGGATGCAAGGGAGGGCTGGAAATGCCAAAGGAGG - 312 (-)</td>
<td>pGChS</td>
</tr>
<tr>
<td></td>
<td>CGTTCCTGACAATTGGAATGCTGATATCCATTATGATACGAG - 1054 (-)</td>
<td>pGChS</td>
</tr>
<tr>
<td>MYMV CP</td>
<td>GAACCATGGGAAACGGAATACG (+)</td>
<td>pGChS</td>
</tr>
</tbody>
</table>
|            | GTATTGGAGTTTAGGAGGATGATGATGAGTGATGATGAGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
N- and C-terminal NLSs direct nuclear targeting of RTBV CP

To identify the RTBV CP domains responsible for nuclear localization, a series of deletion mutants was created (Fig. 2a). Mutated CP sequences were cloned as EGFPChS fusions and expressed in N. plumbaginifolia protoplasts as above. Nuclear localization results are summarized in Fig. 2(a), with representative constructs shown in Fig. 2(c).

When N- and C-terminal parts of the CP were expressed separately in fusion with EGFPChS (Fig. 2a; GChSRTCP1 and GChSRTCP2, respectively), only GChSRTCP2 localized to the nucleus. GChSRTCP1 was not imported into the nucleus, but we noted that this fusion had a tendency to form aggregates, making it impossible to draw any conclusions regarding nuclear import signals from this construct (Fig. 2c). The C-terminal region of RTBV CP contains a large conserved basic region (residues 743–769) followed by a zinc finger motif, which can be considered as the signature of all plant pararetroviral CPs (Rothnie et al., 1994). Several basic motifs present in this region could potentially constitute monopartite and/or bipartite NLSs. Three additional EGFPChS-truncated-CP fusions based on GChSRTCP2 were prepared, in which either a large part of the basic domain (GChSRTCP5) or one of two regions further upstream (GChSRTCP3 and -4) were removed (Fig. 2a). GChSRTCP3 and -5 were still able to localize within the nucleus, but GChSRTCP4 was not, indicating that the domain encompassing residues 713–749 very likely contains residues that could constitute an NLS. A stretch of basic residues (743KKRK) may form part of this signal (Fig. 2b). A similar motif also occurs just outside this region, at position 756 (756KKRRK, Fig. 2b). To investigate further the importance of these two motifs, the C-terminal CP region (735–791) was fused to EGFPChS (GChSRTCP6, Fig. 2a). GChSRTCP6 was able to accumulate very efficiently within the nucleus (Fig. 2c), allowing a minimal domain of the RTBV CP possessing an NLS to be defined (735–791). The substitution of residues 7444KKRK by alanines (GChSRTCP7) did not abolish nuclear import, indicating redundancy within the C-terminal CP portion (735–791; Fig. 2a, b). Nuclear localization was prevented only by combining KRK44AAA with alanine substitution of the 758RRK basic cluster (GChSRTCP8, Fig. 2c). Residues 758RRK are not sufficient for nuclear import (see construct GChRRTCP4). We therefore conclude that both motifs (744–746 and 758–760) are important for nuclear localization. The spacing of these two basic clusters is compatible with the formation of a bipartite-like NLS. However, the presence of both motifs is not strictly necessary for nuclear import (see GChSRTCP5 and GChSRTCP7), suggesting that other basic residues located in this region could play a role in nuclear import, forming, together with residues 744–746 and 758–760, a more complex NLS and thus minimizing the effect of the modifications introduced.

The mutations present in GChSRTCP8 were introduced in the context of the full-length CP in fusion with EGFP

RESULTS

Nuclear localization of RTBV and MYMV CPs expressed in protoplasts

The 37 kDa RTBV CP is part of a large polyprotein encoded by ORF III, from which it is excised by the viral aspartyl protease, yielding a protein that spans aa 477–791 (Hay et al., 1991; Ku et al., 1991; Marrrey et al., 1999). The MYMV CP (735–791) was encoded as a mature protein by component A of the bipartite genome (Karthikeyan et al., 2004).

The coding sequences of the RTBV and MYMV CPs were fused in-frame to the 5′-end of the EGFP ORF in the vector pCK-EGFP. Plasmids expressing the RTBV or MYMV CP-EGFP fusion proteins (pRTCPG1 and pMYCPG1, respectively) were used to transfect N. plumbaginifolia protoplasts. After about 10 h of incubation, transfected protoplasts were fixed, stained with DAPI and examined by epifluorescence microscopy. The green fluorescence seen in protoplasts expressing RTCPG1 or MYCPG1 co-localized with the DAPI signal, indicating nuclear localization of both fusion proteins (Fig. 1a). Confocal microscopy studies confirmed that both RTCPG1 and MYCPG1 accumulate exclusively within the nuclear compartment (Fig. 1b). In contrast, when EGFP was expressed alone, signal accumulation was mainly within the cytoplasm but also within the nucleus due to passive diffusion of EGFP. The larger EGFPChS protein (74 kDa – a fusion between EGFP and the cytosolic enzyme ChS; Haasen et al., 1999) is clearly excluded from the nucleus when expressed in protoplasts (Fig. 1a, b). Some sublocalization of RTCPG1 and MYCPG1 within nuclei was observed – whereas both proteins were visualized as a diffuse fluorescence in the nucleoplasm, RTCPG1 accumulated strongly in the nucleolus, and MYCPG1 at its periphery. Small aggregates of variable size could also be observed throughout the nucleus.

The molecular masses of RTCPG1 and MYCPG1, 64 and 55 kDa, respectively, are close to the exclusion limit (~60 kDa) generally ascribed to the nuclear pore (Mattaj & Englmeier, 1998). To confirm that the RTBV and MYMV CPs contain active nuclear targeting signals, either the complete CP (MYMV) or parts of it (RTBV) were expressed as tripartite fusions with EGFP and ChS (see Figs 2 and 3). In contrast to EGFPChS, GChSMYCP1, and GChSRTCP2 and -6 (see also below) were able to accumulate within the nucleus, indicating that they do indeed contain signals directing active nuclear import.
Nuclear targeting of RTBV and MYMV CPs

(RTCPG2), and a CP–EGFP fusion missing the C-terminal extremity 735–791 was also analysed (RTCPG3). RTCPG2 and -3 were still able to localize within the nucleus (Fig. 2c), indicating the presence of one or more additional NLS(s) elsewhere in the CP sequence. RTCPG2 localizes mostly around the nucleolus, but as a discontinuous signal that contrasts with the continuous signals observed in the nucleolus with the wild-type protein RTCPG1 or with RTCPG3 (Figs 1b and 2c). The N-terminal part of the RTBV CP (477–499) also contains stretches of basic amino acids that might constitute a bipartite NLS (Fig. 2b). Mutations within the N-terminal CP domain were introduced in the

---

**Fig. 1.** Subcellular localization of RTBV and MYMV CP–EGFP fusion proteins in fixed N. plumbaginifolia protoplasts. (a) Epifluorescence microscopy images of protoplasts expressing RTBV CP fused to EGFP (RTCPG1), MYMV CP fused to EGFP (MYCPG1), EGFP alone or EGFP fused to chalcone synthase (EGFPChS). Non-transfected protoplasts (mock) served as a control. Protoplasts stained with DAPI were photographed in the GFP, DAPI and bright-field channels. (b) Confocal microscopy images of protoplasts expressing RTCPG1, MYCPG1, EGFP or EGFPChS. The subcellular localizations reported here and in Figs 2 and 3 were found in more than 90% of the cells observed. Bars, 10 μm.
Fig. 2. Subcellular localization in N. plum- 
aginifolia protoplasts of RTBV CP and its 
derivatives fused to EGFP or EGFPChS. 
(a) Schematic representation of the RTBV 
CP and truncated CP versions fused to 
EGFP (RTCPG1–7) or EGFPChS 
(GChSRTCP1–8). CP and truncated CP 
forms (open boxes), EGFP and ChS (inter-
rupted dark and light grey boxes, respec-
tively), CP basic domain (hatched box), zinc 
finger motif (black dot), deletions (V) 
and replacements by alanines (asterisks). 
Accumulation of the EGFP fusion proteins 
within the nucleus (+), cytoplasmic localiza-
tion (–). (b) N- and C-terminal RTBV CP 
residues. Basic amino acids that were 
mutated to alanines are shown in bold. 
Numbers indicated above open boxes (a) and 
amino acid sequences refer to positions in 
the RTBV ORF III polyprotein. (c) Locali-
zation of the most representative EGFP and 
EGFPChS fusions as analysed by confocal 
microscopy. Bars, 10 μm.
context of the full-length CP construct fused to EGFP. Deletion of the 26 N-terminal amino acids (RTCPG4) did not abolish nuclear localization (Fig. 2a and c), most probably because the C-terminal NLS is active. Therefore, the mutations of the N-terminal domain were combined with mutations of the C-terminal NLS. RTCPG5 did not localize to the nucleus (Fig. 2a, c), showing that the 26 N-terminal amino acids are necessary for nuclear import (Fig. 2b). Furthermore, substitution of either of two basic clusters for alanines (KRPK479AAPA or KRK497AAA; RTCPG6, -7) also abolished nuclear import (Fig. 2c), indicating that these two N-terminal basic motifs could constitute a bipartite NLS (Fig. 2b).

N-terminal amino acids of MYMV CP encompass two nuclear localization signals

To characterize the NLS(s) present within the MYMV CP, the complete CP coding sequence (GChSMYCP1) and a series of mutants (ChSMYCP2 to -15) were cloned in fusion to EGFPChS (Fig. 3a) and expressed in N. plumbaginifolia protoplasts. GChSMYCP1 and a fusion containing the CP N-terminal region (GChSMYCP2) showed an intense signal in the nucleolus, especially at its periphery (Fig. 3c). In contrast, GChSMYCP3, which contains a large C-terminal part of the CP, was excluded from the nucleus. These results indicate that the first 68 aa of MYMV CP contain
a signal(s) necessary for its nuclear localization. Inspection of the N-terminal region revealed the presence of four clusters of basic residues (Fig. 3b). These basic residues were changed to alanines in each of these motifs individually (GChSMYCP4, -5, -6, -7). All of these modified fusions were still localized in the nucleus of transfected protoplasts (GChSMYCP4, Fig. 3c). The arrangement of these clusters resembles two possible bipartite NLSs. For this reason, the N-terminal 68 aa region was divided into two parts and each was fused separately to EGFPChS. GChSMYCP8 contains the first two clusters (residues 3 and 4, 18–20) and GChSMYCP12 contains the remaining two (residues 41–44, 52–55). Alanine substitutions were introduced in the clusters of basic residues, either separately (GChSMYCP9, -10, -13 and -14) or in combination (GChSMYCP11 and -15). GChSMYCP8 and -12 were able to direct the fusion protein into the nucleus but this localization was abolished when the first cluster of basic residues in each part (GChSMYCP9 and -13) or when both clusters (GChSMYCP11 and -15) were changed to alanines (Fig. 3). Fusions GChSMYCP10 and -14, with mutation of the second cluster of basic residues, both localized to the nucleus of transfected protoplasts (not shown).

Our results indicate that the basic clusters at position 3, 4 and 41–44 are both important for nuclear localization of the EGFPChS fusions, with each possibly constituting a monopartite-type NLS.

**RTBV and MYMV CPs interact with importin α**

To investigate the nuclear import pathway(s) used by RTBV and MYMV CPs, we studied their interaction with the nuclear import factor importin α. *In vitro* GST pull-down experiments were performed as described in Methods. GST and GST fusions to either rice or pepper importin α (GST–Rimpα and GST–Pimpα) (Matsuki et al., 1998; Shoji et al., 1998; Szurek et al., 2001) were expressed in *E. coli* and purified using glutathione-Sepharose 4B beads. Fractions of the different proteins bound to the beads were analysed by SDS-PAGE (Fig. 4a). GST–Pimpα was present in lower amounts compared with GST or GST–Rimpα. 35S-labelled RTBV and MYMV CPs (RTCP1 and MYCP1, respectively) were produced by *in vitro* translation in wheat germ extract or reticulocyte lysate. The labelled proteins were either analysed directly by SDS-PAGE (Fig. 4b, left panel) or incubated with GST, GST–Rimpα or GST–Pimpα immobilized on glutathione-Sepharose beads. After removal of supernatant and extensive washing, proteins present in the glutathione-Sepharose 4B bead complex fractions were eluted, separated by SDS-PAGE and visualized by autoradiography (Fig. 4c).

**Fig. 4.** *In vitro* binding of RTBV and MYMV CPs to importin α. (a) Expression and purification of rice (GST–Rimpα, 84 kDa) and pepper (GST–Pimpα, 85 kDa) importin α fused to GST. GST (26 kDa) and GST fusion proteins were expressed in *E. coli* and isolated by binding to glutathione-Sepharose 4B beads. Fractions bound to beads were subjected to SDS-PAGE and visualized by Coomassie blue staining. Arrowheads indicate the positions of GST and GST fusion proteins. (b) Interaction of RTBV and MYMV CPs with importin α. Left panel: autoradiograph of SDS-PAGE gel after separation of 35S-labelled MYMV (MYCP1) and RTBV (RTCP1) CPs produced in wheat germ extract. Right panel: autoradiograph of SDS-PAGE gel after separation of labelled CPs co-precipitated with GST, GST–Rimpα or GST–Rimpα, immobilized on glutathione-Sepharose 4B beads. (c) Interaction of modified RTBV CPs with importin α. Left panel: autoradiograph of SDS-PAGE gel after separation of 35S-labelled RTBV CP (RTCP1) and its derivatives (RTCP2, -3 and -5) produced in wheat germ extract. Modifications in RTCP2, -3 and -5 are identical to those present in EGFP fusion proteins RTCPG2, -3 and -5 (Fig. 2a). The positions of RTCP1 (37 kDa), RTCP2 (37 kDa), RTCP3 (31 kDa) and RTCP5 (34 kDa) are indicated on the left. Right panel: autoradiograph of SDS-PAGE gel after separation of labelled RTBV CP and its derivatives co-precipitated with GST or GST–Rimpα. (d) Interaction of modified MYMV CPs to importin α. Left panel: autoradiograph of SDS-PAGE gel after separation of 35S-labelled MYMV CP (MYCP1) and its derivatives (MYCP3 and MYCPNmut) produced in the Tnt Coupled Reticulocyte Lysate system (Promega). MYCP3 is a truncated version corresponding to the C-terminal portion of the MYMV amino acid sequence found in GChSMYCP3. MYCPNmut corresponds to the full-length CP with substitution of residues 3 and 4, and 41–44 by alanines. Right panel: autoradiograph of SDS-PAGE gel after separation of labelled MYMV CP and its derivatives co-precipitated with GST or GST–Rimpα.
autoradiography. MYCP1 and RTCP1 were detected in the GST–Pimp and GST–Rimp fractions but not in the GST fraction (Fig. 4b, right panel). These observations indicate a specific interaction of both CPs with rice and pepper importin α.

C-terminal region of RTBV CP is involved in the interaction with importin α

To identify the region of RTBV CP involved in binding to importin α, RTBV CP mutants (RTCP2, -3 and -5) containing the same mutations as those introduced in RTCPG2, -3 and -5 (Fig. 2a), were synthesized in vitro (Fig. 4c, left panel) and tested for interaction with rice importin α. 35S-labelled proteins were incubated with either GST or GST–Rimpα bound to glutathione-Sepharose beads. The pull-down assays showed that mutants RTCP2 and RTCP5 can still interact with GST–Rimpα, whereas RTCP3 does not bind. These results indicate that the interaction between RTBV CP and importin α occurs through the C-terminal domain of the CP and that the motifs 744–746 and 758–760 are not crucial for binding.

MYMV CP possesses multiple importin α-binding sites

To investigate whether basic residues 3, 4 and 41–44 of MYMV CP (Fig. 3b), which are important for nuclear import in protoplasts, also mediate the interaction with importin α, full-length MYMV CP with those residues mutated for alanines was produced in vitro (MYCPNmut, Fig. 4d, left panel) and tested for interaction with rice importin α. The binding capacity of MYCPNmut to Rimpα was greatly reduced but not completely abolished (Fig. 4d, right panel). MYCP3, in which the N-terminal 68 residues are absent, interacts with Rimpα, albeit more weakly than full-length CP (Fig. 4d). These results indicate that the basic residues 3, 4 and 41–44 are important for efficient binding to importin α but that other binding site(s), located between aa 68 and the C terminus, also exist.

DISCUSSION

In this study, we have characterized the signals involved in the nuclear import of two plant DNA virus CPs. The RTBV CP possesses at least two NLSs, located at both the N- and C-terminal ends of the protein. For the MYMV CP, we identified two distinct NLSs, both located in the N-terminal region. The NLSs characterized here resemble classic mono- and bipartite NLSs. However, they show peculiarities in the number and position of basic residues. The RTBV CP NLSs (Fig. 2b) could be considered as bipartite-like, corresponding to two groups of at least three basic residues separated by 11–14 aa (479–482/497–499 and 744–746/758–760). The MYMV CP NLSs resemble monopartite signals (Fig. 3b). The N-terminal MYMV CP NLS consists of only two basic residues (3 and 4) and the second NLS is a stretch of 4 aa (41–44). The position and number of these crucial basic clusters differ from those forming the NLSs characterized in the CPs of TYLCV and ACMV (Kunik et al., 1998; Unseld et al., 2001).

When expressed in protoplasts, RTBV- and MYMV-CP GFP fusion proteins (RTCPG1 and MYCPG1) were detected in the nucleoplasm and accumulated in the nucleolar subcompartment (Fig. 1b). In the case of MYCPG1, a marked fluorescence at the periphery of the nucleolus was observed. Both NLSs characterized in the RTBV and MYMV CPs are each able to act independently and can target a large fraction of the GFP fusion proteins to the nucleolar or perinucleolar subcompartment (Figs 2 and 3). The nucleolar localization of CaMV CP (p44) has also been reported (Leclerc et al., 1999). Accumulation of CaMV CP within the nucleolus is abolished when the C-terminal region, which is rich in basic residues, is deleted. In geminiviruses, TYLCV CP accumulates in the nucleolus, and nucleolar abnormalities have been observed in geminivirus-infected cells (Rojas et al., 2001). The biological significance of the nucleolar localization of plant DNA pararetro- and geminivirus CPs remains unclear and will require further investigation.

Both RTBV and MYMV CPs interact with importin α from rice and pepper (Merkle, 2001). Interaction of MYMV CP with pepper importin α was also detected using the yeast two-hybrid system (E. Herzog, unpublished data). Our results suggest that the RTBV and MYMV CPs are targeted to the nucleus via the importin α-based pathway, a highly conserved process in eukaryotic cells.

We found no direct correlation between the nuclear localization of EGFP fusions RTCPG2, -3 and -5 (Fig. 2a) and interaction of the corresponding RTBV CP versions with importin α (Fig. 4c). Deletion of the RTBV CP C-terminal domain (RTCP3) abolished association with importin α, underlining the central role of this region in establishing the interaction. However, the effect of mutating the basic amino acids that constitute, at least in part, the C-terminal NLS, alone (RTCP2), or in combination with the deletion of the N-terminal domain (RTCP5), was only to weaken the CP-importin α interaction (Fig. 4c), suggesting that uncharacterized residues located within the C-terminal region (735–791) play a role in binding to importin α. The results also suggest that the N-terminal NLS could be involved in RTBV CP nuclear import in an importin α-independent manner. Other possibilities include that this NLS is either being recognized by importin α isoforms other than those tested in our experiments or is only properly exposed after post-translational modifications of the CP that occur in plant cells.

Proteins with exposed basic domains tend to aggregate and precipitate in the presence of polyanionic molecules like nucleic acids. The aggregation of basic domain proteins can be inhibited by interaction with specific importin isoforms (α or β) (Jäkel et al., 2002). It is therefore also possible that the association of importin α with RTBV CP could prevent aggregation of the viral protein.
The NLSs characterized on MYMV CP are important for both nuclear localization and interaction with importin α. However, we cannot exclude the presence of other NLSs; it was recently reported that the ACMV CP contains additional NLSs located within the central and C-terminal parts (Unseld et al., 2001). When fused to EGFP and ChS, a large C-terminal region (aa 68 to the C terminus) was unable to target the fusion (GChSMYCP3) to the nucleus (Fig. 3). However, the same domain was able to interact with importin α in vitro (Fig. 4d), indicating that MYMV might contain more than two NLSs.

For plant DNA viruses, translocation of the viral genome into the nuclear compartment is a key step in the replication cycle. This process involves recognition by factors of the nuclear import machinery, transport of the virions to the nuclear periphery and binding to NPCs. The CP is the major component of the virus capsid shell in both caulimoviruses and geminiviruses and is believed to play a central role during nuclear import. Our findings that both RTBV and MYMV CPs contain several NLSs reinforce this idea.

Our comparative study reveals that plant DNA viruses might share common cellular nuclear import pathway(s). However, genome amplification and encapsidation processes differ for plant pararetroviruses and geminiviruses, in terms both of molecular mechanism and localization (Rothnie et al., 1994; Gafni & Epel, 2002). The shape and size (two potential constraints for nuclear import) of the virions are also very different. The size of geminate caulimovirus (20 × 30 nm) (Zhang et al., 2001; Harrison et al., 2002) or bacilliform RTBV (30 × 130 nm) (Hull, 1996) particles are close to or above, respectively, the estimated 39 nm capacity of NPC pores (Pante & Kann, 2002). Because of their specificities, plant DNA viruses belonging to different families could have developed specialized mechanisms to control virion docking and/or CP targeting to the nucleus.

Leclerc et al. (1999) demonstrated that the NLS located within the N-terminal domain of the CaMV CP is exposed on the outside of mature viral particles. This NLS facilitates docking of viral particles to the nucleus, possibly through interaction with importin α (Karsies et al., 2002). Disassembly of CaMV virions (50 nm in diameter) very likely occurs before the viral genome is transported through the nuclear pore. In RTBV, the positioning of an NLS at the N terminus of the CP, and the size and shape of the virions, are suggestive of a similar scenario. Nevertheless, our results indicate that the N-terminal NLS is not involved, at least in vitro, in direct interaction with the importin α clones tested. RTBV docking to the nucleus could require proper exposure of the CP N-terminal NLS, under the control of post-translational modifications, as shown for CaMV (Chapdelaine et al., 2002) and HBV (Kann et al., 1999), or interaction with other cellular or viral protein(s). Moreover, our investigations revealed the presence of an additional NLS within the RTBV CP C-terminal basic domain. According to the encapsidation model of pararetroviruses, this region, together with the conserved zinc finger motif, should be located within the mature viral particles, in close contact with the viral genome. Therefore, it seems unlikely that this NLS plays a role during the initial step of docking to the nuclear pore. However, it could participate in nuclear entry of the genome after either partial or complete virion decapsidation.

In geminiviruses, the complete replication cycle and encapsidation of the newly synthesized viral DNA occur within the nucleus (Gafni & Epel, 2002). The NLSs characterized here could therefore facilitate proper localization of the CP within this cellular compartment at two crucial steps of the MYMV life cycle: docking of virus particles at the nuclear pore (at least in the first infected cell) and encapsidation. It has not yet been determined if geminivirus virions enter the nucleus intact or after complete or partial disassembly. The size of geminate particles could, in theory, allow direct import through the NPC without disassembly. In MSV, the CP N-terminal domain (which encodes a NLS) is masked during viral DNA recognition and virion assembly (Zhang et al., 2001). Similarly, the N-terminal NLS of the MYMV CP could also be masked and thus necessitating the presence of other NLSs for nuclear import. The BV1 protein encoded by the B component of the begomoviruses (bipartite geminiviruses) is a nuclear shuttling protein (Sanderfoot et al., 1996). BV1 participates in the import and export processes of the viral DNA necessary for replication and movement through plant cells. In monopartite geminiviruses, the CP has taken over these two functions (Gafni & Epel, 2002). The CP is dispensible for movement of bipartite geminiviruses in their natural hosts (Pooma et al., 1996; Gafni & Epel, 2002). However, it is required for infection in hosts to which they are not well adapted. The presence of NLSs, as described here and elsewhere, and of nuclear export signals (Rhee et al., 2000; Unseld et al., 2001) suggests that, in coordination with BV1, the CP of bipartite geminiviruses could also actively participate in the docking of the viral genome to the nuclear pore and later, after replication, in its export. The participation of CP could be crucial, especially during the primary infection of a newly infected plant, i.e. just after transmission of the viral particles by the insect vector.

ACKNOWLEDGEMENTS

We are grateful to Helen Rothnie and Mikhail Pooggin for critical reading of the manuscript and Aleta Karsies, Karin Wiebauer and the other members of our laboratories for helpful discussions. We thank Thomas Merkle (Center for Biotechnology, Bielefeld, Germany), Akiko Baba (National Institute of Agrobiological Resources, Tsukuba, Japan) and Ulla Bonas (Institute of Genetics, Martin-Luther University, Halle, Germany) for providing the parsley chalcone synthase, the rice and pepper importin clones, respectively. We also thank Mathias Müller for the preparation of prototypes. This work was partially supported by the Novartis Research Foundation, the Indo-Swiss Collaboration in Biotechnology programme, the Centre National de la Recherche Scientifique and by the Université Louis Pasteur (Strasbourg, France).
REFERENCES


