Characterization of nuclear localization signals of the prototype foamy virus integrase

Dog Gn An, Usok Hyun and Cha-Gyun Shin

Department of Biotechnology, Chung-Ang University, Ansung, Kyungki 456-756, Republic of Korea

To analyse the potential karyophilic activity of prototype foamy viruses (PFVs), we expressed the PFV integrase (IN) and its mutants as fusion proteins with enhanced green fluorescence protein. The subcellular localization of the fusion proteins was investigated by fluorescence microscopy. The PFV IN was found to be karyophilic and targeted the fusion protein to the nucleus. Mutational analyses demonstrated that the PFV IN contains a potent but non-transferable nuclear localization signal (NLS) in its C-terminal domain and contains five arginine and lysine residues between amino acids 308 and 329 that are critical for its NLS function.

Foamy viruses (FVs), also called spumaviruses, are members of the retroviral family Retroviridae. The best-known FV is prototype foamy virus (PFV), previously referred to as human foamy virus (HFV), which was initially isolated from lymphoblastoid cells of a Kenyan patient with a nasopharyngeal carcinoma (Achong et al., 1971). Recent studies indicated that FVs are unconventional retroviruses and their particles have large amounts of functionally relevant DNA (Linial, 1999; Moebes et al., 1997; Rethwilm, 1996; Weiss, 1996; Yu et al., 1999). An active foamy virus integrase (IN) was found to be absolutely required for virus replication in infected cells (Enssle et al., 1999).

The retroviral IN, one of the constituents of the preintegration complex (PIC), has at least two important roles in the viral life cycle. First, it catalyses the integration of the viral cDNA into the cellular genomic DNA in the nuclei of infected cells (Bushman et al., 1990; Engelman et al., 1991); biochemical studies of these catalytic reactions have been well documented (Asante-Appiah & Skalka, 1997; Brown, 1997; Farnet & Haseltine, 1991). Secondly, the retroviral IN mediates transport of the PIC from the cytoplasm to the nucleus (Craigie, 2001). The human immunodeficiency virus type 1 (HIV-1) IN has two nuclear localization signals (NLS) of the basic bipartite type at residues 186–188 and 211–219 (Gallay et al., 1997). However, a separate study reported an additional NLS located in the central domain of the HIV-1 IN (residues 161–173) (Bouyac-Bertoia et al., 2001). The karyophilic determinant of the feline immunodeficiency virus IN has been mapped to the highly conserved N-terminal zinc-binding HHCC motif (Woodward et al., 2003). A functional NLS of the avian sarcoma virus IN was mapped at residues 206–235 (Kukolj et al., 1997).

Multiple studies have pursued the characterization and functional analyses of oncoretroviral and lentiviral INs, whereas few studies on the foamy viral IN have been reported (Pahl & Flugel, 1993, 1995). We recently characterized the functional domains and residues in the PFV IN using chimeric proteins and domain swapping with the HIV-1 IN (Lee et al., 2005). Previously, Imrich et al. (2000) reported a karyophilic activity of the PFV IN using IN-specific monoclonal antibodies. However, a karyophilic determinant of the PFV IN has yet to be identified. To characterize the putative karyophilic activity, we expressed wild-type and mutant PFV INs as fusion proteins with enhanced green fluorescence protein (EGFP) and maltose-binding protein (MBP) and analysed the subcellular localization of the fusion proteins by fluorescence microscopy. Here, we confirmed the karyophilic property of the PFV IN. Mutational analysis demonstrated that the PFV IN harbours a potent but non-transferable NLS in its C-terminal domain and contains five arginine and lysine residues between amino acids 308 and 329 that are critical for conferring this NLS activity.

Expression vectors were constructed by the ligation of either full-length PFV IN or its truncated mutant in the HindIII/KpnI site of the pEGFP C3 vector (Clontech) (Fig. 1a). We aligned the PFV IN amino acid sequences with the well-characterized HIV-1 IN to define the N-terminal, central, and C-terminal domains of the PFV IN (Lee et al., 2005). The 43 kDa MBP was inserted between the EGFP and PFV IN sequences to increase the size of the fusion protein and prevent non-specific, passive diffusion across the nuclear envelope. MBP has previously been demonstrated to successfully increase the size of fusion proteins without changing the karyophilic properties of the original proteins (Fassati et al., 2003). COS-1 and 293T cells plated on eight-well plates (Nunc) were transfected with DNA constructs for fusion proteins using the SuperFect transfection reagent (Qiagen). Twenty-four hours post-transfection, subcellular localization of the fusion proteins...
was analysed by fluorescence microscopy. Since MBP lacks an intrinsic nuclear localization sequence (Fassati et al., 2003), EGFP–MBP (68 kDa) served as a control for the nuclear exclusion of fusion proteins that are incapable of nuclear import and showed the expected cytoplasmic fluorescence pattern (Fig. 1b, panel i). EGFP–MBP containing the SV40 NLS served as a positive control for nuclear import (Fig. 1b, panel ii). The fluorescent signal of the full-length, wild-type PFV IN localized predominantly to the nucleus, suggesting that the PFV IN is karyophilic and can confer nuclear localization to the large GFP fusion protein in transfected cells (Fig. 1b, panel v). Western blotting analysis using both nuclei and cytosol isolated from transfected cells confirmed that the full-length, wild-type PFV IN localizes primarily to nuclei (data not shown). Two truncated PFV IN mutants fused with EGFP and MBP

<table>
<thead>
<tr>
<th>DNA Constructs</th>
<th>Size (kDa)</th>
<th>Localization</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) EGFP–MBP</td>
<td>68</td>
<td>Cyt (1.3±1.2)</td>
</tr>
<tr>
<td>(ii) SV40NLS</td>
<td>69</td>
<td>Nuc (98.6±1.5)</td>
</tr>
<tr>
<td>(iii) M1–88</td>
<td>78</td>
<td>Cyt (8.7±2.3)</td>
</tr>
<tr>
<td>(iv) M1–288</td>
<td>100</td>
<td>Nuc (70.7±3.1)</td>
</tr>
<tr>
<td>(v) M1–371</td>
<td>111</td>
<td>Nuc (96.5±3.4)</td>
</tr>
<tr>
<td>(vi) M89–288</td>
<td>90</td>
<td>Cyt (10.6±2.7)</td>
</tr>
<tr>
<td>(vii) M89–371</td>
<td>101</td>
<td>Nuc (92.1±3.6)</td>
</tr>
<tr>
<td>(viii) M289–371</td>
<td>79</td>
<td>Nuc (88.6±4.2)</td>
</tr>
</tbody>
</table>

Fig. 1. Subcellular localization of the EGFP–MBP–PFV IN fusion proteins in COS-1 cells. (a) Schematic representation of the DNA constructs. Full-length, wild-type (construct v) or truncated (constructs iii, iv and vi–viii) PFV INs were cloned into the HindIII/KpnI site of the pEGFP C3 vector. The 43 kDa maltose binding protein (MBP) was inserted between EGFP and the PFV IN to increase the size of the fusion proteins. EGFP–MBP fusion protein alone (construct i) and EGFP–MBP fused with the SV40 large T antigen NLS (MPKKRKKVEDPGT) (construct ii) were used as negative and positive controls for subcellular localization, respectively. The amino acid positions for the PFV IN are shown at the top of the figure. The expected molecular mass (in kDa) and subcellular localization are indicated to the right of the corresponding construct. Transfected cells were noted as exhibiting localization primarily to the cytoplasm (Cyt) and nucleus (Nuc). The numbers in parentheses indicate nuclear localization efficiency as a percentage, which is the ratio of the number of cells showing fluorescence in the nucleus to the number of cells showing fluorescence in either the nucleus or cytoplasm. At least 60–80 cells were measured for one replicate of each independent experiment. The data represent the mean ± SEM and are representative of three to five independent experiments. (b) Subcellular localization. DNA (0.5 μg) was transfected into 1.6×10⁴ COS-1 cells using the SuperFect reagent (Qiagen). Twenty-four hours post-transfection, the subcellular localization of the fusion proteins was analysed by fluorescence microscopy. Top panels (i–viii) depict superimposed visible and fluorescent images. Bottom panels (i’–viii’) show fluorescent images.

http://vir.sgmjournals.org 1681
(M1–88 and M89–288) localized primarily to the cytoplasm (Fig. 1b, panels iii and vi). In contrast, the fluorescent signals of the other mutants (M1–288, M89–371 and M289–371) localized primarily to nuclei (Fig. 1b, panels iv, vii and viii). Together, these results indicate that the N-terminal (1–88) and central (89–288) domains contain a weak NLS, which alone are not able to mediate nuclear import. However, the cooperative activities of their NLS were capable of directing nuclear import (Fig. 1b, panel iv). This result can be explained by the fact that M1–288 has a cooperative and additive NLS effect as the two domains are fused. In addition, the fusion of the two domains may provide additional binding sites for cellular factors that are known to be interactive proteins for the nuclear import of the PIC. The C-terminal domain (289–371) is highly karyophilic and harbours a potent NLS that is capable of independently promoting nuclear localization (Fig. 1b, panel viii). The percentage of nuclear localization efficiency in the transfected cells is presented as a number within parentheses in the localization column of Fig. 1(a). The subcellular localization patterns of the fusion proteins were identical in both COS-1 and 293T cell lines (data not shown).

Based on the localization patterns from our truncated mutants (Fig. 1), we concluded that a strong karyophilic determinant(s) is present in the C-terminal domain of the PFV IN and, therefore, pursued the identification of which residues are responsible for conferring karyophilic activity. We analysed the amino acid sequence of the C-terminal domain and searched for a region enriched with basic amino acids, a characteristic feature of canonical NLS (Dingwall & Laskey, 1991; Gorlich & Kutay, 1999). Although the PFV IN does not have a region with consecutive basic amino acids, its C-terminal domain has a region (305–329) which contains seven basic amino acids: RVARPSLRPRWHKPSTVLKVLPVR

In order to pinpoint residues that contribute to the karyophilic activity, we generated seven point mutants of the C-terminal domain where the lysine or arginine at residues 305, 308, 313, 315, 318, 324 and 329 was changed to a threonine or proline residue, respectively (Fig. 2a). In transfected cells expressing the PFV IN C-terminal domain (M289–371) fused to MBP and EGFP, mutations at residues 305, 308, 313, 315, 318, 324 and 329 was changed to a threonine or proline residue, respectively (Fig. 2a). In transfected cells expressing the PFV IN C-terminal domain (M289–371) fused to MBP and EGFP, mutations at residues 305, 308, 313, 315, 318, 324 and 329 had no impact on nuclear localization, indicating that these arginines do not significantly contribute to PFV IN karyophilic activity (Fig. 2b, panels ii and v). In contrast, the other mutants (M308, M313, M318, M324 and M329) all displayed cytoplasmic fluorescence (Fig. 2b, panels iii, iv, vi, vii and viii). These results indicate that the arginine or lysine present at these residues critically contribute to the nuclear localization activity of the C-terminal domain. As a short peptide comprised of HIV-1 IN residues 161–173 was reported to be karyophilic and mediate active nuclear import of covalently attached BSA (Armon-Omer et al., 2004), we tested whether a short peptide comprised of PFV IN residues 306–334 could mediate the nuclear import of the EGFP–MBP fusion protein. The nuclear localization efficiency was approximately 5.2%, indicating that the karyophilic activity of the short peptide was not sufficiently strong to mediate the nuclear import of the fusion protein and that the NLS of the short peptide was not transferable to other proteins. We therefore conclude that a potent but non-transferable NLS of the PFV IN is located within the C-terminal domain and spans five basic amino acids between residues 308 and 329.

The NLS of the PFV IN that we report here is not a classical NLS. It has five discontinuous, functional basic amino acids over a region of 26 aa. Recent studies revealed that the interactions of viral INs and cellular factors are essential for the nuclear import of the PIC as well as viral DNA integration (Busschots et al., 2007; Rijck et al., 2007). Although a short peptide of discontinuous, functional amino acids was unable to mediate the nuclear import of the fusion proteins, it was found to be karyophilic in the entire conformation of the C-terminal domain when there was an effective interaction between the cellular factors and the other region of the C-terminal domain. Recently, the mechanism of HIV-1 IN nuclear import has been well described (Busschots et al., 2007; Emiliani et al., 2005). The amino acids that contribute to nuclear localization appear to be important for binding to the lens epithelial cell line growth factor (LEDGE/p75), which is a cellular co-factor for HIV-1 replication. The role of LEDGE/p75 is to target the HIV-1 IN to chromosomes (Emiliani et al., 2005). The interaction of LEDGE/p75 with the HIV-1 IN was confirmed by at least two different groups (Emiliani et al., 2005; Turlure et al., 2004) and is lentivirus-specific (Busschots et al., 2005; Cherepanov, 2007; Llano et al., 2004). Therefore, it will be interesting to see whether the PFV IN interacts with LEDGE/p75 or its analogue in simian cells for viral replication in future studies.

The avian sarcoma virus (ASV) IN was previously reported to have a functional NLS spanning amino acids 206–235, which contains six basic amino acids compared with the five basic residues that we mapped in the functional NLS of the PFV IN (Kukolj et al., 1997). The functional basic amino acids in both sequences are present discontinuously, but not in short tandem sequences that are characteristic of well-known NLS (Boulikas, 1993; Dingwall & Laskey, 1991; Silver, 1991). However, the HIV-1 IN has a functional NLS composed of a short tandem sequence as well as a discontinuous array of basic amino acids, such as KRK and KELQKQITK (Gallay et al., 1997). The discontinuous sequence in the HIV-1 IN is shorter than that of the PFV IN that we mapped in this study. In contrast with the ASV, HIV-1 and PFV INs, the nucleophilic determinant of the feline immunodeficiency virus (FIV) IN has been mapped to the highly conserved HHCC motif in the N-terminal domain (Woodward et al., 2003). The karyophilic property of the FIV IN was postulated to require multimerization mediated by the HHCC motif, which in turn directly promotes the interaction between the FIV IN and the cellular proteins involved in nuclear import.
No conserved amino acid motif of the NLS has been identified among the retroviral INs. Studies on the intrinsic nuclear localization activity of INs by the identification of the critical amino acids for nuclear import will provide important insight into the mechanism of retroviral replication.

Acknowledgements
We particularly thank Dr Samson Chow and Cora Woodward for the generous gift of a SV40 NLS plasmid that was slightly modified in this study. We thank Da-Young Youk for assistance with cell culture. This work was supported in part by the Chung-Ang University Research Grants.

References


