Identification of a novel functional nuclear localization signal in the protein encoded by open reading frame 47 of Bombyx mori nucleopolyhedrovirus

ARTICLE *in* ARCHIVES OF VIROLOGY · DECEMBER 2010

Impact Factor: 2.39 · DOI: 10.1007/s00705-010-0782-1 · Source: PubMed

Arch Virol (2010) 155:1943–1950
DOI 10.1007/s00705-010-0782-1

ORIGINAL ARTICLE

Identification of a novel functional nuclear localization signal in the protein encoded by open reading frame 47 of *Bombyx mori* nucleopolyhedrovirus

Zhong-Jian Guo · Dian-Xuan Wang · Qin Yao · Ke-Ping Chen · Chuan-Xi Zhang

Received: 29 April 2010 / Accepted: 30 July 2010 / Published online: 13 August 2010

Abstract BM47 is encoded by open reading frame 47 of *Bombyx mori* nucleopolyhedrovirus (BmNPV). BM47 was localized in the nucleus of BmNPV-infected cells. In the present study, we investigated a novel nuclear localization signal (NLS) for BM47 transport and accumulation in the nucleus. By expressing various regions of BM47 fused to enhanced green fluorescent protein (EGFP), we demonstrated that residues 117–148 are necessary for mediating nuclear localization of BM47. Site-directed mutation analysis showed that the two basic residue clusters at positions 117–120 (^{117}RKRR) and 144–148 (^{144}RKR-K) constitute an authentic NLS for BM47 localization. Finally, we observed that two clusters of basic residues were conserved in BM47 homologues of group-I nucleopolyhedroviruses.

Electronic supplementary material The online version of this article (doi:10.1007/s00705-010-0782-1) contains supplementary material, which is available to authorized users.

Z.-J. Guo (✉) · Q. Yao · K.-P. Chen
Institute of Life Science, Jiangsu University, 301 Xuefu Road, Zhenjiang 212013, Jiangsu, People's Republic of China
e-mail: gzh762677@ujs.edu.cn

D.-X. Wang
School of Food Science and Technology, Engineering Research Center of Grain Storage and Security of Ministry of Education, Henan University of Technology, Zhengzhou 450052, China

C.-X. Zhang (✉)
Institute of Insect Science, Zhejiang University, 268 Kaixuan Road, Hangzhou 310029, Zhejiang, People's Republic of China
e-mail: chxzhang@zju.edu.cn

Introduction

An essential feature of eukaryotic cells is the segregation of genetic material and transcriptional machinery of the nucleus from the translational machinery of the cytoplasm via a nuclear envelope. This separation facilitates regulation of cellular processes such as gene expression, signal transduction, and cell cycle progression through selective regulation of bidirectional transport between the nucleus and the cytoplasm [20]. The transport of proteins is mediated by the nuclear pore complex. This complex allows passive diffusion of small proteins up to a molecular mass of approximately 40 kDa between the nucleus and the cytoplasm. In contrast, proteins with a molecular mass of more than 40 kDa require the assistance of importins or exportins to be transported into or out of the nucleus [14]. For the process of importing proteins, importins bind to the cargo protein, then shuttle the protein complex into the nucleus through direct interactions with the nuclear pore complex. The importins must recognize an element called the nuclear localization signal (NLS) within the cargo protein for binding. Hence, the NLS plays an important role in transporting cargo into the nucleus.

Large numbers of studies have shown that most proteins undergoing nuclear import contain a NLS. The prototypical monopartite NLS, as exemplified by SV40 large T antigen (PP**KKKRK**V) [18], consists of a single, short stretch of several basic amino acids. The bipartite NLS, as exemplified by nucleoplasmin (**KR**PAATKKAGQA**KKKK**) [31], contains two clusters of basic residues separated by a spacer of 9–12 amino acids [7]. However, many NLSs do not conform to the prototypical NLS consensus sequence, such as the glycine-rich M9 domain distinguished from basic residues found in heterogeneous nuclear ribonucleoprotein (hnRNP) A1 protein [35], KNS in hnRNP K protein [26], ankyrin repeats in IκBα [33] and the Nab2p NLS [37].

Nonetheless, basic amino acid motifs that conform to classic NLS consensus sequences may not be functional, as observed in some non-nuclear proteins [4].

The BmNPV is a well-characterized virus belonging to the *Baculoviridae*, a family of DNA viruses that have a large circular, supercoiled and double-stranded DNA genome within a rod-shaped nucleocapsid [36]. BmNPV genome replication occurs within the nucleus of infected cells, necessitating timely nuclear import of viral proteins involved in transcription, DNA replication, and virus particle assembly. Many baculovirus proteins that must enter the insect cell nucleus to perform their functions have been shown to contain an NLS. The protein BM47, which is encoded by the ORF47 gene of BmNPV, contains a putative bipartite NLS at residues 132–148 (^{132}**KK**IPIGKVVSTP**RKRLK**). A homologue of BM47, Se101, encoded by *Spodoptera exigua* multicapsid nucleopolyhedrovirus (SeMNPV) is present in the nucleocapsid of occlusion-derived virus (ODV), and distributed in both the cytoplasm and the nucleus [22]. Another homologue, Sl52 of *Spodoptera litura* multicapsid nucleopolyhedrovirus (SpltMNPV), is also a structural component of ODV. Interestingly, Sl52 is expressed as 23- and 26-kDa doublets in infected *S. litura* cells. The 23-kDa form is distributed in both the cytoplasm and the nucleus, whereas the 26-kDa form is only present in the nucleus [23]. However, our earlier protein composition work showed that BM47 is not a structural component of ODV [24], indicating that BM47 may function specifically in the BmNPV life cycle. In this paper, we studied the subcellular distribution of BM47, which was localized exclusively in the nucleus of infected cells, and identified two basic residue clusters that constitute a novel functional NLS. This NLS is a strong one that can efficiently localize a small protein to the nucleus for its special characteristics such as compensation for mutation and duplication of basic residues in either cluster.

Materials and methods

BmNPV bacmid and cells

The BmNPV bacmid system was constructed by Motohashi and colleagues [27]. The *B. mori* cell line BmN derived from ovary of silkworm [11], was cultured at 27°C in TC-100 insect medium (Sigma–Aldrich, St. Louis, MO, USA) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (Gibco-BRL, Gaithersburg, MD, USA).

Immunofluorescence microscopy analysis

A monolayer (about 1.0×10^6, 70–80% confluence) of BmN cells was infected with 5.0 $TCID_{50}$/cell of BmNPV (T3 strain) budded virus (BV). At 48 h post-infection (p.i.), cells were incubated with the 1:5,000-diluted polyclonal BM47 antiserum [39], followed by incubation with FITC-conjugated goat anti-rabbit immunoglobulin G (IgG) (Sigma–Aldrich, St. Louis, MO, USA) diluted 1:5,000. Infected cells incubated with pre-immune serum were used as a negative control. Stained cells were examined using a Leica confocal laser microscope for fluorescence detection.

Plasmid construction

We constructed BM47 truncations fused to enhanced green fluorescent protein (EGFP). Polymerase chain reaction (PCR) was performed as follows: preheating at 94°C for 5 min, 30 cycles of 45 s at 94°C, 30 s at 50°C, 90 s (for EGFP amplification) or 45 s (for ORF47 truncation amplifications) at 72°C, extension at 72°C for 10 min. To create an EGFP-fused expression vector, a *Bam*HI–*Xho*I fragment containing EGFP ORF was amplified with primers 5′-A GGATCC ATG GTG AGC AAG GGC GAG GAG-3′ and 5′-A CTCGAG CTT GTA CAG CTC GTC CAT GCC-3′, and the plasmid pEGFP-C1 (BD Biosciences Clontech, San Jose, CA, USA) as the template. The fragment was inserted into *Bam*HI–*Xho*I-digested pFastBac1 (Invitrogen Life Technologies, Carlsbad, CA, USA) to produce pFast-EGFP, which was confirmed by sequencing with primers Seq-F (5′-ATG ATA ACC ATC TCG CAA AT-3′; nt 3922–3941 of pFastBac1) and Seq-R (5′-AAT ACA AAG TCC AAG TCC CC-3′; nt 4243–4224 of pFastBac1). The BM47-encoding region was amplified with primers BM47-F (5′-C CTCGAG ATG TAT CAA ATT CCC GAT-3′) and BM47-R (5′-C AAGCTT TTA ATA GTT GTA ATA ATT-3′) and cloned into the corresponding restriction sites of pFast-EGFP to obtain pFast-EGFP-$BM47_{1–171}$. BM47 deletion fragments expressing the amino acid residues 1–148, 1–130, and 1–74 were amplified with primers $BM47_{1–148}$ (5′-G AAGCTT TTA TTT TAA TCG TTT ACG CGG CG-3′), $BM47_{1–130}$ (5′-G AAGCTT TTA TTT CGA ATT AAA CGC C -3′) and $BM47_{1–74}$ (5′-A AAGCTT TTA GTC GTC CCT AAC TAT C -3′) coupled with the forward primer BM47-F. These fragments were cloned into the *Xho*I/*Hin*dIII sites of pFast-EGFP to produce plasmids pFast-EGFP-$BM47_{1–148}$, pFast-EGFP-$BM47_{1–130}$, and pFast-EGFP-$BM47_{1–74}$. The DNA sequence encoding amino acid residues 131–171 was obtained with primers $BM47_{131–171}$ (5′-C CTCGAG TCC AAA AAA ATT CCT ATC GGC -3′) and BM47-R, and introduced into pFast-EGFP to obtain vector pFast-EGFP-$BM47_{131–171}$. The fragment for expressing amino acid residues 117–148 was amplified with primers $BM47_{117–148}$ (5′-C CTCGAG CGT AAA CGG CGA GTC GTC G-3′) and $BM47_{1–148}$ to produce the plasmid pFast-EGFP-$BM47_{117–148}$. Fragments for expressing amino acid residues

121–148 and 117–143 were obtained with primer pairs $BM47_{1–148}$/$BM47_{121–148}$ (5′-C CTCGAG GTC GTC GCT AAG CGG GCG-3′), and $BM47_{117–148}$/$BM47_{117–143}$ (5′-A AAGCTT TTA CGG CGT CGA CAC CAC-3′) to produce pFast-EGFP-$BM47_{121–148}$ and pFast-EGFP-$BM47_{117–143}$ plasmids, respectively. The identities of these plasmids were verified by sequencing with primer Seq-R.

Site-directed mutagenesis

Vector pFast-EGFP-$BM47_{1–171}$ and complementary primers (shown as the plus-sense sequence) that contained the desired mutations (Table S1, boldface letters) were used for mutagenesis. Site-directed mutagenesis was conducted using a QuikChange® XL site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA) according to the manufacturer's instructions. The basic amino acids of the NLS were mutated to alanine. These mutations were verified by DNA sequencing with primer Seq-R.

Cell transfection, microscopy and fluorescent cell counting

Truncations and mutations were transferred to *Escherichia coli* strain BmDH10Bac, which contained the BmNPV genome bacmid. The purpose of this transfer was to transpose the EGFP ORF and related constructs or mutants into the BmNPV genome by Tn7-mediated transposition [25]. Positive bacmids were screened by PCR with M13 forward and reverse primers. About 1×10^6 BmN cells were transfected with bacmid according to the instruction manual of BAC-TO-BAC™ Baculovirus Expression Systems (Invitrogen Life Technologies, Carlsbad, CA, USA) to obtain BV, whose titer was determined by endpoint dilution and expressed as $TCID_{50}$ units/ml [19]. A monolayer (about 1.0×10^6, 70–80% confluence) of cells was infected with BV at an MOI of 1 $TCID_{50}$ unit/cell. At 48 h p.i., cells were visualized using a Leica confocal laser scanning microscope and scored for subcellular localization analysis.

Results

Subcellular localization of BM47 in infected cells

The subcellular localization of BM47 protein was investigated by immunofluorescence microscopy using BM47 antiserum. The results revealed that the fluorescence signal was strictly concentrated within the nucleoplasm. BmNPV-infected cells that were used as negative control were incubated with pre-immune serum and FITC-conjugated goat anti-rabbit IgG and did not show detectable fluorescence signals (Fig. 1). The results showed that BM47 was localized exclusively in the nucleus of infected cells.

Identification of a functional NLS in BM47

To define the BM47 NLS that is necessary and sufficient for nuclear accumulation, several recombinant vectors were produced to express various segments of BM47 fused to the C-terminus of EGFP (Fig. 2A). The localization of different EGFP-fusion proteins in BmN cells was determined by fluorescence microscopy. As expected, EGFP alone showed diffuse distribution in both the cytoplasm and the nucleus (Table 1; Fig 2B, panel i). Similar to the

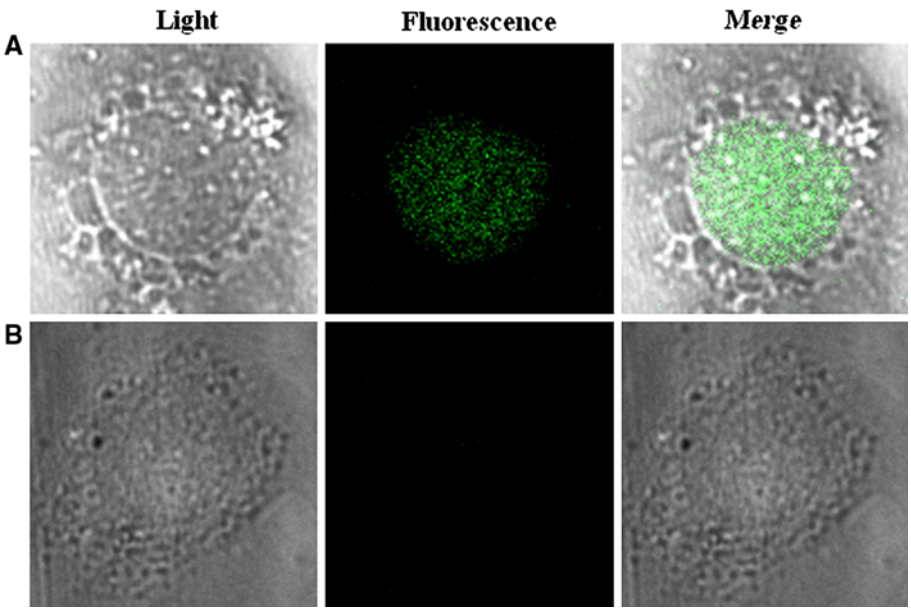

Fig. 1 Immunofluorescence analysis of the subcellular localization of BM47 in BmNPV-infected cells. Cells were collected at 48 h p.i. After being treated with paraformaldehyde and Triton X-100, cells were incubated with BM47 antiserum and FITC-conjugated goat anti-rabbit IgG (**A**). In the control experiment, pre-immune serum was used as the primary antibody (**B**). Cells were viewed using confocal laser fluorescence microscopy

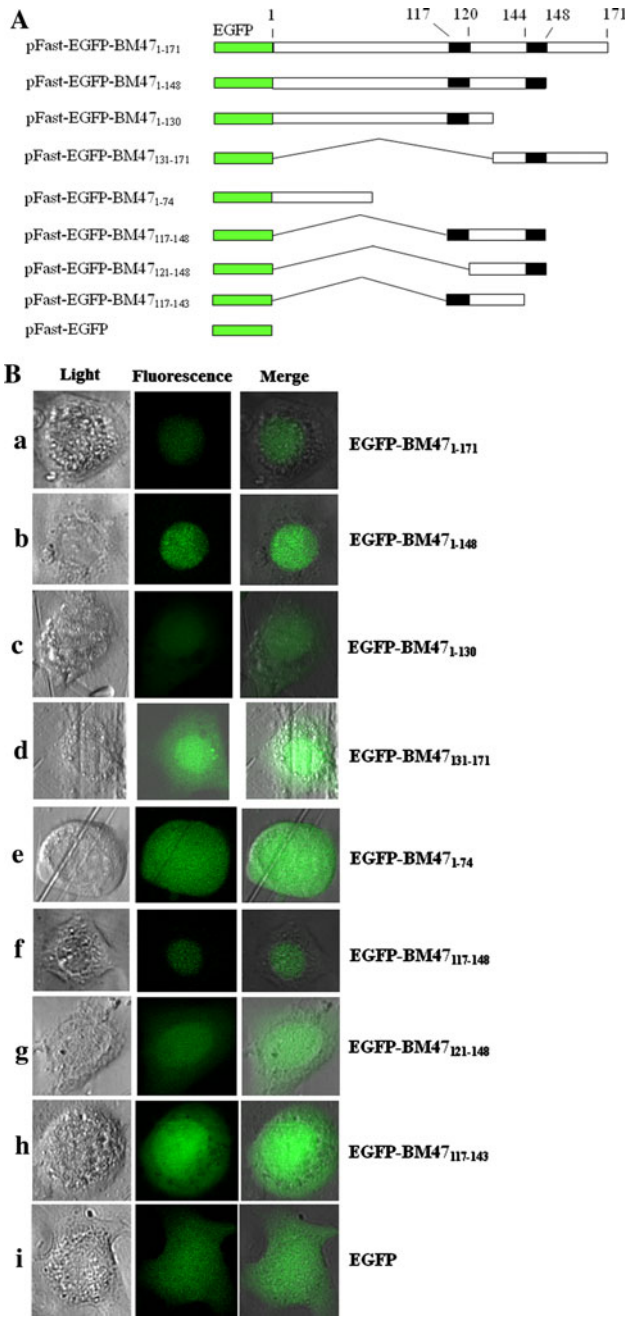

Fig. 2 Identification of a functional BM47 NLS sequence. **A** Construction of vectors to express various regions of BM47 fused to the EGFP C-terminus. *Black boxes* correspond to the basic residue clusters ^{117}RKRR and ^{144}RKRLK, and green boxes correspond to EGFP. Subscript numbers refer to BM47 amino acids fused to EGFP. **B** Fluorescence microscopy of BM47 expression in regions fused to EGFP in BmN cells. *E. coli* strain BmDH10Bac was transformed with the vectors in (**A**) to produce recombinant BmNPV bacmids, which were used to transfect BmN cells to produce BVs. A monolayer of cells was infected with BVs at 1 $TCID_{50}$ unit/cell and visualized at 48 h p.i. under a Leica confocal laser scanning microscope

subcellular localization of EGFP-BM47$_{1–171}$ (Table 1; Fig 2B, panel a), EGFP-BM47$_{1–148}$ accumulated exclusively in the nucleus (Table 1; Fig. 2B, panel b). Chimeric

Table 1 Subcellular localization of BM47 truncations in infected BmN cells

Truncation	Localization[a]		
	N (%)	N(+)/C (%)	N/C (%)
EGFP-BM47$_{1–171}$	97.2	2.8	–
EGFP-BM47$_{1–148}$	96.1	3.9	–
EGFP-BM47$_{1–130}$	–	73.4	26.6
EGFP-BM47$_{131–171}$	–	78.4	21.6
EGFP-BM47$_{1–74}$	–	8.3	91.7
EGFP-BM47$_{117–148}$	98.3	1.7	–
EGFP-BM47$_{121–148}$	–	79.5	20.5
EGFP-BM47$_{117–143}$	–	74.9	25.1
EGFP	–	3.5	96.5

[a] Infected cells were analyzed at 48 h p.i. For each treatment, 67–90 cells expressing EGFP or EGFP-fused protein were counted. The truncated BM47 was localized exclusively in the nucleus (N), predominantly but not exclusively in the nucleus (N(+)/C), or evenly between the nucleus and the cytoplasm (N/C). Each value refers to the percentage of each accumulation and is the average of three independent infections

– No accumulation was observed

EGFP-BM47$_{1–74}$ was distributed throughout the cell (Table 1; Fig. 2B, panel e), whereas EGFP-BM47$_{1–130}$ and EGFP-BM47$_{131–171}$ accumulated predominantly in the nucleus of infected cells, although not exclusively (Table 1; Fig. 2B, panels c and d). Residues 117–148 were necessary and sufficient to localize the fusion protein completely in the nucleus, as demonstrated by EGFP-BM47$_{117–148}$, which displayed exclusive nuclear localization (Table 1; Fig. 2B, panel f). Deletion of either the first four residues or the last five residues of this segment showed similar signal distribution to that of EGFP-BM47$_{1–130}$ (Table 1; Fig. 2B, panels g and h). Taken together, these results suggested that the BM47 NLS is located between residues 117 and 148.

Analysis of the functional basic residues for the BM47 NLS

Examination of the amino acid sequence between residues 117 and 148 showed that the 32-amino acid sequence is enriched with basic amino acids. To identify critical residues for BM47 NLS function, two complementary primers were used to generate mutants in the basic residues of NLS fused to EGFP (Table S1). The localization of mutant proteins was then examined by fluorescence microscopy. Strictly nuclear localization was observed at the N-terminal (^{117}RKRR^{120}) and C-terminal (^{144}RKR-K^{148}) arm of the NLS with the following mutations: individual-residue mutations (Table 2, Mutants 1–8), double-residue mutations (Table 2, Mutants 9–17) and triple-residue mutations

Table 2 Determination of functional basic residues required for efficient nuclear accumulation of BM47

Mutants	Sequences and mutations[a]	Subcellular localization[b]
Wt	RKRRVVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 1	**A**KRRVVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 2	R**A**RRVVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 3	RK**A**RVVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 4	RKR**A**VVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 5	RKRRVVAKRAFNSKSKKIPIGKVVSTP**A**KRLK	N
Mutant 6	RKRRVVAKRAFNSKSKKIPIGKVVSTPR**A**RLK	N
Mutant 7	RKRRVVAKRAFNSKSKKIPIGKVVSTPRK**A**LK	N
Mutant 8	RKRRVVAKRAFNSKSKKIPIGKVVSTPRKRL**A**	N
Mutant 9	**AA**RRVVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 10	R**AA**RVVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 11	RK**AA**VVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 12	**A**KR**A**VVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 13	RKRRVVAKRAFNSKSKKIPIGKVVSTP**AA**RLK	N
Mutant 14	RKRRVVAKRAFNSKSKKIPIGKVVSTP**A**K**A**LK	N
Mutant 15	RKRRVVAKRAFNSKSKKIPIGKVVSTP**A**KRL**A**	N
Mutant 16	RKRRVVAKRAFNSKSKKIPIGKVVSTPR**AA**LK	N
Mutant 17	RKRRVVAKRAFNSKSKKIPIGKVVSTPRK**A**L**A**	N
Mutant 18	**AAA**RVVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 19	R**AAA**VVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 20	**A**K**AA**VVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 21	**AA**R**A**VVAKRAFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 22	RKRRVVAKRAFNSKSKKIPIGKVVSTPR**AA**L**A**	N
Mutant 23	RKRRVVAKRAFNSKSKKIPIGKVVSTP**A**K**A**L**A**	N
Mutant 24	RKRRVVAKRAFNSKSKKIPIGKVVSTP**AA**RL**A**	N
Mutant 25	RKRRVVAKRAFNSKSKKIPIGKVVSTP**AAA**LK	N
Mutant 26	RKRRVVA**AA**AFNSKSKKIPIGKVVSTPRKRLK	N
Mutant 27	RKRRVVAKRAFNS**A**SKKIPIGKVVSTPRKRLK	N
Mutant 28	RKRRVVAKRAFNSKS**AA**IPIGKVVSTPRKRLK	N
Mutant 29	RKRRVVAKRAFNSKSKKIPIG**A**VVSTPRKRLK	N
Mutant 30	**AAAA**VVAKRAFNSKSKKIPIGKVVSTPRKRLK	N(+)/C
Mutant 31	**AAAA**VVAKRAFNSKSKKIPIGKVVSTP**A**KRLK	N/C
Mutant 32	**AAAA**VVAKRAFNSKSKKIPIGKVVSTPR**A**RLK	N/C
Mutant 33	**AAAA**VVAKRAFNSKSKKIPIGKVVSTPRK**A**LK	N/C
Mutant 34	**AAAA**VVAKRAFNSKSKKIPIGKVVSTPRKRL**A**	N/C
Mutant 35	RKRRVVAKRAFNSKSKKIPIGKVVSTP**AAA**L**A**	N(+)/C
Mutant 36	**A**KRRVVAKRAFNSKSKKIPIGKVVSTP**AAA**L**A**	N/C
Mutant 37	R**A**RRVVAKRAFNSKSKKIPIGKVVSTP**AAA**L**A**	N/C
Mutant 38	RK**A**RVVAKRAFNSKSKKIPIGKVVSTP**AAA**L**A**	N/C
Mutant 39	RKR**A**VVAKRAFNSKSKKIPIGKVVSTP**AAA**L**A**	N/C

[a] Amino acid sequence from residues 117 to 148 and mutations of basic residue(s) to alanines. The basic residues in clusters (^{117}RKRR^{120} and ^{144}RKR-K^{148}) are *underlined*. Mutations are indicated in *bold letters*

[b] The mutant proteins were localized exclusively in the nucleus (N), predominantly but not exclusively in the nucleus (N(+)/C), or equally between the nucleus and the cytoplasm (N/C)

(Table 2, Mutants 18–25). Conversely, the quadruple-residue mutation at ^{117}RKRR^{120} or ^{144}RKR-K^{148} cluster showed a nucleus-dominated signal distribution (Table 2, Mutants 30 and 35), similar to localization of chimeric EGFP-BM47$_{121–148}$ and EGFP-BM47$_{117–143}$ (Fig. 2B, panels g and h). When any one basic residue of one cluster and four basic residues of the other cluster were mutated to alanines, the BM47 NLS lost its nuclear localization function and exhibited pancellular dispersal (Table 2, Mutants 31–34 and 36–39). As described above, chimeric EGFP-BM47$_{1–130}$ (Fig. 2B, panel c) and EGFP-BM47$_{121–148}$ (Fig. 2B, panel g) accumulated predominantly, but not exclusively, in the nucleus, indicating that residue(s) between two clusters may play a role in nuclear

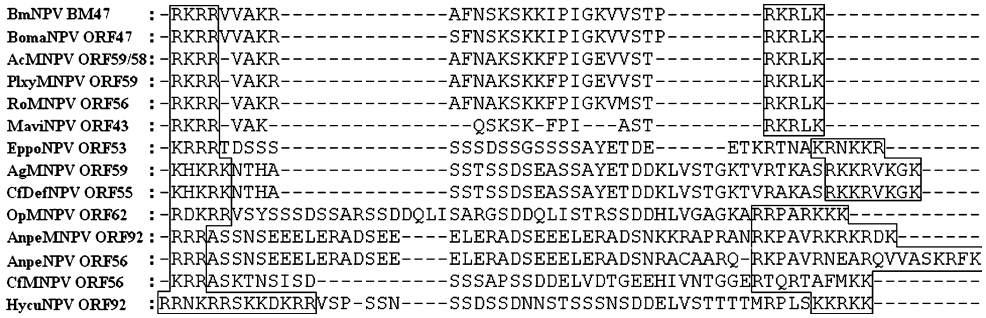

Fig. 3 Two clusters of basic residues in the C-terminal region of BM47 and homologues in group-I nucleopolyhedroviruses. Alignment was performed using ClustalW software on the website http://www.ebi.ac.uk/Tools/clustalw2/index.html. The clusters enriched in basic residues are *boxed*. Viruses abbreviations are as follows: BmNPV, *Bombyx mori* nucleopolyhedrovirus [10]; BomaNPV, *Bombyx mandarina* nucleopolyhedrovirus [40]; AcMNPV, *Autographa californica* multiple nucleopolyhedroviruses [3]; PlxyMNPV, *Plutella xylostella* multiple nucleopolyhedrovirus [13]; RoMNPV, *Rachiplusia ou multiple* nucleopolyhedrovirus [12]; MaviNPV, *Maruca vitrata* nucleopolyhedrovirus [6]; AgMNPV, *Anticarsia gemmatalis* multiple nucleopolyhedrovirus [29]; CfDefNPV, *Choristoneura fumiferana* DEF multiple nucleopolyhedrovirus [21]; EppoNPV, *Epiphyas postvittana* nucleopolyhedrovirus [15]; OpMNPV, *Orgyia pseudotsugata* multiple nucleopolyhedrovirus [1]; HycuNPV, *Hyphantria cunea* nucleopolyhedrovirus [16]; AnpeMNPV, *Antheraea pernyi* multiple nucleopolyhedrovirus [9]; CfMNPV, *Choristoneura fumiferana* multiple nucleopolyhedrovirus [8]; AnpeNPV, *Antheraea pernyi* nucleopolyhedrovirus [28]

localization. However, conversion of basic residues into alanines did not affect nuclear import function at the following positions: 124 and 125 (Table 2, Mutant 26), 130 (Table 2, Mutant 27), and 132 and 133 (Table 2, Mutant 28). Identical results were also obtained when the residue at position 138 was mutated (Table 2, Mutant 29). These observations demonstrated that the two basic residue clusters at the N-terminus (^{117}RKRR) and C-terminus (^{144}RKR-K) constituted the NLS required for efficient nuclear accumulation of BM47.

Basic clusters in BM47 homologues of group-I nucleopolyhedroviruses

Database searches with BLASTP revealed that BM47 homologues were found in the sequenced nucleopolyhedroviruses belonging to the genus *Alphabaculovirus*, but not in viruses of the genera *Betabaculovirus, Deltabaculovirus* and *Gammabaculovirus.* At present, there are 14 group-I and 24 group-II nucleopolyhedroviruses in the genus *Alphabaculovirus* whose sequences are available in GenBank. Interestingly, two clusters of basic amino acids were not found in homologues of group-II viruses, but they were found in those of group-I viruses. These two clusters of basic residues are aligned in Fig. 3, showing conservation of the two clusters of basic residues in the group-I nucleopolyhedroviruses.

Discussion

ORF47 is a conserved gene whose homologues are found in all of the sequenced lepidopteran nucleopolyhedroviruses. These homologues contain a putative ChaB domain in the N-terminus. In *E. coli*, ChaB may regulate the function of ChaA, a major transport system involved in sodium ion extrusion [34]. However, the exact role of the ChaB-containing homologues in the individual viral life cycle remains unknown, although the characteristics of the homologues Se101 [22] and Sl52 [23] have been investigated.

In this study, as a first step in the characterization of BM47, we investigated the sequence involved in nuclear localization of BmNPV BM47 and demonstrated the presence of a functional NLS sequence. This NLS shares features with the known bipartite NLS because it contains two basic residue clusters separated by a spacer of 23 amino acids, and both clusters are essential for efficient nuclear localization. However, it represents a novel bipartite NLS because of its special characteristics. There are differences between the NLS of BM47 and bipartite short- or long-type NLSs reported previously [32]. The BM47 NLS is rich in basic residues, and single-, double- or triple-residue mutations in either cluster do not abolish the NLS function, suggesting that other basic residue(s) in the NLS can compensate for mutations, and duplication of basic residues in either cluster may lead to more efficient transport of BM47 into the nucleus. The compensation and duplication phenomenon has also been observed in a nuclear localization domain of late expression factor 3 (LEF-3) of AcMNPV. The N-terminal 56 residues are required for nuclear transport of LEF-3 [5], and in this region, LEF-3 carries a duplication of basic residues in two domains that can act separately as NLS sequence [2]. However, the NLS of BM47 is different from that of LEF-3 in that the two basic clusters are not independent, and both are essential for the function of the NLS.

Theoretically, BM47 is small enough to diffuse into the nucleus through the nuclear pore complex; however, this would be a relatively inefficient mechanism. Compensation for mutations and enrichment in basic residues suggest that BM47 NLS is strong enough to transport a cargo protein into the nucleus, even though the cargo protein is small. This signal probably facilitates the nuclear localization of BM47 by allowing it to participate in a more efficient receptor-mediated transport mechanism.

To date, several baculovirus proteins destined to perform their functions in the nucleus have been shown to contain an NLS. The first characterized NLS is the sequence KRKK, located between residues 32 and 35 of polyhedrin, which directs the nuclear localization of polyhedrin [17]. The NLS of BV/ODV-C42, ^{357}KRKK^{360}, mediates nuclear entry of P78/83 [38]. The basic residues R537 and R538 of AcMNPV IE1 constitute a novel nuclear localization element, and upon IE1 dimerization, two nuclear localization elements function as NLS [30]. LEF-3 carries two separate basic clusters, and each can act as a NLS to transport the helicase protein P143 into the nucleus [2]. These NLSs resemble classical monopartite ones. However, the BM47 NLS is different from these NLSs for its special characteristics, such as compensation for mutation, duplication of the basic residues and dependence of the two clusters.

Database searches have revealed that homologues of BM47 are encoded by all of the sequenced lepidopteran nucleopolyhedroviruses. Interestingly, in the C-terminal region, the homologues encoded by group-II NPVs are rich in acidic residues (Asp and/or Glu), whereas these of group-I NPVs have a high content of basic residues (Lys and/or Arg). These basic residues constitute two clusters, separated by a spacer. Sequence alignment analysis also shows that two basic clusters are conserved in all homologues of group-I NPVs. It is inferred that homologues of group-I viruses, as exemplified by BM47 in this study, show different subcellular localization from those of group-II NPVs, as exemplified by Sl52 [23] encoded by SpltMNPV and Se101 [22] of SeMNPV. Differences in acidic and basic residues in the C-terminal region maybe account for the distinct subcellular localization.

Although the exact role of BM47 in the BmNPV life cycle is still unclear, the present study will lead to a better understanding of the two basic residue clusters that constitute a novel functional NLS in the C-terminal region of BM47. These clusters are conserved in homologues of all sequenced group-I NPVs. This NLS is strong enough to localize a small protein to the nucleus efficiently for its characteristics of compensation for mutation and duplication of basic residues.

Acknowledgments We would like to thank Professor Enoch Y. Park, Department of Applied Biological Chemistry, Faculty of Agriculture, Shizuoka University, for kindly providing the BmNPV BAC-TO-BAC system. This project was supported by grants from the High Talent Fund of Jiangsu University (No. 05JDG048), the College Natural Science Foundation of Jiangsu Province (No. 07KJB180013), the China Postdoctoral Science Foundation funded project (No. 20100471387) and the National Basic Research Program of China (No. 2005CB121000).

References

1. Ahrens CH, Russell RL, Funk CJ, Evans JT, Harwood SH, Rohrmann GF (1997) The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. Virology 229:381–399
2. Au V, Yu M, Carstens EB (2009) Characterization of a baculovirus nuclear localization signal domain in the late expression factor 3 protein. Virology 385:209–217
3. Ayres MD, Howard SC, Kuzio J, Lopez-Ferber M, Possee RD (1994) The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virology 202:586–605
4. Boulikas T (1994) Putative nuclear localization signals (NLS) in protein transcription factors. J Cell Biochem 55:32–58
5. Chen Z, Carstens EB (2005) Identification of domains in *Autographa californica* multiple nucleopolyhedrovirus late expression factor 3 required for nuclear transport of P143. J Virol 79:10915–10922
6. Chen YR, Wu CY, Lee ST, Wu YJ, Lo CF, Tsai MF, Wang CH (2008) Genomic and host range studies of *Maruca vitrata* nucleopolyhedrovirus. J Gen Virol 89:2315–2330
7. Cokol M, Nair R, Rost B (2000) Finding nuclear localization signals. EMBO Rep 1:411–415
8. de Jong JG, Lauzon HAM, Dominy C, Poloumienko A, Carstens EB, Arif BM, Krell PJ (2005) Analysis of the *Choristoneura fumiferana* nucleopolyhedrovirus genome. J Gen Virol 86:929–943
9. Fan Q, Li S, Wang L, Zhang B, Ye B, Zhao Z, Cui L (2007) The genome sequence of the multinucleocapsid nucleopolyhedrovirus of the Chinese oak silkworm *Antheraea pernyi*. Virology 366:304–315
10. Gomi S, Majima K, Maeda S (1999) Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. J Gen Virol 80:1323–1337
11. Grace TDC (1967) Establishment of a line of cells from the silkworm *Bombyx mori*. Nature 216:613
12. Harrison RL, Bonning BC (2003) Comparative analysis of the genomes of *Rachiplusia ou* and *Autographa californica* multiple nucleopolyhedroviruses. J Gen Virol 84:1827–1842
13. Harrison RL, Lynn DE (2007) Genomic sequence analysis of a nucleopolyhedrovirus isolated from the diamondback moth, *Plutella xylostella*. Virus Genes 35:857–873
14. Hawkins J, Davis L, Bodén M (2007) Predicting nuclear localization. J Proteome Res 6:1402–1409
15. Hyink O, Dellow RA, Olsen MJ, Caradoc-Davies KMB, Drake K, Herniou EA, Cory JS, O'Reilly DR, Ward VK (2002) Whole genome analysis of the *Epiphyas postvittana* nucleopolyhedrovirus. J Gen Virol 83:957–971
16. Ikeda M, Shikata M, Shirata N, Chaeychomsri S, Kobayashi M (2006) Gene organization and complete sequence of the *Hyphantria cunea* nucleopolyhedrovirus genome. J Gen Virol 87:2549–2562

17. Jarvis DL, Bohlmeyer DA, A Garcia Jr (1991) Requirements for nuclear localization and supramolecular assembly of a baculovirus polyhedrin protein. Virology 185:795–810
18. Kalderon D, Roberts BL, Richardson WD, Smith AE (1984) A short amino acid sequence able to specify nuclear location. Cell 39:499–509
19. King LA, Possee RD (1992) The baculovirus expression system: a laboratory guide. Chapman and Hall, London
20. Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH (2007) Classical nuclear localization signals: definition, function, and interaction with importin α. J Biol Chem 282:5101–5105
21. Lauzon HAM, Jamieson PB, Krell PJ, Arif BM (2005) Gene organization and sequencing of the *Choristoneura fumiferana* defective nucleopolyhedrovirus genome. J Gen Virol 86:945–961
22. Li L, Li Z, Chen W, Liu C, Huang H, Yang K, Pang Y (2006) Characterization of *Spodoptera exigua* multicapsid nucleopolyhedrovirus ORF100 and ORF101, two homologues of *E. coli* ChaB. Virus Res 121:42–50
23. Li Z, Li L, Yu H, Li S, Pang Y (2006) Characterization of two homologues of ChaB in *Spodoptera litura* multicapsid nucleopolyhedrovirus. Gene 372:33–43
24. Liu X, Chen K, Cai K, Yao Q (2008) Determination of protein composition and host-derived proteins of *Bombyx mori* nucleopolyhedrovirus by 2-dimensional electrophoresis and mass spectrometry. Intervirology 51:369–376
25. Luckow VA, Lee SC, Barry GF, Olins PO (1993) Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. J Virol 67:4566–4579
26. Michael WM, Eder PS, Dreyfuss G (1997) The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. EMBO J 16:3587–3598
27. Motohashi T, Shimojima T, Fukagawa T, Maenaka K, Park EY (2005) Efficient large-scale protein production of larvae and pupae of silkworm by *Bombyx mori* nuclear polyhedrosis virus bacmid system. Biochem Biophys Res Commun 326:564–569
28. Nie ZM, Zhang ZF, Wang D et al (2007) Complete sequence and organization of *Antheraea pernyi* nucleopolyhedrovirus, a dr-rich baculovirus. BMC Genomics 8:248–261
29. Oliveira JV, Wolff JL, Garcia-Maruniak A, Ribeiro BM, de Castro ME, de Souza ML, Moscardi F, Maruniak JE, Zanotto PM (2006) Genome of the most widely used viral biopesticide: *Anticarsia gemmatalis* multiple nucleopolyhedrovirus. J Gen Virol 87:3233–3250
30. Olson VA, Wetter JA, Friesen PD (2002) Baculovirus transregulator IE1 requires a dimeric nuclear localization element for nuclear import and promoter activation. J Virol 76:9505–9515
31. Robbins J, Dilworth SM, Laskey RA, Dingwall C (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell 64:615–623
32. Romanelli MG, Morandi C (2002) Importin α binds to an unusual bipartite nuclear localization signal in the heterogeneous ribonucleoprotein type I. Eur J Biochem 269:2727–2734
33. Sachdev S, Hoffmann A, Hannink M (1998) Nuclear localization of IκBα is mediated by the second ankyrin repeat: the IκBα ankyrin repeats define a novel class of *cis*-acting nuclear import sequences. Mol Cell Biol 18:2524–2534
34. Sakuma T, Yamada N, Saito H, Kakegawa T, Kobayashi H (1998) pH dependence of the function of sodium ion extrusion systems in *Escherichia coli*. Biochim Biophys Acta 1363:231–237
35. Siomi H, Dreyfuss G (1995) A nuclear localization domain in the hnRNP A1 protein. J Cell Biol 129:551–560
36. Theilmann DA, Blissard GW, Bonning B, Jehle JA, O'Reilly DR, Rohrmann GF, Thiem S, Vlak JM (2005) Baculoviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy: the eighth report of the International Committee on Taxonomy of Viruses. Elsevier, San Diego, pp 177–185
37. Truant R, Fridell RA, Benson RE, Bogerd H, Cullen BR (1998) Identification and functional characterization of a novel nuclear localization signal present in the yeast Nab2 poly(A) + RNA binding protein. Mol Cell Biol 18:1449–1458
38. Wang Y, Wang Q, Liang C, Song J, Li N, Shi H, Chen X (2008) *Autographa californica* multiple nucleopolyhedrovirus nucleocapsid protein BV/ODV-C42 mediates the nuclear entry of P78/83. J Virol 82:4554–4561
39. Wu C, Guo Z, Chen K, Shen H (2008) High-level expression, polyclonal antibody preparation and bioinformatics analysis of *Bombyx mori* nucleopolyhedrovirus *orf47* encodes protein. Int J Ind Entomol 16:87–92
40. Xu YP, Ye ZP, Niu CY, Bao YY, Wang WB, Shen WD, Zhang CX (2010) Comparative analysis of the genomes of *Bombyx mandarina* and *Bombyx mori* nucleopolyhedrovirus. J Microbiol 48:102–110