Yeast-Plant Coupled Vector System for Identification of Nuclear Proteins\textsuperscript{1}[OA]

Adi Zaltsman, Bu-Young Yi, Alexander Krichevsky, Yedidya Gafni, and Vitaly Citovsky* 

Department of Biochemistry and Cell Biology, State University of New York, Stony Brook, New York 11794–5215 (A.Z., B.-Y.Y., A.K., V.C.); Department of Environmental Horticulture, University of Seoul, Tongdaemoon, Seoul 130–743, Korea (B.-Y.Y.); and Department of Genetics, Agricultural Research Organization, Volcani Center, Bet Dagan 50250, Israel (Y.G.)

Nuclear proteins are involved in many critical biological processes within plant cells and, therefore, are in the focus of studies that usually begin with demonstrating that the protein of interest indeed exhibits nuclear localization. Thus, studies of plant nuclear proteins would be facilitated by a convenient experimental system for identification of proteins that are actively imported into the cell nucleus and visualization of their nuclear accumulation in vivo. To this end, we developed a system of vectors that allows screening for cDNAs coding for nuclear proteins in a simple genetic assay in yeast cells, and verification of nuclear accumulation in planta following one-step transfer and autofluorescent tagging of the identified clones into a multiple cloning site-compatible and reading frame-compatible plant expression vector. In a recommended third experimental step, the plant expression cassette containing the identified clone can be transferred, also by a one-step cloning, into a binary multigene expression vector for transient or stable coexpression with any other proteins.

Nuclear proteins perform a multitude of diverse cellular functions, from structure (e.g. lamins and histones) to maintenance (e.g. DNA repair and replication) to chromatin dynamics and gene regulation (e.g. chromatin modifiers and transcription factors). Thus, different nuclear proteins represent the focus of numerous studies that often begin with a single observation that a protein of interest indeed resides within the cell nucleus. This first experimental step of protein characterization would be facilitated by a convenient experimental system that allows identification of nuclear proteins and demonstration of their nuclear import in vivo.

Protein nuclear import occurs through the nuclear pore complex. While small proteins (40–60 kD) often diffuse through the nuclear pore complex, nuclear entry of larger molecules and even of some small endogenous proteins, such as histones (Breeuwer and Goldfarb, 1990), is an active process mediated by specific nuclear localization signals (NLSs) contained in the transported molecule (Dingwall, 1991; Garcia-Bustos et al., 1991). That NLSs are relatively short, e.g. 7 and 16 amino acid residues for the SV40 large T antigen NLS and Xenopus nucleoplasmin bipartite NLS, respectively (Smith et al., 1985; Goldfarb et al., 1986; Robbins et al., 1991), and often diverse in their amino acid composition (Dingwall, 1991; Garcia-Bustos et al., 1991) makes their identification difficult when based solely on amino acid sequence analysis. Moreover, even apparent consensus NLSs, such as the one present at the C terminus of the cytoplasmic protein GUS (Varagona et al., 1991; Citovsky et al., 1992), may not represent active signals. Thus, the only practical way to identify NLSs that function within living cells is by microinjecting (Kalderon et al., 1984; Goldfarb et al., 1986; Guralnick et al., 1996) or expressing the protein of interest in eukaryotic cells (Varagona et al., 1991; Citovsky et al., 1992; Michael et al., 1995). Although these approaches represent a definitive identification of a specific nuclear protein, they are poorly suitable for large-scale screens for karyophilic proteins.

Here, we describe a system of vectors that allows (1) identification of nuclear proteins from products encoded by large numbers of cDNA clones, or even entire libraries, via a simple genetic assay in yeast cells, and (2) verification of nuclear accumulation in planta following one-step transfer and autofluorescent tagging of the identified clones into a multiple cloning site (MCS)-compatible and open reading frame (ORF)-compatible plant expression vector. In a recommended third experimental step, the plant expression cassette containing the identified clone can be transferred, also by a one-step cloning, into a binary multigene expression vector for transient or stable coexpression—following biolistic delivery or Agrobacterium-mediated genetic transformation, respectively—with any other
proteins, such as subcellular localization markers or potential interactors.

RESULTS

Design of the Assay

We designed a simple two-stage assay for identification of proteins containing an active NLS, with a third recommended stage for transgenic expression of the identified proteins (Fig. 1). In the first stage, a cDNA library is constructed in (or transferred to) a yeast nuclear import assay (NIA) expression vector, pNIA-C, which allows specific induction of a reporter gene if the expressed protein reaches the yeast cell nucleus. For this library construction, the cDNA can be either inserted into one of the pNIA-C vectors, or it can be cloned into the mixture of all three vectors, i.e. pNIA-Ca, pNIA-Cb, and pNIA-Cc (Fig. 2), which allows better representation of all three reading frames in the library. The resulting library is introduced into a yeast strain with an inducible selectable heterotrophic marker; the same strain also carries an inducible β-galactosidase marker that allows verification of the reporter gene induction in the selected colonies. This experimental step allows initial selection of NLS-containing protein products from a cDNA expression library. Alternatively, it is also very useful for rapid and inexpensive testing of individual proteins of interest for the presence of active NLS sequences.

For plant nuclear proteins, it is important to verify their nuclear accumulation directly in planta. Thus, the second stage of our assay is one-step transfer of the selected cDNA clones into a modular plant expression vector, pSAT6-EGFP-C (Fig. 1). Subcellular localization of GFP-tagged proteins transiently expressed from this vector is easily determined directly in living plant tissues by confocal or epifluorescence microscopy. In pNIA-C and pSAT6-C vectors, the proteins are tested as N-terminal fusions with the corresponding reporter protein. Although C-terminal fusions, i.e. fusions between the C terminus of the tested protein and the N terminus of the reporter, can also be used for determination of protein localization (see, for example, Tzfira et al. [2005] for pSAT6-N vector series), NLS sequences are often found within proteins (e.g. Dingwall and Laskey, 1991), rather than at their termini, making the location of the reporter tag less critical for detection of the NLS activity.

Finally, it is often useful to augment the initial identification of a nuclear protein by comparing its localization to that of known subcellular markers or by analyzing potential colocalization with its interactors. Furthermore, it may also be useful to express the identified clones stably, in transgenic plants, for example, for studies of its nuclear import during plant development and morphogenesis, or in response to biotic and abiotic factors. These goals are achieved in the third stage of our assay (Fig. 1), during which the expression cassette is transferred—also by a single-step cloning—to a binary multigene expression vector. This vector, pPZP-RCS2, does not contain promoter and terminator sequences and is specifically designed to accept expression cassettes based on pSAT vectors (or comparable constructs; Goderis et al., 2002; Tzfira et al., 2005), rather than for direct cloning of the tested genes; furthermore, it can also accept up to six additional expression cassettes for other proteins. The

![Figure 1](https://example.com/figure1.png)
resulting binary construct can be used for both transient and stable expression.

Detection of Protein Nuclear Import in Yeast

We produced three vectors, pNIA-Ca, pNIA-Cb, and pNIA-Cc, which constitutively produce translational fusions of the tested protein to the C terminus of the mLexA-Gal4AD chimera. A, Three pNIA-C vectors for translational fusions in three reading frames. ORFs of Gal4AD are indicated; mLexA and Gal4AD sequences are indicated by a dark and a light blue box, respectively; two point mutations that produce mLexA are indicated; nucleotide residues inserted to produce three different reading frames are indicated in red; stop codons are indicated by black boxes; MCS restriction sites are indicated; and restriction sites that are not unique are indicated by asterisks. B, Positive and negative controls. mLexA and Gal4AD are indicated by dark and light blue boxes, respectively; VirD2 and VirD2NL are indicated by gray boxes; and NLS is indicated by a red box, and its sequence is indicated in one-letter amino acid code with the basic residues corresponding to the small and large domains of the bipartite NLS shown in capital letters. In the mutated NLS sequence (white box), the deleted residues are indicated by dashes. prom., Promoter; term., terminator.

![Figure 2](image_url)

**Figure 2.** Structural features of the pNIA-C series of vectors. The plasmids produce in-frame fusions of the protein of interest to the C terminus of the mLexA-Gal4AD chimera. A, Three pNIA-C vectors for translational fusions in three reading frames. ORFs of Gal4AD are indicated; mLexA and Gal4AD sequences are indicated by a dark and a light blue box, respectively; two point mutations that produce mLexA are indicated; nucleotide residues inserted to produce three different reading frames are indicated in red; stop codons are indicated by black boxes; MCS restriction sites are indicated; and restriction sites that are not unique are indicated by asterisks. B, Positive and negative controls. mLexA and Gal4AD are indicated by dark and light blue boxes, respectively; VirD2 and VirD2NL are indicated by gray boxes; and NLS is indicated by a red box, and its sequence is indicated in one-letter amino acid code with the basic residues corresponding to the small and large domains of the bipartite NLS shown in capital letters. In the mutated NLS sequence (white box), the deleted residues are indicated by dashes. prom., Promoter; term., terminator.
with an unknown subcellular localization. Figure 3A shows that mLexA-Gal4AD-VIP1 and mLexA-Gal4AD-At1g31350 promoted cell growth on a His-deficient medium (lanes 1 and 2), indicating nuclear import of the fusion protein that is required for induction of the HIS3 reporter gene. This was then confirmed by induction of a different reporter gene, lacZ, encoding the β-galactosidase activity (Fig. 3B, lanes 1 and 2). As expected, the positive control, mLexA-Gal4AD-VirD2, was imported into the yeast cell nucleus, allowing His prototrophy (Fig. 3A, lane 3) and inducing the lacZ gene expression (Fig. 3B, lane 3), whereas the negative control, mLexA-Gal4AD-VirD2ΔNLS, failed to elicit induction of either of the reporter genes (Fig. 3, A [lane 4] and B [lane 4]), indicating the lack of nuclear import that is consistent with the absence of the active NLS in this VirD2 mutant (Howard et al., 1992). Without selection, i.e. in the presence of His, cells harboring all mLexA-Gal4AD fusions displayed vigorous and essentially similar growth (Fig. 3C), indicating that the protein hybrids did not adversely and nonspecifically affect cell physiology.

**Detection of Protein Nuclear Import in Planta**

Once plant proteins with active NLS sequences are identified in yeast, their nuclear accumulation in plant tissues is tested using three plant constitutive expression vectors, pSAT6-EGFP-Ca, pSAT6-EGFP-Cb, and pSAT6-EGFP-Cc, which allow fusion of the tested proteins to the C terminus of the autofluorescent tag GFP in all three ORFs (Fig. 4A). The MCSs and reading frames of the pSAT6-EGFP-C vectors are identical to those of the corresponding pNIA-C vectors; thus, the tested protein can be transferred from the yeast nuclear localization assay system to the plant nuclear localization detection system by a single cloning step. For positive and negative controls of nuclear accumulation, we inserted VirD2 and VirD2ΔNLS sequences into pSAT6-EGFP-C, producing GFP-VirD2 and GFP-VirD2ΔNLS fusions (Fig. 4B).

In addition to their stand-alone use for transient expression, the pSAT6-EGFP-C vectors can serve as modules for the pPZP-RCS2-based binary plasmids (Goderis et al., 2002; Tzfira et al., 2005) capable of multigene expression and stable expression of integrated transgenes (Fig. 4C). In this approach, expression cassettes from the pSAT6-EGFP-C-based constructs are cloned into the PI-PstI rare-cutting intron-encoded nuclease recognition site of pPZP-RCS2 or its derivatives. For multigene expression, additional expression cassettes from other pSAT vectors, i.e. pSAT1-pSAT5 and pSAT7 (Tzfira et al., 2005), can be inserted into the corresponding sites of intron-encoded nucleases indicated in Figure 4C. For production of transgenic plants, the expression cassettes should be mounted onto pPZP-RCS2 derivatives, i.e. pRCS2-apt, pRCS2-ntpII, or pRCS2-bar (Tzfira et al., 2005), carrying different selection markers (Fig. 4C).

As proof of concept, we transferred VIP1 and At1g31350 from the pNIA-C constructs to pSAT6-EGFP-C and transiently expressed the resulting GFP-VIP1 and GFP-At1g31350 fusions in plant tissues following microbombardment. Figure 5 shows that both fluorescently tagged VIP1 and At1g31350 accumulated in the plant cell nucleus (A and B). Interestingly, another F-box protein, the Agrobacterium VirF, which functions in plants, also has been shown to accumulate predominantly in the plant cell nucleus (Tzfira et al., 2004). The nuclear localization pattern, determined by confocal laser scanning microscopy with optical sections through the cell nucleus, matched that of the positive control, GFP-VirD2 (Fig. 5C), whereas the negative control, GFP-VirD2ΔNLS, remained clearly cytoplasmic (Fig. 5D).

Next, we coexpressed our tested and control proteins from a multigene expression vector pPZP-RCS2 with free DsRed2 (Fig. 4C). This reporter, as well as any other small autofluorescent protein, such as cyan fluorescent protein, is very useful for studies of nuclear import because it is known to partition between the cell cytoplasm and the nucleus, conveniently visualizing and identifying both of these cellular compartments (Dietrich and Maiss, 2002; Goodin et al., 2002; Schultheiss et al., 2003; Tzfira et al., 2004; Lacroix et al., 2005; Krichevsky et al., 2007).

Figure 6 shows that GFP-tagged VIP1 (panel A), At1g31350 (panel D), and VirD2 (panel G) accumu-
lated in the plant cell nucleus. As expected, in the same cells, DsRed2 was found both in the cytoplasm and in the nucleus (Fig. 6, B, E, and H, respectively). Combined images of GFP and DsRed2 fluorescence showed overlapping signal (yellow color) within the cell nucleus, confirming the nuclear localization of these GFP-tagged proteins (Fig. 6, C, F, I, and L). In contrast, GFP-VirD2ΔNLS colocalized mainly with the cytoplasmic DsRed2 (Fig. 6, J–L).

**DISCUSSION**

Ideally, an assay for identification of nuclear proteins would combine an easy, high-throughput, and...
the protein within the cell nucleus, confirming the results of the genetic assay, as well as determination of whether this protein is also found in any other cellular compartments. Furthermore, the capability of multiple pSAT vector-derived expression cassettes to be mounted onto a binary vector for coexpression (Tzfira et al., 2005) makes it easy to express the nuclear protein of interest with subcellular localization markers.

To couple both parts of the vector system together, pNIA-C and pSAT6-EGFP-C plasmids were generated with identical MCSs and reading frames, allowing one-step transfer of the tested sequences from one vector to another. For proof of concept, we put through the entire pNIA-C/pSAT6-EGFP-C protocol four unrelated proteins: Arabidopsis nuclear factor VIP1 and Arabidopsis F-box protein with an unknown localization (At1g31350), as tested proteins, and Agrobacterium VirD2 protein and its mutant VirD2ΔNLS as positive and negative controls, respectively. All these proteins exhibited their expected functionalities in respect to subcellular localization both in the yeast and in planta assays. Importantly, the transfer of the DNA sequences encoding the tested and control proteins from pNIA-C to pSAT6-EGFP-C and then to pFZP-RCS2 vectors was very simple and efficiently achieved in a single cloning step. Further, the identification of nuclear import was unambiguous in both assays, detected as either vigorous yeast cell growth or substantial intranuclear accumulation of the GFP signal. Identification of the latter step was facilitated by coexpression of a small diffusible autofluorescent marker, DsRed2, which clearly delineated the cell nucleus and cytoplasm.

In summary, we described the construction and experimental application of two sets of vectors for identification of functional NLS sequences in yeast and subsequent demonstration of their nuclear import capabilities in planta. We hope that these vectors will further facilitate experimentation in plant cell biology that often requires identification of proteins with nuclear import capabilities. Furthermore, in the future, this assay system can be expanded to incorporate pSAT vectors adapted for Gateway cloning and for use with other spectral variants of GFP (Tzfira et al., 2005).

**MATERIALS AND METHODS**

**Construction of pNIA-C and pSAT6-EGFP-C Vectors**

Standard DNA amplification and cloning methods were used to construct all vectors. pEGFP-C1 (CLONTECH) was used as the source of the MCS and SV40 terminator, which were amplified by PCR using 5‘AACCGGTGTA-
CAAGTCCGGACTCAGATCTCG3, 5‘AACCGGTGTTACAAAGTCCGGAC-
TCAGATCTCG3, and 5‘AACCGGTGTCAGTCCGGACTCAGATCTCG3 as forward primers for each of the three reading frames, and 5‘TTGCC-
GCCCAGCCTGTTACAATTTACGCTG3 as reverse primer. Each of the amplification products was digested with AgeI and Nael and inserted into the Xmal/Nael sites of pNIA (Rhee et al., 2000), resulting in pNIA-Ca, pNIA-Cb, and pNIA-Cc (GenBank accession nos. EF587311, EF587312, and EF587313, respectively). The MCS-terminator cassette was then excised from each of the pNIA-C vectors by digesting them with BsgI and Nofl and inserted into the Xmal/Nael sites of pNIA (Rhee et al., 2000), resulting in pNIA-Ca, pNIA-Cb, and pNIA-Cc (GenBank accession nos. EF587311, EF587312, and EF587313, respectively). The MCS-terminator cassette was then excised from each of the pNIA-C vectors by digesting them with BsgI and Nofl and inserted into the same site of pSAT6-EGFP-C1 (Tzfira et al., 2005), producing pSAT6-EGFP-Ca, pSAT6-EGFP-Cb, and pSAT6-EGFP-Cc (GenBank accession nos. EF587314,
EF587315, and EF587316, respectively). VIP1 cDNA (Tzfira et al., 2001) was cloned as a PCR-amplified fragment into the SalI-BamHI sites of pNIA-Cb and then transferred into the SalI-BamHI sites of pSAT6-EGFP-Cb. Similar cloning strategy was employed for At1g31350 and VirD2 (or VirD2ΔNLS), except that they were cloned into the XhoI-PstI and BamHI sites of the expression vectors, respectively. Note that while we demonstrate nuclear import using the pNIA-Cb and pSAT6-EGFP-Cb constructs, the functionality of vectors expressing in all three reading frames has been tested (data not shown).

For coexpression of a tested GFP-tagged protein with free DsRed2 from a single plasmid, we first produced pSAT5-DsRed2 by cloning the DsRed2 expression cassette as a AgeI-NotI fragment from pSAT6-DsRed2-C1 (Tzfira et al., 2005) into a pSAT5-based vector (Tzfira et al., 2005), producing pSAT5-DsRed2-C1. The DsRed2 expression cassette was then removed by digestion with I-CeuI and inserted into the I-CeuI site of pPZP-RCS2, producing pRCS2-DsRed2 (Tzfira et al., 2005). Finally, the EGFP-tested protein expression cassette was removed from the pSAT6-EGFP-Cb vector by digestion with PspI-PspI and inserted into the PspI-PspI site of pRCS2-DsRed2 (Tzfira et al., 2005). pRCS2-hpt, pRCS2-nptII, and pRCS2-bar have been described before (Tzfira et al., 2005). All constructs were verified by DNA sequencing.

Nuclear Import in Yeast Cells

For yeast nuclear import, the Saccharomyces cerevisiae strain L40 cells (Hollenberg et al., 1995) were transformed with the indicated pNIA-C constructs and grown for 2 d at 30°C on a Leu- and Trp-deficient medium in the presence or absence of His; His prototrophy indicated functional nuclear import (Rhee et al., 2000). Alternatively, cells were grown on a Leu- and Trp-deficient medium, and the β-galactosidase activity was assayed on nitrocellulose filters (Hollenberg et al., 1995).

Nuclear Import in Plant Tissues

DNA (50 μg) was adsorbed onto 10 mg of 1-μm gold particles (Bio-Rad) and bombarded at 150 to 200 psi into the leaf epidermis of greenhouse-grown Nicotiana benthamiana plants using a Helios gene gun (PDS-1000/He; Bio-Rad), followed by incubation for 24 h at 25°C. For detection of fluorescently tagged proteins, the bombarded tissues were directly viewed under a Zeiss LSM 5 Pascal confocal laser scanning microscope. Experiments were repeated at least four times, with 10 to 20 expressing cells examined in each experiment.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers EF587311 (pNIA-Ca), EF587312 (pNIA-Cb), EF587313 (pNIA-Cc), EF587314 (pSAT6-EGFP-Ca), EF587315 (pSAT6-EGFP-Cb), and EF587316 (pSAT6-EGFP-Cc).

ACKNOWLEDGMENT

We thank Dr. Tzvi Tzfira for stimulating discussions.

Received July 26, 2007; accepted August 13, 2007; published August 17, 2007.

Figure 6. Assay for nuclear import in planta: multi-gene expression from the pPZP-pRCS2 vector. A to C, GFP-VIP1 + free DsRed2. D to F, GFP-At1g31350 + free DsRed2. G to I, GFP-VirD2 + free DsRed2. J to L, GFP-VirD2ΔNLS + free DsRed2. A, D, G, and J show GFP signal in green; B, E, H, and K show DsRed2 signal in red; and C, F, I, and L show merged GFP and DsRed2 images with colocalizing signals in yellow. Plastid autofluorescence was filtered out. Images are projections of several confocal sections.
LITERATURE CITED


