Plant Physiol. (1993) 101: 353–361

Nuclear Targeting of the Maize R Protein Requires Two Nuclear Localization Sequences[1]

Mark W. Shieh, Susan R. Wessler, and Natasha V. Raikhel*

Michigan State University Department of Energy Plant Research Laboratory, East Lansing, Michigan 48823-1312 (M.W.S., N.V.R.); and Department of Botany, University of Georgia, Athens, Georgia 30602 (S.R.W.)

Previous genetic and structural evidence indicates that the maize *R* gene encodes a nuclear transcriptional activating factor. In-frame carboxyl- and amino-terminal fusions of the *R* gene to the reporter gene encoding β-glucuronidase (GUS) were sufficient to direct GUS to the nucleus of the transiently transformed onion (*Allium cepa*) epidermal cells. Further analysis of chimeric constructs containing regions of the *R* gene fused to the *GUS* cDNA revealed three specific nuclear localization sequences (NLSs) that were capable of redirecting the GUS protein to the nucleus. Amino-terminal NLS-A (amino acids 100–109, GDRRAAPARP) contained several arginine residues; a similar localization signal is found in only a few viral proteins. The medial NLS-M (amino acids 419–428, MSERKRREKL) is a simian virus 40 large T antigen-type NLS, and the carboxyl-terminal NLS-C (amino acids 598–610, MISESLRK-AIGKR) is a mating type α2 type. NLSs M and C are independently sufficient to direct the GUS protein to the nucleus when it is fused at the amino terminus of GUS, whereas NLS-A fused to GUS partitioned between the nucleus and cytoplasm. Similar partitioning was observed when localization signals NLS-A and NLS-C were independently fused to the carboxy-terminal portion of GUS. A sequential deletion of the localization signals indicated that the amino-terminal and carboxyl-terminal fusions of R and GUS were redirected to the nucleus only when both NLS-A and -M, or NLS-C and -M, were present. These results indicate that multiple localization signals are necessary for nuclear targeting of this protein. The conservation of the localization signals within the alleles of *R* and similar proteins from other organisms is also discussed.

In eukaryotic cells, proteins can be targeted to a variety of subcellular compartments such as the ER, mitochondrion, chloroplast, peroxisome, glyoxisome, or nucleus. The import of proteins into the nucleus, which has been examined extensively in mammalian, amphibian, and yeast systems, can be distinguished from transport into other organelles because proteins and small molecules traverse the nuclear envelope through a macromolecular complex known as the nuclear pore (for reviews see Wagner et al., 1990 and Nigg et al., 1991). The nuclear pore complex forms a large channel across the nuclear membrane that allows diffusion of small molecules, yet tightly regulates the movement of larger molecules (for review see Dingwall and Laskey, 1986; Newmeyer et al., 1986). Unlike the amino-terminal signal sequences that direct proteins from the cytoplasm to the ER, mitochondrion, and chloroplast, the import of nuclear proteins is mediated by NLSs that may be located at any position within a protein (Garcia-Bustos et al., 1991). In addition, NLSs are not proteolytically cleaved from the protein, which allows nuclear proteins to reenter the nucleus after cell division.

There is no consensus sequence for NLSs; however, they are characterized as short aa regions that are rich in basic residues (Garcia-Bustos et al., 1991). The known NLSs can be categorized into three classes based upon their composition and structure: those resembling the SV40 large T type antigen (Kalderon et al., 1984a, 1984b; Lanford and Butel, 1984), those resembling MAT α2 (Hall et al., 1984), and those with a bipartite signal structure (nucleoplasmin; Dingwall and Laskey, 1991). Recently, several NLSs have been identified in plants, and these are similar to the mammalian and yeast NLSs (see Raikhel [1992] for review).

For our localization studies in higher plants, we have chosen to utilize the maize R protein. Prior genetic analysis indicates that the R protein controls where and when the anthocyanin biosynthetic pathway is expressed in plant tissues (Ludwig and Wessler, 1990). Consistent with a proposed regulatory role was the finding that the *R* gene encodes a protein with the structural features of a transcriptional activator, including large acidic and basic regions and a basic helix-loop-helix domain (Ludwig et al., 1989). As a transcriptional activator, the R protein should localize to the nucleus. However, the predicted molecular mass of the R protein is 66 kD, which exceeds the size limit for the diffusion of gold particles through the nuclear pore complex (Paine et al., 1975). Thus, the R protein is a reasonable choice for the study of nuclear protein import in higher plants because it should possess at least one NLS.

The goal of this study was to identify NLSs in the R protein and to determine whether or not they were sufficient and necessary for nuclear transport. To facilitate the localization of the protein within plant cells, the reporter gene *GUS* was fused to the cDNA of an allele of the *R* gene called *Lc* (leaf color). The gene fusions were transiently expressed in onion (*Allium cepa*) epidermal cells following introduction of the DNA by particle bombardment. Using this system, three

[1] Research supported by U.S. Department of Energy Grant DE-FG02-90ER20021 and U.S. Department of Agriculture Grant 92-37301-7709 to N.V.R.

* Corresponding author; fax 1–517–353–9168.

Abbreviations: aa, amino acid; DAPI, 4′,6′ diamidino-2-phenylindoledihydroxychloride; DEL, delila; GUS, β-glucuronidase; Lc, leaf color; MAT α2, mating type α2; MS, Murashige and Skoog; NLS, nuclear localization sequence; nt, nucleotide; SV40, simian virus 40; X-gluc, 5-bromo-4-chloro-3-indoyl glucuronide.

NLSs were identified in the maize R protein. We have also determined that at least two of the NLSs are necessary and sufficient to target the R:GUS fusion protein to the nucleus in onion cells. These results may be of broad significance because they constitute the first reported instance where multiple NLSs are required for competent transport of a plant regulatory protein.

MATERIALS AND METHODS

Materials

White onions (*Allium cepa*) were purchased locally, stored at 4°C in the dark, and used within 2 weeks. Oligonucleotides were synthesized by the Michigan State University Macromolecular Facility or by CIBA-GEIGY Biotechnology (Research Triangle Park, NC). The enzymes used in the restriction digests were purchased from Boehringer Mannheim Biochemicals (Indianapolis, IN), and enzymes used for other molecular manipulations were purchased from New England Biolabs (Beverly, MA). The supplies for the helium biolistic gun transformation system (Du Pont, Wilmington, DE) were from Bio-Rad (Richmond, CA).

Constructs

All standard recombinant DNA protocols were obtained from Sambrook et al. (1989). The protocol for site-directed mutagenesis was performed as described by Kunkel et al. (1987). After mutagenesis, constructs were sequenced to verify their integrity, and completed constructs were subcloned into the expression vector pGA643 (An et al., 1988), except for the *R:GUS* 598–610 construct, which was ligated into the pMF6 expression vector (Goff et al., 1990). Expression vectors pGA643 and pMF6 expressed the gene fusions at the same relative level as determined by histochemistry (data not shown). The allele of the *R* gene used in this study was *Lc* (Ludwig et al., 1989).

R:GUS

A *Sac*I restriction site was inserted before the stop codon (nt 1830), and a *Sma*I restriction site was inserted after the stop codon of the *Lc* cDNA by site-directed mutagenesis. The *GUS* cDNA (pBI101.3, Jefferson, 1987) was then subcloned in front of the stop codon of the *Lc* cDNA.

GUS:R

Lc cDNA was modified to include an *Xho*I and a *Sma*I site in frame before the first initiating AUG codon by site-directed mutagenesis. Also an *Xho*I restriction site was inserted in frame in front of the stop codon in *GUS* (nt 1807) by site-directed mutagenesis. The modified *GUS* gene was then subcloned in front of the *Lc* cDNA.

R:GUS 82–610, 598–610

Restriction enzymes *Bgl*II and *Sac*I were used to construct the *R:GUS* gene fusions encoding aa's 82–610 and 598–610 from the *R:GUS* construct.

GUS:R 1–109

Restriction enzymes *Nae*I and *Sca*I were used to construct *GUS:R* 1–109 from the *GUS:R* construct.

To facilitate subcloning of the *Lc* cDNA deletion constructs, a set of restriction enzyme digest sites was constructed that includes a *Kpn*I site followed by an ATG and an *Xho*I site and was introduced into both *R:GUS* and *GUS:R*. By adding this set of restriction sites at positions before nt's encoding aa 411 (nt 1231), 457 (nt 1368), and 512 (nt 1533), it was possible to subclone the fragments as *Kpn*I and *Eco*RI *(R:GUS)* fragments, thereby making the constructs *R:GUS* 411–610, *R:GUS* 457–610, and *R:GUS* 512–610. To construct *GUS:R* 1–411, *GUS:R* 1–457, and *GUS:R* 1–512, the same set of restriction sites was added. However, these *GUS:R* constructs were subcloned as *Xba*I (5′ of *GUS*) and *Kpn*I fragments into *Xba*I and *Kpn*I sites situated so that there is a stop codon in frame after the *Kpn*I site.

When two sets of the restriction sites (*Xho*I-ATG-*Kpn*I) were added at nt's encoding aa's 411 and 457 (nt's 1231 and 1368) or aa's 411 and 512 (nt's 1231 and 1533), they allowed the isolation and cloning of an *Xho*I (5′) to *Kpn*I (3′) fragment. To subclone the *R Xho*I (5′) to *Kpn*I (3′) fragment into *GUS*, an additional *Kpn*I site was added to either *R:GUS* at nt 1831 of *R* or *GUS:R* at nt 1831 of *R*; in subcloning, the *R* gene is removed before the fragments encoding aa's 411–457 or 411–512 are inserted.

R:GUS 128–411

Utilizing the construct *R:GUS* 1–411, an *Nae*I and *Kpn*I fragment (encoding aa's 128–411) was inserted into a *Sma*I- and *Kpn*I-cut *R:GUS* construct. (The *Kpn*I site was added by site-directed mutagenesis at nt 1831; the *Sma*I and *Kpn*I cut drops out the *R* gene.)

R:GUS 1–109 and 82–109

R:GUS, with the additional *Kpn*I site at nt 1831, was cut with *Nae*I and *Kpn*I restriction enzymes. T4 DNA polymerase was then used to make blunt ends, which were ligated (*R:GUS* 1–109). The *R:GUS* 1–109 construct was then digested with *Bgl*II (leaving the first ATG at nt 246, aa 82) and *Eco*RI and ligated into pUC118 at *Bam*HI and *Eco*RI sites to construct *R:GUS* 82–109.

To construct the *R:GUS* and *GUS:R* fusions that encoded aa's 100–109, a *Kpn*I site was introduced at nt 300 of *R* in the *R:GUS* 1–109 construct encoding nt's 1–327. Then the fragment encoding aa's 100–109 was subcloned into pUC118 as a *Kpn*I and *Eco*RI fragment. The constructs encoding aa's 419–428 were constructed by adding a *Kpn*I site after the codon for aa 428 (nt 1284) to the *R:GUS* 411–457 and *GUS:R* 411–457 constructs. The nucleotides encoding aa's 429–457 were then dropped from the clone as a *Kpn*I fragment.

The deletion constructs outlined in Figure 4 were also constructed using site-directed mutagenesis on the *R:GUS* and *GUS:R* constructs. When each NLS-encoding region was deleted, a specific restriction site was inserted or created for confirmation that the sequence was deleted. The site-directed mutagenesis removed nt's 300–327 for NLS-A (*Nae*I), nt's 1257–1284 for NLS-M (*Acc*I), or nt's 1794–1830 for NLS-C

(*Kpn*I). By utilizing three, two, or one of the deletion mutations in a single construct, the different combinations of NLSs could be deleted.

Transformation of Onion Cells

Onion epidermal layers were placed inside up on a Petri plate containing MS basal media (per liter: 4.2 g of MS salts [Gibco-BRL], 1 mg of thiamine, 10 mg of *myo*-inositol, 180 mg of KH_2PO_4 [Miller I], 30 g of sucrose, pH 5.7) (Murashige and Skoog, 1962) with the antifungal agent amphotericin B (2.5 mg/L; Sigma) and 6% agar. Plasmid DNAs were prepared using either $CsCl_2$ gradient purification (Sambrook et al., 1989) or column purification (Qiagen, Chatsworth, CA). The plasmids (2.5 μg) were precipitated onto 1.6-μm gold particles (1.25 μg) as described by the manufacturer (DuPont, Wilmington, DE). DNA-coated particles were washed with 180 μL of 100% ethanol and then resuspended in 30 μL of 100% ethanol. Vortexing and then sonication (cup horn probe, 60% power, 5 s) were used to resuspend the particles before loading 10 μL of the suspension onto each of three particle delivery discs. Petri plates of onion epidermal cells were transformed with the three particle delivery discs (two discs on one plate and one disc on another plate) via the helium biolistic gene transformation system. Rupture discs of 1300 p.s.i. were optimal for onion cell transformation. Transformed cells were incubated at 28°C in the dark for 24 or 48 h.

Histochemical Analysis

The colorimetric substrate X-gluc was used to determine the location of the enzymic activity of the R:GUS and GUS:R fusion proteins. The protocol for the addition of substrate to the onion cells was described in Varagona et al. (1992). The DNA-specific nuclear stain DAPI was included in the mounting solution for each sample (Varagona et al., 1991). Intracellular localization of the blue precipitate was determined using a Zeiss Axiophot microscope with Nomarski optics. Location of the blue precipitate was compared with the location of DAPI-stained nuclei using fluorescence optics. The subcellular localization of each fusion protein was determined from two to four separate transformations. The minimum number of cells analyzed for each construct was 3 and the maximum was 30.

RESULTS

The R Protein Redirects GUS to the Nucleus

To determine whether the R protein is imported to the nucleus, the *R* (*Lc* allele) and *GUS* cDNAs were ligated to form a gene fusion. Because an active GUS enzyme might sterically alter the R protein, the coding region of *R* was ligated both 5′ and 3′ of the *GUS* gene to increase the probability that putative targeting signals would be properly exposed for recognition by the nuclear targeting apparatus. The fusion constructs were then ligated into the expression vector pGA643 between the cauliflower mosaic virus 35S promoter and the nopoline synthetase terminator sequences. The constructs *R:GUS* (*R* 5′ of *GUS*), *GUS:R* (*R* 3′ of *GUS*), and *GUS* were then transformed into a monolayer of onion epidermal cells by particle-gun bombardment. Subcellular localization of the fusion proteins was determined with the histochemical substrate X-gluc, which, when processed by GUS, forms a blue precipitate. When the GUS protein was expressed in onion cells, the blue dye remained in the cytoplasm (results not shown, Varagona et al., 1992). However, when *R* is fused to *GUS* (*R:GUS* or *GUS:R* fusion constructs) and expressed in onion cells, *GUS* activity was redirected to

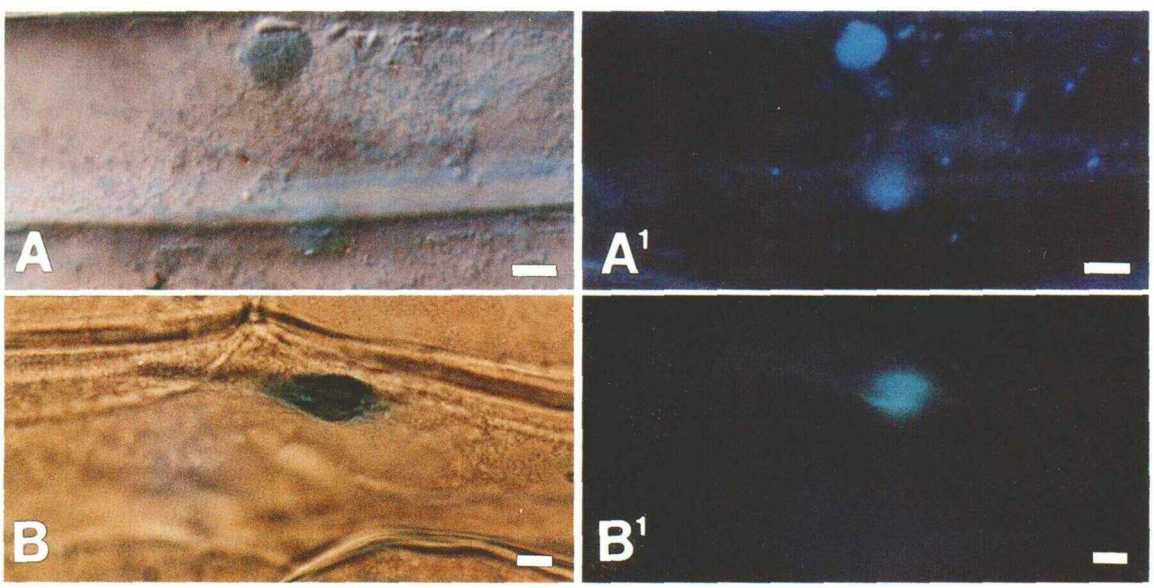

Figure 1. Histochemical localization of R:GUS (A) and GUS:R (B) fusion proteins in onion epidermal cells. Tissues were simultaneously analyzed using both X-gluc histochemical staining (A and B) and nuclei-specific DAPI staining (A^1 and B^1). Nomarski optics were used in A and B and fluorescence optics in A^1 and B^1. Bars = 10 μm.

the nucleus (Fig. 1, A and B, respectively). The conclusion from these experiments was that the R protein was sufficient to redirect the reporter protein GUS to the nucleus, indicating that the R protein contained at least one NLS.

R Protein Contains Three NLSs

The strategy used to identify the NLSs in the R protein was to construct gene fusions in which coding regions from either the 5′ or 3′ end of the *R* gene were deleted (Fig. 2, A and B). Thus, putative NLS could be identified by a process of elimination. Initially, the *R:GUS* construct was modified with deletions at the 5′ terminus (Fig. 2A), and the *GUS:R* construct was modified with 3′ deletions (Fig. 2B). In addition, constructs were specifically designed around aa's 411–457 because this region is enriched in basic aa's, characteristic of NLSs, and contains the helix-loop-helix region (aa's 420–462; Ludwig et al., 1989). Upon completion, the constructs were ligated into expression vectors as described in "Materials and Methods."

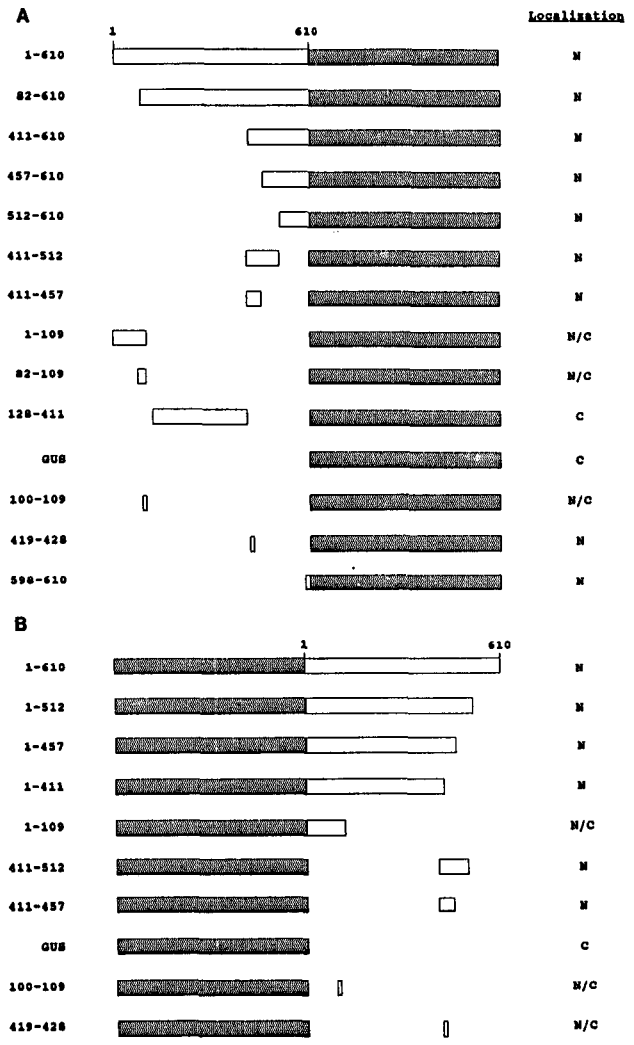

Figure 2. Cloning strategy for preparing R:GUS (A) and GUS:R (B) fusions and results of localization experiments. The upper construct in A represents the amino-terminal fusion of coding sequences of the *R* cDNA clone (open box) and the *GUS* cDNA clone (wavy-lined box), and, in B, the carboxyl-terminal fusion of *GUS* cDNA clone to the *R* cDNA clone. The positions of first and last deduced aa's in the *R* cDNA clone are indicated above the constructs in A and B. The aa's of the R protein used to prepare amino- (A) and carboxyl- (B) terminal fusions to GUS are indicated on the left. The results of subcellular localizations determined by histochemical assays for GUS activity are indicated on the right. N, Nuclear; C, cytoplasmic; N/C, both nuclear and cytoplasmic.

The *R:GUS* deletion constructs were expressed in onion epidermal cells and the subcellular locations of the resulting proteins were determined by assaying for GUS activity (Fig. 2A). The series of deletions from the amino terminus contained aa's 82–610, 411–610, 457–610, 512–610, and 598–610 and revealed NLS-C (NLS in the carboxyl terminus). The 13 aa's encoded at position 598–610 (NLS-C) of R were sufficient to redirect GUS to the nucleus (Fig. 3). The localization of GUS by NLS-C was exclusively to the nucleus and exhibited subcellular localization similar to that of the intact R protein fused to GUS (Fig. 1A).

The deletion constructs were also used to examine the amino terminus of the GUS:R fusions (Fig. 2B). The series of deletions from the carboxyl terminus contained aa's 1–512, 1–457, 1–410, and 1–109 and revealed NLS-A (NLS in the amino terminus). NLS-A was further defined by constructs containing aa's 82–109 and 100–109 (Fig. 2, A and B). GUS activity of the fusion protein NLS-A + GUS (aa's 100–109) partitioned between the nucleus and cytoplasm (Fig. 3).

Because the *R:GUS* and *GUS:R* deletion constructs described could not distinguish any NLSs in the region of aa's 109–598, a second set of constructs was designed (Fig. 2, A and B). The central region of the R protein (aa's 109–598) was subdivided into constructs containing the basic helix-loop-helix motif (aa's 411–457 and 419–428) and the nonbasic residue-rich region (aa's 128–411). aa's 128–411 were unable to redirect GUS to the nucleus and remained in the cytoplasm (Fig. 2A); this was not analyzed in the GUS-R orientation. However, aa's 411–512, 411–457, and 419–428 (NLS-M) were sufficient to redirect GUS to the nucleus (Figs. 2A and 3). NLS-M was located in the amino terminus of the helix-loop-helix motif and, unlike NLS-A, was as efficient as NLS-C in localizing GUS activity exclusively to the nucleus. The GUS-NLS-M fusion protein (GUS:R orientation) resulted in GUS activity partitioned in the cytoplasm and nucleus (Fig. 2B). Therefore, in this study, the amino-terminal GUS fusions displayed stronger redirection of GUS activity to the nucleus. In conclusion, the R protein contained three NLSs (A, M, C), each of which was sufficient to redirect the reporter protein GUS to the nucleus of onion epidermal cells (Fig. 3).

Two NLSs Are Necessary for Transport of R:GUS to the Nucleus

The identification of three NLSs in the R protein that were sufficient to redirect the GUS reporter protein to the nucleus prompted our investigation of the role of these NLSs in the full-length protein. To determine which NLSs were functional and necessary for the import of intact R protein, site-directed mutagenesis was used to delete the NLSs from the fusion constructs of *R:GUS* and *GUS:R*. This strategy resulted

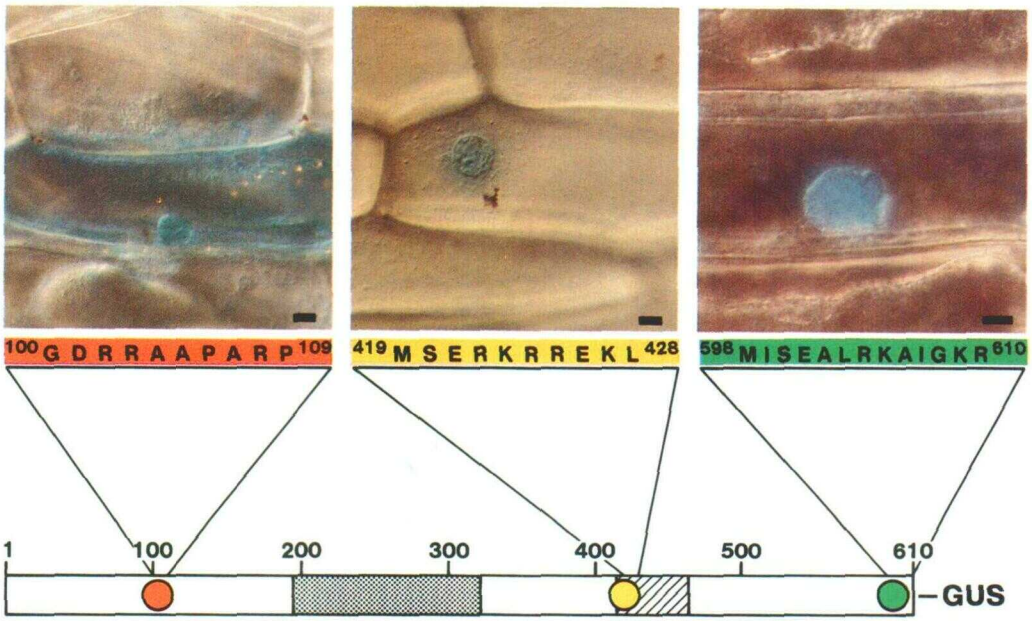

Figure 3. Histochemical localization of three NLS regions of the R protein fused to GUS (above) and schematic representation of R:GUS fusion protein showing localization of three NLSs (below). Positions of aa's are indicated above the construct; the acidic domain of R protein (striped box), helix-loop-helix domain (stippled box), and three NLSs (NLS-A, orange circle; NLS-M, yellow circle; and NLS-C, green circle) are indicated. aa sequences of three NLSs of R protein are shown under corresponding photomicrographs. Tissues were stained using X-gluc histochemical staining and analyzed with Nomarski optics. The brown particles on pictures with NLS-A and NLS-M result from gold precipitation. Bars = 10 μm.

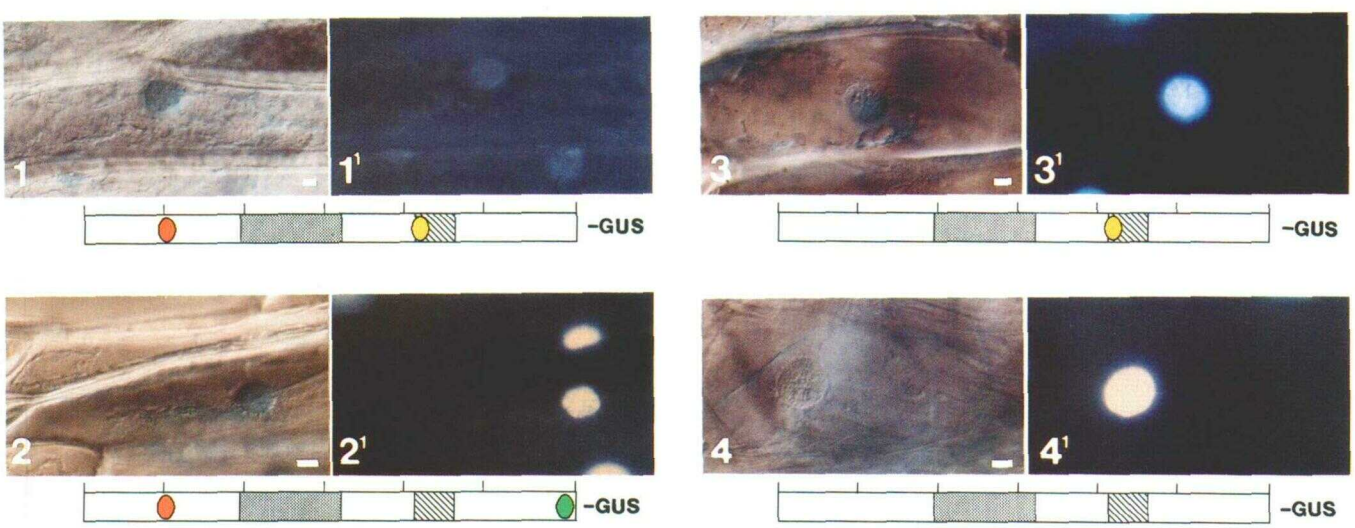

Figure 4. Effect of deletion of different NLSs on the histochemical localization of R:GUS fusion proteins. Deletion of different NLSs from the intact R protein fused to GUS showed that NLS-A and -M, or NLS-M and -C, are required for nuclear targeting. Several examples of the histochemical localizations for R-GUS fusion proteins are shown. The main features of the R protein are the same as in Figure 3, except intact R protein was fused to GUS with deletions of specific NLSs. 1, R protein containing NLS-A (orange circle) and NLS-M (yellow circle) is indicated. 2, R protein containing NLS-A and NLS-C (green circle) is indicated. 3, R protein containing only NLS-M. 4, All three NLSs deleted from R protein. Tissues were simultaneously analyzed using both X-gluc histochemical staining (1–4) and nucleus-specific DAPI staining (1^1–4^1). Tissues were stained and analyzed as in Figure 1. Bar = 10 μm.

$A_{100-109}$ NLS	$M_{419-428}$ NLS	$C_{598-610}$ NLS	LOCALIZATION R-GUS
-	-	-	C
+	-	-	N/C
-	+	-	N/C
-	-	+	N/C
+	+	-	N
+	-	+	N/C
-	+	+	N

Figure 5. Summary of histochemical analysis of R:GUS fusion proteins, which identified NLSs that were necessary for nuclear localization.

in either none, one, or two NLSs in the R protein (Fig. 4). The constructs were then subcloned into expression vectors and transiently expressed in onion epidermal cells, as in the previous experiments.

When all three NLSs (A, M, C) were deleted from R:GUS and GUS:R fusion proteins, GUS activity was retained within the cytoplasm (Fig. 4[4]). This indicated that all NLSs in the R protein were identified, although the formal possibility exists that the deletion of the NLS could sterically hinder an unidentified signal. These results also showed that the strongest determinants of each targeting signal were within the identified NLSs.

To determine whether or not any single NLS was capable of targeting the fusion protein, two of the three NLSs were deleted from R:GUS and GUS:R fusion proteins in each of three possible combinations (Fig. 5). NLS-A, in the intact R protein was inefficient as a signal and resulted in GUS activity in both the nucleus and cytoplasm (Fig. 5). Therefore, in both the intact R:GUS and NLS-A:GUS protein, NLS-A was an inefficient NLS (Figs. 3 and 5). Both NLS-M and NLS-C were inefficient in the intact R protein and conferred partitioned localization. However, their expression in the nucleus was visibly greater than in the cytoplasm (Figs. 4[3] and 5).

The constructs that retained two of the three NLSs displayed different subcellular locations depending upon the orientation of R protein to GUS (Fig. 5). Because the R:GUS fusions exhibited stronger nuclear localization than the GUS:R fusions (Fig. 5), conclusions were drawn from the R:GUS fusion proteins. If NLS-A (Fig. 5) or NLS-C (Figs. 4[1] and 5) were deleted, the fusion protein localized to the nucleus (Fig. 5). Therefore, combinations of either NLS-A and NLS-M or NLS-C and NLS-M were sufficient for nuclear localization. However, if NLS-M was deleted, the fusion protein partitioned between the nucleus and cytoplasm (Fig. 4[2]). Our conclusion from these data was that two NLSs, one of which must be NLS-M, were sufficient and necessary for the transport of R:GUS protein to the nucleus.

DISCUSSION

To identify the NLSs of the maize R protein, a transient expression system was developed utilizing onion cells. Onion epidermal cells were used because their large size facilitated subcellular localization and provided a useful transformation system for particle gun bombardment (Klein et al., 1987). Furthermore, the results of subcellular localization in onion cells were shown to correlate with the localizations determined by stable transformation of Opaque2-GUS fusion proteins in tobacco plants (Varagona et al., 1992). In that study, cellular fractionation and histochemical analysis of the transgenic tobacco cells was used to determine the location of the fusion proteins. It was shown that the subcellular locations of the GUS enzymic activities correlated with those determined by the transient expression assays in onion epidermal cells. Therefore, transformation of onion cells by particle bombardment is a rapid and efficient system for studying nuclear localization.

The full-length R protein fused to GUS yielded nuclear localization in both amino- and carboxyl-terminal orientations. However, only amino-terminal fusion proteins were efficiently transported to the nucleus when smaller regions of the R protein were fused to GUS, indicating that the position of the NLS in the transported protein is important.

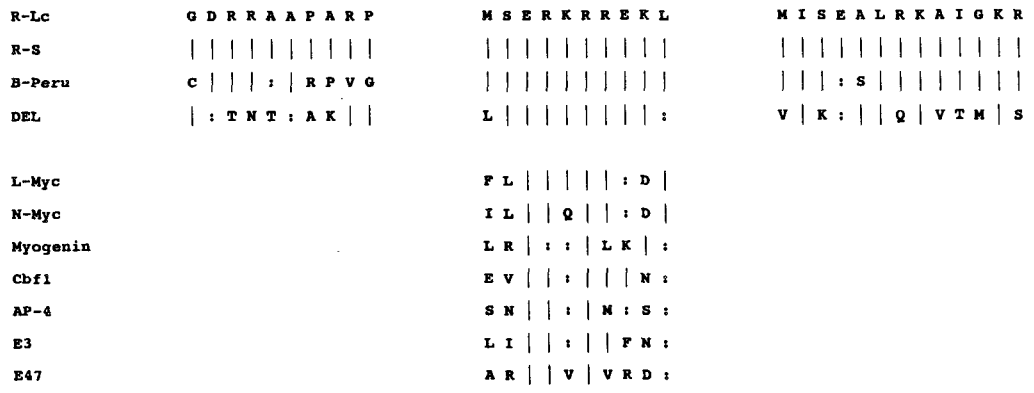

Figure 6. Amino acid composition of R-Lc to other homologous regulatory proteins. Alignments are made to maximize homology with the NLSs of R. Identical aa's are marked by vertical lines and the conservative substitutions by two dots. The sequences shown are for maize R-Lc (Ludwig et al., 1989); maize R-S (Perrot and Cone, 1989); maize B-Peru (Radicella et al., 1991); *Antirrhinum* DEL (Goodrich et al., 1992); L-Myc (DePhino et al., 1987); N-Myc (Kohl et al., 1986); myogenin (Edmundson and Olson, 1989); CBF-1 (Cai and Davis, 1990); AP-4 (Hu et al., 1990), human E3 (Beckman et al., 1990); and human E47 (Voronova and Baltimore, 1990).

A similar conclusion was drawn when the bipartite NLS of Opaque2 protein was analyzed (Varagona et al., 1992).

Three nuclear localization signals were identified in the R protein (NLS-A, -M, and -C) utilizing the onion system. Two of the NLSs, NLS-M (419–428) and NLS-C (598–610), are intact signals, because they redirected GUS activity exclusively to the nucleus (Fig. 3). The third signal, NLS-A (100–109), partially redirected GUS to the nucleus, partitioning the fusion protein between the nucleus and the cytoplasm. Because several larger constructs including NLS-A, encoding aa's 82–109 and 1–109 (Fig. 2), also partially redirected GUS, this inefficient targeting may be due to intrinsic weakness of the targeting signal, or it is possible that aa's following aa 109 are part of the signal, but this was not analyzed. The identification of the three NLSs was confirmed by the finding that the gene fusion constructs containing the full-length R protein with the three NLSs (A, M, and C) deleted were retained in the cytoplasm. Deletion of two of the three NLSs revealed that all three NLSs were not required for nuclear localization and that each signal could function independently. However, localization of R-GUS or GUS-R constructs containing individual NLSs was less efficient than localization of constructs containing all three or two of the three signals.

The NLSs of R were dissimilar in their aa composition and may confer different specificities to the nuclear import machinery. NLS-A had the most intriguing composition because it contained Arg's and no Lys's. This is a characteristic of some viral NLSs. Examples of viral proteins with NLSs containing no lysines are influenza nucleoprotein and NS1 (Davey et al., 1985), adenovirus pTP (Zhao and Padmanabhan, 1988), and human immunodeficiency virus REV (Malim et al., 1989). NLS-C was enriched in hydrophobic aa's, which were interspersed within its basic residues. One of the few NLSs that has a high content of hydrophobic aa's is the yeast MAT α2 protein (KIPIK; Hall et al., 1984), which is similar to NLS-C (MISESLRKAIGKR).

NLS-M, located within the amino terminus of the helix-loop-helix homologous motif, contained more basic aas than NLSs A or C, with 5 Arg's and 1 Lys within the 10-aa signal. The high concentration of basic aa's in NLS-M is similar to the SV40 large T antigen NLS (Kalderon et al., 1984a), in which 5 of the 7 aa's in the signal are basic. Another transcription factor, myoD1, which shares homology with the helix-loop-helix domain, also contains an NLS in this motif. However, the NLS was defined to 34 aa's of the helix-loop-helix, and it is not known if the NLS of myoD1 is in the amino terminus (the first 10 aa's of the 34-aa signal identified) of the helix-loop-helix domain (Tapscott et al., 1988). A comparison of NLS-M to the amino terminus of other DNA-binding helix-loop-helix domains revealed a conserved region (Fig. 6). It is logical in evolutionary terms to retain an NLS within an essential domain of a transcriptional activator, and it would be interesting to determine whether or not the import and DNA-binding functions are separable.

Two NLSs were required for efficient transport of the R:GUS fusion proteins. Combinations of NLS-A and NLS-M or NLS-M and NLS-C conferred exclusive nuclear localization to the fusion proteins. This requirement of two NLSs for efficient transport to the nucleus is known to occur in other nuclear proteins and was proposed to be a consensus structure, termed bipartite, by Dingwall and Laskey (1991). Bipartite signals contain two regions enriched in basic aa's separated by more than four aa's (Robbins et al., 1991).

The NLSs of the R protein are bipartite, but they do not fit the model proposed by Dingwall and Laskey (1991). First, unlike the model signal, in which both basic regions are required for efficient targeting of a reporter protein to the nucleus, two NLSs of the R protein, NLSs M and C, independently and efficiently redirected the reporter protein GUS to the nucleus. Second, although two NLSs are necessary for targeting of R-GUS protein to the nucleus, the spacing between NLSs A, M, and C (at least 170 aa's) is greater than the spacing found in the NLSs examined by Dingwall and Laskey (1991). Also, the potyviral protein NIa (Carrington et al., 1991) has a longer spacer (32 aa's) separating the two basic regions that are involved in nuclear localization. Although the significance of NLS repetition is not understood, this phenomenon has been reported in many proteins: glucocorticoid steroid hormone receptor (Picard and Yamamoto, 1987), *Agrobacterium* VirE2 (Citovsky et al., 1992), and *Zea mays* O2 (Varagona et al., 1992) are examples. One study examined the effect of multiple NLSs upon the import of peptide-coated gold particles (Dworetsky et al., 1988). Increasing amounts of SV40 large T antigen NLS were covalently linked to coat gold particles, which were microinjected into *Xenopus laevis* oocytes. The results showed that larger-diameter gold particles require several NLSs to enter the nucleus.

To determine which aa's of the NLSs might be important for function, we searched for conserved aa's in members of the *R* gene family (*Lc, R-s, B-peru*) and an *R* homolog from *Antirrhinum majus* (DEL, Fig. 6). Two of the alleles (*R-Lc, R-S*) of the *R* gene are cloned and they are 95% homologous in their aa's. Therefore, the regions corresponding to R NLSs are equally conserved (Fig. 6). However, the maize B (Radicella et al., 1991) and *Antirrhinum* DEL (Goodrich et al., 1991) proteins share 78 and 25% aa homology to R-Lc and encode sequences similar to the NLSs of R (*R-Lc* was used in this study). The greatest homology was retained for NLS-M, which was associated with the helix-loop-helix domain (Fig. 6). NLS-A was the least conserved and was the weakest of the signals that we have identified. Although NLS-C was not highly conserved (Fig. 6), the presence of two Lys's and the overall hydrophobic content of the carboxyl terminus is retained. Because the R:GUS fusion protein required two NLSs for exclusive nuclear localization, the conservation of NLS-M and NLS-C indicates that they may be the NLSs utilized in the R protein.

NLS-M represents a second function for the helix-loop-helix domain: to serve in both DNA binding and nuclear targeting. Because NLS-M is absolutely necessary for efficient targeting of R and is also the most conserved region among transcriptional activators carrying helix-loop-helix motifs, the dual function of the R protein's basic helix-loop-helix may be conserved in other transcriptional activators with helix-loop-helix domains. A similar hypothesis was proposed for the b-ZIP proteins (Raikhel, 1992; Varagona et al., 1992) and for steroid hormone receptors that contain zinc-finger motifs (Picard and Yamamoto, 1987).

Possible functions for the multiple NLSs of R could be to act as developmentally regulated or tissue-specific signals. Recently, a developmentally regulated NLS was identified in the adenovirus type 5 E1a protein (Standiford and Richter, 1992). Standiford and Richter (1992) identified the second of two NLSs in E1a, termed drNLS, which is not constitutively utilized as a signal for nuclear transport. It has been shown using developing *Xenopus* oocytes that the drNLS alone resulted in transport to the nucleus until oocytes reach the late gastrula stage, when the drNLS E1a protein is retained in the cytoplasm. The composition of the drNLS is unusual: it contains no basic aa's and is enriched in hydrophobic aa's. None of the R protein's NLSs share homology with the drNLS of E1a. However, the possibility exists that these multiple NLSs function at different developmental stages. Another possibility is that multiple NLSs are required in the R protein to regulate tissue-specific expression because different alleles of the *R* gene are expressed in different tissues (Styles et al., 1973; Coe, 1985). The *Lc* allele used in this study is expressed in a number of tissues, including pericarp, ligule, midribs, coleoptiles, anthers, silks, and brace roots; whereas another allele of *R*, *R-nj*, is expressed only in the scutellum, coleoptiles, and brace roots. One proposal is that tissue specificity is regulated by different promoters. However, it is possible that the NLSs of R may function differentially, with each NLS providing different efficiencies for transport in a tissue-specific manner.

The most striking feature of the different NLSs of the R protein was their varying compositions. NLS-A contained no Lys residues, a characteristic that has been observed only in viral proteins. NLS-M possessed the greatest density of charged residues, with 7 of the 10 aa's being basic. NLS-C was enriched with hydrophobic residues, which also affect the charge density of the NLS. Although it is not surprising that the compositions of the signals are different, since NLSs lack a consensus sequence, it is obvious that the import machinery has to recognize some general features of the NLSs. Therefore, signals that are as divergent in charge and hydrophobicity as those in the R protein could be useful in the identification of different NLS binding proteins.

ACKNOWLEDGMENTS

We would like to thank Drs. Glenn Hicks, Marguertie Varagona, and Susannah Gal for many helpful discussions and critical reading of the manuscript.

Received October 14, 1992; accepted November 16, 1992.

LITERATURE CITED

An G, Ebert PR, Mitra A, Ha SB (1988) Binary vectors. Plant Mol Biol Manual **A3:** 1–19

Beckmann H, Su LK, Kadesch T (1990) TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev **4:** 167–179

Cai M, Davis RW (1990) Yeast centromere binding protein CBF1, of the helix-loop-helix protein family, is required for chromosome stability and methionine prototrophy. Cell **61:** 437–446

Carrington J, Freed DD, Leinicke A (1991) Bipartite signal sequences mediates nuclear translocation of the plant potyviral NIa protein. Plant Cell **3:** 953–962

Citovsky V, Zupan J, Warnick D, Zambryski P (1992) Nuclear localization of *Agrobacterium* VirE2 protein in plant cells. Science **256:** 1802–1805

Coe EH Jr (1985) Phenotypes in corn: control of pathways by alleles, time and place. *In* M Freeling, ed, Plant Genetics, UCLA Symposia on Molecular and Cellular Biology, Vol 35. New York, Alan R Liss, pp 509–521

Davey J, Dimmock NJ, Colman A (1985) Identification of the sequence responsible for the nuclear accumulation of the influenza virus nucleoprotein in *Xenopus* oocytes. Cell **40:** 667–675

DePhino RA, Hatton KS, Tesfaye A, Kohl NE, Yancopoulos GD, Alt FW (1987) The human *myc* gene family: structure and activity of L-*myc* and an L-*myc* pseudogene. Genes Dev **1:** 1311–1326

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

Dingwall C, Laskey RA (1991) Nuclear targeting sequences—a consensus? Trends Biochem Sci **16:** 478–481

Dworetsky SI, Lanford RE, Feldherr CM (1988) The effect of variations in the number and sequence of targeting signals on nuclear uptake. J Cell Biol **107:** 1279–1287

Edmundson DG, Olson N (1989) A gene with homology to *myc* similarity of MyoD1 is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes Dev **3:** 628–640

Garcia-Bustos J, Heitman J, Hall MN (1991) Nuclear protein localization. Biochim Biophys Acta **1071:** 83–101

Goff SA, Klein TM, Roth BA, Fromm ME, Cone KC, Radicella JP, Chandler VL (1990) Transactivation of anthocyanin biosynthetic genes following transfer of *B* regulatory genes into maize tissue. EMBO J **9:** 2517–2522

Goodrich J, Carpenter R, Coen ES (1992) A common gene regulates pigmentation pattern in diverse plant species. Cell **68:** 955–964

Hall MN, Hereford L, Herskowitz I (1984) Targeting of *E. coli* β-galactosidase to the nucleus in yeast. Cell **36:** 1057–1065

Hu Y-F, Luscher B, Admon A, Mermod N, Tijan R (1990) Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. Genes Dev **4:** 1741–1752

Jefferson RA (1987) Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep **5:** 387–405

Kalderon D, Richardson WD, Markham AF, Smith AE (1984a) Sequence requirements for nuclear location of simian virus 40 large T antigen. Nature **311:** 33–38

Kalderon D, Roberts BL, Richardson WD, Smith AE (1984b) A short amino acid sequence able to specify nuclear location. Cell **39:** 499–509

Klein TM, Wolf ED, Wu R, Sanford JC (1987) High-velocity microprojectiles for delivering nucleic acids into living cells. Nature **327:** 70–73

Kohl NE, Legouy E, DePhino RA, Nisen PD, Smith RK, Gee CE, Alt FW (1986) Human N-*myc* is closely related in organization and nucleotide sequence to C-*myc*. Nature **319:** 73–77

Kunkel TA, Roberts JD, Zakour RA (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol **154:** 367–382

Lanford RE, Butel JS (1984) Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. Cell **37:** 801–813

Ludwig SR, Habera LF, Dellaporta SL, Wessler SR (1989) *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the *myc*-homology region. Proc Natl Acad Sci USA **86:** 7092–7096

Ludwig SR, Wessler SR (1990) Maize *R* gene family tissue-specific helix-loop-helix proteins. Cell **62:** 849–851

Malim MH, Böhnlein S, Hauber J, Cullen BR (1989) Functional dissection of the HIV-rev trans-activator—derivation of a trans-dominant repressor of rev function. Cell **58:** 205–214

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio-assays with tobacco tissue culture. Physiol Plant **15:** 473–497

Newmeyer DD, Finlay DR, Forbes DJ (1986) *In vitro* transport of a fluorescent nuclear protein and exclusion of non-nuclear proteins. J Cell Biol **103:** 2091–2102

Nigg EA, Baeuerle PA, Luhrmann R (1991) Nuclear import-export: in search of signals and mechanisms. Cell **66:** 15–22

Paine PL, Moore LC, Horowitz SB (1975) Nuclear envelope permeability. Nature **254:** 109–114

Perrot GH, Cone KC (1989) Nucleotide sequence of the maize *R-S* gene. Nucleic Acid Res **17:** 8003

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

Radicella PJ, Turks D, Chandler VL (1991) Cloning and nucleotide sequence of a cDNA encoding *B-Peru*, a regulatory protein of the anthocyanin pathway of maize. Plant Mol Biol **17:** 127–130

Raikhel NV (1992) Nuclear targeting in plants. Plant Physiol **100:** 1627–1632

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

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning: A Laboratory Manual, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Standiford DM, Richter JD (1992) Analysis of a developmentally regulated nuclear localization signal in *Xenopus*. J Cell Biol **118:** 991–1002

Styles DE, Ceska O, Seah K (1973) Developmental differences in action of *R* and *B* alleles in maize. Can J Genet Cytol **15:** 59–72

Tapscott SJ, Davis RL, Thayer MJ, Cheng P, Weintraub H, Lassar AB (1988) *MyoD1*: a nuclear phosphoprotein requiring a *Myc* homology region to convert fibroblasts to myoblasts. Science **242:** 405–411

Varagona MJ, Schmidt RJ, Raikhel NV (1991) Monocot regulatory protein Opaque-2 is localized in the nucleus of maize endosperm and transformed tobacco plants. Plant Cell **3:** 105–113

Varagona MJ, Schmidt RJ, Raikhel NV (1992) Nuclear localization signal(s) required for nuclear targeting of the maize regulatory protein, Opaque-2. Plant Cell **4:** 1213–1227

Voronova A, Baltimore D (1990) Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. Proc Natl Acad Sci USA **87:** 4722–4726

Wagner P, Knuz J, Koller A, Hall MN (1990) Active transport of proteins into the nucleus. FEBS **275:** 1–5

Zhao L, Padmanabhan R (1988) Nuclear transport of adenovirus DNA polymerase is facilitated by interaction with preterminal protein. Cell **55:** 1005–1015