

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

**THE IDENTIFICATION AND CHARACTERIZATION
OF THE NUCLEAR LOCALIZATION SEQUENCES OF
THE MAIZE R PROTEIN**

By

Mark William Shieh

A DISSERTATION

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

DOCTOR OF PHILOSOPHY

Department of Botany and Plant Pathology

1994

ABSTRACT

THE IDENTIFICATION AND CHARACTERIZATION OF THE NUCLEAR LOCALIZATION SEQUENCES OF THE MAIZE R PROTEIN

By

Mark William Shieh

Previous genetic and structural evidence indicates that the maize R gene encodes a nuclear transcriptional activating factor. In-frame carboxy- and amino-terminal fusions of the R gene to the reporter gene β -glucuronidase (GUS) were sufficient to direct GUS to the nucleus of transiently transformed onion 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 (aa 100-109, GDRRAAPARP) contained several arginine residues; a similar localization signal is found in only a few viral proteins. The medial NLS M (aa 419-428, MSERKRREKL) is an SV40-type NLS and the carboxyl-terminal NLS C (aa 598-610,

MISEALRKAIGKR) is a Mat α 2 type. A deletion analysis of the three localization signals indicated that the amino-terminal and carboxyl terminal fusions of R and GUS were redirected to the nucleus only when NLSs A and M, or C and M, were both present. These results indicate that multiple localization signals are necessary for nuclear targeting of this protein.

NLS-C is similar to the Mat α 2-type NLS because it contains several hydrophobic residues and the basic amino acids are spaced equally apart. In addition, when the conserved region of the Mat α 2 NLS (KIPIK) is substituted into the KAIGK region in NLS-C, the hybrid NLS can still redirect GUS activity to the nucleus in onion epidermal cells. To identify the essential amino acids for NLS-C function, mutant NLS:GUS fusions were constructed. A mutant NLS with the hydrophobic amino acids in NLS-C substituted with non-polar hydrophilic residues partially redirected GUS activity to the nucleus. Mutated NLSs, with either the two lysines of NLS-C substituted with non-basic amino acids or the order of the amino acids in NLS-C reversed, resulted in GUS activity in the cytoplasm. Therefore, in NLS-C, the hydrophobic amino acids are important and the two lysines are necessary for its targeting function. In addition, reversing the

order of the amino acids of NLS-C negated its ability to function as a NLS, indicating a NLS requires more than the basic amino acid charge density to function.

DEDICATION

**For my mother Rosemary
and in memory of my father William**

ACKNOWLEDGEMENTS

I would like to thank Phil McCrea, my high school advisor and Biology club sponsor who encouraged my interest in science. I would also like to thank Natasha Raikhel for her guidance and support. Her lab has been an excellent environment to do research. I would like to thank the members of my guidance committee; Dr. Natasha Raikhel, Dr. Pamela Green, Dr. Ronald Patterson and Dr. John Wang as well as Tracey Reynolds for their critical review of this manuscript and the helpful discussions during my dissertation.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
ABBREVIATION	xii
CHAPTER 1	1
INTRODUCTION	1
REFERENCES	28
CHAPTER 2 NUCLEAR TARGETING OF THE MAIZE R PROTEIN REQUIRES TWO NUCLEAR LOCALIZATION SEQUENCES	36
ABSTRACT	37
INTRODUCTION	39
MATERIALS AND METHODS	42
RESULTS	48
DISCUSSION	63
ACKNOWLEDGEMENTS	74
REFERENCE	75

CHAPTER 3	CHARACTERIZATION OF THE CARBOXY-TERMINAL NUCLEAR LOCALIZATION SEQUENCE OF THE MAIZE R PROTEIN	80
	INTRODUCTION	81
	MATERIALS AND METHODS	84
	RESULTS	87
	DISCUSSION	96
	REFERENCES	103
CHAPTER 4	FUTURE RESEARCH PROSPECTIVES	105
	REFERENCES	110

LIST OF TABLES

Table 1.1 Summary of plant nuclear localization sequences identified and the known function of their proteins. 25

Table 3.1 Summary of the histochemical analysis for the mutated NLS-C:GUS fusion protein. 92

LIST OF FIGURES

Figure 1.1 Nuclear import of proteins in <i>Xenopus</i> oocytes is a two step process, involving binding and translocation across the nuclear envelope.	6
Figure 2.1. Histochemical localization of R:GUS (A) and GUS:R (B) fusion proteins in onion epidermal cells.	50
Figure 2.2. Cloning strategy for preparing R:GUS (A) and GUS:R (B) fusions and results of localization experiments.	52
Figure 2.3. 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.	54
Figure 2.4. 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 M and C are required for nuclear targeting.	58
Figure 2.5. Summary of histochemical analysis of R:GUS fusion proteins which identified NLSs that were necessary for nuclear localization.	61
Figure 2.6. Amino acid comparison of R-Lc to other homologous regulatory proteins.	67
Figure 3.1. Amino acid sequences of the mutated NLS-C polypeptides.	89

Figure 3.2. Histochemical localization of GUS activity of each of the mutated NLS-C GUS fusion constructs. 90

ABBREVIATIONS

a.a.	amino acid
ATP	adenosine triphosphate
β-gal	β-galactosidase
C	Celsius
CaMV	Cauliflower mosaic virus
cDNA	complementary deoxyribonucleic acids
DAPI	4',6' diamidino-2-phenylindole dihydrochloride
DNA	deoxyribonucleic acids
DEL	delila
FGF	fibroblast growth factor
G1	growth phase 1
GTP	guanosine triphosphate
GUS	β-glucuronidase
HBV	hepatitis virus
HSP 90	heat shock protein 90
I kappa β2	Inhibitor kappaβ
kDa	kilodalton
Matα2	mating factor α2
MS	Murashige and Skoog
NEM	N-ethylmaleimide
NF kappaβ	nuclear factor kappa β2
NLS	nuclear localization sequence
ηm	nanometer
NPC	nuclear pore complex
NOS	nopaline synthetase
nt	nucleotide
NUP	nuclear pore protein
P10	10 kDa factor
PRnls	progesterone receptor nuclear localization sequence

RNA	ribonucleic acids
RNase	ribonuclease
S	synthesis (growth phase)
SMHC-29	myosin heavy chain from rabbit
ssDNA	single stranded deoxyribonucleic acids
SV40	Simian virus 40
SW15	switch 15
μg	microgram
μm	micrometer
WGA	wheat germ agglutinin

CHAPTER 1

INTRODUCTION

INTRODUCTION

The increase in genome size from prokaryotes to eukaryotes has led to the development of the nucleus. The nuclear compartment organizes the genomic DNA and effectively separates the processes of transcription and translation. RNA must be exported to the cytoplasm to be translated and proteins such as transcription factors require import to the nucleus to function. Therefore, regulation of transport across the nuclear envelope serves as an additional level of control over vital cellular processes.

The nuclear envelope consists of two lipid bilayer membranes. The outer lipid bilayer is contiguous with the membrane of the endoplasmic reticulum. On the nucleoplasmic side of the inner membrane, there is a lattice of lamin intermediate filaments which provide rigidity and shape to the nucleus. Spanning the inner and outer membrane are large protein complexes termed nuclear pore complexes (NPC). NPCs are approximately 120 nm in diameter and are a round structure formed by eight protein complexes located on both the nucleoplasmic and cytoplasmic sides of the envelope, with a central protein complex in the center (Rout and Wente, 1994).

Extending from the pore complex are fibrial networks on the inside of the nucleus. This interlaced pattern of the fibrials has led them to being coined "fish baskets" (Jarnik and Aeby, 1991; Goldberg and Allen, 1992). The nuclear pore complexes form aqueous channels, 9 nm in diameter, which connect the nucleoplasm to the cytoplasm. Through the aqueous channels, small molecules are able to freely diffuse (dextrans up to 50 kDa; for review, Forbes, 1992; Paine et. al., 1975) and proteins and RNAs are transported.

Nuclear pores complexes were first identified as the site of protein transport by using gold particles coated with nuclear targeting signals which accumulated at the pore complex (for review, Rout and Wentz, 1994). Later studies demonstrated that both protein import and RNA export occur through the same pores, because gold particles coated with either RNA or NLS localized to a single pore (Dworetzky and Feldherr, 1988a). Addition of wheat germ agglutinin (WGA), a lectin with specificity for N-acetylglucosamine, to *in vitro* protein transport systems in animal cells blocked transport (Finlay et. al., 1987; for review, Forbes, 1992). Immunocytochemical analysis revealed that WGA was bound to the pore complexes, indicating that proteins of the pore complex are

modified in the cytoplasm with N-acetylglucosamine. This carbohydrate modification was used to purify and isolate a number of proteins which reside at the pore complex. Recently, N-acetylglucosamine modified proteins have been discovered on nuclear envelope proteins in plants, most are localized at the pore complex. In addition, carbohydrate analysis of these plant proteins reveals that, unlike the single O-linked carbohydrate N-acetylglucosamine modification in animal cells, plants can have an oligosaccharide at a single O-linked site (A. Heese-Peck, Cole, Hart and N. Raikhel, unpublished).

The function of proteins located at the pore complex is difficult to ascertain, although functions for some have been proposed because they contain regions with homology to nucleotide binding domains. The nuclear pore proteins, NUP 149, NUP 100, and NUP 116 contain domains homologous to RNA binding motifs (Fabre et. al., 1994) and these proteins may be involved in the export of RNA to the cytoplasm. The NUP 149 protein binds poly-guanosine (ssDNA) and a temperature-sensitive mutant of NUP 149 accumulates polyadenylated messages in the nucleus (Fabre et. al., 1994). Though the specificity of nucleotide binding for NUP 149 *in*

vitro does not match the phenotype of the mutant, the NUP 149 mutant does interfere with RNA export. The pore protein NUP 153 contains a zinc finger motif and may function as an anchor for chromatin (Sukegawa and Blobel, 1993). Our understanding of the nuclear pore is increasing rapidly and utilization of techniques such as synthetic lethality are expanding our knowledge of the interactions which occur between nuclear pore proteins and the factors with which they associate (Fabre et. al., 1994).

Molecular Mechanism of Nuclear Transport

In general, the import of proteins into the nucleus is both energy and signal dependent, however, small macromolecules do diffuse through the pore complexes (for review see Forbes, 1992; Paine et. al., 1975). The molecular mechanism of nuclear transport can be separated into two steps in *Xenopus* oocytes (Fig. 1.1). First, the protein binds to the nuclear envelope; this is mediated by its nuclear localization signal (NLS) and cytosolic factors. Then the protein is translocated across the envelope. The transported protein binds to the nuclear envelope with its NLS and this requires

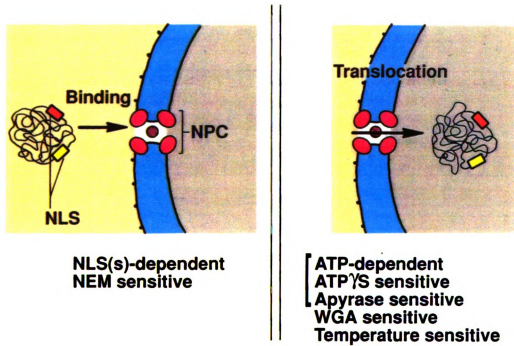


Figure 1.1 Nuclear import of proteins in *Xenopus* oocytes is a two step process, involving binding and translocation across the nuclear envelope.

cytosolic factors (Fig. 1.1; Richardson et. al., 1988, Newmeyer and Forbes, 1988b). A NLS-containing protein will not bind to the nucleus if the cytosol is treated with the sulfhydryl modifying agent N-ethylmaleimide (NEM) (Fig. 1.1; Newmeyer and Forbes, 1990). The second step, translocation, is an energy dependent process requiring hydrolysis of ATP and, possibly, GTP (Fig. 1.1). The requirement for ATP hydrolysis has been shown by blocking transport with the non-hydrolyzable analog ATP γ S and because transport is inhibited when endogenous ATP is depleted with apyrase, in *in vitro* reconstituted and semi-*in vivo* (microinjection) transport systems (Newmeyer and Forbes, 1988; Richardson et. al., 1988). The requirement for GTP is implicated from studies analyzing the cytosolic proteins necessary for transport. Blobel and coworkers separated the cytosol into two fractions by ion-exchange chromatography; one fraction was necessary for binding and the other for translocation (Moore and Blobel, 1992). The cytosolic protein in the translocation fraction was isolated and identified as *Ran/TC4*, a GTP-binding protein with roles in DNA replication, cell cycle checkpoint control, RNA synthesis, processing and export (Moore and Blobel, 1993; Melchior et. al., 1993). Prior to these

studies, the role of GTP in nuclear transport had not been thoroughly studied in *Xenopus* oocytes and no additional GTP is required to reconstitute transport *in vitro*. The non-hydrolyzable analog GTP γ S inhibits but does not block nuclear import *in vitro*, therefore, the function of GTP in nuclear transport is not clear (Melchior et. al., 1993). Recently, a 10 kDa factor (p10) in the translocation fraction of the cytosol was identified. The p10 protein, with the *Ran/TC4* protein, functions as well as the completed translocation fraction of cytosol when combined with the binding fraction in *in vitro* transport systems (Moore and Blobel, 1994). Translocation across the envelope is also inhibited at lower temperatures (4°C). Microinjection of cells with nucleoplasm protein at 4°C results in no nuclear import; if the temperature is raised to 37°C, transport resumes (Richardson et. al., 1988). In addition, microinjected WGA, which binds to nuclear pore proteins modified with N-acetylglucosamines, blocks translocation of proteins and export of RNA across the envelope (Fig. 1.1; Newmeyer and Forbes, 1988b; for review Rout and Wentz, 1994).

Nuclear transport in systems other than the well studied *Xenopus* oocytes differ in their behavior. As an example, in somatic

animal cells, there is no requirement for cytosol and nuclear transport can be reconstituted *in vitro* with purified nuclei alone (Markland et. al., 1987; Dean and Kasamatsu, 1994).

Nuclear Localization Sequences

NLSs are short polypeptides that are the sole determinant for protein import to the nucleus. They are identified by their ability to direct a reporter protein, normally localized in the cytoplasm, to the nucleoplasm. The properties of the reporter proteins are that they must be a soluble cytoplasmic-localized protein and larger than the predicted exclusion limit for diffusion through the nuclear pore (40-50 kDa; Garcia-Bustos, 1991). The reporter proteins used for yeast and animal cells are human (bovine or chicken) serum albumin (Dworetsky et. al., 1988), pyruvate kinase (Dingwall et. al., 1988; Kalderon et. al., 1984) and β -galactosidase (Hall et. al., 1990; Picard and Yamamoto, 1987). β -glucuronidase (GUS) is a common reporter protein in plants (Shieh et. al., 1993; Raikhel, 1992).

In general, the ability of a polypeptide to function as a NLS has been assessed by two methods. First, a gene fusion of the

putative NLS to a reporter cDNA is stably or transiently expressed in the cells of interest. Localization of the reporter protein is determined by immunocytochemistry or histochemical (GUS and β -gal) assays (Opaque 2; Varagona et. al., 1992, R protein Shieh et. al., 1993). Second, putative NLSs are fused to a reporter protein as either synthetic peptides chemically crosslinked to a reporter protein, or gene fusions are *in vitro* translated proteins or *in vitro* transcribed RNA (SV40-NLS + pyruvate kinase DNA; Kalderon et. al., 1984b) and are microinjected into cells. Localization, in general, is determined by immunofluorescence as the reporter proteins are generally labeled with fluorescent compounds.

Utilizing these methods, over a hundred NLSs have been identified from plant, animal and yeast systems (for review see Boulikas, 1993; Raikhel, 1992). The NLSs vary in amino acid composition and length with no consensus sequence or structure and therefore they must be identified empirically. Unlike targeting signals to other organelles, the NLSs are found throughout a protein and multiple targeting signals are frequently found in a single protein (for review, Garcia-Bustos et. al., 1991). All NLSs contain the basic amino acid(s) lysine and/or arginine (Garcia-Bustos et. al.,

1991; Boulikas, 1993). Based on amino acid composition and size, NLSs can be separated into three groups: SV40 large T antigen-like (Kalderon et. al., 1984a,b; Lanford and Butel, 1984), bipartite (nucleoplasmin; Dingwall and Laskey, 1991) and Mat α 2-like (Hall et. al., 1984). SV40-like signals are 7-20 amino acids (a.a.) in length and enriched in basic amino acids, whereas bipartite signals contain two basic amino acid enriched regions separated by 10-30 amino acids. The NLS of SV40 large T antigen has been shown to function in animal, yeast and plant cells (Lanford and Butel, 1984; Nelson and Silver, 1989; Varagona and Raikhel, 1994; van der Krol and Chua, 1991). Mat α 2-like (mating factor α 2) NLSs contain several hydrophobic residues and a single basic amino acid enriched region. The polypeptide motif KIPK in the amino terminal Mat α 2 NLS is conserved in several other yeast nuclear proteins, and is considered to be a Mat α 2-like type of NLS. In addition, unlike other NLSs which can function as targeting signals in animal, plant and fungal systems, the yeast Mat α 2 NLS does not function as a targeting signal in animal cell lines (Chelsky et. al., 1989; Lanford et. al., 1990).

Many of the NLSs contain a region enriched in basic amino acids such as the SV40 large T antigen where five out of seven

residues are basic (Kalderon et. al., 1984a and for review see Boulikas, 1993). However, a highly basic amino acid stretch does not function as a NLS since a number of cytoplasmic proteins such as SMHC-29 (rabbit myosin heavy chain; Nagai et. al., 1988), 2-5A dependent RNase (Murine; Zhou et. al., 1993) and porcine p11 (calcium binding protein; Mastakowski and Shooter, 1988) contain basic a.a. enriched regions. Boulikas (1993) proposed that a polypeptide containing at least four basic amino acids in a hexapeptide would function as a NLS. This hypothesis was tested by analyzing 117 transcription factors and 109 non-nuclear proteins for the basic rich regions. The hypothesis was false as a number of non-nuclear proteins contained regions with four of six basic amino acids (Boulikas, 1994). The author proposed that the basic rich regions may not function as a NLS because many of the proteins were sequestered from the nuclear transport machinery in an organelle or by association with a membrane. Also, the soluble non-nuclear cytoplasmic proteins were thought to limit the exposure of the basic rich regions on the surface of the protein therefore, hiding the region from the nuclear transport machinery.

Studies with the SV40 large T antigen NLS and androgen

steroid hormone receptor indicate that the location of the NLS within a protein has a strong influence on its function. To determine if the SV40 NLS can function in different locations within a protein, transcribed message from gene fusion constructs encoding the SV40 NLS in different locations in the pyruvate kinase cDNA were expressed in animal cells and their subcellular localization determined. The NLS functions in four of the five regions tested within pyruvate kinase and if inserted into amino acid position 125-136, the SV40 NLS:pyruvate kinase remains cytoplasmic (Roberts et. al., 1987). Deletion analysis of the androgen receptor indicated that its bipartite NLS will not function when one of the two DNA binding domains is deleted (Zhou et. al., 1994). Therefore, the context in which a NLS is presented is crucial to its function.

Recently, it has been proposed that NLSs may also function in the export of proteins. This is based upon the observation that, cells microinjected with PRnls- β -gal protein (progesterone receptor NLS fused to β -galactosidase) will import the fusion protein into nucleus. However, upon energy depletion, induced by low concentrations of sodium azide, the PRnls- β -gal protein accumulated in the cytoplasm. When the same experiment is performed with β -

gal microinjected into the nucleus, the β -gal is not exported to the cytoplasm. The effect of the energy depletion is reversible and results in the nuclear import and nuclear localization of the export PRnls- β -gal protein (Guichon-Mantel et. al., 1994). These experiments indicate that retention of proteins in the nucleus may be an active process or that nuclear proteins may shuttle between the two compartments in a NLS-dependent manner.

Dual Localization in the Nucleus and Secretory Pathway

Several proteins contain two different targeting signals, one is a signal peptide and the other a NLS. These proteins are found in the secretory pathway (Golgi and extracellular space) and also in the nucleus. The FGF3 (mouse fibroblast growth factor) protein is an example of a protein with two targeting signals. When translation is initiated from the ATG codon, the FGF3 protein is secreted from the cell (Keifer et. al., 1994). However, it was determined that translation of the FGF3 protein is normally initiated from an upstream CUG codon and this additional 29 a.a. results in FGF3 protein localized to the nucleus and extracellular space (via the

secretory pathway). The NLS in FGF3 is not located in the additional amino acids but is contained within the amino-terminus of the mature protein (Keifer et. al., 1994). Therefore, this may represent a mechanism which regulates the subcellular localization of FGF3 by the selection of different initiation sites. Another protein which is dually-localized but not regulated in its localization, is the hepatitis virus (HBV) precore P22 protein. P22 contains a 19 amino acid secretory signal sequence and a NLS. The p22 is processed to its mature form by a signal peptidase and localizes to both the Golgi (30%) and nucleus (70%, Ou et. al., 1989). The authors propose that after cleavage of the signal sequence from P22 that some of the protein is released back to the cytoplasm where it is imported to the nucleus. However, the mature protein may have been secreted and then reintroduced to the cells as part of a viral particle. Then without a secretory signal the NLS on the mature P22 protein targets it for import to the nucleus. NLSs can also reside in normally secreted proteins without being utilized. The secreted protein of simian sarcoma virus *V-sis* contains both targeting signals but unlike FGF3 and P22, the *V-sis* protein is secreted. If the secretory signal sequence is destroyed, the *V-sis* protein is then

imported to the nucleus (Lee et. al., 1987). These findings raise a number of questions concerning the evolution of different targeting signals and how selection of NLSs verses other signals occurs. Is there an advantage to a protein containing different targeting signals as opposed to having a gene family with a single type of signal? Alternatively, are these examples of genes which, by selection, will develop into gene families, each member with a single type of targeting signal?

Modification of NLSs and Regulation of Transport

Regulation of the cell cycle, development and transcriptional activation can occur by controlling the transport of proteins into the nucleus. Two mechanisms have been identified which control nuclear transport: modification of the protein to change the efficiency of the NLS and retention of proteins in the cytoplasm by a cytoplasmic-anchoring protein.

The efficiency of NLSs can be altered by the phosphorylation state of the protein. The cell cycle is regulated by a number of

protein phosphorylation/dephosphorylation events that can change the rate or time of nuclear transport. The SW15 protein of *Saccharomyces cerevisiae* is a transcription factor which regulates the HO gene, the protein involved in mating type switches. SW15 message is synthesized in the S phase of growth and transcription stops before entering G1. The protein product accumulates in the cytoplasm until the cell enters G1, then SW15 is transported into the nucleus (Nasmyth et. al., 1990). Near the NLS are three CDC28 kinase sites that are phosphorylated when the cell enters the S phase (Moll et. al., 1991). Mutation of the three phosphorylated serine residues to non-phosphorylatable alanines results in nuclear localization of the SW15 protein throughout the cell cycle. This indicates that the unphosphorylated form of SW15 is nuclear transport competent (Moll et. al., 1991). Therefore, nuclear import of SW15 is prevented by phosphorylation and stimulated by the unphosphorylated form of SW15.

In another protein, phosphorylation can increase the rate of import to the nucleus. Rihs and coworkers (1991) identified two putative casein kinase II sites within 15 amino acids of the SV40 large T antigen NLS. Substitution of either of the casein kinase II

site, with non-phosphorylatable amino acids, decreased the rate of nuclear accumulation. This was done by microinjection of the modified SV40 NLS β -gal fusion protein into cells (Rihs et. al., 1991). Therefore, the phosphorylated form of SV40 large T antigen NLS is imported at a much greater rate (15 fold) than the non-phosphorylatable mutant; this increase in transport rate may be due to conformational changes in the phosphorylated NLS (Rihs et. al., 1991).

Phosphorylation of proteins also regulates protein-protein interactions of the *rel* (proto)-oncogene family, NF kappaB (nuclear factor) and dorsal. NF kappa B is a transcription factor for the synthesis of immunoglobulin Kappa-light chain. The dorsal protein is a putative transcription factor that is essential for the *Drosophila* embryo to develop a dorsal/ventral axis. Dorsal creates this developmental polarity by establishing a gradient of nuclear accumulation where the ventral pole has the highest level of dorsal in the nucleus (Govind and Steward, 1991). Both NF kappaB and dorsal are retained in the cytoplasm until, through a signal transduction pathway from the plasma membrane, they are phosphorylated and enter the nucleus. The proposed mechanism is

that the *rel* proteins are anchored in the cytoplasm by interaction with another protein. It is known by genetic and *in vitro* binding assays that the dorsal protein binds to a protein called cactus. Cells with a mutant cactus protein have weakly ventralized embryos with dorsal protein in more nuclei of the dorsal pole (Isoda et. al., 1991). Cactus protein resides in the cytoplasm and will not bind phosphorylated dorsal, (Whalen and Steward, 1993) suggesting that the cactus/dorsal complex is nuclear import incompetent and that cactus, not phosphorylation of the dorsal NLS, is retaining dorsal in the cytoplasm. A similar interaction is found between NF kappaB and I kappaB; the binding affinity between these two proteins is also disrupted by phosphorylation of NF kappaB (Beg et. al., 1993). Sequence analysis of cactus and I kappaB showed that both proteins contain ankryin-like repeats which are known to be involved in protein-protein interaction. Therefore, the ankyrin-repeat is thought to bind NF kappa B and dorsal or anchor the heterodimer complex to the cytoskeletal proteins spectrin or ankyrin (Blank et. al., 1992). The mechanism by which cactus and I kappaB block the import of the *rel* proteins is not understood, but it is hypothesized that they bind and mask the region containing the NLS in NF

kappaB and dorsal (Beg et. al., 1993). This hypothesis is supported by studies which show that a protein which does not contain a NLS can enter the nucleus when carried by another nuclear protein and no cactus or I kappaB is found in the nucleus (Kang et. al., 1994). An alternative hypothesis is that cactus and IkappaB are anchored to the cytoskeleton and that the NLSs of dorsal or NF kappaB though recognized by the transport machinery cannot facilitate the nuclear import of the protein complex.

Another example of cytoplasmic retention is with members of the steroid hormone receptors such as the glucocorticoid, progesterone and estrogen receptors. In the absence of a steroid hormone, they reside in the cytoplasm and addition of hormone results in their import to the nucleus and transcriptional activation. The receptors are known to bind to HSP90 (heat shock protein) in the absence of hormone and it is believed that HSP90 retains the hormone receptors in the cytoplasm. Once the receptor binds the hormone, it is then imported into the nucleus. Recently, it was shown that the glucocorticoid and progesterone receptors bind to HSP90 *in vivo*; when the two NLSs are deleted from the hormone receptor and a NLS added to HSP90, the two proteins are localized

in the nucleus (Kang et. al., 1994). This indicates that HSP90 and the hormone receptor are bound together in the cytoplasm. Since HSP90 with a NLS is able to redirect the cytoplasmic form of the hormone receptor to the nucleus, the HSP90 protein bound to the hormone receptor must block the two NLSs on the hormone receptor preventing the nuclear import of the heterodimeric protein.

The regulation of nuclear import adds an additional level of control for transcriptional activity and it is interesting to speculate why such a mechanism has developed. The advantage of retaining transcriptional proteins in the cytoplasm until needed suggests that a response to a stimuli should occur quickly and be faster than initiating transcription and translation of the transcription factor. Alternatively, the nuclear import process can be modulated, varying the amount of transcription induced as in the case of dorsal which is differentially imported along the axis of embryo development. Perhaps irreversible developmental changes in cells, such as the formation of the dorsal/ventral axis or initiation of immunoglobulin production, require a unique form of regulation where the transcription factors are produced at a single stage of cell development and reside in the cytoplasm until activated.

Nuclear Transport in Plants

The study of nuclear transport in plants has only recently begun and already differences are being discovered. The identification of oligosaccharide modifications on nuclear envelope proteins is unique to plants and N-acetylglucosamine residues have not yet been discovered on yeast nuclear pore proteins. To begin, the identification and characterization of the molecular mechanism in plants, NLSs of plant proteins were identified. There have been a number of reviews covering the NLSs of animal and yeast proteins (Boulikas, 1993, Garcia-Bustos, 1991) but they do not include plant NLSs, therefore a list of plants NLSs has been included in Table 1.1. Similar to the targeting signals identified in other eukaryotes, plant NLSs contain basic amino acids and they can be classified under the three types of NLSs (SV40-like, bipartite, and Mat α 2-like). However, plant NLSs have a stronger binding affinity for plant nuclear envelope proteins (putative receptors) than mammalian NLSs. When the binding affinity of NLSs for NLS-binding proteins on isolated nuclei from tobacco or maize suspension cells the plant Opaque 2 NLS had a stronger binding affinity and was a better

competitor than the SV40 large T antigen NLS (Hick and Raikhel, 1993). The nuclear NLS-binding affinity for peptides of the Opaque 2 bipartite NLS is 200 μ M and the binding activity is proteinaceous (Hicks and Raikhel, 1993).

When I began my thesis research virtually nothing was known about nuclear transport in plants. Therefore, to study the molecular mechanism of nuclear transport in plants, we chose to identify the nuclear localization signals in the maize R protein; a proposed regulator of anthocyanin biosynthesis. Anthocyanins are pigments (red or purple) expressed in a tissue-specific manner in maize. Since anthocyanins are an easily detectable phenotype, the genetics of its biosynthetic pathway has been thoroughly studied and R expression was found to be modulated by a number of genes and it may be regulated at the level of import to the nucleus.

To identify the NLS in the R protein, gene fusion constructs with the R cDNA or short segments of the R cDNA were fused to the reporter gene GUS. These gene fusions were transiently expressed in onion epidermal cells and the subcellular location of the GUS activity was determined with a histochemical substrate using Nomarski optics. Three NLSs were identified in the R protein.

Because there are multiple NLSs, those necessary for exclusive nuclear localization of the full length R cDNA fused to GUS were determined. In addition, the carboxy-terminal NLS (NLS-C) of R has an unusual amino acid composition similar to the previously uncharacterized yeast Mat α 2-like NLS. Therefore, NLS-C was further analyzed in detail to determine which of its 13 a.a. are essential for its function.

Table 1.1 Summary of plant nuclear localization sequences identified and the known function of their proteins. The bold letters denote basic amino acids.

Table 1.1

Protein and Function	Nuclear Localization Sequence	Type of NLS	Reference
N1a amino-terminus - Tobacco etch virus, catalyzes proteolysis of the viral polyprotein in five places and is a genomic VPg	GKKNQKHKLK M	SV40-Like	Carrington et. al., 1991
N1a central signal	KRKGTTTRGMG AKSRKFINMYG FDPTDFSXI	Bipartite	Carrington et. al., 1991
Opaque 2 NLS-A - Maize transcriptional regulator of α zein biosynthesis	MEEAVTMAPA AVSSAVVGDP EYNAILRRKLE EDLE	SV40-like	Varagona et. al., 1992
Opaque 2 NLS-B	MPTEERVVRKR KESNRESARRS RYRKA AHLKE L	Bipartite	Varagona et.al, 1992
R NLS-A -Maize transcriptional regulator of anthocyanin biosynthesis	GDRRAAPARP	SV40-like	Shieh et. al., 1993
R NLS-M -As above	MSE RKRREKL	SV40-like	Shieh et. al., 1993
R NLS-C -As above	MISEALRKAIG KR	Matα2-like	Shieh et. al., 1993

Table 1.1 (cont'd)

Protein and Function	Nuclear Localization Sequence	Type of NLS	Reference
TGA-1A -transcription factor with homology to the CREB protein	MAKPVEVLRR LAQNREAARKS RLRKKAYVQQ LENSKLKLIQL EQELEQILERA RKQGMCVGGG VDASQLSYSGT ATRGSPGGQSL	Possible Bipartite	van der Krol and Chua, 1991
TGA-1B -transcription factor with homology to the CREB protein	MAEKKRARLV RNRESAQLSRQ RKKHV-----MS	Bipartite	van der Krol and Chua, 1991
VirD2 amino terminus -Agrobacterium; VirD2 binds at the 5' end of the T-DNA strand and may facilitate the nuclear import of the T-DNA	YISRKGKLEL	SV40-like	Tinland et. al.,1992
VirD2 carboxyl-terminus	KRPREDDDGEP SERKRER	Bipartite	Howard et. al.,1992; Tinland et. al., 1992
VirE2 -Agrobacterium ssDNA binding protein coats the intermediate T-DNA strand of Agrobacterium T-DNA and may facilitate the T-DNA transport to the nucleus	KLRPEDRYIQT EKYGRR-49a.a. spacer-- TKYGSDTEIKL KSK	Proposed Bipartite maybe two SV40-like NLSs	Citovsky et. al., 1992

REFERENCES

- Beg AA, Ruben SM, Scheinman RI, Haskill S, Rosen CA, Baldwin AS Jr. (1992)** I-kappaB interacts with the nuclear localization sequences of the subunits of NF-kappa B: a mechanism for cytoplasmic retention. *Genes and Dev.* 6: 1889-1913
- Blank V, Kourilsky P, Israel A (1992)** NF kappaB and related proteins: Rel /dorsal homologues meet ankyrin-like repeats. *T.I.B.S.* 17: 135-140
- Boulikas T (1993)** Nuclear localization signals. *CRC Crit. Rev. Euk. Gene Expr.* 3(3): 193-227
- Boulikas T (1994)** Putative nuclear localization signals (NLS) in protein transcription factors. *J. Cell. Biochem.* 55: 32-58
- Carrington JC, Freed DD, Leinicke AJ (1991)** Bipartite signal sequence mediates nuclear translocation of the plant potyviral Nla protein. *Plant Cell* 3: 953-962
- Chelsky D, Ralph R, Jonak G (1989)** Sequence requirements for synthetic peptide-mediated translocation to the nucleus. *Mol. Cell. Biol.* 9: 2487-2492
- Citovsky V, Zupan J, Warnick D, Zambryski P (1992)** Nuclear localization of *Agrobacterium* VirE2 protein in plant cells. *Science* 256: 1802-1805
- Dean DA, Kasamatsu H (1994)** Signal- and energy-dependent nuclear transport of SV40 Vp3 by isolated nuclei. Establishment of a filtration assay for nuclear transport. *J. Biol. Chem.* 269: 4910-4916
- Dingwall C, Laskey RA (1991)** Nuclear Targeting sequences- a consensus? *TIBS* 16: 478-481

- Dingwall C, Robbins J, Dilworth SM, Roberts B, Richardson WD (1988)** The nucleoplasm nuclear location sequence is larger and more complex than that of SV40 large T antigen. *J. Cell. Biol.* **107**: 841-849
- Dworetsky SI, Feldherr CM (1988a)** Translocation of RNA-coated gold particles through the nuclear pores of oocytes. *J. Cell. Biol.* **106**: 575-584
- Dworetsky SI, Lanford RE, Feldherr CM (1988b)** The effect of variations in the number and sequence of targeting signals on nuclear uptake. *J. Cell. Biol.* **107**: 1279-1287
- Fabre E, Boelens WC, Wimmer C, Mattaj IW, Hurt EC (1994)** Nup145p is required for nuclear export of mRNA and binds homopolymeric RNA *in vitro* via a novel conserved motif. *Cell* **78**: 275-289
- Finlay DR, Newmeyer DD, Price TM, Forbes DJ (1987)** Inhibition of *in vitro* nuclear transport by a lectin that binds to nuclear pores. *J. Cell. Biol.* **104**: 189-200
- Forbes D (1992)** Structure and function of the nuclear pore complex. *Ann. Rev. Cell Biol.* **8**: 495-527
- Garcia-Bustos J, Heitman J, Hall MN (1991)** Nuclear protein localization. *Biochim. Biophys. Acta.* **1071**: 83-101
- Goldberg MW, Allen TD (1992)** High resolution scanning electron microscopy of the nuclear envelope: demonstration of a new, regular, fibrous lattice attached to the baskets of the nucleoplasmic face of the nuclear pores. *J. Cell. Biol.* **119**: 1429-1440
- Govind S, Steward R (1991)** Dorsoventral pattern formation in *Drosophila*. *Trends In Gen.* **7**: 119-125

- Guiochon-Mantel A, Delabre K, Lescop P, Milgrom E (1994)**
Nuclear localization signals also mediate the outward movement of proteins from the nucleus. *Proc. Natl. Acad. Sci. USA* **91**: 7179-7183
- Hall MN, Craik C, Hiraoka Y (1990)** Homeodomain of yeast repressor $\alpha 2$ contains a nuclear localization signal. *Proc. Natl. Acad. Sci.* **87**: 6954-6958
- Hall MN, Hereford L, Herskowitz I (1984)** Targeting of *E.coli* β -galactosidase to the nucleus in yeast. *Cell* **36**: 1057-1065
- Hick GR, Raikhel NV (1993)** Specific binding of nuclear localization sequences to plant nuclei. *Plant Cell* **5**: 983-994
- Howard EA, Zupan JR, Citovsky V, Zambryski PC (1992)**
The VirD2 protein of *A. tumefaciens* contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells. *Cell* **68**: 109-118
- Isoda K, Roth S, Nusslein-Volhard C (1992)** The functional domains of the *Drosophila* morphogen *dorsal*: evidence from the analysis of mutants. *Genes and Dev.* **6**: 619-630
- Jarnik M, Aeby U (1991)** Toward a more complete 3-D structure of the nuclear pore complex. *J. Structural Biol.* **107**: 291-308
- 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

- Kang KI, Devin J, Cadepond F, Jibard N, Guiochon-Mantel A, Baulieu E, Catelli M (1994)** *In vitro* functional protein-protein interaction: nuclear targeted hsp90 shifts cytoplasmic steroid receptor mutants into the nucleus. *Proc. Natl. Acad. Sci.* **91**: 340-344
- Kiefer P, Acland P, Pappin D, Peters G, Dickson C (1994)** Competition between nuclear localization and secretory signals determines the subcellular fate of a single CUG-initiated form of FGF3. *EMBO J.* **13** (17): 4126-4136
- Lanford RE, Butel JS (1984)** Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. *Cell* **37**: 801-813
- Lanford RE, Feldherr CM, White RG, Dunham RG, Kanda P (1990)** Comparison of diverse transport signals in synthetic peptide-induced nuclear transport. *Exp. Cell Res.* **186**: 32-38
- Lee BA, Maher DW, Hannink M, Donoghue DJ (1987)** Identification of a signal for nuclear targeting in platelet-derived-growth-factor-related molecules. *Mol. Cell. Biol.* **7**: 3527-3537
- Markland W, Smith AE, Roberts BL (1987)** Signal-dependent translocation of simian virus 40 large-T antigen into rat liver nuclei in a cell-free system. *Mol. Cell. Biol.* **7**: 4255-4265
- Masiakowski T, Shooter EM (1988)** Nerve growth factor induces the genes for two proteins related to a family of calcium-binding proteins in PC12 cells. *Proc. Natl. Acad. Sci.* **85**: 1277-1281
- Melchior F, Paschal B, Evans J, Gerace L (1993)** Inhibition of nuclear protein import by nonhydrolyzable analogues of GTP and identification of the small GTPase Ran/TC4 as an essential transport factor. *J. Cell Biol.* **123**: 1649-1659

- Moll T, Tebb G, Surana U, Robitsch H, Nasmyth K (1991)** The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SW15. *Cell* **66**: 743-758
- Moore MS, Blobel G (1992)** The two steps of nuclear import, targeting to the nuclear envelope and translocation through the nuclear pore, require different cytosolic factors. *Cell* **69**: 939-950
- Moore MS, Blobel G (1993)** The GTP-binding protein Ran/TC4 is required for protein import into the nucleus. *Nature* **365** 661-663
- Moore MS, Blobel G (1994)** Purification of a Ran-interacting protein that is required for protein import into the nucleus. *Proc. Natl. Acad. Sci. USA* **91**: 10212-10216
- Nagai R, Larson DM, Persiasamy M (1988)** Characterization of a mammalian smooth muscle myosin heavy chain cDNA clone and its expression in various smooth muscle types. *Proc. Natl. Acad. Sci.* **85**: 1047-1051
- Nasmyth K, Adolf G, Lydall D, Seddon A (1990)** The identification of a second cell cycle control on the HO promoter in yeast: cell cycle regulation of SW15 nuclear entry. *Cell* **62**: 631-647
- Nelson M, Silver P (1989)** Context affects nuclear protein localization in *Saccharomyces cerevisiae*. *Molec. and Cell. Biol.* **9**: 384-389
- Newmeyer DD, Forbes DJ (1988)** Nuclear import can be separated into distinct steps *in vitro*: nuclear pore binding and translocation . *Cell* **52**: 641-653

- Newmeyer DD, Forbes DJ (1990)** An N-ethylmaleimide-sensitive cytosolic factor necessary for nuclear protein import: requirement in signal-mediated binding to the nuclear pore. *J. Cell. Biol.* **110**: 547-557
- Ou JH, Yeh CT, Yen TS (1989)** Transport of hepatitis B virus precore protein into the nucleus after cleavage of its signal peptide. *J. Virol.* **63**: 5238-5243
- Paine PL, Moore LC, Horowitz SB (1975)** Nuclear envelope permeability. *Nature* **254**: 109-114
- Picard D, Yamamoto KR (1987)** Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. *E.M.B.O. J.* **6**: 3333-3340
- Raikhel NV (1992)** Transport of proteins to the nucleus. *Plant Phys.* **100**: 1627-1632
- Restrepo-Hartwig MA, Carrington JC (1992)** Regulation of nuclear transport of a plant potyvirus protein by autoproteolysis. *J. Virol.* **66**: 5662-5666
- Richardson WD, Mills AD, Dilworth SM, Laskey RA, Dingwall C (1988)** Nuclear protein migration involves two steps: rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. *Cell* **52**: 655-664
- Rihs H, Jans DA, Fan H, Peters R (1991)** The rate of nuclear cytoplasmic protein transport is determined by the casein kinase II site flanking the nuclear localization sequence of the SV40 T-antigen. *EMBO J.* **10**: 663-661
- 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

- Roberts BL, Richardson WD, Smith AE (1987)** The effect of protein context on nuclear location signal function. *Cell* **50**: 465-475
- Rout MP, Wente SR (1994)** Pores for thought: nuclear pore complex proteins. *Trends in Cell Biol.* **4**: 357-365
- Shieh MW, Wessler SR, Raikhel NV (1993)** Nuclear targeting of the maize R protein requires two nuclear localization sequences. *Plant Physiol.* **101**: 353-361
- Sukegawa J, Blobel G (1993)** A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm. *Cell* **72**: 29-38
- Tinland B, Koukolikova-Nicola Z, Hall MN, Hohn B (1992)** The T-DNA linked VirD2 protein contains two distinct functional nuclear localization signals. *Proc. Natl. Acad. Sci. USA* **89**: 7442-7446
- van der Krol AR, Chua N-H (1991)** The basic domain of plant B-ZIP proteins facilitates import of a reporter protein into plant nuclei. *Plant Cell* **3**:667-675
- 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
- Whalen AM, Steward R (1993)** Dissociation of the dorsal-cactus complex and phosphorylation of the dorsal protein correlate with the nuclear localization of dorsal. *J. Cell. Biol.* **123**: 523-534
- Zhou A, Hassel BA, Silverman RH (1993)** Expression cloning of 2-5-A-dependent RNase: a uniquely regulated mediator of interferon action. *Cell* **72**: 753-765

Zhou Z, Sar M, Simental JA, Lane MV, Wilson EM (1994) A ligand-dependent bipartite nuclear targeting signal in the human androgen receptor. J. Biol. Chem. 269 (18): 13115-13123

CHAPTER 2

NUCLEAR TARGETING OF THE MAIZE R PROTEIN REQUIRES TWO NUCLEAR LOCALIZATION SEQUENCES

**Reference: Shieh M.W., Wessler S.R., Raikhel N.V. (1993) Plant
Physiol. 101: 353-361**

ABSTRACT

Previous genetic and structural evidence indicates that the maize **R** gene encodes a nuclear transcriptional activating factor. In-frame carboxy- and amino-terminal fusions of the **R** gene to the reporter gene **β -glucuronidase** (GUS) were sufficient to direct GUS to the nucleus of transiently transformed onion 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 (a.a. 100-109, **GDRRAAPARP**) contained several arginine residues, a similar localization signal is found in only a few viral proteins. The medial NLS M (a.a. 419-428, **MSERKRREKL**) is an SV40-type NLS, and the carboxyl-terminal NLS C (a.a. 598-610, **MISEALRKAIGKR**) is a MAT $\alpha 2$ type. NLSs M and C are independently sufficient to direct the GUS protein to the nucleus when fused at the amino-terminus of GUS, while 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

deletion analysis of the three localization signals indicated that the amino-terminal and carboxyl terminal fusions of R and GUS were redirected to the nucleus only when NLSs A and M, or C and M, were both 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 are also discussed.

INTRODUCTION

In eukaryotic cells, proteins can be targeted to a variety of subcellular compartments such as the endoplasmic reticulum, 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 review, see Nigg et. al., 1991; Wagner et. al., 1990). The nuclear pore complex forms a large aqueous 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 endoplasmic reticulum, mitochondrion, and chloroplast, the import of nuclear proteins is mediated by nuclear localization sequences (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 re-enter the nucleus after cell division.

There is no consensus sequence for NLSs, however they are characterized as short amino acid 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: the SV40 large T type antigen (Kalderon et. al., 1984a,b; Lanford and Butel, 1984), MAT α 2 (Hall et. al., 1984) and 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 R protein controls where and when the anthocyanin biosynthetic pathway is expressed in plant tissues (Ludwig et. al., 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 kDa 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, since 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 epidermal cells following introduction of the DNA by particle bombardment. Using this system, three 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, since they constitute the first reported instance where multiple NLSs are required for competent transport of a plant regulatory protein.

MATERIALS AND METHODS

Materials

The white onions were purchased locally, stored at 4°C in the dark and used within two weeks. Oligonucleotides were synthesized by the MSU Macromolecular Facility (#1-3, #9-13) or by CIBA-GEIGY Biotechnology (#4-8, 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 (Dupont, 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 (leaf color; Ludwig et. al., 1989).

R:GUS -A SacI restriction site was inserted before the stop codon (nt 1830) and a SmaI restriction site was inserted after the stop codon of the Lc cDNA by site-directed mutagenesis. The GUS cDNA (pBI101.3, Jefferson et. al., 1987) was then subcloned in front of the stop codon of the Lc cDNA.

GUS:R -Lc cDNA was modified to include a XhoI and SmaI site in frame before the first initiating AUG codon by site-directed mutagenesis. Also, a XhoI 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 BglII and SacI were used to construct the R:GUS gene fusions encoding a.a. 82-610 and

598-610 from the R:GUS construct.

GUS:R 1-109 -Restriction enzymes NaeI and ScaI 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 encoding a KpnI site followed by an ATG and XhoI site and were introduced into both R:GUS and GUS:R. By adding this set of restriction sites at positions before nts encoding a.a. 411 (nt 1231), 457 (nt 1368) and 512 (nt 1533) it was possible to subclone the fragments as a KpnI and EcoRI (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 were added. However these GUS:R constructs were subcloned as XbaI (5' of GUS) and KpnI fragments into a XbaI and KpnI site which had a stop codon in frame after the KpnI site.

When two sets of the restrictions sites (XhoI-ATG-KpnI) were added at nts encoding a.a. 411 and 457 (1231 and 1368) or 411 and 512 (nts 1231 and 1533), they allowed the isolation and cloning of a XhoI

(5') to KpnI (3') fragment. To subclone the R XhoI (5') to KpnI (3') fragment to GUS, an additional KpnI sites 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 a.a. 411-457 or 411-512 are inserted.

R:GUS 128-411 -Utilizing the construct R:GUS 1-411, a NaeI and KpnI fragment (encoding a.a. 128-411) was inserted into a SmaI and KpnI (KpnI site added by site-directed mutagenesis at nt 1831, the SmaI and KpnI cut drops out the R gene) cut R:GUS construct.

R:GUS 1-109 and 82-109 -R:GUS, with the additional KpnI site at nt 1831, was cut with NaeI and KpnI 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 restriction digested with BglII (leaving the first ATG at nt 246, a.a. 82) and EcoRI to and ligated into pUC118 into a BamHI and EcoRI site to construct R:GUS 82-109

To construct the R:GUS and GUS:R fusions which encoded a.a.

100-109, a KpnI site was introduced at nt 300 of R in the R:GUS 1-109 construct encoding nts 1-327. Then, the fragment encoding a.a. 100-109 was subcloned into pUC118 as a KpnI and EcoRI fragment. The constructs encoding a.a. 419-428 were constructed by adding a KpnI site after the codon for a.a. 428 (nt 1284) to the R:GUS 411-457 and GUS:R 411-457 constructs. The nucleotides encoding a.a. 429-457 were then excised from the clone as a KpnI fragment.

The deletion constructs outlined in Figure 2.4 were also constructed by 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 nts 300-327 for NLS-A (NaeI), nts 1257-1284 for NLS-M (AccI), or nts 1794-1830 for NLS-C (KpnI). 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 gm MS salts (Gibco-BRL,

Gaitherburg, MD), 1 mg thiamine, 10 mg myo-inositol, 180 mg KH_2PO_4 (Miller I), 30 gm sucrose, pH 5.7] (Murashige and Skoog, 1962) with the antifungal agent amphotericin B (2.5 mg/L; Sigma, St. Louis, MO) 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 /disc (3 times) of the suspension onto 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 PSI were optimal for onion cell transformation. Transformed cells were incubated at 28°C in the dark for 24 or 48h.

Histochemical Analysis

The colorimetric substrate X-gluc was used to determine the location of the enzymatic 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 three and the maximum was thirty.

RESULTS

The R Protein Redirects GUS to the Nucleus

To determine whether the R protein is imported into the nucleus, the R (Lc allele) and GUS cDNAs were ligated to form a gene fusion.

Since 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 CaMV 35S promoter and the NOS 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 the nucleus (Fig. 2.1A 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.

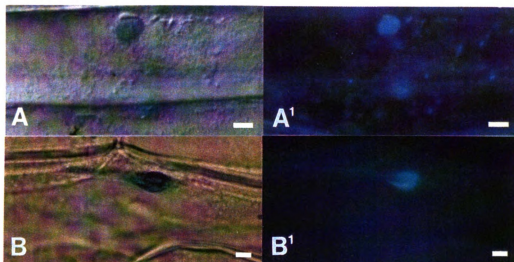


Figure 2.1. Histochemical localization of R:GUS (A) and GUS:R (B) fusion proteins in onion epidermal cells. Tissues were simultaneously analyzed using both X-gluc histochemical staining (A and B) and nuclei-specific DAPI staining (A' B'). Nomarski optics were used in A and B and fluorescence optics in A' and B'. Bars= 10 μ m.

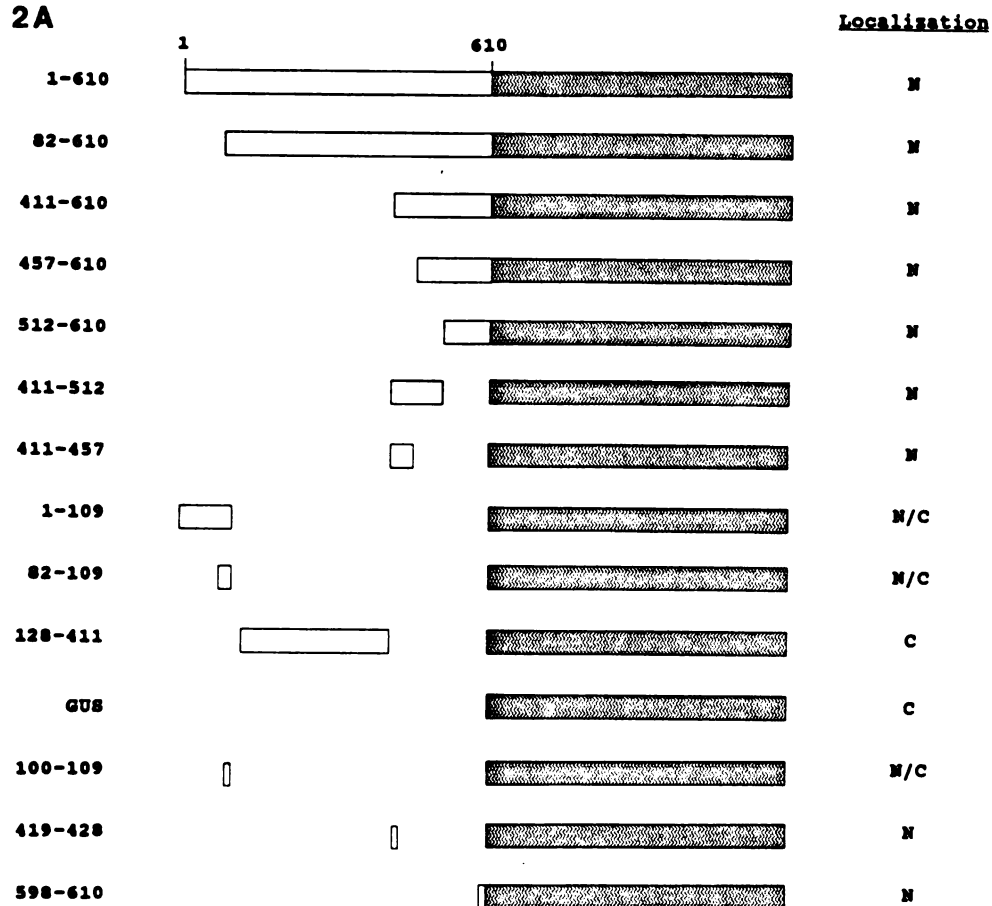
R Protein Contains Three NLSs

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. 2.2A,B). Thus, putative NLSs could be identified by the process of elimination. Initially, the **R:GUS** construct was modified with deletions at the 5' terminus (Fig. 2.2A), and the **GUS:R** construct was modified with 3' deletions (Fig. 2.2B). In addition, constructs were specifically designed around a.a. 411-457 because this region is enriched in basic amino acids, characteristic of NLSs, and contains the helix-loop-helix region (a.a. 420-462; Ludwig et. al., 1989). Upon completion, the constructs were ligated into expression vectors as described in Material and Methods.

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. 2.2A). The series of deletions from the amino-terminus contained a.a. 82-610, 411-610, 457-610, 512-610, and 598-610 and revealed NLS-C (NLS in the carboxyl-terminus). The thirteen amino acids encoded at position 598-610 (NLS-C) of **R** were sufficient to redirect GUS to the nucleus (Fig. 2.3). The localization of GUS by NLS-C was exclusively to the nucleus

Figure 2.2. Cloning strategy for preparing R:GUS (A) and GUS:R (B) fusions and results of localization experiments. The upper construct in Figure 2.2A 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 Figure 2.2B, the carboxyl-terminal fusion of GUS cDNA clone to the R cDNA clone. The position of first and last deduced amino acids in the R cDNA clone are indicated above the constructs in 2.2A and 2.2B. The amino acids of R protein used to prepare amino- (2.2A) and carboxyl- (2.2B) 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.

2A



2B

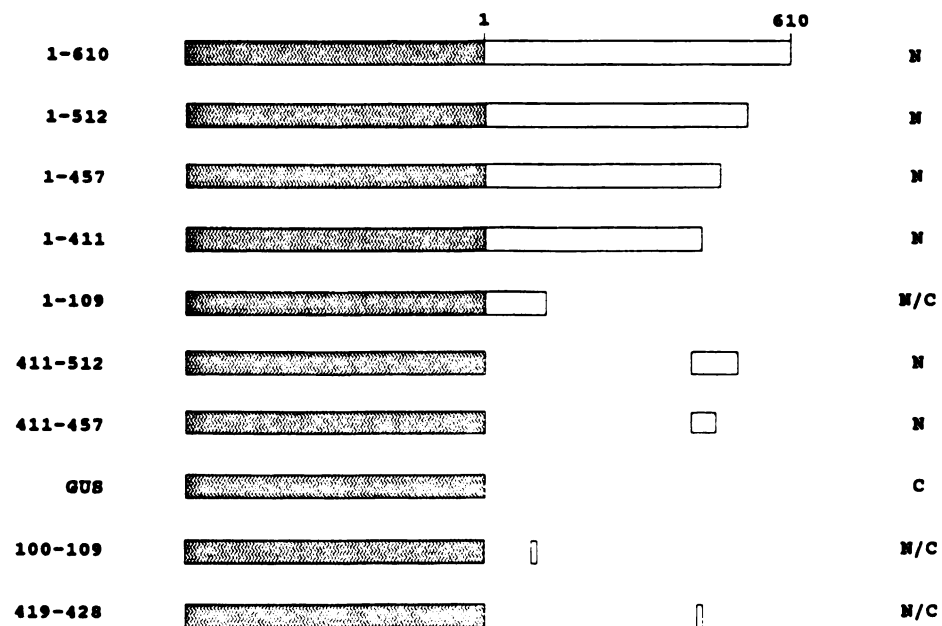
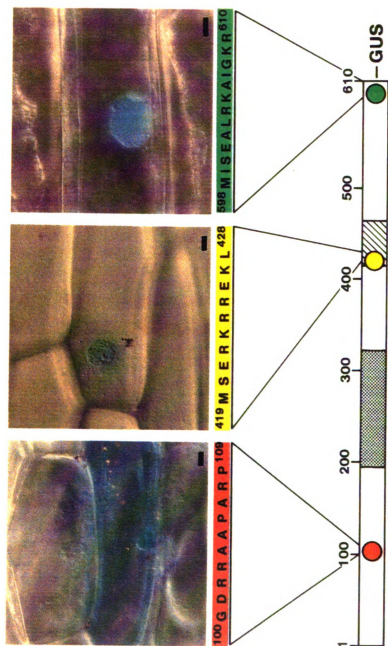


Figure 2.3. Histochemical localization of three NLS regions of the R protein fused to GUS (above) and schematic representation of R:GUS fusion protein showing localization of three NLSs (below). Positions of amino acids are indicated above the construct; the acidic domain of R protein (striped box), helix-loop-helix domain (stippled box) and three NLSs (NLS-A -orange circle, NLS-M -yellow circle and NLS-C -green circle) are indicated. Amino acid sequences of three NLSs of the R protein are shown under corresponding photomicrographs. Tissues were stained using X-gluc histochemical staining and analyzed with the Nomarski optics. The brown particles on pictures with NLS-A and NLS-M result from gold precipitation. Bars = 10 μ m.



and exhibited subcellular localization similar to the intact R protein fused to GUS (Fig. 2.1A).

The deletion constructs were also used to examine the amino-terminus of the GUS:R fusions (Fig. 2.2B). The series of deletions from the carboxyl-terminus contained a.a. 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 a.a. 82-109 and 100-109 (Fig. 2.2A,B). GUS activity of the fusion protein NLS-A+GUS (a.a. 100-109) partitioned between the nucleus and cytoplasm (Fig. 2.3).

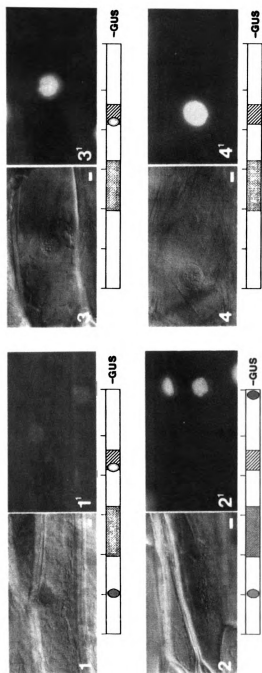
Since the R:GUS and GUS:R deletion constructs described could not distinguish any NLSs in the region of a.a. 109-598, a second set of constructs was designed (Fig. 2.2A,B). The central region of the R protein (a.a. 109-598) was subdivided into constructs containing the basic helix-loop-helix motif (a.a. 411-457 and a.a. 419-428) and the non-basic residue rich region (a.a. 128-411). Amino acids 128-411 were unable to redirect GUS to the nucleus and remained in the cytoplasm (Fig. 2.2A); this was not analyzed in the GUS-R orientation. However, a.a. 411-512, a.a. 411-457 and 419-428 (NLS-M) were sufficient to redirect GUS to the nucleus (Fig. 2.2A, and 2.3). 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. 2.2B). Therefore, for 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 were sufficient to redirect the reporter protein GUS to the nucleus of onion epidermal cells (Fig. 2.3).

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 in either none, one or two NLSs in the R protein (Fig. 2.4). The constructs were then subcloned into expression vectors and transiently expressed in onion epidermal cells, as in the previous experiments.

Figure 2.4. 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 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 2.3, 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 nuclei-specific DAPI staining (1¹-4¹). Tissues were stained and analyzed as in Figure 2.1. Bar = 10 μ m.



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. 2.4 (4)]. This indicated that all NLSs in the R protein were identified. However there is the formal possibility exists that the deletion of the NLSs 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. 2.5). NLS-A, in the intact R protein, was able to function as an NLS but it was inefficient as a signal and resulted in GUS activity in both the nucleus and cytoplasm (Fig. 2.5). Therefore, in both the intact R:GUS and NLS-A:GUS protein, NLS-A was an inefficient NLS (Figs. 2.3 and 2.5). Both NLS-M and NLS-C also retained their functions as NLSs and conferred partitioned localization. However, their expression in the nucleus was visually greater than in the cytoplasm [Figs. 2.4(2.3) and 2.5]. Despite the ability of the polypeptide encoding sequences for NLS-M and NLS-C to redirect GUS activity to the nucleus, when those signals were present in the R

A₁₀₀₋₁₀₉	NLS	M₄₁₉₋₄₂₈	NLS	C₅₉₈₋₆₁₀	NLS	LOCALIZATION
						<u>R-GUS</u>
-		-		-		C
+		-		-		N/C
-		+		-		N/C
-		-		+		N/C
+		+		-		N
+		-		+		N/C
-		+		+		N

Figure 2.5. Summary of histochemical analysis of R:GUS fusion proteins which identified NLSs that were necessary for nuclear localization.

protein minus the two other NLSs, NLS-M and NLS-C were incapable of conferring exclusive nuclear localization to the R:GUS fusion protein. The constructs which retained two of the three NLSs displayed different subcellular locations depending upon the orientation of R protein to GUS [Figs. 2.4(1), 2.2 and 2.5]. Since the R:GUS fusions exhibited stronger nuclear localization than the GUS:R fusions (Fig. 2.5), conclusions were drawn from the R:GUS fusion proteins. If NLS-A (Fig. 2.5) or NLS-C [Figs. 2.4 (1) and 2.5] were deleted, the fusion protein localized to the nucleus (Fig. 2.5). Therefore, either combinations of 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. 2.4(2)]. Our conclusion from this data was that two NLSs, one of which must be NLS-M, were sufficient and necessary for the strong localization 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 enzymatic 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 efficient 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. 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, 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. 2.3). The third signal, NLS-A (100-109), partially redirected GUS to the nucleus, partitioning the fusion protein between the nucleus and cytoplasm. Since several larger constructs including NLS-A, encoding a.a. 82-109 and 1-109 (Fig. 2.2), also partially redirected GUS, this inefficient targeting may be due to intrinsic weakness of the targeting signal or it is possible that amino acids following a.a. 109 are part of the signal, but this was not analyzed. The identification of the three NLSs was confirmed when the gene fusion constructs containing the full-length R protein with the three NLSs deleted (A,M,and C) were retained in the cytoplasm. Deletion of two of the three NLSs revealed that not all three NLSs were 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 amino acid composition and may confer different specificities to the nuclear import machinery. NLS-A had the most intriguing composition because it contained arginines and no lysines. 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 amino acids that were interspersed within its basic residues. One of the few NLSs which has a high content of hydrophobic amino acids is the yeast Mat α 2 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 amino acids than NLSs A or C, with five arginines and one lysine within the ten amino acid signal. The high concentration of basic amino acids in NLS-M is similar to the

SV40 large T antigen NLS (Kalderon et. al., 1984a) in which five of the seven amino acid signal are basic. Another transcription factor, myoD1, which shares homology to the helix-loop-helix domain, also contains an NLS in this motif. However, the NLS was defined to 34 a.a. of the helix-loop-helix and it is unknown if the NLS of myoD1 is in the amino-terminus (the first 10 a.a. of the 34 a.a. 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. 2.6). 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 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 (Dingwall and Laskey, 1991). Bipartite signals contain two regions enriched in basic amino acids separated by more than 4 amino acids

Figure 2.6. Amino acid comparison of R-Lc to other homologous regulatory proteins. Alignments are made to maximize homology with the NLSs of R. Identical amino acids 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 (Edmondson 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).

R-Lc	G D R R A A P A R P	M S E R K R R E K L	M I S E A L R K A I G K R
R-S			
B-Peru	C : R P V G		: S
DEL	: T N T : A K	L	V K : Q V T M S

L-Myc
 N-Myc
 Myogenin
 Cbf1
 AP-4
 E3
 E47

F L : D
I L Q : D
L R : L K :
E V : M :
S N : M : S :
L I : P M :
A R V V R D :

(Robins 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 a.a.) 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 long spacer (32 a.a.) separating the two basic regions which 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 (Dworetzky 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 amino acids of the NLSs might be important for function, we searched for conserved amino acids 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 amino acids. Therefore, the regions corresponding to R NLSs are equally conserved (Fig. 2.6). However, the maize B (Radicella et. al., 1991) and *Antirrhinum* Del (Goodrich et. al., 1991) proteins share 78% and 25% amino acid 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. 2.6). NLS-A was the least conserved and was the weakest of the signals that we have identified. Though NLS-C was not highly conserved (Fig. 2.6), the presence of two lysines and overall hydrophobicity content of the carboxyl-terminus are retained. Since 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. Since 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 (Varagona et. al., 1992; Raikhel, 1992) and for steroid hormone receptors which 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 Richer (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. None of the R protein's NLSs share homology to 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, since 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 lysine residues, a characteristic that has been observed only in viral proteins. NLS-M possessed the greatest density of charged residues, with seven of the ten amino acids being basic. NLS-C was enriched with hydrophobic residues which also affect the charge density of the NLS. Though it is not surprising that the compositions of the signals are different, as NLSs lack a consensus sequence, it is obvious that the import

machinery has to recognize some general features of the NLSs. Therefore, signals which are as divergent in charge and hydrophobicity as those in the R protein could be useful in the identification of different NLS binding proteins.

ACKNOWLEDGEMENTS

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

REFERENCE

- An G, Ebert PR, Mitra A, HA SB (1988) Binary vectors. Plant Mol Biol Man A3: 1-19**
- Beckmann H, Su LK, Kadesch T (1990) TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μ E3 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 Plant Genetics, UCLA Symposia on Molecular and Cellular Biology, M. Freeling, ed (New York: Alan R Liss) Vol 35: 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 psuedogene. Genes Dev. 1:1311-1326**
- Dingwall C, Laskey RA (1986) Protein import into the cell nucleus. Ann Rev Cell Biol 2: 367-390**

- Dingwall C, Laskey RA (1991)** Nuclear Targeting sequences- a consensus? *TIBS* 16: 478-481
- Dworetzky 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: 2517-2801
- 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
- Howard EA, Zupan JR, Citovsky V, Zambryski PC (1992)** The VirD2 protein of *A. tumefaciens* contains a C-terminal bipartite nuclear localization signal: Implications for nuclear uptake of DNA in plant cells. *Cell* 68: 109-118
- Hu Y-F, Luscher B, Admon A, Mermoud 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 Reporter* 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, T.M., Wolf, E.D., Wu, R., and Sanford, J.C. (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, T. A., Roberts, J. D., and Zakour, R. A. (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(#1, pt1): 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. *Nucl 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)** Transport of proteins to the nucleus. *Plant Phys*, 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, E. F., and Maniatis, T. (1989).** *Molecular Cloning: A Laboratory Manual*, Ed 2 Cold Spring Harbor, Cold Spring Harbor NY
- Standiford DM, Richter JD (1992)** Analysis of a developmentally regulated nuclear localization signal in *Xenopus*. *J. Cell Biol.* vol 118 5: 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 1,2: 1-5**
- Zhao L, Padmanabhan R (1988) Nuclear transport of adenovirus DNA polymerase is facilitated by interaction with preterminal protein. *Cell* 55: 1005-1015**

CHAPTER 3

CHARACTERIZATION OF THE CARBOXY-TERMINAL NUCLEAR LOCALIZATION SEQUENCE OF THE MAIZE R PROTEIN

INTRODUCTION

Eukaryotes contain nuclei which organize the genomic DNA and separate the processes of transcription and translation. The separation of transcription and translation by the nuclear envelope necessitates the transport of transcription factors and other regulatory proteins from their site of synthesis in the cytoplasm to the nucleoplasm. Therefore, proteins to be transported to the nucleus must contain a targeting signal, termed a nuclear localization sequence (NLS), recognizable to the nuclear transport machinery. Translocation of proteins across the nuclear envelope occurs at nuclear pore complexes which are aqueous channels connecting the cytoplasm to the nucleoplasm. Though diffusion of small molecules does occur (Paine et. al., 1975), the import of most proteins to the nucleus is both energy and NLS dependent (for review, see Forbes, 1992). NLSs vary from 7 to 40 amino acids in length with no consensus sequence though they are enriched in basic amino acids. Unlike targeting signals to other organelles, NLSs are found in various locations within different nuclear proteins, are not cleaved and can contain multiple targeting signals (for review, see Garcia-Bustos et. al., 1991).

Based on amino acid composition and size, NLSs can be separated into three groups: SV40 large T antigen-like (Kalderon et. al., 1984a,b; Lanford and Butel, 1984), bipartite (nucleoplasmin; Dingwall and Laskey, 1991) and Mat α 2-like (Hall et. al., 1984). SV40-like signals are characterized as being 7-20 amino acids in length and enriched in basic amino acids, whereas bipartite signals contain two basic amino acid enriched regions separated by 10-30 amino acids. Mat α 2-like NLSs contain several hydrophobic amino acids and a single basic amino acid enriched region.

The Mat α 2 protein contains two NLSs, however, the Mat α 2-like type of NLS is based on the targeting signal located in the amino terminus of the protein (Hall et. al., 1984). The polypeptide motif KIIPIK in the amino terminal Mat α 2 NLS is conserved in several other yeast nuclear proteins and is, therefore, considered to be a type of NLSs, Mat α 2-like. However, there are no defined features for the Mat α 2 NLS because it has not been thoroughly analyzed and no other Mat α 2-like NLS has been identified. In addition, unlike other NLSs which can function as targeting signals in animal, plant and fungal systems, the yeast Mat α 2 NLS (amino terminal signal) does not function as a targeting signal in animal cell lines (Chelsky et. al., 1989;

Lanford et. al., 1990). NLSs of other proteins, such as SV40 large T antigen, have been shown to function in animal, yeast and plant cells (Lanford and Butel, 1984; Nelson and Silver, 1989; Varagona and Raikhel, 1994; van der Krol and Chua, 1991). It is unknown why the Mat α 2 NLS does not function in animal systems.

Several NLSs have been identified in plants, similar to the animal and yeast NLSs (see Raikhel, 1992 for review). Recently, we identified three NLSs in the maize R protein, a transcriptional activator in the anthocyanin biosynthesis pathway (Shieh et. al., 1993). The NLS located in the amino terminus of the R protein (NLS-A) is unusual in its amino acid composition because it does not contain the basic amino acid lysine. All eukaryotic NLSs identified contain the basic amino acid lysine and the absence of lysines in NLSs is only found in some viral NLSs. The NLS located in the middle of the R protein (NLS-M) is adjacent to the helix-loop-helix motif and is a SV40-like NLS. The carboxyl terminal NLS (NLS-C) of the R protein is similar to the Mat α 2-like NLS because it contains several hydrophobic amino acids and the basic amino acids are arranged in a similar pattern. Unlike the well studied SV40-like NLSs (NLS-M) and bipartite NLSs, NLSs which are similar to the Mat α 2-like NLSs have not been studied in

detail. Therefore, we choose to identify the essential amino acids within NLS-C. To accomplish this, mutations in NLS-C were designed to examine the role of the hydrophobic amino acids, basic amino acids, NLS-C's similarity to the yeast Mat α 2 NLS, and to determine if amino acid context is important. The ability of the mutated NLSs to redirect GUS to the nucleus was then assayed. Several of these mutations were unable to redirect GUS activity to the nucleus. In addition, because NLS-C is similar to the Mat α 2 NLS, we have made a Mat α 2 NLS β -glucuronidase (GUS) fusion construct and transiently expressed the construct in onion epidermal cells to ascertain if the Mat α 2 NLS can redirect the reporter protein to the nucleus.

MATERIALS AND METHODS

Materials

White onions were purchased locally, stored at 4° C in the dark and used within two weeks. Oligonucleotides were synthesized by the MSU Macromolecular Facility and enzymes for molecular manipulations were obtained from Boehringer Mannheim Biochemicals

(Indianapolis, IN) and New England Biolabs (Beverly, MA). The materials for the helium biolistic gun transformation system (Dupont, 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 pMF6 (Goff et. al., 1990).

The NLS-C:GUS construct was described in Shieh et. al., 1993. These oligonucleotides were designed to mutagenize the NLS-C:GUS construct to create the new sequence;

Mat α 2 NLS:GUS TGCCGTCGTG CCCTGGATCG
ATTCTAGAAT GAACAAGATC CCGATCAAGG ACCTGCTGAA
CCCGCAGAGT GGGTACGGTC AG

NLS-C/Mat α 2 hybrid:GUS CGAGGCTCTT CGCAAGATCC
CGATCAAGCG GAGTGGGTAC

Reverse NLS-C:GUS GCCGTCGTGC CCTGTATCGA
TCATATGCGG AAGGGGATAG CTAAACGCCT TGCTGAGAGC

ATCATGAGTG GGTACGGTCA G

Minus Basic NLS-C:GUS GCCGTCGTGC CCTGTATCGA
TCATATGATC AGCGAGGCTC TGCGCCAGGC TATAGGGCAG
CGGAGTGGGT ACG

Minus Hydrophobic NLS-C:GUS GCCGTCGTGC CCTGTATCGA
TCATATGACC AGCGAGGCTC AGCGCAAAGC TACCGGGAAG
CGGAGTGGG.

Transformation of Onion Cells

Onion epidermal layers and plasmids were prepared as described in Shieh et. al. (1993). The preparation of the gold particles and the transformation conditions are the same as described in Shieh et. al. (1993) except the plasmid DNA concentration was increased from 2.5 to 5.0 μg and the duration of sonication of the gold particles was decreased (cup horn probe, 60% power, 10 s).

Histochemical Analysis

The colorimetric assay for β -glucuronidase (GUS) activity is the same as described in Shieh et. al. (1993) except for the X-gluc buffer

and substrate incubation temperature. The onion cell incubation which was increased to 37°C for 24 hours and the X-gluc buffer was altered to improve the viability of the onion cells (2mM X-gluc, 20 mM NaPO₄, pH 7.0, 0.05 mM K Ferrocyanide, 0.05 mM K Ferricyanide and 0.01% Triton X100). As previous, the intracellular localization of the blue precipitate was determined using a Zeiss Axiophot microscope with Nomarski optics. The subcellular localization of each fusion protein was determined from five to ten separate transformations. The minimum number of cells analyzed for each construct was twenty.

RESULTS

Previously, it was demonstrated that the carboxyl terminus of the maize R protein encoded a nuclear localization sequence (NLS C, 13 a.a.) which was able to redirect GUS protein from the cytoplasm to the nucleus in onion epidermal cells (Shieh et. al. 1993). Based on amino acid preference for helix formation, computer predicted secondary structure and plotting on a helical wheel, NLS-C is predicted to form an amphipathic alpha helix. To identify the amino acids which are

important for NLS-C to function, several constructs were made to study the function of the hydrophobic amino acids, basic amino acids, context of the amino acids and structural similarity between NLS-C and the yeast Mat α 2 NLS.

To construct the GUS:modified NLS gene fusions, the NLS-C:GUS construct was altered by site-directed mutagenesis. The constructs were designed to encode the modified NLS-C at the amino terminus of GUS as this orientation was previously shown to be optimal for targeting of the gene fusion constructs to the nucleus (Shieh et. al., 1993). The constructs were transformed into a monolayer of onion epidermal cells by particle gun bombardment and subcellular localization of the fusion proteins was determined with the histochemical substrate X-gluc. Addition of NLS-C to GUS redirected the GUS activity to the nucleus as indicated by the histochemical stain X-gluc in the nucleus (Fig. 3.1 and 3.2A, Table 3.1).

Context of the Amino Acid Sequence Influences NLS Function

Since all NLSs contain basic amino acids it is hypothesized that they are the sole determinant for a functional NLS (Boulikas. 1993). To ascertain whether or not this hypothesis is correct, a construct was

M I S E A L R K A I G K R	NLS-C
M R K G I A K R L A E S I M	rev-C NLS-C
M I S E A L R Q A I G Q R	minus-basic NLS-C
M <u>T</u> S E A Q R K A <u>T</u> G K R	minus-hydrophobic
	NLS-C
M I S E A L R <u>K I P I K</u> R	NLS-C/Matα2
M N <u>K I P I K</u> D L L N P Q	NLS-Matα2

Figure 3.1. Amino acid sequences of the mutated NLS-C polypeptides.

Figure 3.2. Histochemical localization of GUS activity of each of the mutated NLS-C GUS fusion constructs (A-G). The corresponding locations of the nuclei are shown with the DNA specific dye DAPI (A'-F'). A, NLS-C GUS; B, rev-NLS-C GUS; C, minus hydrophobic NLS-C GUS; D, minus basic NLS-C GUS; E NLS-C/Mat α 2 GUS; F, Mat α 2 GUS

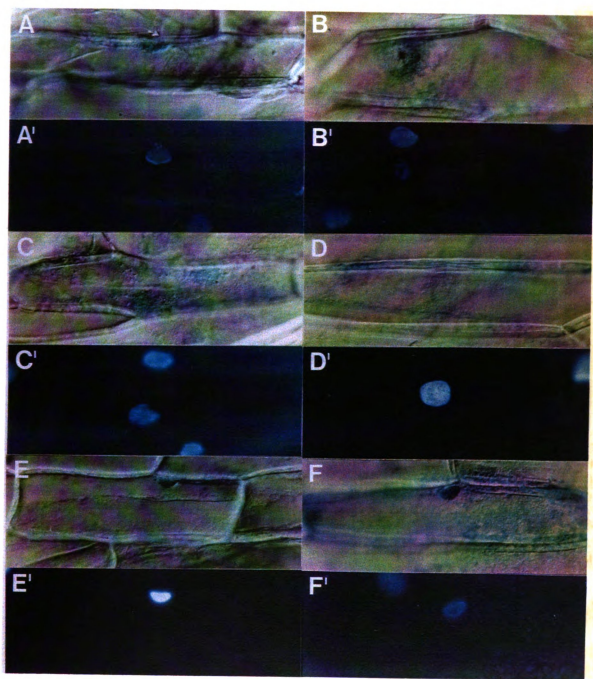


Table 3.1

<u>NLS</u>	<u>Localization</u>
NLS-C	Nuclear
Rev-NLS-C	Cytoplasmic
Minus-Basic	Cytoplasmic
Minus-Hydrophobic NLS-C	Nuclear/Cytoplasmic
NLS-C/Matα2	Nuclear
NLS-Matα2	Nuclear/Cytoplasmic

Table 3.1 Summary of the histochemical analysis for the mutated NLS-C:GUS fusion protein.

designed to reverse the amino acid sequence of NLS-C (Fig. 3.1, rev-NLS-C). The reversed NLS-C retains the spacing of the basic amino acids while changing the context in which the amino acids are presented. The reverse NLS-C polypeptide was unable to redirect GUS activity to the nucleus as indicated by the cytoplasmic localization of the histochemical stain X-gluc (Fig. 3.2B, Table 3.1). The dark spot in the area of the nucleus are the gold particles which carried the reverse NLS-C:GUS DNA into the nucleus and does not represent GUS activity. The micrograph in Figure 3.2B shows that GUS activity remained in the area surrounding the nucleus; this may represent the signal binding at the nuclear envelope without translocation across the nuclear envelope.

The Hydrophobic Amino Acids are not Essential

The role of the hydrophobic amino acids in NLS-C was examined by substituting the hydrophobic amino acids with more charged molecules (Fig. 3.1, minus-hydrophobic NLS-C); this removes the predicted amphipathic nature of the signal. The threonine for isoleucine and glutamine for leucine substitutions were chosen to negate the hydrophobicity while maintaining average surface volume and

normalized frequency of occurrence in an alpha-helix. When expressed in onion epidermal cells, the histochemical stain X-gluc was partitioned between the nucleus and cytoplasm (Fig. 3.2C, Table 3.1), indicating that the minus-hydrophobic NLS:GUS protein was partitioned between the nucleus and cytoplasm. Although the amount of partitioning varied, distinct nuclear (or cytoplasmic) targeting was not observed during any experiment with this construct.

Basic Amino Acids are Essential

The single feature which all NLSs share is the presence of basic amino acids which are considered essential. To determine if the basic amino acids are essential for NLS-C's function, the two lysines were substituted with two polar uncharged glutamine residues (Fig. 3.1, minus-basic NLS-C). This change alters the charge density of the polypeptide while maintaining the average surface volume, hydrophilicity and normalized frequency of occurrence in an alpha-helix. This construct did not redirect GUS activity to the nucleus as indicated by the histochemical stain X-gluc in the cytoplasm (Fig. 3.2D, Table 3.1). Therefore, the lysines are essential for NLS-C's function. However, it is possible that another amino acid, such as an arginine

may substitute for the lysine.

Similarity of NLS-C to the NLS of Mat α 2

To determine if NLS-C and the yeast Mat α 2 NLS are similar, a hybrid NLS was constructed substituting the consensus region of the Mat α 2 NLS (KIPIK; Hall et. al., 1984) into NLS-C (Fig. 3.1, NLS-C/Mat α 2). The two lysines of both NLSs are separated by three amino acids indicating that the two signals are similar. If the hybrid NLS functions as a targeting signal then the implication is that yeast and plant NLSs are similar. The hybrid targeting signal, NLS-C/Mat α 2, redirected GUS activity from the cytoplasm to the nucleus (Fig. 3.2E, Table 3.1) and there was no notable difference between the localization of either NLS-C/Mat α 2 or NLS-C GUS fusion proteins.

Mat α 2 NLS in Plants

NLS-C and the yeast Mat α 2 NLS are similar in the basic amino acid spacing and content of hydrophobic amino acids. Therefore, based on the similarity of NLS-C to Mat α 2, we wanted to determine if the Mat α 2 NLS can function in plant cells. When Mat α 2 NLS was

fused to GUS, the activity was localized to the nucleus and cytoplasm as indicated by the histochemical stain X-gluc in the nucleus and cytoplasm (Fig. 3.2F, Table 3.1). This indicated that Mat α 2 NLS retains partial function in plants.

DISCUSSION

NLS-C of the maize R protein contains several hydrophobic amino acids and the lysines are spaced in the same pattern as those in the Mat α 2-like NLSs. Therefore, NLS-C is more similar to the Mat α 2-like NLS than to the SV40-like NLS. As evidence to indicate that NLS-C and Mat α 2 NLS are similar, a fusion construct of the two signals was constructed. If the two targeting signals are similar, then regions from each should be capable of complementing a similar region in the other. Therefore, the five central amino acids of NLS-C (KAIGK) were substituted with the KIIPIK sequence of the Mat α 2 NLS (construct NLS-C/Mat α 2, Fig. 3.1). The KIIPIK polypeptide is conserved in several yeast nuclear proteins and is therefore considered the core of the Mat α 2 NLS (Hall et. al., 1984). The NLS-C/Mat α 2 hybrid NLS

when fused to GUS redirected GUS activity to the nucleus with no notable difference between the NLS-C and NLS-C/Mat α 2 in their ability to direct GUS activity to the nucleus (Fig. 3.2E, Table 3.1). This indicates that the secondary structure in the basic region of NLS-C and Mat α 2 are homologous and suggests that NLS-C is a Mat α 2-like signal. Since the KIPK substitution in NLS-C contains a proline residue it may be that the predicted alpha-helix structure of NLS-C would be disrupted. However, a proline within four amino acids of the end of a helix can be tolerated in a helix and therefore it may be tolerated within the NLS-C/Mat α 2 NLS.

If NLS-C is a Mat α 2-like NLSs, then the Mat α 2 NLS should function in plants. Therefore, the ability of the Mat α 2 NLS to redirect GUS to the nucleus was assayed. When the Mat α 2 NLS was fused to GUS, the GUS protein was located in both the nuclear and cytoplasmic compartments (Fig. 3.2F, Table 3.1). This indicated that despite its inability to function as an NLS in animal systems, the Mat α 2 NLS can function in plant systems but it is not as strong a signal as NLS-C. In addition, preliminary experiments have been performed to determine if NLS-C functions as a nuclear targeting signal in yeast and animal cells. When overexpressed in yeast, the NLS-C and Mat α 2 NLS

redirected GUS to the nucleus and, like the Mat α 2 NLS, NLS-C was unable to redirect GUS to the nucleus in the *Xenopus* oocytes *in vitro* transport system (preliminary data not shown). This substantiates the hypothesis that NLS-C is a Mat α 2-like NLS.

The amino acids which are important for a Mat α 2-like NLSs to function have not been identified. Therefore, we analyzed NLS-C to determine if the basic amino acids, charge density and hydrophobic amino acids are important for its function. Basic amino acids are the single feature common in all NLSs and their importance is best indicated by the fact that, unlike other amino acids, a single alteration of a basic residue in a targeting signal can completely negate its ability to function as a targeting signal. The best studied of these mutations is the SV40 large T antigen NLS (P₁₂₆KK₁₂₈KRKV) in which the lysine at position 128 was replaced with a threonine (Kalderon et. al., 1984b). Similar studies in plants were performed on bipartite NLSs where a substitution of the basic residues affects the NLS's targeting function (Varagona et. al., 1994).

Therefore, two constructs were designed to determine if the context of the amino acids or the basic amino acids are important in NLS-C. It has been proposed that if four amino acids are basic in a

hexapeptide then that region will constitute a NLS (Boulikas, 1993) as the basic charge is the major factor defining a polypeptide as a targeting signal. To determine if the basic charge density of a NLS is an essential factor for NLS functioning, the amino acid order in NLS-C was reversed (Fig. 3.1, reverse NLS-C), thereby altering the context of the NLS without changing the charge density or hydrophilicity of the signal. When fused to the GUS protein, the reversed NLS-C was unable to redirect GUS activity to the nucleus (Fig. 3.2B, Table 3.1) and, therefore, the charge density does not determine NLS function. Rather, the context of the amino acids was crucial for it to function as a targeting signal. These findings are interesting because the seven amino acids encompassing NLS-C (RKAIGKR) are virtually palindromic (alanine and glycine are very similar in structure). This inversion may have created a subtle change in the signal which negated its function as a targeting signal. Alternatively, residues amino terminal to the basic amino acids may have a strong influence on the structure of the NLS, possibly to initiate the formation of the alpha-helix. Similarly, Adam et. al. (1989) demonstrated that a reversed-order SV40 large T antigen NLS will not compete against a wild type SV40 large T antigen NLS for binding to a putative NLS receptor in

mammalian cells.

To determine if the basic amino acids are essential in NLS-C, the two lysine residues were substituted with glutamic acids (minus-basic NLS-C). Substitution of the two lysine residues negated the ability of the targeting signal to redirect GUS activity to the nucleus (Fig. 3.2D, Table 3.1). Therefore, the lysines were essential for NLS-C to function and this correlates with numerous examples indicating that the basic amino acids are essential in NLSs (Kalderon et. al., 1984; Varagona and Raikhel, 1994).

Multiple hydrophobic residues are not frequently found in NLSs, therefore, the role of the hydrophobic amino acids in NLS-C was investigated. The predicted secondary structure for NLS-C is an amphipathic alpha-helix which exposes the basic amino acids on one side of the helix while the hydrophobic residues are hidden from the surface of the protein. Substitution of the hydrophobic amino acids with polar uncharged amino acids would eliminate the amphipathicity. The minus-hydrophobic NLS-C partially redirected GUS activity to the nucleus (Fig. 3.2C, Table 3.1). Therefore, the hydrophobic amino acids are important but not essential for NLS-C to function. Unlike other constructs tested in this and our previous study (Shieh et. al.

1993), the ratio of GUS activity in the nucleus and cytoplasm varied in different experiments but always partitioned (data not shown). The variability suggests that the minus-hydrophobic NLS does not form a stable structure recognizable by the transport machinery.

In our study, it has been demonstrated that NLS-C is a Mat α 2-like targeting signal because it contains several hydrophobic amino acids, the basic amino acids are similarly spaced and the conserved sequence KIPK of the Mat α 2 can substitute for five amino acids in NLS-C. It was determined that the hydrophobic amino acids are important for NLS-C to function as a targeting signal. In addition, because the Mat α 2 NLS from yeast was able to redirect GUS activity to the plant cell nucleus, the nuclear transport mechanism in plant and yeast systems are closely related.

Since several of the mutations delete the targeting function of NLS-C, future experiments will use these constructs to identify components of the nuclear transport machinery. Other mutated NLSs tested in plants (i.e. SV40 large T antigen mutant and the mutated bipartite signal of the maize O2 protein; Varagona and Raikhel, 1994) were capable of redirecting some GUS activity to the nucleus. Therefore, the minus-basic and reverse NLS-C constructs, which are

restricted to the cytoplasm, are stronger negative controls for transport. In addition, the reverse NLS-C contains the same amino acids and charge density as NLS-C while not functioning as a targeting signal. The reverse-NLS-C is, therefore, a more appropriate control to distinguish non-specific binding of NLSs to NLS-binding proteins which are interacting based upon charge.

REFERENCES

- Adam SA, Lobi TJ, Mitchell MA, Gerace L (1989)** Identification of specific binding proteins for a nuclear location sequence. *Nature* 337: 276-279
- Boulikas T (1993)** Nuclear localization signals. *CRC Crit. Rev. Euk. Gene Expr.* 3(3): 193-227
- Chelsky D, Ralph R, Jonak G (1989)** Sequence requirements for synthetic peptide-mediated translocation to the nucleus. *Mol. Cell. Biol.* 9: 2487-2492
- Dingwall C, Laskey RA (1991)** Nuclear Targeting sequences- a consensus? *TIBS* 16: 478-481
- Forbes D (1992)** Structure and function of the nuclear pore complex. *Ann. Rev. Cell Biol.* 8: 495-527
- 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: 2517-2801
- Hall MN, Hereford L, Herskowitz I (1984)** Targeting of E.coli β -galactosidase to the nucleus in yeast. *Cell* 36: 1057-1065
- 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

Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (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

Lanford RE, Feldherr CM, White RG, Dunham RG, Kanda P (1990) Comparison of diverse transport signals in synthetic peptide-induced nuclear transport. *Exp. Cell Res.* **186**: 32-38

Nelson M, Silver P (1989) Context affects nuclear protein localization in *Saccharomyces cerevisiae*. *Molec. and Cell. Biol.* **9**: 384.389

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

Raikhel NV (1992) Transport of proteins to the nucleus. *Plant Phys.* **100**: 1627-1632

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

Shieh MW, Wessler SR, Raikhel NV (1993) Nuclear targeting of the maize R protein requires two nuclear localization sequences. *Plant Physiol.* **101**: 353-361

van der Krol AR, Chua N-H (1991) The basic domain of plant B-ZIP proteins facilitates import of a reporter protein into plant nuclei. *Plant Cell* **3**:667-675

Varagona MJ, Raikhel NV (1994) The basic domain in the bZIP regulatory protein Opaque2 serves two independent functions: DNA binding and nuclear localization. *Plant J.* **5**: 207-214

CHAPTER 4

FUTURE RESEARCH PROSPECTIVES

Towards a consensus NLS

The structural elements which are fundamental for a NLS are unknown and will require either identification of additional NLSs or more detailed information on the NLSs identified if a consensus is to ever to be discovered. Site-specific mutations of SV40 large T antigen NLS (Kalderon et. al.,1984), bipartite opaque 2 NLS (Varagona and Raikhel, 1994) and the Mat α 2-like maize NLS-C (chapter 3) have all indicated that the basic amino acids are essential for NLS function. However, the amino acid spacing of the basic residues is not the sole factor defining a NLS because rev-NLS-C does not redirect GUS to the nucleus. Analysis of the amino acid sequence of the NLSs identified (Boulikas, 1993) and the mutational NLS studies do not reveal a common structural feature which could denote a consensus sequence. Rather, the variation in NLS amino acid composition and length suggests that there are structural features which are not obvious from the sequence. In addition, NLS function is dependent on the context in which it is presented. As an example, the NLSs (A, M and C) of the maize R protein redirected more GUS protein to the nucleus when fused to the amino- rather than the carboxy-terminus of GUS, see chapter 2 (also see chapter 1; SV40 NLS in pyruvate kinase). The

large variation in NLS amino acid length and context in which they are presented distinguishes them from other organelle targeting signals with consensus structures. Chloroplast transit peptides, mitochondrial signals and signal peptides (secretory) have consensus sequences and are typically found as the amino-terminal residues and are not as variable in size and number as NLSs (Boulikas, 1993). Therefore, it is my assertion that there is not a single consensus sequence for all NLSs. Rather, NLS function is based upon common structural features. Therefore, to identify the NLS consensus structure it will be necessary to derive the structure from X-ray crystallographic data of the NLSs in their native proteins.

The nuclear transport machinery in plants

Having identified the NLSs of the maize R protein, the next objective is to use the NLSs to identify other components of the nuclear transport machinery in plants. Along with the Opaque2 bipartite NLS (Varagona et. al., 1992) we have plant NLSs similar to the three types of NLSs, the SV40-like (R-NLS-M), bipartite (opaque2) and Mat α 2-like (R-NLS-C). The plant bipartite NLS from opaque2 has already been shown to have a higher affinity for binding to tobacco and maize nuclei than the animal SV40 large T antigen NLS (Hick and Raikhel, 1993).

Similar experiments have been performed with R-NLS-C and binding to the nucleus is NLS specific with similar characteristics to those of the bipartite opaque2 NLS (Smith S, Hick GR and Raikhel NV; unpublished results). Chemical-crosslinking of radiolabeled opaque2 bipartite NLS to proteins extracted from the nuclear envelope has labeled three putative NLS-binding proteins which have binding characteristics similar to those of the isolated nuclei (Hick and Raikhel, unpublished). These putative NLS-binding proteins will be purified so that they can be studied in detail.

To determine if the putative NLS-binding proteins are part of the nuclear transport machinery, it will be necessary to develop an *in vitro* nuclear transport system. Reconstituting transport will allow the identification of individual components of the transport machinery. Also, the energy and cellular requirements for transport are unknown and there may be cytosolic NLS-binding proteins required for transport. The mutated NLS, rev-NLS-C, will be a useful control in the characterization of both an *in vitro* transport system and NLS-binding proteins because it maintains the charge density and amino acid content of NLS-C without functioning as a NLS in onion epidermal cells.

Understanding the molecular mechanism of nuclear transport is essential if we are to learn how cellular and developmental processes are regulated. Nuclear transport is typically thought to be constitutive, such that all translated nuclear proteins are immediately transported into the nucleus. However, some nuclear proteins are retained in the cytoplasm until activated for transport. For example, the glucocorticoid steroid hormone receptor is retained in the cytoplasm until it binds hormone and is then transported into the nucleus to initiate transcription. Regulation of nuclear import is of major importance in cellular differentiation. Nuclear localization of the *Rel* related proteins dorsal and NF-Kappa B is the determinant for dorsal-ventral axis formation and immunoglobulin synthesis. Both of these changes are key steps in the final determination of cell type. Since plant cells are fixed in place by the cell wall, they will be a model system to study regulated nuclear transport during cellular differentiation. In addition, due to the totipotency of plant cells, there must be additional levels of regulation unique to plants. Therefore, the study of nuclear transport in plants will give additional insights into the complexity of cellular regulation and differentiation.

REFERENCES

- Boulikas T (1993) Nuclear localization signals. CRC Crit. Rev. Euk. Gene Expr. 3(3): 193-227**
- Hick GR, Raikhel NV (1993) Specific binding of nuclear localization sequences to plant nuclei. Plant Cell 5:983-994**
- 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**
- 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**
- Varagona MJ, Raikhel NV (1994) The basic domain in the bZIP regulatory protein Opaque2 serves two independent functions: DNA binding and nuclear localization. Plant J. 5: 207-214**