This is to certify that the

dissertation entitled

IDENTIFICATION AND CHARACTERIZATION
OF NUCLEAR PORE COMPLEX PROTEINS IN PLANTS

presented by

Antje Heese-Peck

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Botany and
Plant Pathology

Natasha Raikhel
Major professor

Date Nov. 4, 1997

IDENTIFICATION AND CHARACTERIZATION OF NUCLEAR PORE COMPLEX PROTEINS IN PLANTS

By

Antje Heese-Peck

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

1997

ABSTRACT

IDENTIFICATION AND CHARACTERIZATION
OF NUCLEAR PORE COMPLEX PROTEINS IN PLANTS

By

Antje Heese-Peck

Nucleocytoplasmic transport of macromolecules occurs through nuclear pore complexes (NPC), proteinaceous structures with an approximate mass of 125 MD. Only approximately one third of more than 100 different NPC proteins have been identified and characterized mainly in vertebrates and yeast, but no plant NPC protein has been isolated so far. As an initial step to identify plant NPC proteins, we demonstrated that NPC proteins from tobacco suspension-cultured cells are modified by *N*-acetylglucosamine (GlcNAc). Using wheat germ agglutinin (WGA), a lectin that binds to GlcNAc, specific labeling is found associated with or adjacent to NPCs as shown by electron microscopy. To obtain more information about the sugar modification of the nuclear glycoproteins, an *in vitro*-labeling procedure using galactosyltransferase (GalTF) was employed. This enzyme specifically transfers the ^{3}H-galactose moiety of UDP-^{3}H-galactose to terminal, but not internal GlcNAc residues of glycoproteins. At least eight proteins are specifically labeled in the presence of the GalTF. The radioactively labeled sugars were bound to the nuclear proteins via an O-linkage. Interestingly, mass analysis indicated that the labeled glycans of plant origin were different from the single O-GlcNAc of vertebrate NPC proteins; the plant glycans consist of oligosaccharides larger in size than 5 GlcNAc.

The nuclear proteins modified with terminal GlcNAc, referred to as tGlcNAc-proteins, were purified by *Erythrina crystagalli* agglutinin affinity chromatography. Internal amino acid sequence information was obtained for three of the purified tGlcNAc-proteins, designated gp33, gp40 and gp65. A combination of PCR-based amplification, screening a cDNA library, and 5'RACE were used to clone a gene encoding gp40. All three gp40-peptides, identified by peptide sequencing of gp40, were localized in the deduced amino acid sequence of the gene product confirming the identity of the clone. No specific gene products have been obtained for gp33 and gp65 using similar approaches so far. Interestingly, the gene encoding gp40 has 28 to 34% amino acid identity to aldose-1-epimerases from different bacteria. Aldose-1-epimerases catalyze the interconversion of the α and ß configurations at the asymmetric C1-atom of aldoses. To date, no other gene encoding an aldose-1-epimerase has been isolated from animals and plants, and the exact function of aldose-1-epimerases is unknown in higher organisms. To determine the subcellular localization of gp40, monospecific antibodies were produced to recombinant gp40. Organelle fractionation in combination with protein blot analysis confirmed that gp40 was a nuclear protein. Consistent with its isolation as a putative NPC protein, gp40 was mainly present at the nuclear rim as shown by conventional light and confocal laser scanning immunofluorescence microscopy. Future analysis of gp40 as an aldose-1-epimerase-like protein may give new insights into the function of aldose-1-epimerases in eukaryotes.

To Scott,
for his love and support

ACKNOWLEDGEMENTS

Where do I start to thank all the people that have helped in some way or another to finish my Ph.D. thesis? I am very grateful to the members of my thesis committee, Drs. Frans deBruijn, Pam Green, Ken Keegstra, and John Wang, for their interest and helpful advice beyond the committee meetings, prelims and thesis defense. And of course, I would like to thank Natasha to give me (at least most of the time) the freedom to pursue my thesis project the way I felt was the "right" way. Over the past years, Natasha has been a major influence on who I became as a person and what I believed in. Thanks to all the past and present members of the Raikhel laboratory for many protocols, discussions and "entertaining" lab meetings throughout the years. In particular, I want to thank Rita and Susannah for their help and friendship during my first years in the lab; Diane and Tony, not only for fun coffee breaks, but also for their discussions and support during some rough times; and Glenn who has been a great mentor during my first years in graduate school. I feel very lucky to have met several special people in the PRL and in East Lansing; thanks to the members of Science Theatre (in particular, Bill, Matt, Dave Batch, and the rest of the "From Stardust to Life - A Cosmic Journey" group) for showing me that science can be very rewarding outside of the lab; Hans Kende who changed the fate of my life by accepting me to the PRL and of course, for introducing me to Scott; Ambro, Nikki, Tom, Beth, Pauline,

Linda, Esther, Eric, Christiane, Diane, Ronda, Joe, Susannah, Hilton, Christine, Frank, Sharon (I hope I did not forget anybody...) for becoming good friends; Beth and Esther for dragging me swimming (hhm, steam room!); Laura for making me laugh and forget work when coming home from the lab; Gabi, Claudia and Klaus for showing me that friendship can survive long distances; Tom for great game nights and keeping Scott occupied while I was frantically finishing my thesis; ... and of course, Jenny for her deep friendship!

I feel fortunate for having a very loving family who supported me during the past years. Vielen Dank, Mama, Papa und Kerstin, für Eure liebevolle Unterstützung in all meinen Entscheidungen während der letzten Jahre.

No words, however, can describe my thanks to Scott who has always been there for me with his love, support, and incredible patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The nuclear pore complex

1. Structure of the Nucleus

The eukaryotic cell is organized into functionally distinct compartments, called organelles, that are separated from the cytoplasm by lipid bilayers. Though organelles function independently, communication between the different compartments is essential for the eukaryotic cell to operate properly. The nucleus is the central organelle of the cell, in which the genetic material is stored in form of chromosomes. It is surrounded by a double membrane, the nuclear envelope (NE) (for review, see Dingwall and Laskey, 1992; Gerace and Foisner, 1994). The NE provides temporal and spatial separation between the transcription and processing of mRNA in the nucleus, and the translation of mRNA into proteins in the cytoplasm. The outer nuclear membrane (ONM) is contiguous and also biochemically and functionally similar to the endoplasmatic reticulum (ER) membrane. In contrast, the inner nuclear membrane (INM) has a different protein composition than the ONM and ER membranes. The INM is underlaid by the nuclear lamina, a filamentous network that is mainly composed of intermediate filaments (IF), named lamins, and provides the structural framework of the nucleus (for review, see Gerace and Foisner, 1994). The ONM and the INM are connected by a specialized "pore membrane" that has been proposed to serve as an attachment site for nuclear pore complexes (NPCs) via proteins present in the pore membrane domain (for review, see Gerace and Foisner, 1994). The large proteinaceous NPCs form aqueous channels and control communication between the nucleus and other cellular compartments. Though some molecules move through the NPCs by passive diffusion, nucleocytoplasmic transport of most proteins and different RNA species is mediated by a highly selective active transport mechanism. Active transport is energy-

dependent and requires the presence of signal sequences within these molecules (for review, see Görlich and Mattaj, 1996; Görlich, 1997). The NPCs, therefore, play a major control point in the movement of molecules in and out of the nucleus.

2. The structure of the nuclear pore complex

The general structure of the NPC seems to be conserved among eukaryotes, and the architecture and components of this complex have been investigated extensively in amphibian cells using different electron microscopic techniques (for review, see Panté and Aebi, 1993; Davis, 1995; Goldberg and Allen, 1995). Using transmission electron microscopy (TEM) and quantitative scanning transmission electron microscopy (STEM), two- and three dimensional maps of "intact" (i.e. without detergent treatment) and detergent released NPCs have revealed its mass and basic framework (Reichelt *et al.*, 1990; Hinshaw *et al.*, 1992; Akey and Rademacher, 1993). The NPC is a proteinaceous supramolecular structure of about 125 MD that has an overall cylindrical shape with a diameter of 120 nm and a height of about 70 nm. When viewed perpendicular to the plane of the NE (i.e. tangential sections), the NPC shows an eightfold rotational symmetry of eight identical units (Figure 1). In longitudinal sections, it appears as a tripartite structure consisting of a cytoplasmic and a nucleoplasmic ring that are connected to one another via two sets of eight spokes referred to as the spoke complex (Figure 1). The spoke-ring complex forms a central channel with a functional diameter of up to 26 nm and eight smaller channels, each with a diameter of about 9 nm. The smaller channels are proposed to allow passive diffusion between the nucleus and cytoplasm (Hinshaw *et al.*, 1992).

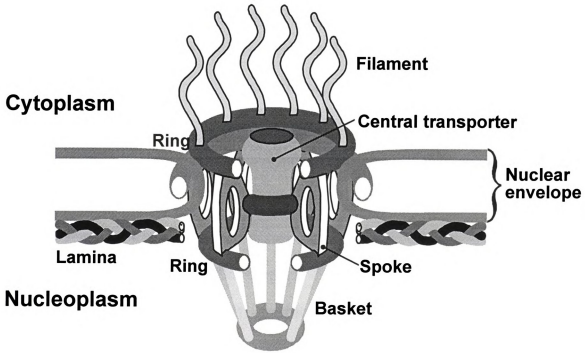


Figure 1.1

Schematic diagram of the nuclear pore complex.

Using non-detergent treated NPCs, other studies indicate that this part of the spoke complex transverses the NE to attach the NPCs to the membrane. Thus, the lumenal channels are partially masked by the membrane and may be too small to allow passive diffusion of molecules with up to 9 nm-diameter (Akey and Rademacher, 1993; Goldberg and Allen, 1996). The hour-glass shaped transporter described by some investigators is found in the central channel complex (Akey, 1995; Akey and Rademacher, 1993; Goldberg and Allen, 1996). This central transporter has been proposed to be involved in the active transport of macromolecules in and out of the nucleus.

Using surface imaging techniques such as field emission in-lens scanning electron microscopy (FEISEM; Goldberg and Allen, 1993, 1996) and metal shadowing in TEM (Jarnik and Aebi, 1991), peripheral structures of the NPCs have been identified. These structures give the NPC an asymmetric appearance and may reflect differences in substrate recognition. Eight cytoplasmic fibrils are attached to the cytoplasmic ring and appear to serve as docking sites for import substrate (Jarnik and Aebi, 1991; Panté and Aebi, 1996). From the outer periphery of the nucleoplasmic ring, eight filaments extend into the nucleoplasm and terminate in a distal ring to form the nuclear basket, and they may be involved in binding of export substrate (Goldberg and Allen, 1993, 1996; Jarnik and Aebi, 1991). The peripheral NPC structures may also connect nucleo- and cytoplasmic cytoskeletal elements to the NPC as suggested by Davis (1995).

Recently, FEISEM studies have revealed other structures including a star ring underlying the cytoplasmic ring and internal filaments in the cyto- and nucleoplasmic

rings (Goldberg and Allen, 1996). The same authors also describe a filamentous structure of 8-10 nm, named the NE lattice, that is different from the nuclear lamina and may connect the distal rings of the nucleoplasmic basket (Goldberg and Allen, 1992, 1996). The NPCs appear to attach to the lamina via the spoke ring complex, and adjacent NPCs seem to be interconnected via small radial arms within the NE lumen (Goldberg and Allen, 1996). Further analysis will be required to understand possible functions of these structures.

Yeast NPCs have a similar morphology to amphibian NPCs, including an eightfold symmetry, and appear to be composed of a similar spoke-ring structure, a central transporter, and peripheral structures (Rout and Blobel, 1993, and see references therein). Mass analysis of highly enriched yeast NPCs isolated by biochemical fractionation suggest that yeast NPCs have a mass of about 66 MD indicating that they are smaller than amphibian NPCs (Rout and Blobel, 1993). These highly enriched NPCs, however, lack components that are detected in thin section TEM using whole yeast cells. Thus, the smaller mass may be due to the loss of components during the fractionation procedure.

To date, plant NPCs have been mainly investigated using thin section transmission electron microscopy (TEM) and freeze-fracture TEM (for review, see Jordon *et al.*, 1980). Consistent with results from amphibians, NPCs from mono- and dicotyledons plants show the typical eightfold symmetry (Franke, 1966; Roberts and Northcote, 1970). They also appear to have the tripartite structure (Roberts and Northcote, 1970), and peripheral components, such as the cytoplasmic fibrils and the nucleoplasmic basket, have been described by many investigators (see references in

Jordan *et al.*, 1980). When isolated nuclei are exposed to detergent, salt, or urea, peripheral structures are released to yield a basic framework as described under similar conditions for amphibian NPCs (Hicks and Raikhel, 1993; see above). The biochemical isolation of plant NPCs has not been reported yet.

3. Identification of nuclear pore complex proteins

Based on quantitative scanning TEM, the amphibian NPC appears to consist of at least 100 different proteins (Reichelt *et al.*, 1990), while the yeast NPC may contain about 80 different NPC proteins (Rout and Blobel, 1993). To date, the genes of approximately one third of the NPC proteins have been cloned, including a dozen of vertebrate and approximately twenty-five of yeast NPC proteins (Tables 1.1 and 1.2). However, genes encoding plant NPC proteins have not yet been isolated. The known NPC components have been identified using a variety of immunological, biochemical and genetic approaches (for review, see Forbes, 1992; Rout and Wente, 1994; Davis, 1995; Doye and Hurt, 1997).

Vertebrate NPC proteins have been identified first by screening for antibodies that reacted specifically with the NPC using monoclonal antibodies raised against NPC/nuclear lamina fractions (for review, see Forbes, 1992; Table 1.1). These antibodies have been useful for cloning several genes encoding vertebrates NPC proteins, including gp210 (Wozniak *et al.*, 1989; Greber *et al.*, 1990) and p62 (D'Onofrio *et al.*, 1988; Cordes *et al.*, 1991, Carmo-Fonesca *et al.*, 1991). p62 is a member of a family of 8-12 NPC proteins that is also recogized by wheat germ agglutinin (WGA), a lectin that binds to *N*-acetylglucosamine (Davis and Blobel, 1986;

Table 1.1

Functional domains, localization and putative functions of vertebrate NPC proteins.
Ab, antibodies; glycosylat., glycosylation; term., terminus; for additional abbreviations, see text.

Table 1.1 Functional domains, localization and putative functions of vertebrate NPC proteins.

Year Protein (Mass)	modifications/ motifs	NPC location	Function / comments
1990			
p62 (62 kD)	O-GlcNAc, FXFG, coiled-coil	spoke-ring complex	form rod-shape homodimer *in vitro*, complex with p54 and p58 (p62-complex), p62-complex binds to importin ß and NTF2 required for protein import and RNA export
gp210 (190 kD)	N-linked glycosyl. cleaved N-terminal signal sequence	pore membrane	one transmembrane domain, large N-terminal in NE lumen, Ab to lumenal domain impair nuclear import
1993			
Nup153 (153 kD)	O-GlcNAc, FXFG Cys_2Cys_2 Zn fingers,	nucleoplasmic basket	C-term. overexpression: blocks poly(A^+) RNA export, but no effect on protein import, N-term.: required for association with NPC
Nup155 (155 kD)	-	cyto- + nucleoplasmic	functional homolog of yeast Nup170p
Pom121 (145 kD)	O-GlcNAc	pore membrane	one transmembrane domain, large C-terminal faces NPC
1994			
Nup214/CAN (214 kD)	O-GlcNAc, FXFG	cytoplasmic rings and fibrils	docking site for import substrate *in vitro* transgenic knock-out mouse: required for protein import and poly$(A)^+$ RNA export may be oncogenic

Table 1.1 (cont'd)

Tpr/Nup265 (265 kD)	coiled-coil	cytopl. or nucleopl. filaments	may be oncogenic (oncogenic fusion with kinases)
Nup107 (107 kD)	Leu zipper	precise localization unknown	homolog of yeast Nup84p
1995			
Nup358 (358 kD)	4 Ran-binding domains 8 Cys_2Cys_2 Zn fingers Leu rich domain cyclophilin A-like domain	cytopl. fibrils	docking site for import receptor complex and Ran
Nup98 (98 kD)	O-GlcNAc, GLFG, RNA-binding motif	nucleopl. basket, nuclear interior	*in vivo*: required for snRNA, mRNA, rRNA, 5S RNA export, but not for protein import and tRNA export *in vitro*: binds to soluble transport factors, indicating function in protein import putative homolog of yeast Nup116
1996			
p58 (58 kD)	O-GlcNAc, FXFG, coiled coil	spoke ring complex	forms complex with p62, p54 and p45 required for protein import (C-term. interacts with importin ß and NTF2
p54 (54 kD)	O-GlcNAc, FXFG, coiled coil	spoke ring complex	forms complex with p62, p58 and p45 required for protein import
p45 (45 kD)	O-GlcNAc, FXFG, coiled coil	spoke ring complex	forms complex with p62, p58 and p54 required for protein import maybe alternative spliced product of *p58*

Finlay *et al.*, 1987; Hanover *et al.*, 1987; Holt *et al.*, 1987; Park *et al.*, 1987; Snow *et al.*, 1987). Further carbohydrate analyses have indicated that these NPC glycoproteins are modified by single O-linked *N*-acetylglucosamine (O-GlcNAc) residues (Holt *et al.*, 1987). In general, the O-GlcNAc modification has served as an excellent tool to isolate O-GlcNAc NPC proteins by WGA affinity chromatography for subsequent peptide analysis and gene cloning of Nup98 (Powers *et al.*, 1995; Radu *et al.*, 1995b), Pom121 (Hallberg *et al.*, 1993), Nup214/CAN (Kraemer *et al.*, 1994) and the related p58, p54, p45 (in complex with p62; Hu *et al.*, 1996). Fractionation of proteins extracted from the NPC/nuclear lamina that did not bind to WGA columns has led to the identification of Nup155 (Radu *et al.*, 1993) and Nup107 (Radu *et al.*, 1994). In the future, biochemical procedures such as the isolation of annulate lamellae, a cytoplasmic organelle consisting ofstacks of flattened membrane cisternae perforated by numerous NPCs (Meier *et al.*, 1995), and vertebrate NPCs (MJ Matunis, G Blobel, Mol Biol Cell Abstr 1996, 7: 95a) may help to identify novel vertebrate NPC proteins. Sequencing of vertebrate genomes may further reveal genes with similarity to those encoding yeast NPC proteins.

In yeast, NPC proteins have been identified initially using antibodies raised against vertebrate NPC proteins or yeast nuclear matrix enriched fractions (Table 1.2). Among these proteins are Nsp1p (Hurt, 1988), Nup1p (Davis and Fink, 1990), Nup2p (Loeb *et al.*, 1991) and the related Nup49p, Nup100p, and Nup116p (Wente *et al.*, 1992). The isolation of yeast genes has allowed the design of genetic screens based on synthetic lethality. Synthetic lethality screens are based on a loss of function as described in the following. Although yeast strains with a mutation in a single NPC gene

are viable, the combination of this mutant gene with a second mutant component results in cell death. The second component may have either overlapping function or physically interacts with the product of the first mutant gene (for review, see Doye and Hurt, 1995). Synthetic lethality screens using temperature sensitive mutants led to the identification of most of the known yeast NPC genes including *NUP49, NUP57, NUP82, NUP84, NUP85, NIC96, NUP116, NUP133, NUP145, NUP170*, and *SRP1*. In addition, these screens have identified several proteins that function in nucleocytoplasmic transport mechanisms and associate with the NPC but are not truly part of the NPC (for review, see Doye and Hurt, 1997). Other genetic screens have been based on specific transport functions such as poly(A)$^+$ RNA export and nuclear protein localization and have identified novel NPC genes as well as several NPC genes that are also found by synthetic lethality screens (see Table 1.2; for review, see Doye and Hurt, 1997). The isolation of highly-enriched yeast NPCs has provided peptide sequence information and allowed the subsequent cloning of genes encoding Nup82p (Hurwitz and Blobel, 1995), Nup120p (Aitchison *et al.*, 1995), Nup159p (Kraemer *et al.*, 1995), Nup188p (Nehrbass *et al.*, 1996). The isolation of yeast NPCs will further contribute to the identification of NPC proteins. In addition, the sequencing of the entire yeast genome will allow identification of novel NPC proteins based on sequence similarity to known NPC proteins.

Table 1.2

Functional domains and other characteristics of yeast NPC proteins.
See text for abbreviations.

Table 1.2 Functional domains and other characteristics of yeast NPC proteins.

Year Protein (Mass)	Motifs	Synthetic lethal with	Physical Interaction	Mutant phenotype
1990				
Nup1p (113 kD)	FXFG	nup2,nup82, nup170, srp1	Srp1p, Nup170p	lethal defects in protein import defects in RNA export NE detachment
Nsp1p (86 kD)	FXFG coiled coil	nup49,nup57,nic96, nup82,nup116,nup145 nup133,nup188,nup84, nup85	Nup49p/Nup57p/Nic96p, Nup82p	lethal defects in protein import (docking step) decrease in NPC density
1992				
Nup49p (49 kD)	GLFG coiled coil	nsp1, nup133	Nsp1p/Nup57p/Nic96p	lethal defects in protein import (translocation step) defects in RNA export
Nup100p (100 kD)	GLFG RNA binding similar to Nup116p+Nup145p	nup116		no phenotype
Nup116p (116 kD)	GLFG	nsp1, nup100, nup145	Kap95	temperature sensitive defects in RNA export membrane seal over NPC
1993				
Nup2p (78 kD)	FXFG Ran-binding	nup1, nsp1, srp1	Srp1p, Ran/TC4	no phenotype

Table 1.2 (cont'd)

Nic96p (96 kD)	coiled coil	nsp1,pom152,nup188	Nsp1p/Nup49p/Nup57p, Pom152p,Nup188p	lethal defects in protein import decrease in pore density
1994				
Pom152p (152 kD)	integral membrane	nic96,nup188,nup170	Nic96p, Nup188p	no phenotype
Nup145p/ Rat10p (145 kD)	GLFG RNA-binding similar to Nup100p+Nup116p	nsp1,nup85,nup100, nup116	Nup84p complex	lethal defects in protein import defects in RNA export NPC herniation+clustering
Nup133p/ Rat3p (133 kD)	-	nsp1,nup49,nup120, nup85	ND	temperature sensitive defects in RNA export NPC clustering
Nup57p (57 kD)	GLFG coiled coil	nsp1	Nsp1p/Nup49p/Nic96p	lethal
1995				
Nup82p/ Nle4p (82 kD)	coiled coil	nsp1,nup1,nup157 nup170	Nsp1p	lethal defects in RNA export
Nup170p/ Nle3p (170 kD)	-	nup157,nup188,nup82 nup1,nup2,pom152	Nup1p	no phenotype
Nup157p (157 kD)	-	nup170,nup1,nup82, nup188	ND	no phenotype
Nup159p/ Rat7p (159 kD)	FXFG	nup85,nup120	ND	lethal defects in RNA export NPC clustering

Table 1.2 (cont'd)

Nup120p/ Rat2p (120 kD)	coiled coil Leu zipper	nup133,nup188,pom152 nup170,nup159	Nup84p subcomplex	temperature sensitive defects in RNA export NPC clustering
1996				
Nup84p (84 kD)	-	nsp1,nup85,nup120	Nup84p subcomplex	temperature sensitive defects in RNA export NPC clustering
Nup85p/ Rat9p (85 kD)	-	nsp1,nup133,nup145 nup159,nup188	Nup84p subcomplex	temperature sensitive defects in RNA export NPC clustering NE detachment
Nup188p (188 kD)	-	nsp1,nic96,pom152 nup170	Nic96p, Pom152p	NE herniation

4. Functions of nuclear pore complex proteins

Considering the structural complexity of the NPC, NPC proteins are likely to play a role in a variety of different cellular mechanisms. Due to our limited knowledge on the NPC, some of these may not be apparent at this point. On the other hand, NPC proteins are likely to function in nucleocytoplasmic transport, NPC assembly and NPC insertion into the NE. Because of the availability of assays, most research has focussed on understanding the role of NPC proteins in transport. To understand this role, it is necessary to summarize the current knowledge on the different transport mechanisms and the soluble factors involved in these processes. Several excellent reviews have been written and provide detailed information on this topic (Hicks and Raikhel, 1995; Görlich and Mattaj, 1996; Görlich, 1997; Nakielny and Dreyfuss, 1997).

4.1. Nucleocytoplasmic transport in vertebrates and yeast

Macromolecules such as proteins and RNA species move not only into, but also out of the nucleus. These active processes require the presence of specific transport signals. In general, active transport is energy dependent and saturable indicating a receptor-mediated process. To date, most is known about the nuclear import machinery of proteins that contain "classical" nuclear localization signals (NLSs) (for review, see Görlich and Mattaj, 1996; Görlich, 1997). The NLSs consist of one or more clusters of basic amino acids that are found within the nuclear protein (for review, see Garcia-Bustos *et al.*, 1991). This import process will be referred to as "NLS-dependent" throughout this manuscript. The development of an animal *in vitro* import system, in which the plasma membrane is selectively permeabilized by digitonin, has aided in the

identification and characterization of the requirements for NLS-dependent import (Adam *et al.*, 1990, 1992). Import of NLS-bearing proteins can be divided into two steps (Figure 1.2). The first step is binding at the NPC, a process dependent on the presence of cytosolic components. The second step is translocation through the NPC, a step that requires energy in form of GTP and is blocked by cold treatment (4°C) or WGA. In addition to several potential regulators, three soluble proteins are required for import, namely importin α, importin ß, and the small nuclear GTPase Ran. Based on studies by many different laboratories, the following model has emerged (for review, see Hicks and Raikhel, 1995; Görlich and Mattaj, 1996; Görlich, 1997). The receptor complex, consisting of importin α and importin ß, binds to import substrate in the cytoplasm and forms a trimeric import complex. While importin α provides the NLS-binding site, importin ß serves as an adaptor protein for docking of the complex at the NPC (Figure 1.2). Movement of the trimeric complex through the NPC requires GTP hydrolysis by Ran and appears to be facilitated by the regulator nuclear transport factor 2. However, the mechanism that drives the trimeric complex through the NPC is not fully understood. Once the trimeric complex reaches the nucleoplasm, importin ß is released due to direct binding of Ran-GTP to importin ß. Because importin α alone has a lower affinity to the NLS than the importin α-ß heterodimer, the import substrate is freed from the complex. Both, importin α and ß are then recycled separately to the cytoplasm to start a new round of protein import.

In yeast, the import mechanism is comparable to that described above, and in general, similar requirements for NLS-dependent protein import have been reported for this organism (Schlenstedt *et al.*, 1993). Furthermore, the yeast proteins

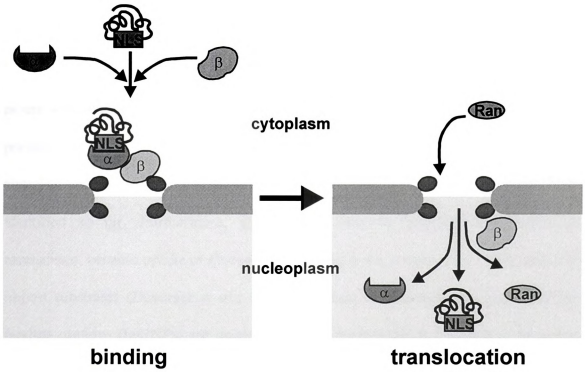


Figure 1.2

Schematic diagram of the NLS-dependent nuclear import machinery.

Srp1p, Kap95p, and Gsp1 have been identified as homologs of importin α, importin ß and Ran, respectively (for references, see Goerlich, 1997), indicating that at least the basic mechanism is conserved between animals and yeast. It should be noted that WGA does not block protein import into yeast nuclei (Kalinich and Douglas, 1989).

It seems likely that a variety of transport pathways exist to accommodate the need to import diverse RNA and proteins into the nucleus. Much less, however, is known about transport of molecules that do not contain classical NLSs. After assembly in the cytoplasm, the import of small nuclear RNA-protein complexes requires the presence of a signal that resides in the RNA as well as the protein portion of the complex (Mattaj *et al.*, 1993). Factors that facilitate this process have not been identified so far. Furthermore, glycoproteins may be imported by a different mechanism, because uptake of glycosylated proteins is not competed by NLS-containing import substrates (Duverger *et al.*, 1995). A subset of heterogeneous nuclear RNA-binding proteins (hnRNPs) are imported into the nucleus due to a stretch of 38 amino acids with no similarity to classical NLSs (Siomi and Dreyfuss, 1995). This amino acid portion is called the M9 domain. While translocation of hnRNP is mediated by the GTPase Ran, as found for the NLS-dependent pathway, the M9 domain is recognized by a different import receptor called transportin (Pollard *et al.*, 1996). Transportin is distantly related to importin ß, and the yeast protein Kap104p appears to be a homolog of the vertebrate transportin (Aitchison *et al.*, 1996) suggesting a conservation of this mechanism in eukaryotes.

Though the above mentioned studies have shed some light on the different nuclear import machineries, little is known about the molecular mechanism of

movement out of the nucleus. The recent identification of nuclear export signals (NES) that are present in shuttling proteins and of receptors that bind to NES and facilitate export may provide insight into the different export pathways (reviewed by Nakielny and Dreyfuss, 1997; Ullman *et al.*, 1997). Several NES-containing proteins associate with different classes of RNA and facilitate export of these RNAs. Interestingly, RNA viruses that replicate in the host nucleus (including HIV and influenza virus) appear to make use of the cellular export machinery to move viral RNA out of the host nucleus (Whittaker *et al.*, 1996). Further analysis of the viral RNA export machinery may provide important information on the molecular mechanism of cellular RNA export. Evidence for a link between the import and export pathways come from a recent report, in which importin α is found to accompany the complex of uridine rich snRNA and cap-binding proteins, that bind to the methylated cap at the 5' end of the U snRNA, during export from the nucleus. In the cytoplasm, importin α is then released from the complex by importin ß, making it available for the NLS-dependent import pathway. Furthermore, Ran has been recently implicated in the export of different classes of RNA (Izaurralde *et al.*, 1997), and therefore this small GTPase appears to be involved in both, nuclear import and export.

4.2. Nucleocytoplasmic transport in plants

In plants, several endogenous proteins are imported into the plant nucleus by virtue of "classical" NLSs that are similar to those identified in other organisms (for reviews, see Raikhel, 1992; Hicks and Raikhel, 1995). Furthermore, pathogenic bacteria and viruses seem to employ the plant NLS-dependent import machinery to

deliver their proteins or nucleic acids to the plant nucleus (for review, see Sheng and Citovsky, 1996; Carrington *et al.*, 1996). So far, no other classes of plant import and export signal have been reported. It is reasonable to assume that other transport pathways may exist in plants, and some of these are likely to be similar to those identified in other eukaryotes.

Because only the classical NLSs have been characterized in plants, research has focussed on the NLS-dependent import machinery. Plant homolog to importin α (Hicks *et al.*, 1996) and to Ran (Ach and Gruissem, 1994; Merkle *et al.*, 1994) have been isolated recently. Consistent with a role in NLS-dependent import, *Arabidopsis* importin α binds NLSs *in vitro* (Smith *et al.*, 1997). Unlike vertebrate importin α, it is not only found as a soluble protein in the cyto- and nucleoplasm, but shows tight interaction with purified nuclei (Smith *et al.*, 1997). The significance of this is observation, however, not understood. A direct involvement of the plant Ran homologs in protein import has not been reported. The plant Ran may play a role in this process because *in vitro* import assays that have been recently developed in plants demonstrate that NLS-dependent protein import requires energy in the form of GTP (Hicks *et al.*, 1996; Merkle *et al.*, 1996; Zupan *et al.*, 1996). Interestingly, plant protein import is slowed significantly at 4°C, but is not completely blocked, as frequently reported in animals (see above). This observation may reflect the adaptation of plants to low temperatures at which they are often exposed in nature. In plants, the lectin WGA does not inhibit protein import of smaller substrates (100-150 kD; Hicks *et al.*, 1996; Merkle *et al.*, 1996), but blocks transport of large import substrate such as an antigen-antibody complex (Harter *et al.*, 1994) or a proteins-single stranded complex of 5000 kD (Zupan *et al.*, 1996). As

discussed in Chapter 2 in detail and proposed for vertebrates, WGA may inhibit transport by simply blocking the NPC, and, thus, its effect may be non-specific.

Taken together, NLS-containing proteins in plants appear to utilize a similar import machinery as described in vertebrates and yeast (see above), but clearly differences in the import process exist.

4.3. Functional domains of NPC proteins and their significance on function

As mentioned above, the genes of approximately one third of at least 100 different NPC proteins have been isolated, mainly in yeast and vertebrates. These studies have permitted to compare the known sequences and to look for structural features within these NPC proteins that may provide clues to their possible roles.

O-GlcNAc modification

A family of vertebrate NPC proteins are modified with single O-GlcNAc residues (Table 1.1). Several studies have demonstrated that the O-GlcNAc NPC proteins play an important role in transport. Both protein import and RNA export are inhibited by agents that bind to these glycoproteins, such as WGA and antibodies made against this protein family (Finlay *et al.*, 1987; Dabauvalle *et al.*, 1988; Featherstone *et al.*, 1988; Newmeyer and Forbes, 1988; Terns and Dahlberg, 1994). Furthermore, reconstituted *Xenopus* oocyte nuclei depleted of O-GlcNAc proteins are not capable of forming transport-competent nuclei and NPCs (Dabauvalle *et al.*, 1990; Finlay and Forbes, 1990; Finlay *et al.*, 1991; Miller and Hanover, 1994). In particular, p62, p56,

p54, and p45 that form the p62 subcomplex are required for proteins import (Dabauvalle *et al.*, 1988; Hu *et al.*, 1996).

The exact role of the O-GlcNAc modification is not known (for reviews, see Haltiwanger *et al.*, 1992; Hart, 1997). It has been shown, however, that the O-GlcNAc residues do not function in nucleocytoplasmic transport or in NPC assembly (Miller and Hanover, 1994), indicating that the protein portion rather than the glycosylation is important for the transport mechanism. Thus, transport inhibition by WGA may be due to blockage of the NPC by sterical hinderance. In addition, yeast nuclei are able to transport substrates across the NPC, although yeast NPC proteins do not appear to be modified by O-GlcNAc (Rout and Wente, 1994).

FXFG and GLFG repeat motifs

All known O-GlcNAc NPC proteins contain a short amino acid sequence motif, called FXFG (i.e. single amino acid code, in which X represents any amino acid; Tables 1.1). This motif is present repetitively within the same subdomain in which the O-GlcNAc modification is also found (for reviews, see Wente and Rout, 1994; Davis, 1995). The FXFG motif is not exclusive to the O-GlcNAc NPC proteins, but is also present in the vertebrate Nup358p and several yeast NPC proteins (Table 1.1 and 1.2). Recent evidence in vertebrate and yeast indicate that this motif may play an important role in transport. Antibodies specific to the FXFG domain block transport *in vivo* in *Xenopus* oocytes (Finlay *et al.*, 1987; Dabauvalle *et al.*, 1988; Featherstone *et al.*, 1988; Newmeyer and Forbes, 1988; Terns and Dahlberg, 1994). Using *in vitro* blot overlay assays, several FXFG-containing NPC proteins from vertebrate and yeast bind

to the importin ß domain of the NLS-dependent import receptor complex (Moroianu *et al.*, 1995; Radu *et al.*, 1995a; Radu *et al.*, 1995b; Rexach and Blobel, 1995). The authors conclude that these NPC proteins function as docking sites for the NLS-dependent protein import machinery. Furthermore, considering the redundancy of FXFG-containing proteins in the NPC, movement of NLS proteins may occur through the NPC via repeated disassociation-reassociation steps of the receptor complex with the FXFG-containing NPC proteins (Rexach and Blobel, 1995).

Another repeat motif, namely GLFG, is present in several yeast NPC proteins and the vertebrate Nup98p (Table 1.2; for reviews, see Wente and Rout, 1994; Davis, 1995; Doye and Hurt, 1997). In yeast, this motif serves as binding site for Kap95p, the yeast homolog of importin ß (Iovine *et al.*, 1995; Iovine and Wente, 1997). In contrast to the proposed role of FXFG motifs in protein import, the GLFG-containing NPC proteins appear to function in recycling Kap95p from the nucleus to the cytoplasm (Iovine and Wente, 1997).

Ran binding sites

The small GTPase Ran is part of the GTPase cycle that provides the energy for transport through the NPC. In search for proteins that interact with Ran, Nup2p and Nup358p have been identified by yeast two-hybrid, biochemical interaction, and screening of cDNA expression library (Dingwall *et al.*, 1995; Wu *et al.*, 1995; Yokoyama *et al.*, 1995; Saitoh *et al.*, 1996). These NPC proteins contain domains that are homologous to the Ran binding protein 1 domain (Table 1.1 and 1.2), and these domains are known to mediate interaction with Ran. Nup358p and Nup2p also interact

with importin ß (Radu *et al.*, 1995a; 1995b) and importin α (Belanger *et al.*, 1994), respectively, and thus, may function as central docking sites for the transport machineries.

DNA and RNA binding sites

Based on sequence comparison, some yeast and vertebrate NPC proteins contain domains that may allow interaction with DNA or RNA (Table 1.1 and 1.2). The vertebrate Nup153p and Nup358p have several Cys_2-Cys_2 type Zinc fingers that are proposed to bind DNA (Table 1.1). Nup153p can bind *E. coli* DNA *in vitro* and, thus, may recognizes specific DNA sequences to organize the genome and to "gate" genes being transcribed to the NPC (Sukegawa and Blobel, 1993). Recent *in vivo* studies also indicate a role of Nup153p in RNA export (Bastos *et al.*, 1996).

The yeast Nup100p, Nup116p, and Nup145p contain a domain that can bind homopolymeric RNA, but not ssDNA or dsDNA *in vitro* (Fabre *et al.*, 1994). A similar domain was observed in the vertebrate Nup98p (Powers *et al.*, 1997). Consistent with the presence of RNA binding domains, the yeast NPC proteins and Nup98p are in involved in RNA export *in vivo* (see below).

Protein-protein interacting domains

Several NPC proteins contain helical coiled-coil domains (Table 1.1 and 1.2). These domains may allow their arrangements into filamental structures similar to those observed in the NPC. In support of this hypothesis, Buss *et al.* (1994) showed that recombinant p62 self-assembles into stiff, rod-shaped structures of 35 nm length *in*

vitro. The coiled-coil domain of p62 also appears to facilitate binding to p54 (Buss and Stewart, 1995) and most likely to other proteins in the p62 subcomplex. A complex similar to the vertebrate p62 subcomplex appears to be present in yeast, and certain coiled-coil regions of Nsp1p appear to allow complex formation with Nup49p and Nup57p *in vitro* (Schlaich *et al.*, 1997).

The putative leucine zipper motifs present in Nup107p, Nup358p, and Nup120p may also be involved in protein-protein interaction (Radu *et al.*, 1994; Aitchison *et al.*, 1995; Wu *et al.*, 1995). This interaction may mediate NPC assembly or function as binding sites for transport substrate.

Transmembrane domains

Three known NPC proteins, gp210 (Greber *et al.*, 1990; Wozniak *et al.*, 1989), Pom121p (Hallberg *et al.*, 1993; Soderqvist and Hallberg, 1994) and Pom152p (Wozniak *et al.*, 1994), are transmembrane proteins present at the pore membrane (Table 1.1 and 1.2). It is conceivable that they facilitate anchoring of the NPC to the NE. In addition, they may function in NPC biogenesis during interphase and after mitosis. gp210 may also have some role in protein import into the nucleus, because microinjection of mRNA encoding antibodies against the ER lumenal domain of gp210 causes reduction of protein import into the nucleus (Greber and Gerace, 1992).

4.4. Localization of NPC proteins and significance on their function

To understand their functional roles in the NPC as a whole, it is necessary to determine the localization of the NPC proteins within the different NPC substructures. While the localization sites of most vertebrate NPC proteins have been determined (Table 1.1), relatively little is known about the localization of yeast NPC proteins. This may be attributable to the general lack of good fixation procedures in yeast necessary for EM studies.

In agreement with their proposed function as docking sites for the importin α/ß receptor complex, the vertebrate Nup214p/CAN and Nup358p are located at the cytoplasmic fibrils (Kraemer *et al.*, 1994; Wilken *et al.*, 1995; Wu *et al.*, 1995; Yokoyama *et al.*, 1995) and the yeast Nup159p is found at the cytoplasmic side of the yeast NPC (Kraemer *et al.*, 1995).

The p62 subcomplex appears to be mainly associated with the cytoplasmic and nucleoplasmic sides of the central spoke/transporter region (for a review, see Forbes, 1992; see Chapter 2 for more details). This is consistent with its proposed function in translocation rather than in initial binding (Forbes, 1992). These findings are supported by recent EM studies, in which nuclear import of an import substrate was followed through the NPC (Panté and Aebi, 1996). Sequential binding was first observed to cytoplasmic fibrils, and then to the cytoplasmic side of the central channel region. Nup98p and Nup153p are constituents of the nucleoplasmic basket (Sukegawa and Blobel, 1993; Powers *et al.*, 1995), and their localization appears to be in agreement with their proposed function in RNA export (Bastos *et al.*, 1996; Powers *et al.*, 1997). Several GLFG NPC proteins from yeast have a role in transport of RNA out of the

nucleus and in recycling Kap95p from the nucleus into the cytoplasm (Iovine *et al.*, 1995, Iovine and Wente, 1997). Localization studies are consistent with these functions, because α-GLFG antibodies localizes this protein family to the nucleoplasmic side of the NPC (Wente *et al.*, 1992). In particular, epitope tagged Nup49p is found in close proximity to the nucleoplasmic face of the NPC (Wente *et al.*, 1992). The yeast Nup188p appears to be located at both the cytoplasmic and the nucleoplasmic side of the NPC (Nehrbass *et al.*, 1996). Based on its localization, it may be one of the core components of the NPC and interacts with Pom152p (Nehrbass *et al.*, 1996). The latter is a transmembrane proteins and found at the pore membrane (Wozniak *et al.*, 1994). The vertebrate transmembrane NPC proteins, gp210 and Pom121p, are also located at the pore membrane and may anchor the NPC to the NE (Gerace *et al.*, 1982; Hallberg *et al.*, 1993).

4.5. Functional *in vivo* and *in vitro* analyses of NPC proteins

Even though localization and sequence comparisons of NPC proteins may provide insight into their possible roles, *in vitro* and *in vivo* assays are necessary to conclusively demonstrate the function of these proteins (see Table 1.1 and 1.2).

Xenopus laevis reconstitution assays have been useful in studying the role of vertebrate NPC proteins in nuclear import and NPC assembly (for review, see Forbes, 1992). After depletion of a subset of NPC proteins by WGA or specific antibodies, nuclei can be reconstituted that contain NPCs lacking these NPC proteins, and nuclear transport of import substrate is then examined. These assays showed that the presence of the p62 subcomplex is required in the NPC for protein import, but not for NPC

assembly (for a review, see Forbes, 1992). When Nup98p, another protein modified by O-GlcNAc, is immunodepleted, NPCs are formed in reconstituted nuclei that are able to import proteins (Powers *et al.*, 1995). Moreover, when α-Nup98p antibodies are microinjected into *Xenopus* oocyte nuclei, export of different RNA classes, but not protein import, is inhibited (Powers *et al.*, 1997; Table 1.1). In contrast, *in vitro* overlay blot assays have shown that Nup98p binds to importin ß, thus implicating Nup98p in protein import (Radu *et al.*, 1995b). Similar contradictory results have been obtained for the vertebrate Nup153p. Overexpression of Nup153p in mammalian cell culture leads to inhibition of RNA export, but no defect in proteins uptake is observed (Bastos *et al.*, 1996). *In vitro* overlay assays, however, suggest a role of Nup153p in protein import. The localization of Nup98p and Nup153p to the nucleoplasmic components of the NPC appear to be in agreement with a function of RNA export rather than protein import (see above). Recently, a gene knock-out of Nup214p/Can in transgenic mice revealed that this protein is required for protein import as well as poly(A)$^+$ RNA export (Van Deursen *et al.*, 1996). This *in vivo* approach may serve as an alternative strategy to study vertebrate NPC proteins in the future. As it becomes obvious from these studies, a combination of different assays may be used to conclusively resolve the functions of NPC proteins.

Yeast genetics provide an excellent tool to investigate the function of NPC proteins *in vivo* (for review, see Doye and Hurt, 1995; 1997). Yeast strains carrying mutations in or deletions of NPC proteins are examined for proper RNA export, protein import, or perturbations of NPC and NE structures.

In situ hybridization with fluorescent labelled oligo(dT) demonstrate that mutations in several yeast NPC genes led to accumulation in poly(A)$^+$ RNA in the nucleus (Table 1.2.). Strikingly, intranuclear RNA accumulation is often associated with morphological changes in yeast NE and distribution of NPC. For example, mutations in *NUP49*, *NUP116* and *NUP145* lead to sealed and herniated NPCs that do not allow export of RNA (Fabre *et al.*, 1994; Wente and Blobel, 1994). In other cases, the RNA export defect appears to coincide with NPC clustering. Several studies, however, have demonstrated that these two defects can be separated under certain conditions. While *nup133/rat3* null mutants show clustering as well as inhibition of RNA export, a partial deletion mutant strain shows clustering of NPCs, but is still capable of exporting RNA (Doye *et al.*, 1994). Further, a *rat7/nup159* mutant strain shifted to 37°C for one hour allows the reversion of NPC clustering, while export of mRNA is still inhibited (Gorsch *et al.* 1995). These aberrant NE and NPC structures may indicate a putative role of these NPC proteins in anchoring the NPC to intranuclear structures or to NPC proteins in the pore membrane.

Mutant strains of *nsp1, nup49*, and *nic96* accumulate NLS-containing import substrate in the cytoplasm implying a role of these NPC proteins in protein import (Nehrbass *et al.*, 1993; Doye *et al.*, 1994; Grandi *et al.*, 1995). The FXFG-containing Nup1p and Nup2p may also function in protein import. Based on yeast two hybrid and synthetic lethality screens, Nup1p and Nup2p interact with Srp1p, the yeast homolog of importin α (Belanger *et al.*, 1994), and *in vitro* assays further show that Nup1p and Nup2p interact with Kap95p, the second component of the import receptor complex (Rexach and Blobel, 1995). Thus, Nup1p and Nup2p may serve as docking sites for the

receptor complex. Taken together, studies on functional domains, localization and functions using *in vivo and in vitro* assays have given some insight into the roles of NPC proteins. More information is needed to resolve the exact function of some NPC proteins, and the development of new assays may be useful.

5. Statement of problem and attribution

The NPC is one of the largest proteinaceous macromolecules within the cell, and only recently, we have started to gain a better understanding of its structure, components, and functions.

When I started this project nearly five years ago, only seven genes encoding NPC proteins had been isolated from vertebrates and yeast, and little was known about their functions (Table 1.1. and 1.2.). In plants, our knowledge on the NPC was limited to morphological and few biochemical studies (see above and Chapter 2). At that time and during the following years, no plant homologs to known NPC proteins were identified in the expressed sequence tags (EST) databases. Further, antibodies against several vertebrate NPC proteins did not specifically crossreact with plant NPC proteins. Thus, we had to consider other means to identify plant NPC proteins.

The O-GlcNAc modification of several vertebrate NPC proteins has provided a useful tool to isolate these glycoproteins. Therefore, as an initial step to identify plant NPC proteins, we were interested in whether plant NPC proteins are also modified by GlcNAc. This sugar modification would enable us to purify the plant glycoproteins by lectin affinity chromatography to obtain peptide sequence information and subsequently to clone the corresponding genes.

References

Ach, R.A. and Gruissem, W. (1994). A small nuclear GTP-binding protein from tomato suppresses a *Schizosaccharomyces pombe* cell-cycle mutant. Proc Natl Acad Sci U S A **91**, 5863-5867.

Adam, S.A., Sterne-Marr, R., and Gerace, L. (1990). Nuclear protein import in permeabilized cells requires soluble cytoplasmic factors. J. Cell Biol. **111**, 807-16.

Adam, S.A., Sterne-Marr, R., and Gerace, L. (1992). Nuclear protein import using-digitonin-permeabilized cells. Meth. Enzym. **219**, 97-110.

Akey, C.W. (1995). Structural plasticity of the nuclear pore complex. J. Mol. Biol. **248**, 273-293.

Akey, C.W. and Rademacher, M. (1993). Architecture of the *Xenopus* nuclear pore complex revealed by three-dimensional cryo-electron microscopy. J. Cell Biol. **122,** 1-19.

Aitchison, J.D., Blobel, G., Rout, M.P. (1995). Nup120: a yeast nucleoporin required for NPC distribution and mRNA export. J. Cell Biol. **131,** 1659-1675.

Aitchison, J.D., Blobel, G., Rout, M.P. (1996). Kap104p: a karyopherin involved in the nuclear transport of messenger RNA binding proteins. Science 274, 624-627.

Bastos, R., Lin, A., Enarson, M., Burke, B. (1996). Targeting and function in mRNA export of nuclear pore complex protein Nup153. J. Cell Biol. **134,** 1141-1156.

Belanger, K.D., Kenna, M.A., Wei, S., and Davis, L.I. (1994). Genetic and physical interaction between Srp1 and nuclear pore complex proteins Nup1p and Nup2p. J. Cell. Biol. **126**, 619-630.

Buss, F., Kent, H., Stewart, M., Bailer, S.M., and Hanover, J.A. (1994). Role of different domains in self-association of rat nucleoporin p62. J. Cell Sci. **107**, 631-638.

Buss, F. and Stewart, M. (1995). Macromolecular interaction in the nucleoporin p62 complex of rat nuclear pores: binding of nucleoporin p54 to the rod domain of p62. J. Cell Biol. **128**, 251-261.

Carmo-Fonseca, M., Kern, H., and Hurt, E.C. (1991). Human nucleoporin p62 and essential yeast nuclear pore protein NSP1 show sequence homology and a similar domain organization. Eur. J. Cell Biol. **55**, 17-30.

Carrington, J.C., Kasschau, K.D., Mahajan, S.K., and Schaad, M.C. (1996). Cell-to-cell and long distance transport of viruses in plants. Plant Cell **8**, 1669-1681.

Cordes, V.I., Waizenegger, I., and Krohne, G. (1991). Nuclear pore complex glycoprotein p62 from *Xenopus laevis* and mouse: cDNA cloning and identification of its glycosylation region. Eur. J. Cell Biol. **55**, 31-47.

Dabauvalle, M.-C., Benevente, R., and Chaly, N. (1988). Monoclonal antibodies to a M_r 68,000 pore complex glycoprotein interfere with nuclear protein uptake in *Xenopus* oocytes. Chromosoma **97**, 193-197.

Dabauvalle, M.-C., Loos, K., and Scheer, U. (1990). Identification of a soluble precursor complex essential for nuclear pore assembly in vitro. Chromosoma **100**, 56-66.

Davis, L.I. (1995). The nuclear pore complex. Annu. Rev. Biochem. **64**, 865-896.

Davis, L.I., and Blobel, G. (1986). Identification and characterization of a nuclear pore complex protein. Cell **45**, 699-709.

Davis, L.I., and Fink, G.R. (1990). The NUP1 gene encodes an essential component of the yeast nuclear pore complex. Cell **61**, 965-978.

Dingwall, C. , Kandelslewis, S., and Seraphin, B. (1995). A family of Ran binding proteins that includes nucleoporins. Proc. Natl. Acad. Sci. USA **92**, 7525-7529.

Dingwall, C. and Laskey, R. (1992). The nuclear membrane. Science **258**, 942-947.

D'Onofrio, M., Starr, C.M., Park, M.K., Holt, G.D., Haltiwanger, R.S., Hart G.W. and Hanover, J.A. (1988). Partial cDNA sequence encoding a nuclear pore protein modified by O-linked N-acetylglucosamine. Proc. Natl. Acad. Sci. U S A **85**, 9595-9599.

Doye, V., and Hurt, E.C. (1995) Genetic approaches to nuclear pore structure and function. Trends in Genet. **11**, 235-241.

Doye, V., and Hurt, E. (1997) From nucleoporins to nuclear pore complexes. Curr. Opin. Cell Biol. **9**, 401-411.

Doye, V., Wepf, R., and Hurt, E.C. (1994). A novel nuclear pore protein Nup133p with distinct roles in poly(A)$^+$ RNA transport and nuclear pore distribution. EMBO J. **13**, 6062-6075.

Duverger, E., Pellerin-Mendes, C., Mayer, R., Roche, A.-C., and Monsigny, M. (1995). Nuclear import of glycoconjugates is distinct from the classical NLS pathway. J. Cell Sci. **108**, 1325-1332.

Fabre, E., Boelens, W.C., Wimmer, C., Mattaj, I.W., Hurt, E.C. (1994). Nup145p is required for nuclear export of mRNA and binds homopolymeric RNA *in vitro* via a novel conserved motif. Cell **78,** 275-289.

Featherstone, C., Darby, M.K., and Gerace, L. (1988). A monoclonal antibody against the nuclear pore complex inhibits nucleocytoplasmic transport of protein and RNA in vivo. J. Cell Biol. **107**, 1289-1297.

Finlay, D.R., Newmeyer, D.D., Price, T.M., and Forbes, D.J. (1987). Inhibition of in vitro nuclear transport by a lectin that binds to nuclear pores. J. Cell Biol. **104**, 189-200.

Finlay, D.R., Meier, E., Bradley, P., Horecka, J., and Forbes, D.J. (1991). A complex of nuclear pore proteins required for pore function. J. Cell Biol. **114**, 169-183.

Finlay, D.R., and Forbes, D.J. (1990). Reconstitution of biochemically altered nuclear pores: Transport can be eliminated and restored. Cell **60**, 17-29.

Forbes, D.J. (1992). Structure and function of the nuclear pore complex. Annu. Rev. Cell Biol. **8**, 495-527.

Franke, W.W. (1966). Isolated nuclear membranes. J. Cell Biol. **31**, 9-623.

Garcia-Bustos, J., Heitman, J. and Hall, M.N. (1991). Nuclear protein localization. Biochem. Biophys. Acta **1071**, 83-101.

Gerace, L. and Foisner, R. (1994). Integral membrane proteins and dynamic organization of the nuclear envelope. Trends in Cell Biology **4**, 127-131.

Gerace, L., Ottaviano, Y., and Kondor-Koch, C. (1982). Identification of a major polypeptide of the nuclear pore complex. J. Cell Biol. **95**, 826-837.

Goldberg, M.W., and Allen, T.D. (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.

Goldberg, M.W. and Allen, T.D. (1993).The nuclear pore complex: three-dimensional surface structure revealed by field emission, in-lens scanning electron microscopy, with underlying structure uncovered by proteolysis. J. Cell Sci. 106, 261-274.

Goldberg, M.W. and Allen, T. (1995). Structural and functional organization of the nuclear envelope. Curr. Opin. Cell Biol. **7**, 301-309.

Goldberg, M.W. and Allen, T.D. (1996). The nuclear pore complex and lamina: three-dimensional structures and interaction determined by field emission in-lens scanning electron microscopy. J. Mol. Biol. **257,** 848-865.

Görlich, D. (1997). Nuclear protein import. Curr. Opin. Cell Biol. **9,** 412-419.

Görlich, D., Mattaj, I.W. (1996). Nucleocytoplasmic transport. Science **271,** 1513-1518.

Gorsch, L.C., Dockendorff, T.C., Cole, C.N. (1995). A conditional allele of the novel-repeat-containing yeast nucleoporin RAT7/NUP159 causes both rapid cessation of mRNA export and reversible clustering of nuclear pore complexes. J. Cell Biol. **129,** 939-955.

Grandi, P., Schlaich, N., Tekotte, H., and Hurt, E.C. (1995). Functional interaction of Nic96p with a core nucleoporin complex consisting of Nsp1p, Nup49p and a novel protein Nup57p. EMBO J. **14**, 76-87.

Greber, U.F. and Gerace, L. (1992). Nuclear protein import is inhibited by an antibody to a lumenal epitope of a nuclear pore complex glycoprotein. J. Cell Biol. **116**, 15-30.

Greber, U.F., Senior, A., and Gerace, L. (1990). A major glycoprotein of the nuclear pore complex is a membrane-spanning polypeptide with a large lumenal domain and a small cytoplasmic tail. EMBO J. **9,** 1495-1502.

Hallberg, E., Wozniak, R.W., and Blobel, G. (1993). An integral membrane protein of the pore membrane domain of the nuclear envelope contains a nucleoporin-like region. J. Cell Biol. **122**, 513-521.

Haltiwanger, R.S., Kelly, W.G., Roquemore, E.P., Blomberg, M.A., Dong, L.-Y.D., Kreppel, L., Chou, T.-Y., and Hart, G.W. (1992). Glycosylation of nuclear and cytoplasmic proteins is ubiquitous and dynamic. Biochem. Soc. Transaction **20**: 264-69.

Hanover, J.A., Cohen, C.K., Willingham, M.C., and Park, M.K. (1987). O-linked *N*-acetylglucosamine is attached to proteins of the nuclear pore. J. Biol. Chem. **262**,-9887-9894.

Hart, G.W. (1997). Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. Annu. Rev. Biochem. **66,** 315-335.

Harter, K., Kircher, S., Frohnmeyer, H., Krenz, M., Nagy, F., and Schäfer, E. (1994). Light-regulated modification and nuclear translocation of cytosolic G-box binding factors in parsley. Plant Cell **6**, 545-559.

Hicks, G.R. and Raikhel, N.V. (1993). Specific binding of nuclear localization sequences to plant nuclei. Plant Cell **5**, 983-994.

Hicks, G.R., and Raikhel, N.V. (1995). Protein import into the nucleus: An integrated view. Annu. Rev. Cell Dev. Biol. **11,** 155-188.

Hicks, G.R., Smith, H.M.S., Lobreaux, S., and Raikhel, N.V. (1996). Nuclear import in permeabilized protoplasts from higher plants has unique features. Plant Cell **8,** 1337-1352.

Hinshaw, J.E., Carragher, B.O., and Milligan, R.A. (1992). Architecture and design of the nuclear pore complex. Cell **69**, 1133-1141.

Holt, G.D., Snow, C.M., Senior, A., Haltiwanger, R.S., Gerace, L., and Hart, G.W. (1987). Nuclear pore complex glycoproteins contain cytoplasmically disposed O-linked N-acetylglucosamine. J. Cell Biol. **104**, 1157-1164.

Hurwitz, M.E. and Blobel, G. (1995). NUP82p is an essential yeast nucleoporin required for poly(A)$^+$ RNA export. J. Cell Biol. **130,** 1275-1281.

Hu, T., Guan, T., Gerace, L. (1996). Molecular and functional characterization of the p62 complex, an assembly of nuclear pore complex glycoproteins. J. Cell Biol. **134,** 589-601.

Hurt, E.C. (1988). A novel nucleoskeletal-like protein located at the nuclear periphery is required for the life cycle of *Saccharomyces cerevisiae.* EMBO J. **7**, 323-4334.

Hurwitz, M.E. and Blobel G. (1995). NUP82 is an essential yeast nucleoporin required for poly(A)+ RNA export. J. Cell Biol. **130**, 1275-1281.

Iovine, M.K., Watkins, J.L., Wente, S.R. (1995). The GLFG repetitive region of the nucleoporin Nup116p interacts with Kap95p, an essential yeast nuclear import factor. J. Cell Biol. **131,** 1699-1713.

Iovine, M.K. and Wente, S.R. (1997). A nuclear export signal in Kap95p is required for both recycling the import factor and interaction with the nucleoporin GLFG repeat regions of Nup116p and Nup100p. J. Cell Biol. **1374,** 797-811.

Izaurralde, E., Kutay, U., Mattaj, I.W., and Görlich, D. (1997). The asymmetric distribution of the constituents of the Ran system is essential for transport in and out of the nucleus. EMBO J. , in press.

Jarnik, M. and Aebi, U. (1991). Towards a more complete 3D structure of the nuclear pore complex. J. Struc. Biol. **107**, 291-308.

Jordan, E.G., Timmis, J.N., and Trewavas, A.J. (1980). The plant nucleus. In Tolbert,N.E. (ed.), The Biochemistry of Plants, Academic Press, Inc., Vol. 1, pp. 490-588.

Kalinich, J.F., and Douglas, M.G. (1989). In vitro translocation through the yeast nuclear envelope: signal dependent transport requires ATP and calcium. J. Cell Biol. **264**,-17979-17989.

Kraemer, D.C., Strambio-de-Castilla, C., Blobel, G., Rout, M.P. (1995). The essential yeast nucleoporin NUP159 is located on the cytoplasmic side of the nuclear pore complex and serves in karyopherin-mediated binding of transport substrates. J. Biol. Chem. **270,** 19017-19021.

Kraemer, D., Wozniak, R.W., Blobel, G., and Radu, A. (1994). The human CAN protein, a putative oncogene product associated with myloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm. Proc. Natl. Acad. Sci. USA **91**, 1519-1523.

Loeb, J.D., Davis, L.I. and Fink, G.R. (1993). NUP2, a novel yeast nucleoporin, has functional overlap with other proteins of the nuclear pore complex. Mol. Biol. Cell **4,** 209-222.

Mattaj, I.W., Boelens, W., Izaurralde, E., Jarmolowski, A., and Kambach, C. (1993). Nucleocytoplasmic transport and snRNP assembly. Mol. Biol. Reports **18**, 79-83.

Meier, E., Miller, B.R., Forbes, D.J. (1995). Nuclear pore complex assembly studied with a biochemical assay for annulate lamellae formation. J. Cell Biol. **129,** 1459-1472.

Merkle, T., Haizel, T., Matsumoto, T., Harter, K., Dallmann, G., and Nagy, F. (1994). Phenotype of the fission yeast cell cycle regulatory mutant pim1-46 is suppressed by a tobacco cDNA encoding a small, Ran-like GTP-binding protein. Plant J 6, 555-565.

Merkle, T., Leclerc, D., Marshallsay, C., and Nagy, F. (1996). A plant *in vitro* system for the nuclear import of proteins. Plant J. **10,** 1177-1186.

Miller, M.W., and Hanover, J.A. (1994). Functional nuclear pores reconstituted with ß-1,4 galactose-modified *O*-linked *N*-acetylglucosamine glycoproteins. J. Biol. Chem. **269**, 9289-9297.

Moroianu, J., Hijikata, M., Blobel, G., and Radu, A. (1995). Mammalian karyopherin α 1 ß and α 2 ß heterodimers: α 1 or α 2 subunit binds nuclear localization signal and ß subunit interacts with peptide repeat-containing nucleoporins. Proc. Natl. Acad. Sci. USA **92,** 6532-6536.

Nakielny S. and Dreyfuss, G. (1997). Nuclear export of proteins and RNAs. Curr. Opin. Cell Biol. **9**, 420-429.

Nehrbass. U., Fabre, E., Dihlman, S., Herth, W., and Hurt, E.C. (1993). Analysis of nucleo-cytoplasmic transport of a thermosensitive mutant of nuclear pore protein Nsp1p. Eur. J. Cell Biol. **62**, 1-12.

Nehrbass, U., Rout, M.P., Maguire, S., Blobel, G., Wozniak, R.W. (1996). The yeast nucleoporin Nup188p interacts genetically and physically with the core structures of the nuclear pore complex. J. Cell Biol. **133,** 1153-1162.

Newmeyer, D.D., and Forbes, D.J. (1988). Nuclear import can be separated into distinct steps in vitro: Nuclear pore binding and translocation. Cell **52**, 641-653.

Panté, N., and Aebi, U. (1993). The nuclear pore complex. J. Cell Biol. **122**, 977-984.

Panté, N., and Aebi, U. (1996). Sequential binding of import ligands to distinct nucleopore regions during their nuclear import. Science **273,** 1729-1732.

Park, M.K., D'Onofrio, M., Willingham, M.C., and Hanover, J.A. (1987). A monoclonal antibody against a family of nuclear pore proteins (nucleoporins): O-linked acetylglucosamine is part of the immunodeterminant. Proc. Natl. Acad. Sci. USA **84**, 6462-6466.

Pollard, V.W., Michael, W.M., Naldalny, B., Slomi, M.C., Wang, P., Dreyfuss, G. (1996). A novel receptor mediated nuclear protein import pathway. Cell **88,** 985-994.

Powers, M.A., Macaulay, C., Masiarz, F.R., and Forbes, D.F. (1995). Reconstituted nuclei depleted of a vertebrate GLFG nuclear pore complex protein, p97, import but are defective in nuclear growth and replication. J. Cell Biol. **128**, 721-736.

Powers, M.A., Forbes, D.J., Dahlberg, J.E. Lund, E. (1997). The vertebrate GLFG nucleoporin, Nup98, is an essential component of multiple RNA export pathway. J. Cell Biol. **136,** 241-250.

Radu, A., Blobel, G. and Wozniak, R.W. (1993). Nup155 is a novel nuclear pore complex protein that contains neither repetitive sequence motifs nor reacts with WGA. J. Cell Biol. **121**, 1-9.

Radu, A., Blobel, G., and Wozniak, R.W. (1994). Nup107 is a novel nuclear pore complex protein that contains a leucine zipper. J. Biol. Chem. **269**, 17600-17605.

Radu, A., Blobel, G., and Moore, M.S. (1995a). Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins. Proc. Natl. Acad. Sci. USA **92**, 1769-1773.

Radu, A., Moore, M.S., and Blobel, G. (1995b). The peptide repeat domain of nucleoporin Nup98 functions as a docking site in transport across the nuclear pore complex. Cell **81**, 215-222.

Raikhel, N.V. (1992). Nuclear targeting in plants. Plant Physiol. **100**, 1627-1632.

Reichelt, R., Holzenburg, A., Buhle, E.L., Jarnik, M., Engel, A., and Aebi, U. (1990). Correlation between structure and mass distribution of the nuclear pore complex, and of distinct pore complex components. J. Cell Biol. **110**, 883-894.

Rexach, M. and Blobel, G. (1995). Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. Cell **83**, 683-692.

Roberts, K., and Northcote, D.H. (1970). Structure of the nuclear pore in higher plants. Nature **228**, 385-386.

Rout, M.P., and Blobel, G. (1993). Isolation of the yeast nuclear pore complex. J. Cell Biol. **123**, 771-783.

Rout, M.P., and Wente, S.R. (1994). Pores for thought: nuclear pore complex proteins. TICB **4**, 357-365.

Saitoh, H., Cooke, C.A., Burgess, W.H., Earnshaw, W.C., and Dasso, M. (1996). Direct and indirect association of the small GTPase Ran with nuclear pore proteins and soluble transport factors: studies in *Xenopus laevis* egg extract. Mol. Biol. Cell **7,** 1319-1334.

Schlaich, N.L., Haner, M., Lustig, A., Aebi, U. and Hurt, E.C. (1997). In vitro reconstitution of a heterotrimeric nucleoporin complex consisting of recombinant Nsp1p, Nup49p, and Nup57p. Mol. Biol. Cell **8**, 33-46.

Schlenstedt, G., Hurt, E., Doye, V. and Silver, P.A. (1993). Reconstitution of nuclear protein transport with semi-intact yeast cells. J. Cell Biol. **123,** 785-798.

Sheng, J. and Citovsky, V. (1996). Agrobacterium-plant cell DNA transport: have virulence proteins, will travel. Plant Cell **8,** 1699-1710.

Siomi, H. and Dreyfuss G. (1995). A nuclear localization domain in the hnRNP A1 protein. J. Cell Biol. **129,** 551-560.

Smith, H.M.S., Hicks, G.R., and Raikhel, N.V. (1997). Importin α from *Arabidopsis thaliana* is a nuclear import receptor that recognizes three classes of import signals. Plant Physiol. **114,** 411-417.

Snow, C.M., Senior, A., and Gerace, L. (1987). Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. J. Cell Biol. **104**, 1143-1156.

Soederqvist, H. and Hallberg, E. (1994). The large C-terminal region of the integral pore membrane proteins, POM121, is facing the nuclear pore complex. Eur. J. Cell Biol. **64**, 184-191.

Sukegawa, J., and Blobel, G. (1993). A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm. Cell **72**, 29-38.

Terns, M.P. and Dahlberg, J.E. (1994). Retention and 5' cap trimethylation of U3 snRNA in the nucleus. Science **264**, 959-961.

Ullman, K.C., Powers, M.A., and Forbes, D.J. (1997). Nuclear export receptors: from Importin to exportin. Cell **90**, 967-970.

Van Deursen, J., Boer, J. Kasper, L., Grosveld, G. (1996). G2 arrest and impaired nucleocytoplasmic transport in mouse embryos lacking the proto-oncogene CAN/Nup214. EMBO J. **15,** 5574-5583.

Wente, S.R. and Blobel, G. (1994). NUP145 encodes a novel yeast glycine-leucine-phenylalanine-glycine (GLFG) nucleoporin required for nuclear envelope structure. J. Cell Biol. **125**, 955-969.

Wente, S.R., Rout, M.P., Blobel, G. (1992). A new family of yeast nuclear pore complex proteins. J. Cell Biol. **119,** 705-723.

Whittaker, G., Bui, M., and Helenius A. (1996). Nuclear trafficking of influenza virus ribonucleoproteins in heterokaryons. J. Virol. **70,** 2743-2756.

Wilken, N., Senecal. J.-L., Scheer, U., Dabauvalle, M.-C. (1995). Localization of the Ran-GTP binding protein Ran-BP2 at the cytoplasmic side of the nuclear pore complex. Eur. J. Cell Biol. **68**, 211-219.

Wozniak, R.W., Bartnik, E., Blobel, G. (1989). Primary structure analysis of an integral membrane glycoprotein of the nuclear pore. J. Cell Biol. **108,** 2083-2092.

Wozniak, R.W., Blobel, G., and Rout, M.P. (1994). POM 152 is an integral protein of the pore membrane domain of the yeast nuclear envelope. J. Cell Biol. **125**, 31-42.

Wu, J., Matunis, M.J., Kraemer, D., Blobel, G., Coutavas, E. (1995). Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. J. Biol. Chem. **270,** 14209-14213.

Yokoyama, N., Hayashi, N., Seki, T., Panté, N., Ohba, T., Nishii, K., Kuma, K., Hayashida, T., Miyata, T., Aebi, U. et al. (1995). A giant nucleopore protein that binds Ran/TC4. Nature **376,** 184-188.

Zupan, J.R., Citovsky, V., and Zambryski P. (1996). Agrobacterium VirE2 protein mediates nuclear uptake of single-stranded DNA in plant cells. Proc Natl Acad Sci U S A **93**, 2392-2397.

Chapter 2

Plant nuclear pore complex proteins are modified by novel oligosaccharides with terminal N-acetylglucosamine

This chapter was originally published in:

Antje Heese-Peck, Robert N. Cole, Olga N. Borkhsenious,

Gerald W. Hart, Natasha V. Raikhel (1995). Plant Cell 7: 1459-1471

ABSTRACT

Only a small number of nuclear pore complex (NPC) proteins, mainly in vertebrates and yeast but none in plants, have been well characterized. As an initial step to identify plant NPC proteins, we examined whether NPC proteins from tobacco are modified by N-acetylglucosamine (GlcNAc). Using wheat germ agglutinin (WGA), a lectin that binds in plants specifically to GlcNAc, specific labeling was often found associated with or adjacent to NPCs. Nuclear proteins containing GlcNAc can be partially extracted by 0.5 M salt as shown by a WGA blot assay, and at least eight extracted proteins are modified by terminal GlcNAc as determined by *in vitro* galactosyltransferase assays. Sugar analysis indicates that the plant glycans with terminal GlcNAc differ from the single O-GlcNAc of vertebrate NPC proteins in that they consist of oligosaccharides that are larger in size than five GlcNAcs. Most of these appear to be bound to proteins via a hydroxyl group. This novel oligosaccharide modification may convey properties to the plant NPC that are different from those of vertebrate NPCs.

INTRODUCTION

The nuclear pore complex (NPC) is the site for nucleocytoplasmic transport of macromolecules such as proteins, T-DNA and different RNA species. The architecture and the major components of this complex have been extensively investigated in amphibians using different electron microscopic (EM) techniques (for review see Panté and Aebi, 1993). The NPC is a supramolecular structure of about 124 mD (Reichelt *et al.*, 1990) and exhibits an eightfold rotational symmetry. It consists of a cytoplasmic and a nucleoplasmic ring, and these rings are connected to the central channel complex via sets of spokes. It is thought that the spokes form eight aqueous channels of approximately 9 nm that allow passive diffusion of some ions and small molecules (Hinshaw *et al.*, 1992; Akey and Rademacher, 1993), whereas the central channel complex with a functional diameter of up to 26 nm is proposed to be involved in the specific transport of macromolecules (Feldherr *et al.*, 1984). Recently, other morphological peripheral structures of the NPC have been observed such as cytoplasmic fibrils that are attached to the cytoplasmic ring and the nuclear basket that extends from the nucleoplasmic ring into the nucleoplasm (for reviews see Akey, 1992; Panté and Aebi, 1993). In some species, the nuclear basket has been shown to be attached to a fibrous lattice termed the nuclear envelope lattice (Goldberg and Allen, 1992).

Based on scanning transmission EM mass analysis, the NPC is proposed to consist of at least 100 different proteins (Reichelt *et al.*, 1990). To date, only a small number of NPC proteins have been identified and characterized at the biochemical and molecular level. These are mainly from vertebrates and yeast, and none have been identified in plants. Little is known about the function of the NPC proteins in

nucleocytoplasmic transport, NPC organization, and assembly in any system studied (for reviews see Forbes, 1992; Fabre and Hurt, 1994; Rout and Wente, 1994; Hicks and Raikhel, 1995b). Of particular interest with respect to protein import is a family of vertebrate NPC glycoproteins that are modified by the addition of single O-linked *N*-acetylglucosamine (O-GlcNAc) residues (for reviews see Hart *et al.*, 1989; Forbes, 1992). This posttranslational modification is distinct from other protein sugar modifications; the sugar is an unmodified monosaccharide moiety whose addition occurs in the cytoplasm (for reviews see Hart *et al.*, 1989; Haltiwanger *et al.*, 1992). In vertebrate nuclei, the O-GlcNAc NPC proteins have been identified by their ability to bind to wheat germ agglutinin (WGA), a lectin that binds to GlcNAc and sialic acids (Goldstein and Hayes, 1978), and by monoclonal antibodies raised against nuclear fractions (Davis and Blobel, 1986; Finlay *et al.*, 1987; Hanover *et al.*, 1987; Holt *et al.*, 1987; Park *et al.*, 1987; Schindler *et al.*, 1987; Snow *et al.*, 1987). The most abundant O-GlcNAc bearing NPC protein, denoted p62, has been sequenced from several vertebrate species and its glycosylation studied in detail (Starr *et al.*, 1990; Carmo-Fonseca *et al.*, 1991; Cordes *et al.*, 1991; Cordes and Krohne, 1993).

Several studies have demonstrated that the O-GlcNAc NPC proteins are involved in protein import. Agents that bind to these glycoproteins such as WGA, antibodies that recognize this protein family and antibodies specific for p62 block nucleocytoplasmic transport (Finlay *et al.*, 1987; Yoneda *et al.*, 1987; Dabauvalle *et al.*, 1988; Featherstone *et al.*, 1988; Newmeyer and Forbes, 1988; Wolff *et al.*, 1988). Moreover, depletion of O-GlcNAc proteins from NPCs reconstituted in *Xenopus* extracts inhibits transport *in vitro* by preventing the formation of transport-competent nuclei and NPCs

(Dabauvalle *et al.*, 1990; Finlay and Forbes, 1990; Finlay *et al.*, 1991; Miller and Hanover, 1994). Studies by Sterne-Marr *et al.* (1992) suggest that a cytosolic factor essential for nuclear import binds to Nup153p, a mammalian O-GlcNAc NPC protein. The WGA binding protein, Nup98p, has also been implicated to function as a docking protein for cytosol-mediated binding of import substrate *in vitro* (Radu *et al.*, 1995a; 1995b). Although the family of O-GlcNAc proteins has been shown to be involved in import, the exact role of the O-GlcNAc sugar modification in nuclear transport is not fully understood (Miller and Hanover, 1994). Interestingly, the O-GlcNAc modification has not yet been observed in any yeast NPC protein (Rout and Wente, 1994). In addition, WGA has no effect on nuclear import in yeast (Kalinich and Douglas, 1989). These observations indicate that differences exist in nuclear transport and NPCs of different organisms.

In plants, little is known about the nuclear transport machinery (for review see Raikhel, 1992; Hicks and Raikhel, 1995b) or about NPC proteins and their function in nuclear transport. Although the plant NPC has been morphologically described for several decades (Roberts and Northcote, 1970, for review see Jordan *et al.*, 1980), almost no further characterization of plant NPC components has been accomplished. Scofield *et al.* (1992) reported that a 100 kD protein, identified using antibodies against the yeast NPC protein Nsp1p (Hurt, 1988), is enriched in nuclear matrix preparations of carrot suspension-cultured cells and located at the plant NPC. Recently, binding of nuclear localization signals (NLSs) was observed to a low affinity binding site that is proteinaceous and firmly associated with the nuclear envelope (NE) and NPCs of

isolated dicot and monocot nuclei (Hicks and Raikhel, 1993; Hicks *et al.*, 1995). Crosslinking studies under conditions similar to those used for NLS binding identified several polypeptides that bind specifically to functional NLSs. The affinities and the biochemical properties of these polypeptides are similar to those of the NLS binding site (Hicks and Raikhel, 1995a). These data indicate that in plants, components that recognize functional NLSs may be located at the plant NPC. Using an immunofluorescence *in vitro* nuclear import system, we have found that WGA does not inhibit protein import into the plant nucleus, although the lectin binds to the nuclear surface (Hicks and Raikhel, 1995b).

In the current study, we investigate the WGA binding sites at the nuclear surface in plants and demonstrate that proteins at the NPC of tobacco suspension-cultured cells are detected by WGA and therefore are modified by GlcNAc. The plant glycoproteins have unusual oligosaccharide modifications that are discussed and compared to those of vertebrate NPC proteins modified by a single O-GlcNAc.

METHODS

Materials

All chemicals were purchased from Sigma (St. Louis, MO) unless otherwise noted.

Cell culture and nuclear purification

Nicotiana tabacum suspension-cultured cells were maintained and subcultured, and protoplasts were prepared from 6- to 7-day cell cultures essentially as described by Bednarek *et al.* (1990). The purification of tobacco nuclei was based upon the methods of Wilmitzer and Wagner (1981) and Saxena *et al.* (1985). Isolation of nuclei in the presence of 0.01% Triton X-100 was performed as described by Hicks and Raikhel (1993), and the isolation of nuclei in the presence of 0.6% Triton X-100 was performed as follows: Protoplasts from two 250-mL cultures were pelleted at 50x*g* for 4 min at room temperature. All subsequent steps were performed at 4°C. All buffers containing Percoll were adjusted to a pH of 5.6. The following concentrations of protease inhibitors were present during all steps: 5 μg/mL pepstatin, 5 μg/mL leupeptin, 5 μm/mL aprotinin, 5 μm/mL aminocaproic acid, 0.4 mM phenylmethylsulfonyl fluoride. The pellet was resuspended in 50 mL of 0.01% Triton X-100 (Boehringer Mannheim, Indianapolis, IN) in nuclear isolation buffer (NIB; 10 mM 2(*N*-morpholino)-ethanesulfonic acid-KOH, pH 5.6, 0.2 M sucrose, 10 mM NaCl, 10 mM KCl, 2.5 mM EDTA, 2.5 mM DTT, 0.1 mM spermine, 0.5 mM spermidine) and placed on ice for 10 min. Protoplasts were ruptured by passage through a 21-gauge needle six times. The lysate was centrifuged for 4 min onto a cushion of 67% Percoll in NIB at 1000x*g*, and the interface material was diluted with 100 mL 0.6% Triton X-100 in NIB. The lysate

was incubated on ice for 20 min and filtered through a 20 μm nylon mesh, then divided and layered onto 6 step gradients consisting of 1 mL of 67% Percoll in NIB and 5 mL of 7.5% Percoll in NIB. The gradients were centrifuged at 1000x*g* for 4 min, and the interfaces between the Percoll phases were combined, diluted with 100 mL 0.6% Triton X-100 in NIB, and incubated on ice for 10 min. The step gradient was repeated, the interfaces combined and diluted with 50 mL 0.6% Triton X-100 in NIB and loaded onto another step gradient. The interfaces containing the purified nuclei were washed in nuclear storage buffer (20% glycerol in NIB) and pelleted at 750x*g* for 5 min. After another washing and pelleting step, the nuclei were quantitated as described in Hicks and Raikhel (1993) and stored at -80°C. Yields were 4 to $6x10^8$ nuclei.

Isolation and fractionation of nuclear proteins

Isolated nuclei were incubated with DNase I (Boehringer Mannheim, Indianapolis, IN) at a concentration of 30 units/10^6 nuclei in the presence of 5 mM Tris-HCl, pH 7.5, 1 mM $MgCl_2$ for 30 min at room temperature and pelleted at 12,000x*g* for 5 min. The DNase I treated nuclear pellet was resuspended in SDS-sample buffer or subjected to further extraction by incubation in a low salt + detergent buffer (LS/TR; 2% Triton X-100, 10% sucrose, 20 mM Tris-ethanolamine, pH 7.4, 20 mM KCl, 5 mM $MgCl_2$, 1 mM DTT) or in a high salt buffer (HS; 0.5 M NaCl, 5 mM $MgCl_2$, 20 mM Tris, pH 7.5) for 30 min on ice. Solubilized proteins were separated from the insoluble fraction by a 15 min centrifugation step (12,000x*g*) at 4°C and were precipitated with 10% TCA followed by an acetone wash. The samples were resuspended in SDS-sample buffer (Laemmli, 1970).

Protein blot analysis using WGA probes

Proteins solubilized in SDS-sample buffer were electrophoresed through SDS-PAGE and transferred to Immobilon-P membranes (Amersham, Arlington Heights, IL) according to standard methods (Harlow and Lane, 1988). Unless otherwise noted, blots were blocked overnight in 5% nonfat dry milk in Tris-buffered saline + 0.1% Tween 20 (TBST), incubated for 2-1/2 hrs with 1 μl/mL WGA-alkaline phosphatase (WGA-AP; EY-Laboratories, San Mateo, CA) or 1μg/mL WGA-horseradish peroxidase (WGA-HRP; EY-Laboratories, San Mateo, CA) in TBST, washed 2x in TBST and 2x in TBS for 10 min each, and developed according to standard methods (Harlow and Lane, 1988). Control blots were treated identically except that 2 mM chitotriose or 250 mM mannose were added during the incubation with the WGA probes. For increased sensitivity, WGA-HRP in combination with the chemiluminescence kit (Boehringer Mannheim, Indianapolis, IN) was used to detect proteins. In these experiments, blots were blocked overnight in 3% BSA in TBST, incubated with 0.5 μg/mL WGA-HRP in TBST, washed as described above, and developed according to the manufacturer's specifications. The monoclonal antibody mAb414 was a generous gift from Dr. L.I. Davis (Dept. of Genetics, Box 3646, Duke Univ. Medical Center, Durham, NC 27710, USA).

Electron microscopy

Purified tobacco nuclei ($5x10^5$) were diluted with binding buffer (50 mM Tris, pH 7.3, 25 mM KCl, 2.5 mM $MgCl_2$, 3 mM $CaCl_2$, 20% glycerol) to 10^5 nuclei and incubated with WGA-conjugated colloidal gold particles (16 μg/mL; 10 nm diameter; EY-

Laboratories, San Mateo, CA) for 20 min at room temperature. Control samples contained 5 mM chitotriose (EY-Laboratories, San Mateo, CA). The nuclei were briefly pelleted, fixed in 2% paraformaldehyde, 1% glutaraldehyde, 50 mM sodium phosphate buffer, pH 7.2, 500 mM sucrose for 1 hr at room temperature and then postfixed in 1% OsO_4 in 10 mM sodium phosphate buffer, pH 7.2, 50 mM sucrose for 1 hr at room temperature. Sections were stained and visualized as described in Hicks and Raikhel (1993).

Galactosyltransferase labeling assays

Bovine milk GlcNAc ß-1,4 galactosyltransferase (GalTF) was purchased from Sigma (St. Louis, MO) or Oxford GlycoSystem (Rosedale, NY). Prior to use, the GalTF from Sigma was autogalactosylated (Roquemore *et al.*, 1994) and stored at -20°C. The GalTF from Oxford GlycoSystem was used as provided by the manufacturer. Equivalent results were obtained with GalTFs from either source. DNase I treated tobacco nuclei were incubated with 0.5 M NaCl, 50 mM Hepes, pH 6.8 for 1 hr on ice and centrifuged at 12,000x*g* for 5 min at 4 °C. *in vitro* GalTF labeling of solubilized proteins was essentially done as described by Roquemore *et al.* (1994) using 100 mU of GalTF and 3 μCi UDP-[^{3}H]-galactose (1.0 mCi/mL; 17.3 Ci/mmol; Amersham, Arlington Heights, IL) in a 100 μL reaction. After incubation at 37°C for 3 hrs, aliquots were taken for protein size analysis on 7.5% SDS-PAGE. The gel was treated with Fluoro-Hance (Research Products International Corp., Mt. Prospect, IL) according to the manufacturer's specification and dried. Labeled proteins were visualized by fluorography using ReflectionA film (NEN Research Products, Boston, MA) at -80 °C.

The remaining samples were used for further carbohydrate analysis and separated from unincorporated label by desalting over a Sephadex G-50 (3 mL) gel filtration column. Radiolabeled proteins were pooled, lyophilized, resuspended in a small volume of deionized water, aliquoted for further analysis, and precipitated with 8 volume acetone for at least 6 hrs at -20°C.

For *on blot* GalTF labeling assay, proteins were solubilized as described for *in vitro* GalTF labeling assay. To lower the salt concentration, the sample was diluted with 50 mM Hepes, pH 6.8, and concentrated by Centriprep 10 Concentrator (Amicon, Beverly, MA) at 4 °C. SDS-PAGE sample buffer was added, and the sample was treated as described for protein blot analysis. Blots were blocked overnight with 3% BSA in TBS and equilibrated for 4x 10 min in labeling buffer (10 mM Hepes-NaOH, pH 7.3, 10 mM galactose, 5 mM $MnCl_2$). After incubation with 50 mU GalTF and 10 μCi UDP-[^{3}H]-galactose in labeling buffer (1 mL reaction) for 3 hrs at 37°C, blots were rinsed for 5 min in 10 mM EDTA, 1% SDS, washed 4x 10 min in labeling buffer, dried, sprayed with EN^{3}HANCE (NEN Research Products, Boston, MA), and dried. Radiolabeled proteins were visualized by fluorography as described above.

PNGase F treatment

Peptide:N-glycosidase F (PNGase F) treatment and Sephadex G-50 column chromatography of salt extracted, [^{3}H]galactose-labeled nuclear proteins and [^{3}H]galactose-labeled ovalbumin (used as a positive control) were performed as described in Roquemore *et al.* (1994) with the addition of the following protease inhibitors to the PNGase F reaction buffer: 0.2 mM phenylmethylsulfonyl fluoride,

2 μg/mL antipain, 10 U/mL aprotinin, 10 μg/mL benzamidine, 1 μg/mL leupeptin, 1 μg/mL pepstatin. 0.5% of fractions from Sephadex G-50 column were assayed for radioactivity by liquid scintillation counting. PNGase F resistant labeled peripheral nuclear proteins in the void volume (Vo) were lyophilized, resuspended in <300 μL ddH_2O and precipitated in 80% cold acetone.

ß-elimination of PNGase F resistant nuclear proteins

Acetone precipitated PNGase F resistant [^{3}H]galactose-labeled nuclear proteins were subjected to alkaline borohydride for 48 hrs and fractionated over a Sephadex G-50 column as described by Roquemore *et al.* (1994). One percent of the fractions from the Sephadex G-50 column were assayed for radioactivity by liquid scintillation counting. Released [^{3}H]galactose-labeled saccharides in the inclusion volume (Vi) were lyophilized and resuspended in 1 mL distilled H_2O. The SDS in the resuspended samples was precipitated by adding 300 μL of 20% KCl, incubating on ice for 15 min and centrifuging at 10,000x*g* for 15 min at 4 °C. The resulting supernatant was fractionated on a 1.5x 200 cm Toyopearl HW 40 (TosoHaas, Philadelphia, PA) column equilibrated with 200 mM ammonium acetate and 10% ethanol at 55°C. One percent dextran (40K) and 1% galactose were added to the samples for internal standards. The Vo (dextran) and the Vi (galactose) were determined by phenol-sulfuric acid assay (Dubois *et al.*, 1956). The Toyopearl HM 40 column was calibrated with [^{3}H]galactose-$(GlcNAc)_n$-GlcNAcitol oligosaccharides, where n=0-3. Five percent of fractions from the Toyopearl column were assayed for radioactivity by liquid scintillation.

The [^{3}H]galactose-labeled oligosaccharides released by ß-elimination were analyzed for

the presence of peptide fragments or single amino acids. The labeled oligosaccharides were concentrated by lyophilization, passed over a Sep-Pak C18 column (Waters, Milford, MA), equilibrated with 0.1% trifluoroacetic acid (TFA) and eluted with 60% acetonitrile in 0.1% TFA. Both, the oligosaccharides that passed freely through the Sep-Pak column and those that were eluted with acetonitrile, were treated with pronase according to Fukuda (1989) and re-fractionated on the 1.5 x 200 cm Toyopearl HW 40 column. These two groups of oligosaccharides were also dansylated to detect primary amines according to Tapuhi *et al.* (1981), and the products were analyzed by reverse phase HPLC on a Microsorb-MV C18 column (Rainin Instrument Comp., Worburn, MA) according to Kaneda *et al.* (1982).

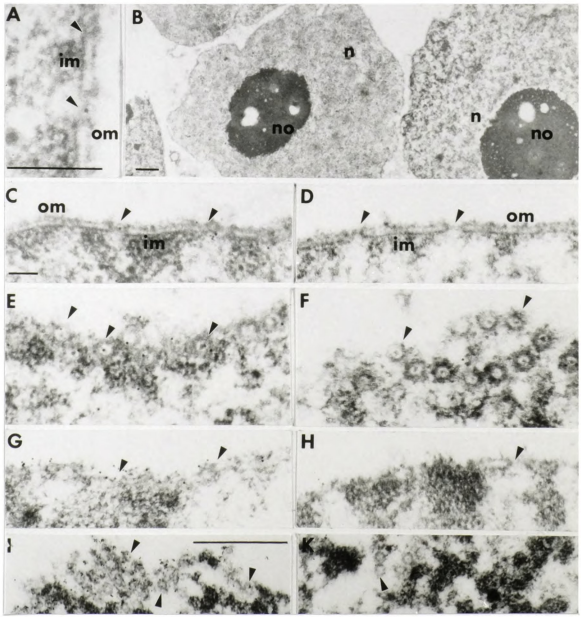
RESULTS

Proteins at the NPC are modified by GlcNAc

Nuclei were purified from tobacco suspension-cultured cells in the presence of either 0.01% or 0.6% Triton X-100 by multiple Percoll gradients. As shown by EM, the structural integrity of the purified nuclei was preserved (Figure 2.1B). Nuclei treated with 0.01% detergent were surrounded by two distinct membranes, the outer and inner membrane of the NE (Figure 2.1A to 1D; longitudinal sections). In longitudinal sections, the NPCs were visible as channels embedded in the NE (Figure 2.1A, C, and D). The typical cylindrical shape of the NPCs was apparent in tangential sections (Figure 1E and F), and the central transporter was present as a dark central granule in some NPCs. Since some other organelles were present in the nuclear fraction isolated with 0.01% detergent, 0.6% Triton X-100 was included in the nuclei isolation buffer. As previously reported by Willmitzer and Wagner (1981), the addition of high detergent concentration eliminates contaminating organelles by lysis. Since the nuclear integrity does not rely on the NE, nuclei stay intact under these conditions (Aaronson and Blobel, 1974). Based on light microscopy and EM, the nuclear preparation treated with 0.6% Triton X-100 appeared to be free of contaminating organelles such as vesicles, chloroplasts and mitochondria and were of equivalent purity as reported in other studies (Wilmitzer and Wagner, 1981; Saxena *et al.*, 1985). Treatment with 0.6% detergent did not appear to change the ultrastructure of the nuclei (Figure 2.1G to 1J). As expected, most of the outer and inner NE were removed by the detergent (Figure 2.1G and H). The absence of the NE made it difficult to observe the NPCs in longitudinal sections (Figure 2.1G and H).

Figure 2.1

Ultrastructural localization of NPC proteins modified by GlcNAc in isolated tobacco nuclei. Nuclei were isolated in the presence of either 0.01% (A to F) or 0.6% Triton X-100 (G to J). Prior to fixation, purified nuclei were incubated with WGA-conjugated colloidal gold particles in the absence (A, B, C, E, G, I) and in the presence of chitotriose (D, F, H, J). (A to D, G and F) longitudinal sections, (E, F, I, J) tangential sections. Arrowheads indicate NPCs. im, inner membrane; om, outer membrane; n, nucleus; no, nucleolus; bars: A to H, 0.1 μm; I and J, 0.5 μm.


A
im
om
B
n
no
n
no
C
om
im
D
im
om
E
F
G
H

However, the NPCs were visible in tangential sections (Figure 2.1I and J), and their overall morphology was preserved. This confirms observations from studies that examined animal nuclei isolated in the presence of high detergent (Aaronson and Blobel, 1974; Reichelt *et al.*, 1990). It is possible, however, that some peripheral components of the NPC may be lost during this isolation procedure.

To examine the WGA binding sites, purified nuclei were incubated with a WGA probe. In contrast to animals, plants do not appear to contain sialic acid moieties (Corfield and Schauer, 1982). Therefore, in plants, the WGA probe is specific for GlcNAc and its oligomers such as chitotriose. The tobacco nuclei were incubated with WGA-conjugated colloidal gold particles, fixed, and examined by EM (Figure 2.1). Most of the WGA labeling was associated with the periphery of nuclei treated with low or high detergent (Figure 2.1A, C, and G; longitudinal sections). Little binding of WGA was observed to internal nuclear structures. As specifically apparent in nuclei surrounded by the NE, WGA binding was often found at or adjacent to NPCs (Figure 2.1C), and an enlarged NPC labeled by WGA is shown in Figure 2.1A. In particular, the rings and the central regions of the NPCs appeared to be labeled by WGA using either nuclear preparation (Figure 2.1E and I; tangential sections). Individual NPCs were often labeled by several gold particles suggesting multiple binding sites at each NPC. Association of WGA to the nuclear periphery and in particular, to the NPCs seemed to be higher in nuclei treated with 0.6% Triton X-100 (Figure 2.1G and I). Higher concentration of detergent may have solubilized some components of the NPC and thereby exposed proteins modified by GlcNAc that were not accessible to the probe

following low detergent isolation. The WGA binding was specific, because it was reduced in the presence of chitotriose, a specific competitor of WGA binding (Figure 2.1D, F, H and J). These results indicate that proteins modified by GlcNAc are present at the periphery of the tobacco nucleus, and some of these glycoproteins are components of the NPC. (Electron microscopy was done with the technical help of Dr. Olga N. Borkhsenious).

Nuclear proteins modified by GlcNAc are extracted by salt

Nuclei isolated in the presence of 0.6% Triton X-100 were used as material for all biochemical studies for two reasons. As shown by EM, these nuclei were devoid of contaminating organelles. In addition, their NPCs appeared structurally preserved and contained proteins modified by GlcNAc. To gain more information about these glycoproteins, nuclei were treated with DNase I, and proteins were subjected to protein blot analysis (Figure 2.2). Proteins of similar mass were detected by two different WGA probes: WGA-alkaline phosphatase (WGA-AP; Figure 2.2, lane 1) and WGA-horseradish peroxidase (WGA-HRP, Figure 2.2, lane 3). The WGA binding was greatly reduced in the presence of 2 mM chitotriose (Figure 2.2, lanes 2 and 4), but it was unaffected by 250 mM mannose (data not shown), indicating that binding was specific. These results demonstrate that several nuclear proteins are modified by GlcNAc.

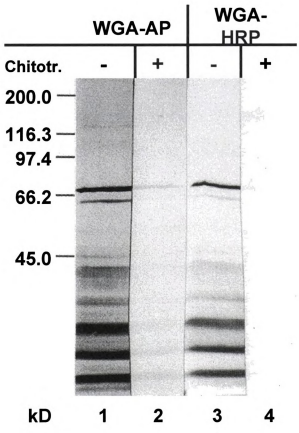


Figure 2.2

Detection of tobacco nuclear proteins modified by GlcNAc.
Nuclear proteins were treated with DNase I and subjected to proteins blot analysis using WGA-AP and WGA-HRP. Equal amounts of nuclei (1.5 x 10^6) were loaded in each lane. (1) WGA-AP, (2) WGA-AP plus chitotriose, (3) WGA-HRP, (4) WGA-HRP plus chitotriose. Mass standards (kD) are as indicated.

The nuclear glycoproteins were further characterized by the following treatment. Nuclear proteins were solubilized by either a low salt/Triton X-100 buffer (LS/TR; 0.02 M KCl, 2% Triton X-100) or a high salt buffer (HS; 0.5 M NaCl) and separated into soluble and insoluble fractions by centrifugation. The protein equivalent of $5x10^6$ nuclei per treatment were analyzed by protein blot analysis using WGA-AP as a probe (Figure 2.3). Seven of eight detected proteins were partially extracted by the HS buffer (Figure 2.3, lane 3), although the majority of the proteins with lower apparent mass (approximately 20 to 30 kD) were still associated with the insoluble fraction (Figure 2.3, lane 4). None of the proteins were extracted by the LS/TR buffer (Figure 2.3, lane 1). These fractionation data show that exposure to salt destabilized the interaction of these glycoproteins with other nuclear components, whereas the exposure to high nonionic detergent had no effect. A similar HS buffer containing 0.5 M salt is used to extract the O-GlcNAc NPC proteins from vertebrates. These O-GlcNAc proteins are also only partially solubilized under these conditions (Davis and Blobel, 1986; Snow *et al.*, 1987) with a similar efficiency as the nuclear glycoproteins from plants. The basis of this extraction is not understood (Davis and Blobel, 1986), and a higher salt concentration does not seem to solubilize a much higher percentage NPC glycoproteins (Davis and Blobel, 1986; Snow *et al.*, 1987). Overall, the results obtained by extraction and EM indicate that at least some of the plant nuclear proteins modified by GlcNAc are located at the NPC.

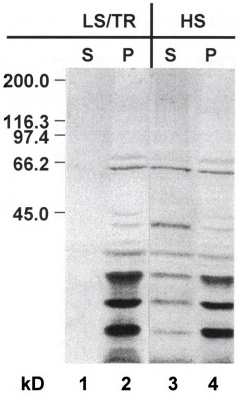


Figure 2.3

Biochemical characterization of tobacco nuclear proteins modified by GlcNAc.
Nuclear proteins treated with DNase I were solubilized in either a 0.02 M KCl/2% Triton X-100 (LS/TR) or a 0.5 M NaCl (HS) buffer and separated into a soluble (S) and an insoluble (P) fraction. The equivalent of 5 x 10^6 nuclei per treatment were subjected to protein blot analysis using WGA-AP. (1 + 2) LS/TR treatment, (3 + 4) HS treatment.
Proteins were resolved by 10% SDS-PAGE. Mass standards (kD) are as indicated.

Nuclear proteins are modified by terminal GlcNAc

Since WGA can bind to terminal as well as internal GlcNAcs (Ebisu *et al.*, 1977), a different approach was used to gain more information about the sugar modification. We employed an *in vitro* labeling procedure using galactosyltransferase (GalTF; Roquemore *et al.*, 1994). This enzyme specifically transfers the [^{3}H]galactose moiety of UDP-[^{3}H]galactose to terminal GlcNAc residues of glycoproteins (Whiteheart *et al.*, 1989; Roquemore *et al.*, 1994). Nuclear proteins were extracted by the HS buffer and labeled *in vitro* with [^{3}H]galactose by GalTF (+) (Figure 2.4, lanes 2 and 4). A control reaction in the absence of GalTF (-) was performed to show that labeling was specific to GalTF, and not to endogenous GalTF activity associated with the nuclei (Figure 2.4, lanes 1 and 3). The radiolabeled proteins were visualized by fluorography, and at least eight proteins were specifically labeled by the GalTF (Figure 2.4, lane 4). In particular, two proteins having apparent masses of 60 kD and 90 kD (after [^{3}H]galactose incorporation) did not appear to be highly abundant proteins by Coomassie Blue-stained protein profile (Figure 2.4, lane 2) suggesting the presence of multiple terminal GlcNAc moieties on these proteins.

Due to the incorporation of [^{3}H]galactose, the mobility of proteins decreases on SDS-PAGE after *in vitro* GalTF labeling (Snow *et al.*, 1987; Miller and Hanover, 1994). Hence, it was not possible to directly align glycoproteins detected by the *in vitro* GalTF labeling assay with proteins detected by WGA blot analysis. We therefore established an *on blot* GalTF labeling assay in which proteins immobilized on membranes were incubated with UDP-[^{3}H]galactose and GalTF under conditions that were similar to those used in the *in vitro* GalTF labeling assay (Figure 2.5).

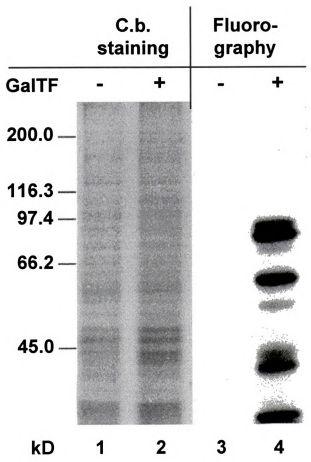
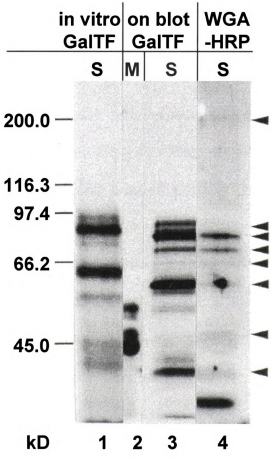


Figure 2.4

Detection of salt extracted nuclear proteins modified by terminal GlcNAc using *in vitro* GalTF labeling assay.
Nuclear proteins were treated with DNase I and solubilized with 0.5 M NaCl. Equal amounts of solubilized proteins (30 μg) were incubated in the presence or absence of GalTF with UDP-[^{3}H]galactose. Radiolabeled proteins were visualized by fluorography for 12 hrs. Labeling in the absence of the GalTF: (1) Coomassie-Blue stained gel, (3) fluorography. Labeling in the presence of the GalTF: (2) Coomassie-Blue stained gel, (4) fluorography. Mass standards are indicated in kilodalton (kD).

Figure 2.5

Detection of salt extracted nuclear proteins modified by terminal GlcNAc using GalTF labeling assays and WGA blot assay.
Nuclear proteins were solubilized as described in (A) and treated as follows.
(1) Proteins were subjected to *in vitro* GalTF labeling assay (as in 1.4).
(2) *On blot* GalTF labeling assay on protein mass standards. Proteins were immobilized on Immobilon-P membranes. One half of the lane was subjected to *on blot* GalTF labeling assay in the presence of GalTF and then visualized by fluorography for 43 hours. The other half of the lane was subjected to *on blot* GalTF labeling assay in the absence of GalTF; no labeling of ovalbumin was observed (data not shown). Ovalbumin (which runs as a doublet on SDS-PAGE; positive control), 45 kD; bovine serum albumin (negative control), 66.2 kD.
(3) *On blot* GalTF labeling assay on salt extracted nuclear proteins. Solubilized proteins of 2 x 10^6 nuclei were treated as described in (2). The entire lane was subjected to the *on blot* GalTF labeling assay in the presence of GalTF and visualized as in (2).
(4) WGA blot analysis of salt extracted nuclear proteins. Solubilized proteins [equal amount as in (3)] were subjected to protein blot analysis and detected by WGA-HRP using chemiluminescence.
Arrowheads indicate proteins detected by both, *on blot* GalTF labeling assay and WGA blot analysis. Proteins were resolved by 7.5% SDS-PAGE. Mass standards (kD) are as indicated.

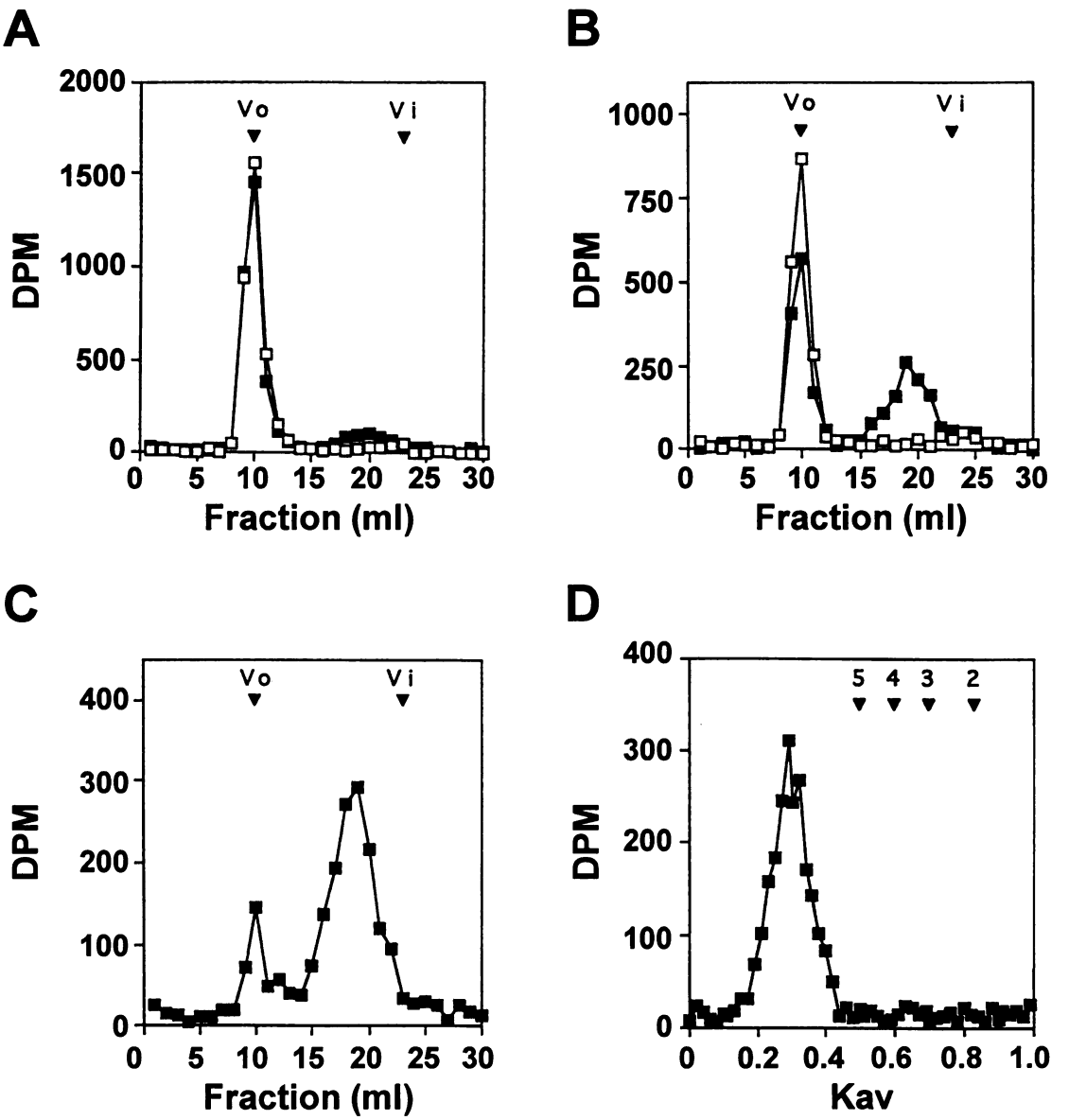

in vitro GalTF
on blot GalTF
WGA -HRP
S
M
S
S
200.0
116.3
97.4
66.2
45.0
kD
1
2
3
4

Under these conditions, [^{3}H]galactose was incorporated after immobilization of proteins on the blot, thus the apparent mass of the radiolabeled proteins did not change. Proteins present in the mass standards served as positive and negative controls for the *on blot* GalTF labeling assay (Figure 2.5, lane 2). Ovalbumin is a 45 kD protein modified by N-linked terminal GlcNAc moieties and functioned as a positive control (see also Holt and Hart, 1986, Hanover *et al.*, 1987). Bovine serum albumin (BSA; 66.2 kD) is not modified by terminal GlcNAc and served as a negative control. The pattern of the galactosylation products using the *on blot* GalTF labeling assay (Figure 2.5, lane 3) was similar to that of the *in vitro* GalTF labeling assay (Figure 2.5, lane 1). These results demonstrate that the *on blot* GalTF labeling assay is an easy and fast method to detect proteins modified by terminal GlcNAc. More importantly, it allowed a direct comparison of proteins containing terminal GlcNAc (Figure 2.5, lanes 3) with the glycoproteins detected by the WGA blot analysis (Figure 2.5, lane 4). The majority of the proteins were detected by both methods but with different intensities (marked with arrows). The glycoproteins with a lower apparent mass (approximately 20 to 30 kD) and detected by the WGA probes were also labeled in the *on blot* GalTF assay, although some with low intensity (data not shown). These correlative studies indicate that the majority of the nuclear proteins detected by WGA are modified by terminal GlcNAc, and some of these proteins with terminal GlcNAc are likely to be NPC proteins.

Salt extracted nuclear proteins are modified by oligosaccharides with terminal GlcNAc

The terminal GlcNAc detected on the nuclear proteins could be attached via the hydroxyl group of a serine, threonine, tyrosine, or hydroxyproline (O-linked) or via the amido group of an asparagine (N-linked). To examine these alternatives, nuclear proteins extracted with the HS buffer were used for the subsequent sugar analysis because of the following reasons. Seven of eight glycoproteins were present in the solubilized fraction, even though their extraction was not complete. In addition, the protein sample has to be fully solubilized and of an ionic strength of less than 0.2 M as an requirement for subsequent *in vitro* GalTF labeling assay (Roquemore *et al.*, 1994). The extracted proteins were radiolabeled in the *in vitro* GalTF labeling assay, and the protein sample was divided into equal aliquots and subjected to the following treatments. One aliquot was digested with peptide:N-glycosidase F (PNGase F; Figure 2.6A; filled squares), an enzyme that releases the common classes of N-linked glycans (Tarentino *et al.*, 1985). The second aliquot was treated similarly except without PNGase F (Figure 2.6A; open squares) to control for proteolysis during the PNGase F experiment. The elution profile of these samples over a Sephadex G-50 column were identical indicating that virtually no radioactivity associated with the proteins was removed by the PNGase F (Figure 2.6A). Parallel experiments using [^{3}H]galactose-labeled ovalbumin, a protein containing N-linked terminal GlcNAc, were performed as a positive control for PNGase F activity. The PNGase F removed 75% of the radioactivity associated with [^{3}H]galactose-labeled ovalbumin (data not shown).

Figure 2.6
Characterization of carbohydrates with terminal GlcNAc on salt extracted nuclear glycoproteins.
(A) Carbohydrate linkages were resistant to PNGase F. Salt extracted, [^{3}H]galactose-labeled nuclear proteins (5 x 10^5 DPM) were incubated with (filled squares) or without (open squares) PNGase F and fractionated over a Sephadex G-50 column. 0.5% of fractions were assayed for radioactivity by liquid scintillation counting.
(B) Plant protein samples did not inhibit PNGase F activity. A 1:1 ratio (by radioactivity) of [^{3}H]galactose-labeled ovalbumin and [^{3}H]galactose-labeled nuclear proteins was incubated with (filled squares) or without (open squares) PNGase F, and the products were analyzed as in (A).
(C) ß-elimination released saccharides from PNGase F resistant glycoproteins. PNGase F resistant labeled nuclear proteins [fractions 8-13 from (A)] were pooled and subjected to alkaline induced ß-elimination. Products were fractionated over a Sephadex G-50 column. One percent of fractions were assayed for radioactivity.
(D) Saccharides released from glycoproteins by ß-elimination were larger than five GlcNAc residues. Released labeled saccharides [fractions 14-23 from (D)] were pooled and fractionated on a Toyopearl HW 40 column. Five percent of fractions were assayed for radioactivity.
Numbers above solid triangles represent migration positions of oligosaccharide standards: 2=Gal-GlcNAcitol, 3=Gal-(GlcNAc)-GlcNAcitol, 4=Gal-$(GlcNAc)_2$-GlcNAcitol, 5=Gal-$(GlcNAc)_3$-GlcNAcitol.
Vo, void volume; Vi, inclusion volume; DPM, disintegration per minute; Kav, volume fraction (Ve-Vo/Vi-Vo); Ve, elution volume.


A
2000
1500
1000
500
0
DPM
Vo
Vi
0 5 10 15 20 25 30
Fraction (ml)
B
1000
750
500
250
0
DPM
Vo
Vi
0 5 10 15 20 25 30
Fraction (ml)
C
400
300
200
100
0
DPM
Vo
Vi
0 5 10 15 20 25 30
Fraction (ml)
D
400
300
200
100
0
DPM
5 4 3 2
0 0.2 0.4 0.6 0.8 1.0
Kav

Furthermore, PNGase F activity was not inhibited by any component in the plant sample because the enzyme released 32% of radioactivity associated with a 50:50 mixture (by radioactivity) of [^{3}H]galactose-labeled ovalbumin and salt extracted nuclear proteins (4B; filled squares). This value was consistent with a 64% cleavage of the ovalbumin glycans in this control group.

O-glycosidic linkages are sensitive to alkaline induced ß-elimination (Spiro, 1972). Radiolabeled proteins resistant to PNGase F treatment (Figure 2.6A) were subjected to alkaline borohydride to release O-linked sugars from the proteins (Figure 2.6C). Approximately 80% of the [^{3}H]galactose-labeled sugars were released from the nuclear proteins as judged by Sephadex G-50 chromatography (Figure 2.6C, fractions 14-23). Twenty percent of the radioactivity remained in the void volume (Figure 2.6C; Vo), indicating that these labeled sugars were on large polysaccharides or still associated with the proteins. These labeled sugars may belong to a small group of plant N-glycans that appear to be resistant to PNGaseF treatment (Tretter *et al.*, 1991). The same profile of ß-eliminated sugars was observed, whether the samples were subjected to ß-elimination with or without prior digestion with PNGase F. These results indicate that at least 80% of the [^{3}H]galactose-labeled sugars are attached to the salt extractable proteins via an O-linkage.

To determine the size of these labeled glycans, ß-eliminated sugars (Figure 2.6C, fractions 14-23) were fractionated on a Toyopearl HW 40 size exclusion column (Figure 2.6D). Oligosaccharide standards composed of Gal-GlcNAc$_{(1-4)}$ were used as an assessment of size (but not as an indication of structure). The radiolabeled sugars migrated at a position larger than five saccharides consisting of Gal-GlcNAc$_3$-

GlcNAcitol. None of the radioactivity comigrated with the disaccharide Gal-GlcNAcitol as would be expected, if the sugars were a single O-GlcNAc modification typical of vertebrate NPC proteins. The elution profile appeared to consist of more than one peak, suggesting the presence of multiple size populations of [^{3}H]galactose-labeled oligosaccharides. Forty percent of these labeled oligosaccharides adsorbed to a Sep-Pak (C18) cartridge indicating the presence of a hydrophobic group (data not shown). The remaining 60% were not retained on the column and were resistant to pronase digestion and dansylation indicating the absence of peptides or primary amines (data not shown). These results are consistent with these labeled oligosaccharides being associated to the nuclear glycoproteins via an O-linkage. (The carbohydrate linkage and size analyses were performed in collaboration with Dr. Robert N. Cole in laboratory of Dr. Gerald W. Hart, University of Alabama at Brimingham, AL).

Overall, these studies indicate that the proteins extracted by high salt from tobacco nuclei are modified by oligosaccharides larger in size than five GlcNAcs including the terminal GlcNAc. Most of these are bound to the proteins via an O-linkage. Fractionation and EM data indicate that at least some of these glycoproteins are located at the NPC.

DISCUSSION

We have identified and characterized a subset of nuclear proteins from tobacco suspension-cultured cells that contain novel carbohydrate modifications. Probing isolated nuclei with WGA revealed the presence of GlcNAc-containing proteins that are associated with the rings or central structures of plant NPCs. Seven of eight glycoproteins detected by WGA are partially extracted by 0.5 M salt, and the majority of these glycoproteins are modified by terminal GlcNAc(s) as shown in GalTF labeling assays. Carbohydrate analysis demonstrates that the glycans with terminal GlcNAc consist of sugars that are larger in size than five GlcNAcs, and most of these oligosaccharides appear to be attached to the proteins via an O-linkage.

Proteins modified by GlcNAc are located at the NPC

In vertebrates, WGA is a potent inhibitor of protein import into the nucleus (Finlay *et al.*, 1987; for review see Forbes, 1992). This lectin binds to the single O-GlcNAc present on several NPC proteins and was successfully used to identify and purify some of these NPC glycoproteins (Finlay *et al.*, 1991; Hallberg *et al.*, 1993; Sukegawa and Blobel, 1993; Kita *et al.*, 1993; Powers *et al.*, 1995; Radu *et al.*, 1995b). Interestingly, WGA does not appear to bind to yeast NPCs (Davis and Fink, 1990; Carmo-Fonseca *et al.*, 1991; Rout and Blobel, 1993) and so far, none of the identified NPC proteins in yeast have been shown to be modified by O-GlcNAc. In plants, the nature of sugar modification on NPC proteins is unknown.

Using isolated nuclei from tobacco cells, several nuclear proteins were detected by WGA. At least some of these nuclear glycoproteins are probably NPC proteins by

the following criteria: a) the WGA labeling was associated with or adjacent to the NPC with little binding to other structural components of the nucleus, b) the proteins detected by WGA were enriched in the nuclear fraction, but were not detectable in the cytoplasmic fraction (A. Heese-Peck, N.V. Raikhel; unpublished results), and c) the plant nuclear glycoproteins were solubilized under the same salt conditions and with a similar efficiency as the O-GlcNAc NPC proteins from vertebrate nuclei (Davis and Blobel, 1986; Snow *et al.*, 1987).

Based on EM using WGA as a probe, the plant NPC glycoproteins were mainly found at the rings and the center of the NPCs. As apparent in longitudinal sections, these glycoproteins appeared to be present at the cytoplasmic and, though in lower amounts, the nucleoplasmic side of the NPCs. Using a similar embedding/thin sectioning preparation, WGA labels the cytoplasmic and the nucleoplasmic center of rat NPCs (Hanover *et al.*, 1987) and almost exclusively the cytoplasmic center of *Xenopus* oocyte NPCs (Finlay *et al.*, 1991). Quick freezing/freeze drying/rotary metal shadowing of *Xenopus* oocyte nuclei showed WGA labeling at the terminal ring of the nuclear baskets as well as the center and inner annulus of the cytoplasmic face (Panté *et al.*, 1994). Two O-GlcNAc proteins, Nup153p and Nup98p, are constituents of nucleoplasmic filaments and the terminal ring of the nucleoplasmic baskets (Cordes *et al.*, 1993; Panté *et al.*, 1994). Further, a putative O-GlcNAc protein, p250, has been localized to the cytoplasmic filaments of the NPC using a polyclonal antibody (Panté *et al.*, 1994) and may correspond to the rat Nup214p/CAN protein (Kraemer *et al.*, 1994). It is possible that several of the plant glycoproteins are homologs of these vertebrate NPC proteins, although the apparent mass of the plant proteins are different

from these O-GlcNAc NPC proteins. The mass of the 55 kD plant GlcNAc-protein (60 kD after [^{3}H]galactose incorporation) raises the possibility that it may be the plant homolog of the vertebrate p62. The p62 is found in multiple copies on the cyto- and/or the nucleoplasmic faces of the central NPC regions (Davis and Blobel, 1986; Dabauvalle *et al.*, 1990; Cordes *et al.*, 1991; Wilken *et al.*, 1993). However, a monoclonal antibody against the rat p62 (mAb414; Davis and Blobel, 1986) did not detect plant NPC proteins of similar mass (A. Heese-Peck, N.V. Raikhel; unpublished results). In fact, mAb414 recognized an antigen associated with tobacco chromatin, but not with the NPC (A. Heese-Peck, O.N. Borkhsenious, N.V. Raikhel; unpublished results). This monoclonal antibody has also been shown to detect NPC as well as non-NPC proteins in yeast (Aris and Blobel, 1989; Davis and Fink, 1990). A study in higher plants reported the detection of a 100 kD protein by mAb414 in nuclear matrix fractions of carrot suspension-cultured cells after extensive extraction with 1 M salt (Scofield *et al.*, 1992), however, no EM data using mAb414 were provided. We investigated whether other antibodies against known NPC proteins recognized plant nuclear proteins and specifically the nuclear proteins modified by terminal GlcNAc. However, no specific labeling of the tobacco NPC was observed on EM level (A. Heese-Peck, O.N. Borkhsenious, N.V. Raikhel; unpublished results).

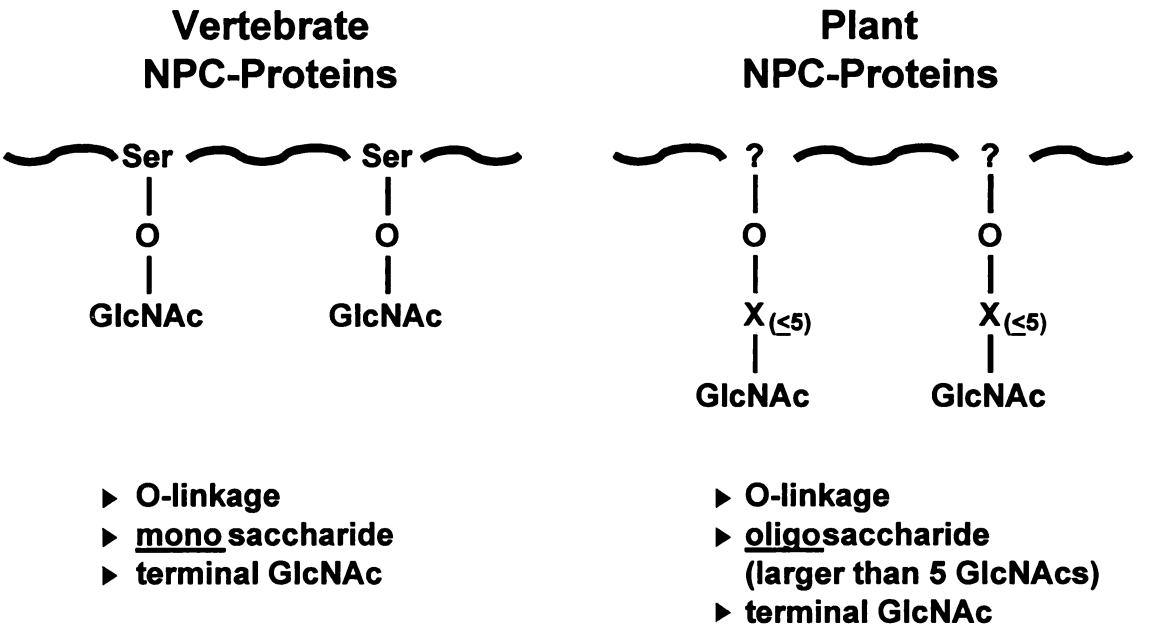


Figure 2.7

Comparison of O-GlcNAc from vertebrate NPC proteins and O-linked oligosaccharides with terminal GlcNAc from plant nuclear and NPC proteins.
?, unknown amino acid (probably serine, threonine or hydroxyproline); X, unknown saccharides (in size $\geq$ 5 GlcNAcs).

Oligosaccharides with terminal GlcNAc

We demonstrated that the O.5 M salt extracted nuclear proteins from tobacco contain a novel sugar modification that is different from the O-GlcNAc modification found on vertebrate NPC proteins (Figure 2.7). The vertebrate NPC proteins are solubilized under similar salt conditions and are modified by single O-GlcNAc(s) attached to the proteins via an O-linkage (Holt *et al.*, 1987; Hanover *et al.*, 1987). Sugar analysis indicates that the salt extractable plant nuclear proteins are modified by oligosaccharides consisting of saccharides larger in size than five GlcNAcs, and most of these oligosaccharides are also attached to the proteins via an O-linkage. At least two size populations of oligosaccharides possessing terminal GlcNAc were attached to the nuclear proteins. At present, it is not known whether these oligosaccharides share a common core structure. To perform such detailed carbohydrate analysis, the purification of individual proteins in large quantities are required and will be pursued in the future. Thus, probing plant nuclear proteins with the GalTF indicates that the single O-GlcNAc modification present on vertebrate NPC proteins is not found on salt extracted nuclear proteins from plants.

We can only speculate about the function of the oligosaccharides on plant NPC proteins. The O-linked oligosaccharides may have similar function(s) as proposed for the O-GlcNAc found in vertebrate NPC proteins. These include the assembly of multimeric protein complexes, resistance to proteolysis, and regulation of protein function (Hart *et al.*, 1989; Haltiwanger *et al.*, 1992). In addition, the O-GlcNAc modification of vertebrate NPC proteins has been proposed to function in nucleocytoplasmic transport because WGA inhibits transport of protein and RNA by

binding to O-GlcNAc (Finlay *et al.*, 1987; Yoneda *et al.*, 1987; Dabauvalle *et al.*, 1988; Featherstone *et al.*, 1988; Newmeyer and Forbes, 1988; Wolff *et al.*, 1988). Interestingly, Miller and Hanover (1994) have shown in reconstitution assays that NPC proteins that are altered by the addition of galactose to O-GlcNAc residues are still able to form NPCs. These NPCs are morphologically normal and are competent to import proteins into the nucleus. Thus, it has been suggested that a specific recognition of the GlcNAc moiety may not play a direct role in nuclear protein import and in NPC assembly. This implies that the protein portion rather than the carbohydrate moiety of these glycoproteins may function in nuclear import. Furthermore, yeast nuclei are able to import proteins, although the NPC proteins do not appear to be modified by O-GlcNAc. In plants, Harter *et al.* (1994) reported that the nuclear import of endogenous G-box binding factors (GBFs) is inhibited in the presence of WGA in permeabilized parsley cells. In these studies, imported GBFs are detected by an antibody cotranslocation assay that is indirect and is based on the detection of protease resistant GBF antibody which is associated with the nuclei. An *in vitro* import system was developed in our laboratory in which the accumulation of import substrate is visualized by immunofluorescence. Using this import assay, WGA (0.7 mg/mL) does not block nuclear import *in vitro* (Hicks and Raikhel, 1995b; G.R. Hicks, S. Lobreaux, N.V. Raikhel; unpublished results). However, as demonstrated in the current study proteins at the plant NPC were modified by GlcNAc. These NPC proteins appeared to be modified by carbohydrates that are larger than the single O-GlcNAc of vertebrate NPC proteins. The larger carbohydrate moieties may extend further away from the NPC center, so that the binding of WGA to the GlcNAc does not hinder protein import into

the plant nucleus, though import of larger substrate such as immunoglobin (Harter *et al.*, 1994) could be affected. Although the carbohydrates with terminal GlcNAc do not appear to be involved in nuclear import, it cannot be excluded that the proteins themselves may play a role in the import process. At this point, the significance of the GlcNAc modification is not understood in any system studied. However, this modification has been proven to be a useful tool to identify and isolate NPC proteins that are involved in nucleocytoplasmic transport.

The oligosaccharides with terminal GlcNAc may also aid in purification of the plant NPC glycoproteins. Preliminary results indicate that several of the plant glycoproteins can be purified by lectin affinity chromatography. The purification of the glycoproteins will enable us to produce specific antibodies and to localize the glycoproteins *in situ*. We will also investigate whether any of the plant NLS-binding proteins that may be localized to the plant NPC (Hicks and Raikhel, 1993; 1995a) correspond to one of the glycoproteins identified in this study.

REFERENCES

Aaronson,R.P. and Blobel,G. (1974) The attachment of the nuclear pore complex. J. Cell Biol. **62**, 746-754.

Akey,C.W. (1992). The nuclear pore complex. Curr. Opin. Struct. Biol. **2**, 258-263.

Akey,C.W., and Rademacher,M. (1993). Architecture of the *Xenopus* nuclear pore complex revealed by three-dimensional cryo-electron microscopy. J. Cell Biol. **122**, 1-19.

Aris,J.P., and Blobel,G. (1989). Yeast nuclear envelope proteins cross react with an antibody against mammalian pore complex proteins. J. Cell Biol. **108**, 2059-2067.

Bednarek,S.Y., Wilkins,T.A., Dombrowski,J.E., and Raikhel,N.V. (1990). A carboxyl-terminal propeptide is necessary for proper sorting of barley lectin to vacuoles of tobacco. Plant Cell **2**, 1145-1155.

Carmo-Fonseca,M., Kern,H., and Hurt,E.C. (1991). Human nucleoporin p62 and essential yeast nuclear pore protein NSP1 show sequence homology and a similar domain organization. Eur. J. Cell Biol. **55**, 17-30.

Cordes,V.I., and Krohne,G. (1993). Sequential O-glycosylation of nuclear pore complex protein p62 *in vitro*. Eur. J. Cell Biol. **60**, 185-195.

Cordes,V., Reidenbach,S., Köhler,A., Stuurmann,N., van Driel,R., and Franke,W.W. (1993). Intranuclear filaments containing a nuclear pore complex protein. J. Cell Biol. **123**, 1333-1344.

Cordes,V.I., Waizenegger,I., and Krohne,G. (1991). Nuclear pore complex glycoprotein p62 from *Xenopus laevis* and mouse: cDNA cloning and identification of its glycosylation region. Eur. J. Cell Biol. **55**, 31-47.

Corfield,A.P., and Schauer,R. (1982). Occurrence of sialic acids. In Schauer,R. (ed.) *Sialic acid: Chemistry, Metabolism and Function*, Springer-Verlag, New York, Vol. 10, pp. 1-50.

Dabauvalle,M.-C., Benevente,R., and Chaly,N. (1988). Monoclonal antibodies to a M_r 68,000 pore complex glycoprotein interfere with nuclear protein uptake in *Xenopus* oocytes. Chromosoma **97**, 193-197.

Dabauvalle,M.-C., Loos,K., and Scheer,U. (1990). Identification of a soluble precursor complex essential for nuclear pore assembly *in vitro*. Chromosoma **100**, 56-66.

Davis,L.I., and Blobel,G. (1986). Identification and characterization of a nuclear pore complex protein. Cell **45**, 699-709.

Davis,L.I., and Fink,G.R. (1990). The NUP1 gene encodes an essential component of the yeast nuclear pore complex. Cell **61**, 965-978.

Dubois,M., Gilles,K.A., Hamilton,J.K., Rebers,P.A., and Smith,F. (1956). Colorimetric method for determination of sugars and related substances. Anal. Chem. **28**, 350-356.

Ebisu,S., Lonngren,J., and Goldstein,I.J. (1977). Interaction of pneumococcal S 14 polysaccharides with lectins from *Ricinus communis*, *Triticum vulgares*, and *Bandeiraea simplicifolia*. Carbohydrate Res. **58**, 187-191.

Fabre,E., and Hurt,E.C. (1994). Nuclear transport. Curr. Opin. Cell Biol. **6**, 335-342.

Featherstone,C., Darby,M.K., and Gerace,L. (1988). A monoclonal antibody against the nuclear pore complex inhibits nucleocytoplasmic transport of protein and RNA in vivo. J. Cell Biol. **107**, 1289-1297.

Feldherr,C.M., Kallenbach,E., and Schultz,N. (1984). Movement of a karyophilic protein through the nuclear pores of oocytes. J. Cell Biol. **99**, 2216-2222.

Finlay,D.R., Newmeyer,D.D., Price,T.M., and Forbes,D.J. (1987). Inhibition of *in vitro* nuclear transport by a lectin that binds to nuclear pores. J. Cell Biol. **104**, 189-200.

Finlay,D.R., Meier,E., Bradley,P., Horecka,J., and Forbes,D.J. (1991). A complex of nuclear pore proteins required for pore function. J. Cell Biol. **114**, 169-183.

Finlay,D.R., and Forbes,D.J. (1990). Reconstitution of biochemically altered nuclear pores: Transport can be eliminated and restored. Cell **60**, 17-29.

Forbes,D.J. (1992). Structure and function of the nuclear pore complex. Annu. Rev. Cell Biol. **8**, 495-527.

Fukuda,M. (1989). Characterization of O-linked saccharides from cell surface glycoproteins. Methods in Enzymology **179**, 17-29.

Goldberg,M.W., and Allen,T.D. (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.

Goldstein,I.J., and Hayes,C.E. (1978). The lectins: Carbohydrate-binding proteins of plants and animals. Adv. Carbohydr. Chem. Biochem. **35**, 120-340.

Hallberg,E., Wozniak,R.W., and Blobel,G. (1993). An integral membrane protein of the pore membrane domain of the nuclear envelope contains a nucleoporin-like region. J. Cell Biol. **122**, 513-521.

Haltiwanger,R.S., Kelly,W.G., Roquemore,E.P., Blomberg,M.A., Dong,L.-Y.D.,

Kreppel,L., Chou,T.-Y., and Hart,G.W. (1992). Glycosylation of nuclear and cytoplasmic proteins is ubiquitous and dynamic. Biochem. Soc. Transaction **20**: 264-69.

Hanover,J.A., Cohen,C.K., Willingham,M.C., and Park,M.K. (1987). *O*-linked *N*-acetylglucosamine is attached to proteins of the nuclear pore. J. Biol. Chem. **262**, 9887-9894.

Harlow,E., and Lane,D. (1988). Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Hart,G.W., Haltiwanger,R.S., Holt,G.D., and Kelly,W.G. (1989). Glycosylation in the nucleus and cytoplasm. Annu. Rev. Biochem. **58**, 841-874.

Harter,K., Kircher,S., Frohnmeyer,H., Krenz,M., Nagy,F., and Schäfer,E. (1994). Light-regulated modification and nuclear translocation of cytosolic G-box binding factors in parsley. Plant Cell **6**, 545-559.

Hicks,G.R., and Raikhel,N.V. (1993). Specific binding of nuclear localization sequences to plant nuclei. Plant Cell **5**, 983-994.

Hicks,G.R., and Raikhel,N.V. (1995a). Nuclear localization signal binding proteins in higher plant nuclei. Proc. Natl. Acad. Sci. USA **92**, 734-738.

Hicks,G.R., and Raikhel,N.V. (1995b). Protein import into the nucleus: an integrated view. Annu. Rev. Cell Biol. 11, in press.

Hicks,G.R., Smith,H.M.S., Shieh,M., and Raikhel,N.V. (1995). Three classes of nuclear import signals bind to plant nuclei. Plant Physiol., **107**, 1055-1058.

Hinshaw,J.E., Carragher,B.O., and Milligan,R.A. (1992). Architecture and design of the nuclear pore complex. Cell **69**, 1133-1141.

Holt,G.D., and Hart,G.W. (1986). The subcellular distribution of terminal N-acetylglucosamine moieties. J. Cell Biol. **261**, 8049-8057.

Holt,G.D., Snow,C.M., Senior,A., Haltiwanger,R.S., Gerace,L., and Hart,G.W. (1987). Nuclear pore complex glycoproteins contain cytoplasmically disposed O-linked N-acetylglucosamine. J. Cell Biol. **104**, 1157-1164.

Hurt,E.C. (1988). A novel nucleoskeletal-like protein located at the nuclear periphery is required for the life cycle of *Saccharomyces cerevisiae*. EMBO J. **7**, 323-4334.

Jordan,E.G., Timmis,J.N., and Trewavas,A.J. (1980). The plant nucleus. In Tolbert,N.E. (ed.), The Biochemistry of Plants, Acad.Press, Inc., Vol. 1, 490-588.

Kalinich,J.F., and Douglas,M.G. (1989). *in vitro* translocation through the yeast nuclear envelope: signal dependent transport requires ATP and calcium. J. Cell Biol. **264**, 17979-17989.

Kaneda,N., Sato,M., and Yagi,K. (1982). Analysis of danyl amino acids by reverse-phase high performance liquid chromatography. Anal. Biochem. **127**, 49-54.

Kita,K., Omata,S., and Horigome,T. (1993). Purification and characterization of a nuclear pore glycoprotein complex containing p62. J. Biochem. **113**, 377-382.

Kraemer,D., Wozniak,R.W., Blobel,G., and Radu,A. (1994). The human CAN protein, a putative oncogene product associated with myloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm. Proc. Natl. Acad. Sci. USA **91**, 1519-1523.

Laemmli,U.K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature (Lond.) **227**, 680-685.

Miller,M.W., and Hanover,J.A. (1994). Functional nuclear pores reconstituted with ß-1,4 galactose-modified *O*-linked *N*-acetylglucosamine glycoproteins. J. Biol. Chem. **269**, 9289-9297.

Newmeyer,D.D., and Forbes,D.J. (1988). Nuclear import can be separated into distinct steps *in vitro*: Nuclear pore binding and translocation. Cell **52**, 641-653.

Panté,N., and Aebi,U. (1993). The nuclear pore complex. J. Cell Biol. **122**, 977-984.

Panté,N., Bastos,R., McMorrow,I., Burke,B., and Aebi,U. (1994). Interactions and three-dimensional localization of a group of nuclear pore complex proteins. J. Cell Biol. **126**, 603-617.

Park,M.K., D'Onofrio,M., Willingham,M.C., and Hanover,J.A. (1987). A monoclonal antibody against a family of nuclear pore proteins (nucleoporins): O-linked acetylglucosamine is part of the immunodeterminant. Proc. Natl. Acad. Sci. USA **84**, 462-6466.

Powers,M.A., Macaulay,C., Masiarz,F.R., and Forbes,D.F. (1995). Reconstituted nuclei depleted of a vertebrate GLFG nuclear pore complex protein, p97, import but are defective in nuclear growth and replication. J. Cell Biol. **128**, 721-736.

Radu,A., Blobel,G., and Moore,M.S. (1995a). Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins. Proc. Natl. Acad. Sci. USA **92**, 1769-1773.

Radu,A., Moore,M.S., and Blobel,G. (1995b). The peptide repeat domain of nucleoporin Nup98 functions as a docking site in transport across the nuclear pore complex. Cell **81**, 215-222.

Raikhel,N.V. (1992). Nuclear targeting in plants. Plant Physiol. **100**, 1627-1632.

Reichelt,R., Holzenburg,A., Buhle,E.L., Jarnik,M., Engel,A., and Aebi,U. (1990). Correlation between structure and mass distribution of the nuclear pore complex, and of distinct pore complex components. J. Cell Biol. **110**, 883-894.

Roberts,K., and Northcote,D.H. (1970). Structure of the nuclear pore in higher plants. Nature **228**, 385-386.

Roquemore,E.P., Chou,T.-Y., and Hart,G.W. (1994). Detection of O-linked *N*-acetylglucosamine (O-GlcNAc) on cytoplasmic and nuclear proteins. Methods in Enzymology **230**, 443-460.

Rout,M.P., and Blobel,G. (1993). Isolation of the yeast nuclear pore complex. J. Cell Biol. **123**, 771-783.

Rout,M.P., and Wente,S.R. (1994). Pores for thought: nuclear pore complex proteins. TICB **4**, 357-365.

Saxena,P.K., Fowke,L.C., and King,J. (1985). An efficient procedure for isolation of nuclei from plant protoplasts. Protoplasma **128**, 184-189.

Schindler,M., Hogan,M., Miller,M., and DeGaetano,D. (1987). A nuclear specific glycoprotein representative of a unique pattern of glycosylation. J. Biol. Chem. **262**, 254-1260.

Scofield,G.N., Beven,A.F., Shaw,P.J., and Doonan,J.H. (1992). Identification and localization of a nucleoporin-like protein component of the plant nuclear matrix. Planta **187**, 414-420.

Snow,C.M., Senior,A., and Gerace,L. (1987). Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. J. Cell Biol. **104**, 1143-1156.

Spiro,R.G. (1972). Study of the carbohydrates of glycoproteins. Methods in Enzymology **28**, 3-43.

Starr,C.M., D'Onofrio,M., Park,M.K., and Hanover,J.A. (1990). Glycosylation of nuclear pore protein p62. J. Biol. Chem. **265**, 6868-6873.

Sterne-Marr,R., Blevitt,J.M., and Gerace,L. (1992). O-linked glycoproteins of the nuclear pore complex interact with a cytosolic factor required for nuclear protein import. J. Cell Biol. **116**, 271-280.

Sukegawa,J., and Blobel,G. (1993). A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm. Cell **72**, 29-38.

Tapuhi,Y., Schmidt,D.E., Lindner,W., and Karger,B.L. (1981). Dansylation of amino acids for high-performance liquid chromatography analysis. Anal. Biochem. **115**, 123-129.

Tarentino,A.L., Gómez,C.M., and Plummer Jr.,T.H. (1985). Deglycosylation of asparagine-linked glycans by peptide:N-Glycosidase F. Biochem. **24**, 4665-4671.

Tretter,V., Altmann,F., and März,L. (1991). Peptide-N^4-(*N*-acetyl-ß-glucosaminyl) asparagine amidase F cannot release glycans attached to α1-3 to the asparagine-linked *N*-acetylglucosamine residue. Eur. J. Biochem. **199**, 647-652.

Whiteheart,S.W., Passaniti,A., Reichner,J.S., Holt,G.D., Haltiwanger,R.S., and Hart,G.W. (1989). Glycosyltransferase probes. Methods in Enzymology **179**, 82-95.

Wilken,N., Kossner,U., Senécal,J.-L., Scheer,U., and Dabauvalle,M.-C. (1993). Nup180, a novel nuclear pore complex protein localizing to the cytoplasmic ring and associated fibrils. J. Cell Biol. **123**, 1345-1354.

Willmitzer,L., and Wagner,K.G. (1981). The isolation of nuclei from tissue-cultured plant cells. Exp. Cell Res. **135**, 69-77.

Wolff,B., Willingham,M.C., and Hanover,J.A. (1988). Nuclear protein import: Specificity for transport across the nuclear pore. Exp. Cell Res. **178**, 318-334.

Yoneda,Y., Imamoto-Sonobe,N., Yamaizumi,M., and Uchida,T. (1987). Reversible inhibition of protein import into the nucleus by wheat germ agglutinin injected into cultured cells. Exp. Cell Res. **173**, 586-595.

Chapter 3

Purification of nuclear glycoproteins modified with terminal *N*-acetylglucosamine by lectin affinity chromatography

ABSTRACT

The nuclear pore complex (NPC) is the site of transport of macromolecules in and out of the nucleus. It has been previously demonstrated that several glycoproteins are present at the NPC of tobacco suspension-cultured cells that are modified with terminal *N*-acetylglucosamine (GlcNAc) residues. In the current studies, we designed a purification scheme to isolate these glycoproteins on the basis of their sugar modification. *In vitro* galactosylation of the glycoproteins generated proteins with terminal galactosyl-ß1,4-GlcNAc, thus permitting their isolation by *Erythrina crystagalli* agglutinin affinity chromatography. Peptide sequence information was obtained from three of these GlcNAc-proteins with apparent molecular masses of ~33 kD, ~40 kD, and ~65 kD, named gp33, gp40, and gp65, respectively.

INTRODUCTION

Molecules, such as proteins, different RNA species, T-DNA and ions, move in and out of the nucleus through nuclear pore complexes (NPCs) that form aqueous channels across the NE (Davis, 1995). The overall morphology of NPCs appears to be highly conserved in different organisms, and its architecture has been extensively investigated in amphibian cells using electron microscopy (Panté and Aebi, 1993; Goldberg and Allen, 1995; see Chapter 1).

The little information we have on plant NPCs has come from microscopic and biochemical analyses (for a review, see Jordan *et al.*, 1980; Chapter 1 and 2). Using wheat germ agglutinin (WGA), a lectin that binds specifically to *N*-acetylglucosamine (GlcNAc), as a probe, proteins modified by GlcNAc have been found at the plant nuclear periphery (Heese-Peck *et al.*, 1995; Hicks *et al.*, 1996; Merkle *et al.*, 1996). As shown by electron microscopy (EM) and biochemical analyses, at least some of these glycoproteins are located at the NPC of tobacco suspension-cultured cells and are modified by terminal GlcNAc (Chapter 2; Heese-Peck *et al.*, 1995). Most glycans with terminal GlcNAc are attached to the proteins via an O-linkage and are larger in size than five GlcNAc residues (Chapter 2; Heese-Peck *et al.*, 1995). Thus, this plant sugar modification is different from the single O-linked GlcNAc (O-GlcNAc) found in vertebrate NPC proteins (for review, see Hicks and Raikhel, 1995; Hart, 1997).

Although the function of the O-GlcNAc modification is not fully understood (Miller and Hanover, 1994), the glycosylation has provided a useful tool in purifying vertebrate NPC proteins and studying their function in nucleocytoplasmic transport and NPC assembly (for reviews, see Forbes, 1992; Davis, 1995; Hicks and Raikhel, 1995).

We were interested in isolating tobacco NPC proteins modified by O-linked oligosaccharides with terminal GlcNAc, which will be referred to as tGlcNAc-proteins throughout this manuscript. In the current studies, a purification scheme was established to obtain large quantities of the tGlcNAc-proteins by lectin affinity chromatography. Tryptic digestion and subsequent internal amino acid analysis generated peptide sequence information for three tGlcNAc-proteins with apparent molecular masses of ~33 kD, ~40 kD, and ~65 kD, named gp33, gp40, and gp65, respectively. Peptide sequence similarity to known proteins are discussed.

METHODS

Nuclear Protein Isolation

Nuclei were isolated in the presence of 0.6% Triton X-100 from 6-7-day *Nicotiana tabacum* suspension-cultured cells, as described in Chapter 2 (Heese-Peck *et al.*, 1995) with the following modifications. Protoplasts were incubated in nuclear isolation buffer (NIB) plus 0.01% Triton X-100 (Boehringer Mannheim) for 10 min, lysed and centrifuged for 5 min onto a cushion of 50% Percoll in NIB at 1000x*g*. The dense yellow-brown band on top of the Percoll cushion was designated the crude nuclear fraction and the clear yellowish fraction overlaying the band was designated the cytosol I fraction. Crude nuclei were incubated with NIB plus 0.6% Triton X-100 for 20 min and filtered twice through 20 μm nylon mesh. The filtrate was then loaded onto percoll gradients consisting of 1 mL of 50% Percoll in NIB and 7 mL 10% Percoll in NIB. A centrifugation step (1000x*g*, 10 min) yielded a yellow-gray band directly on top of the 50% percoll cushion (purified nuclear fraction) that was overlaid by a clear light yellow cytosolic fraction II. In general, 4-6 x 10^8 nuclei were obtained. Aliquots of each cell fraction were immediately frozen in liquid N_2 and stored at -80°C. For further nuclear protein isolation, purified nuclei (N) were incubated with DNase I (Boehringer Mannheim), separated into a soluble (S1) and insoluble (P1) fraction, and P1 was then incubated in 0.5 M NaCl, 50 mM Hepes, pH 6.8 (2 μL per 10^6 nuclei) followed by a centrifugation step to yield a soluble (S2) and insoluble (P2) fraction (Heese-Peck *et al.*, 1995). Protein concentrations were determined using Bradford Protein Assay (BioRad). In general, 3-3.5 μg protein/μL were found to be present in S2. Proteinase inhibitors were present during all steps (Heese-Peck *et al.*, 1995).

Protein Blot Analysis

Proteins were separated on SDS-polyacrylamide gels, transferred to Immobilon-P membranes (Millipore), blocked overnight in 3% BSA in Tris-buffered saline (TBS) plus 0.1% Tween 20 (TBST), and further processed according to standard procedures (Harlow and Lane, 1988). Where indicated, 1μL/mL *Erythrina cristagalli* agglutinin-alkaline phosphatase (ECA-AP; EY Laboratories, San Mateo, CA) was used as a probe and processed according to standard procedures (Harlow and Lane, 1988).

Galactosyltransferase labeling assays

In vitro GalTF labeling of salt extracted proteins (S2) was performed essentially as described by Roquemore *et al.* (1994), using 100 milliunits of GalTF and either 4 μCi of UDP-^{3}H-galactose (1.0 mCi/mL; 17.3 Ci/mmol; Amersham) or 500 μM of UDP-galactose (Sigma) in a 100-μL reaction. After incubation at 37°C for 2-3 hrs, samples were cleared from precipitated proteins by a centrifugation step at 12,000x*g* for 5 min at room temperature into a soluble (S3) and an insoluble (P3) fraction. Proteins were separated on 9% SDS-polyacrylamide gels, stained by Coomassie blue-G or transferred to Immobilon-P membranes (Amersham) according to standard methods (Harlow and Lane, 1988; see above). Blots containing ^{3}H-galactose labeled proteins were dried, sprayed with EN^{3}HANCE (New England Research Products, Boston, MA), and then dried. Radiolabeled proteins were visualized by fluorography using ReflectionA film (New England Research Products) at -80°C. Blots containing proteins labeled with non-radioactive galactose were probed with ECA-AP and processed as described above.

***Erythrina cristagalli* agglutinin (ECA) affinity chromatography**

Salt extracted nuclear proteins (S2; ~ 500 μg) were galactosylated *in vitro* in the presence of non-radioactive UDP-galactose and GalTF as described above. In a control reaction, proteins were incubated with non-radioactive UDP-galactose but in the absence of the GalTF. ECA affinity chromatography was based on Miller and Hanover (1994) with the following modifications. Prior to loading onto the lectin columns, cleared supernatants were diluted 1:3 with 20 mM Hepes-KOH, pH 7.3, 5 mM $MgCl_2$, 2 mM ß-mercapthoethanol to lower the salt concentration in the samples. The following steps were performed at 4°C. All buffers contained a proteinase inhibitor mix as described for nuclear protein isolation. Diluted protein samples were batch-bound for 1-2 hr to 0.5 mL ECA-agarose columns (EY-Laboratories, San Mateo, CA) which were first equilibrated with 10 bed volumes of column wash buffer (CWB; 20 mM Hepes-KOH, pH 7.3, 5 mM $MgCl_2$, 100 mM NaCl, 2 mM ß-mercapthoethanol). After unbound proteins were collected, the columns were washed extensively with CWB (15-20 bed volumes) followed by two batch washes with CWB containing 0.1 M mannose (2 bed volumes) for 20 min each. Bound proteins were batch-eluted twice with elution buffer (0.05 M lactose in CWB; 2 bed volumes) followed by batch-elution with 20 mM Hepes-KOH, pH 7.3, 1 M NaCl (2 bed volumes) for 30 min each. Proteins from different column fractions were precipitated with 10% trichloroacetic acid followed by an acetone wash and analyzed by protein blot analysis as described. In some cases, protein blots were incubated with diluted Coomassie Brilliant Blue-R250 stain.

Peptide sequence analysis

For peptide sequencing, proteins were extracted from 3.5 x 10^9 nuclei, galactosylated *in vitro* in the presence of non-radioactive UDP-galactose, and galactosylated proteins were isolated by ECA affinity chromatography using 16 columns (1 mL bed volume each) as described above. Proteins eluted in the presence of lactose and 1 M NaCl were separated on 7.5 - 15% SDS-polyacrylamide gels and transferred to Immobilon P^{SQ} membranes (Millipore) in 25 mM Tris base, 192 mM glycine, 0.01% SDS, 10% methanol by stepwise electroblotting (0.6 mA/cm^2, 1 mA/cm^2, 2 mA/cm^2, 15 min each) and using a semi-dry protein electroblotting apparatus (Model 6000, E&K Scientific Products, Saratoga, CA). Blots were washed in ddH_2O, and proteins were detected in 0.1% Amido black (Sigma), 45% methanol, 2.5% acetic acid. Horizontal strips corresponding to gp33, gp40, and gp65 were excised and washed extensively in ddH_2O. The following procedure was performed by Dr. William S. Lane and personnel of the Harvard Microchemistry Facility for HPLC, mass spectrometry and peptide sequencing (Harvard University, MA). Strips containing proteins were subjected to *in situ* digestion with trypsin (Fernandez *et al.*, 1994). The resulting peptide mixture was separated by microbore high performance liquid chromatography using a Zorbax C18 1.0 x 150 mm reverse-phase column on a Hewlett-Packard 1090 HPLC/1040 diode array detector. Optimum fractions from the chromatogram were chosen based on differential UV absorbance at 205 nm, 277 nm and 292 nm, peak symmetry and resolution. Peaks were further screened for length and homogeneity by matrix-assisted laser desorption time-to-flight mass spectroscopy (MALDI-MS) on a Finnigan Lasermat 2000 (Hemel England). Selected fractions were submitted to automated Edman degradation on an Applied

Biosystems 494A, 477A (Foster City CA) or Hewlett Packard G1005A (Palo Alto CA). Details of strategies for the selection of peptide fractions and their microsequencing have been previously described (Lane *et al.*, 1991).

RESULTS

As an initial step to identify plant NPC proteins, we have previously demonstrated that glycoproteins modified by GlcNAc are present at the NPC of *Nicotiana tabacum* suspension-cultured cells. Most of these glycoproteins are partially extracted with 0.5 M salt and thus display similar properties as described for vertebrate NPC proteins. As determined by *in vitro* galactosyltransferase (GalTF) assays, the majority of these glycoproteins are modified by O-linked oligosaccharides with terminal GlcNAc (Chapter 2; Heese-Peck *et al.*, 1995). In the *in vitro* GalTF assays, the enzyme GalTF adds galactose (Gal) specifically to terminal, but not internal GlcNAc residues to generate glycoproteins with terminal galactosyl-ß1,4-*N*-acetylglucosamines (Gal-ß1,4-GlcNAc; Roquemore, *et al.*, 1994). This assay is therefore specific for detection of proteins with terminal GlcNAc residues and was used to purify tGlcNAc-proteins present at the tobacco NPC.

Nuclei were isolated from tobacco suspension-cultured cells in the presence of 0.6% Triton X-100 by Percoll gradients. High-detergent concentration leads to the lysis of contaminating organelles including plastids, mitochondria and vesicles (Willmitzer and Wagner, 1981; Heese-Peck *et al.*, 1995). Under these conditions, the endoplasmic reticulum (ER) and the NE are also removed. Nevertheless, the nuclei stay intact because the nuclear integrity does not rely on the NE (Heese-Peck, *et al.*, 1995; Masuda *et al.*, 1997). Purified nuclei were treated with DNase I and then incubated with 0.5 M salt buffer to extract tGlcNAc-proteins as described by Heese-Peck *et al.* (1995). This solubilized fraction containing unmodified tGlcNAc-proteins (Figure 3.1; Lanes 1 and 5) was incubated with radiolabeled UDP-^{3}H-Gal in the presence of GalTF

(Figure 3.1, Lanes 2 and 6). As previously shown, after separation of the proteins on SDS-polyacrylamide gels and subsequent transfer to membranes, at least eight tGlcNAc-proteins were labeled with ^{3}H-Gal and detected by fluorography (Figure 3.1, Lane 6; Heese-Peck *et al.*, 1995). The *in vitro* GalTF labeling assay was performed at 37°C, at which some proteins precipitated out of solution. The precipitated proteins (Figure 3.1, Lanes 3 and 7) were separated from the soluble proteins (Figure 3.1, Lanes 4 and 8) by centrifugation, and nearly all tGlcNAc-proteins were found in the soluble fraction (Figure 3.1; Lane 8). Thus, incubation at 37°C provided a convenient step to enrich for tGlcNAc-proteins, as it was particularly obvious when the insoluble (containing the precipitated proteins) and soluble protein pattern were compared using Coomassie Brilliant Blue R 250 staining (Figure 3.1; Lanes 3 and 4, respectively).

Galactosylation of the tGlcNAc-proteins with radiolabeled ^{3}H-Gal was not practical for large scale protein purification. Therefore, the tGlcNAc-proteins were labeled with non-radioactive Gal as previously described by Miller and Hanover (1994). We then examined whether the galactosylated tGlcNAc-proteins could be detected by *Erythrina crystagalli* agglutinin (ECA), a lectin shown to be specific for Gal-ß1,4-GlcNAc (Lis *et al.*, 1985). Galactosylated tGlcNAc-proteins were visualized using ECA coupled to alkaline phosphatase (ECA-AP; Figure 3.1, Lanes 9-12), and proteins with similar apparent molecular mass to those visualized by fluorography were detected by the ECA-probe (Figure 3.1; compare Lanes 6 and 10 as well as 8 and 12, respectively). One additional glycoprotein with an apparent molecular mass of 35 kD was only detected by the ECA-probe (Figure 3.1, Lanes 9-12). Because this protein was also detected by ECA-AP prior to *in vitro* GalTF labeling (Figure 3.1, Lane 9; see below),

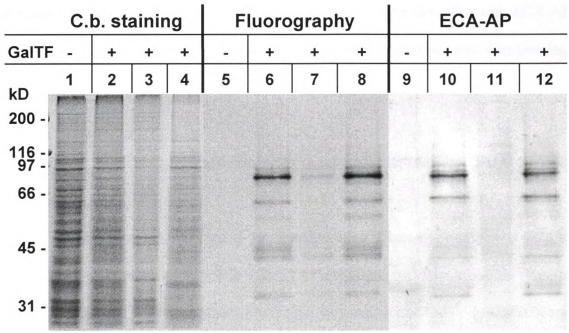


Figure 3.1

Detection of galactosylated tGlcNAc-proteins using fluorography and ECA blot analysis.
tGlcNAc-proteins were extracted with 0.5 M salt from DNase I treated tobacco nuclei. Extracted proteins were either not further modified (GalTF -; Lanes 1, 5, and 9) or subjected to GalTF assays in the presence of UDP-^{3}H-galactose (GalTF +; Lanes 2-4, 6-8) and non-radioactive UDP-galactose (GalTF +; Lanes 10-12). Following GalTF assay, total proteins (Lanes 2, 6, 10) were separated into an insoluble (Lanes 3, 7, 11) and a soluble (Lanes 4, 8, 12) fraction by centrifugation. After separation on SDS-polyacrylamide gels, total proteins were detected by Coomassie-blue (Lanes 1-4), or transferred to Immobilon-P (Lanes 5-12), and galactosylated tGlcNAc-proteins were visualized by fluorography (Lanes 5-8) or by protein blot analysis using ECA-AP (Lanes 9-12). Molecular mass standards are indicated at left in kilodaltons (kD). C.b., Coomassie blue; GalTF, galactosyltransferase.

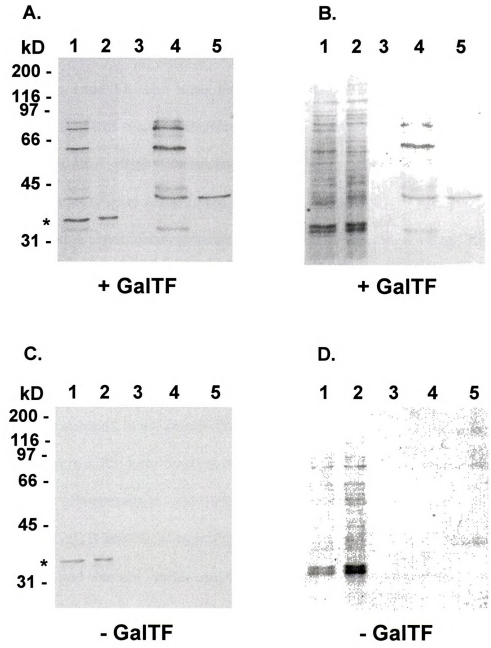
it was not considered a tGlcNAc-protein. As defined above (see Introduction), tGlcNAc-proteins required the addition of galactose in the *in vitro* GalTF assay for subsequent detection with ECA-AP. Taken together, these results indicated that labeling of tGlcNAc-proteins with non-radioactive Gal and subsequent detection with ECA-AP provided an excellent alternative method for the detection of glycoproteins with terminal GlcNAc.

Isolation of nuclear glycoproteins with terminal GlcNAc by ECA affinity chromatography

The ability of ECA to bind to the galactosylated tGlcNAc-proteins was used to purify these glycoproteins by lectin affinity chromatography. Salt-extracted proteins were galactosylated with non-radiolabeled Gal in the presence of GalTF (+GalTF), and soluble proteins were loaded onto ECA-agarose columns (Figure 3.2A; Lane 1). Different column fractions were collected, and the presence of the tGlcNAc-proteins was determined by protein gel blot analysis using ECA-AP as a probe (Figure 3.2A). With the exception of a 35-kD glycoprotein (Figure 3.2A; astrix; see below), the tGlcNAc-proteins bound to the ECA-agarose columns and only small portions of the tGlcNAc-proteins were found in the eluate (Figure 3.2A, Lane 2). The tGlcNAc-proteins were eluted with the specific sugar lactose (Gal-ß1,4-Glu; Figure 3.2A; Lane 4), a disaccharide closely related to Gal-ß1,4-GlcNAc and shown to be a competitor ofGal-ß1,4-GlcNAc binding (Lis *et al.*, 1985). These glycoproteins were not eluted with the unrelated sugar mannose (Man; Figure 3.2A; Lane 3) demonstrating the specific interaction of the galactosylated tGlcNAc-proteins with the ECA-agarose.

Figure 3.2

Purification of galactosylated tGlcNAc-proteins by ECA affinity chromatography. Salt extracted nuclear proteins were subjected to nonradioactive GalTF assay in the presence (A, B) or absence (C, D) of GalTF and separated into insoluble and soluble fractions. The soluble fractions (Lane 1) were loaded onto ECA agarose columns, and column fractions were collected as Lane 2, eluate; Lane 3, 0.1 M mannose wash; Lane 4, 0.05 M lactose eluate; Lane 5, 1 M NaCl eluate. All fractions were subjected to protein gel blot analysis using ECA-AP (A, C). After detection with ECA-AP, the total proteins were visualized by immersing the blots in Coomassie blue (B, D). Molecular mass standards are indicated in kilodaltons (kD). GalTF, galactosyltransferase.


A.
kD 1 2 3 4 5
200 -
116 -
97 -
66 -
45 -
*
31 -
+ GalTF
B.
1 2 3 4 5
+ GalTF
C.
kD 1 2 3 4 5
200 -
116 -
97 -
66 -
45 -
*
31 -
- GalTF
D.
1 2 3 4 5
- GalTF

Tight interaction between glycoproteins and the lectin was further disrupted by 1 M salt. As shown in Figure 3.2A (Lane 5), one GlcNAc-protein with an apparent molecular mass of ~40 kD, designated gp40, was not only present in the lactose but also in the 1 M salt eluate indicating that gp40 bound tightly to the ECA-agarose (Figure 3.2A; Lane 5). The same blot was subsequently stained with Coomassie Brilliant Blue R 250 to visualize all proteins in each column fractions (Figure 3.2B) and confirmed that ECA affinity chromatography served as an efficient means to purify tGlcNAc-proteins (Figure 3.2B, Lanes 4 and 5).

In control experiments, proteins were treated as described, except that GalTF was omitted (-GalTF; Figure 3.2C and D). No proteins were present in the lactose and 1 M salt eluted fractions (Figure 3.2C and D; Lanes 4 and 5, respectively). These results further demonstrated that the binding of galactosylated tGlcNAc-proteins to the ECA-agarose columns was specific. Only one protein with an apparent molecular mass of 35 kD was detected in the protein fraction loaded onto the ECA agarose column by ECA-AP (Figure 3.2C; Lane 1; astrix), but it did not bind to the column (Figure 3.2C; Lane 2). This glycoprotein is most likely the 35 kD-protein that was detected by ECA-AP prior to GalTF labeling assays (Figure 3.1; Lane 9 and Figure 3.2A; Lanes 1 and 2; see astrix) and was not further pursued.

Large scale purification of tGlcNAc-proteins for internal amino acid analysis

To obtain sufficient material for internal amino acid sequence analysis, large scale purification of tGlcNAc-proteins was performed using ECA-affinity column chromatography. A total of 3.5 x 10^9 isolated nuclei were used as starting material.

DNase I treated nuclei were incubated with 0.5 M salt buffer to extract tGlcNAc-proteins. The solubilized fraction was then subjected to non-radioactive *in vitro* GalTF assay followed by ECA affinity chromatography as described above. Proteins eluted in the presence of lactose as well as 1 M salt were separated on 7.5-15% gradient SDS-polyacrylamide gels, transferred to membranes, and visualized by amido-black staining.

Peptide sequences of gp33, gp40 and gp65

For internal amino acid analysis, sufficient material was recovered from three tGlcNAc-proteins with apparent molecular masses of ~33 kD, ~40 kD, and ~65 kD, referred to as gp33, gp40, and gp65, respectively. More specifically, gp33 and gp65 were isolated from the lactose eluate, and gp40 from the 1 M salt eluate. The subsequent tryptic digestion and peptide analysis were performed by Dr. William Lane and personnel at the Harvard Microchemistry Facility.

For each of these tGlcNAc-protein, amino acid sequence information was obtained from three peptides as shown in Figure 3.3. The peptides were compared to sequences of known proteins and expressed sequence tags from eukaryotes and prokaryotes present in available databases. Initial database searches did not reveal convincing information on the identity of the tGlcNAc-proteins (see below). Therefore, we employed polymerase chain reaction (PCR) to obtain clones encoding these proteins. Degenerate oligonucleotide primers were designed to the different peptide sequences in either sense or antisense directions (data not shown; for gp40-peptides, see Chapter 4).

protein	peptide	amino acid sequence
gp33	PT35	T T G T V A S F E T R
	PT84	T R P L P A D
	PT118	V P P A A V V A A F N S Q L P G T Q S I A T
gp40	PT48	I G G A Q F T L - G T H[Y]
	PT64	N T P Y D F L K[P][R]
	PT72	I G D V V L G Y[D](T)
gp65	PT56	T S L (S/D) E Y E P[L][P]/(D) R
	PT63	V D L I N G L K
	PT90	T N V V P F Y P D T M I R

Figure 3.3

Amino acid sequences of peptides obtained from gp33, gp40, and gp65.
All amino acids were determined with high confidence except those indicated by the following symbols: -, no amino acid determination; [ ], probable/reasonable confidence; ( ) possible/low confidence; /, relative assignment to major or minor residue was not feasible.

The primers were then used in all possible permutations to amplify specific PCR-fragments using a cDNA library made from poly(A)$^+$ RNA of tobacco suspension-cultured cells (Facchini and Chappell, 1992) as template. The PCR approach did not lead to the isolation of specific PCR-fragments encoding gp33 and gp65 (data not shown). As described in detail in Chapter 4, this approach was successful in identifying a specific PCR fragment encoding a portion of gp40.

DISCUSSION

In the present studies, we successfully developed a purification scheme to isolate nuclear glycoproteins modified with terminal GlcNAc. Labeling these tGlcNAc-proteins with Gal *in vitro* yielded glycoproteins with terminal Gal-ß1,4-GlcNAc. Due to the specific affinity of ECA to Gal-ß1,4-GlcNAc, lectin affinity chromatography provided an efficient means to purify the modified tGlcNAc-proteins in large enough quantities and to obtain amino acid sequence information. Thus, this procedure may be employed in the isolation of other tGlcNAc-proteins in the future. Amino acid sequence information allowed us to compare the peptide sequences to known proteins and ESTs in available databases and to design experiments to isolate genes encoding the tGlcNAc-proteins.

gp33

Two peptides obtained from gp33, designated gp33-PT35 and gp33-PT84, were 100% identical to regions of *Erythrina corallodendron* lectin (ECorL), a protein closely related to ECA (Arango *et al.*, 1990; Figure 3.4A). This lectin has a similar apparent molecular mass (36 kD) to gp33. It is likely that ECA may have not been properly linked to the ECA agarose (obtained from EY-Laboratory) and contaminated the isolated gp33 fraction.

The third peptide, gp33-PT118, did not show any sequence similarity to the deduced amino acid sequence of ECorL, demonstrating the presence of a second protein (gp33) other than ECorL. Initial analysis of gp33-PT118 revealed only low sequence similarity to a variety of different proteins (data not shown).

A.

```
gp33-PT35                1   T T G T V A S F E T R   11
                             * * * * * * * * * * *
ECorL                   89   T T G T V A S F E T R   99


gp33-PT84                1   T R P L P A D     7
                             * * * * * * *
ECorL                  109   T R P L P A D   115
```

B.

```
gp33-PT118           1  V P A A V V A A F N S Q L P G T Q S I A T   21
                        * * + + * + * * * * * * * * * * * * * *
germin-like prot.  148  K P A S V I S A F N S Q L P G T Q S I A A  167
(U95036)
```

Figure 3.4

A. Amino acid similarity between gp33-PT35 and gp33-PT84 peptides and *Erythrina corallodendron* lectin (ECorL).
B. Amino acid similarity between gp33-PT118 and one of the germin-like protein from *A. thaliana* (Accession number: U95036).
* and + , identical and conserved amino acids, respectively.

Searches of the *Arabidopsis* database performed in the past year showed that 16 of 21 amino acids were identical to a region in germin-like proteins from *Arabidopsis* (Figure 3.4B). In general, germin-like proteins are found in a variety of different plant species and have an apparent molecular mass of 25-30 kD (Ono *et al.*, 1996, and references within). As the name indicates, germin-like proteins show sequence similarity to germins, proteins originally identified during seed germination in cereals (Lane, 1994). So far, the function of germin-like proteins is unknown. Examining different organisms, these proteins do not share common expression pattern (Berna and Bernier, 1996 and references within). Some are expressed during germination, others in roots, leaves, or embryos. While some germin-like proteins are produced in response to auxin, osmotic stress, or pathogen invasion, others are regulated during the circadian cycle. In future experiments, genes encoding germin-like proteins and antibodies against α-germin-like protein can be used as heterologous probes to determine whether gp33 is indeed a germin-like protein.

gp40

The three peptide fragments (gp40-PT48, gp40-PT64, gp40-PT72) obtained from gp40 had no similarity to any known protein in the database indicating that gp40 may be a novel protein. The isolation of the gene encoding gp40 and further characterization of this protein will be described in Chapter 4.

gp65

As shown in Figure 3.5, eight of 13 amino acids of gp65-PT90 were identical to a region in the A6 gene from *B. napus* and *A. thaliana* (Hird *et al.*, 1993). The A6 gene encodes a protein with a predicted mass of 53 kD that is specifically expressed in anthers. The function of this protein is unknown, but it appears to have low sequence similarity to ß-1,3-exoglucanases. These exoglucanases have a much smaller apparent molecular mass (36 kD), and so far, no glucanase activity has been observed for A6 (Hird *et al.*, 1993). gp65-PT90 showed low similarity to ß-1,3-exoglucanases (20-25%, data not shown). Whether gp65 is indeed a A6-protein remains to be determined because the other peptides, gp65-PT56 and gp65-PT63, did not show any similarity to the A6 gene. Further, no sequence similarity of these peptides was observed to other known sequences in the databases.

```
gp65-PT90              1  T N V V P F Y P D T M I R   13
                          * * + + * + * *   *   * *
At A6                116  T N I L P Y Y P Q T Q I R  128
```

Figure 3.5

Amino acid similarity between gp65-PT90 and A6 from *A. thaliana* (At A6). The corresponding sequence of A6 from *B. napus* is identical to that shown for At A6. * and + , identical and conserved amino acids, respectively.

Since initial database searches did not allow us to obtain convincing information on the tGlcNAc-proteins, we employed polymerase chain reaction (PCR) amplification approaches to obtain genes encoding these proteins. The isolation of a specific PCR fragment encoding gp40 (see Chapter 4) demonstrated the validity of this approach. During the course the cDNA library screening (see Chapter 4), we discovered that the cDNA library was biased towards inserts smaller than 1 kb. Thus, it was likely that this library did not contain the complete genome of tobacco suspension-cultured cells. This observation may explain why we were not able to obtain specific PCR-fragments encoding gp33 and gp65 (data not shown). It should be mentioned, however, that this library was the only cDNA library made from tobacco suspension-cultured cells available to us. The construction of a new tobacco cDNA libraries may be necessary to isolate genes encoding gp33 and gp65. In addition, 5' rapid amplification of cDNA ends may provide another way to identify these genes.

References

Arango, R., Rozenblatt, S., and Sharon, N. (1990). Cloning and sequence analysis of the *Erythrina corallodendron* lectin cDNA. FEBS **264,** 109-111.

Bednarek, S.Y., Wilkins, T.A., Dombrowski, J.E., and Raikhel, N.V. (1990). A carboxyl-terminal propeptide is necessary for proper sorting of barley lectin to vacuoles of tobacco. Plant Cell **2**, 1145-1155.

Berna, A., and Bernier, F. (1997). Regulated expression of a wheat germin gene in tobacco: oxalate oxidase activity and apoplastic localization of the heterologous protein. Plant Mol. Biol. **33,** 417-429.

Davis, L. (1995). The nuclear pore complex. Annu. Rev. Biochem. **64,** 865-896.

Fernandez, J., Andrews L., and Mische, S.M. (1994). An improved procedure for enzymatic digestion of polyvinylidiene difluoride-bound proteins for internal sequence analysis. Anal. Biochem. **218,** 113-117.

Forbes, D.J. (1992). Structure and function of the nuclear pore complex. Annu. Rev. Cell Biol. **8**, 495-527.

Goldberg, M.W., and Allen, T.D. (1995). Structural and functional organization of the nuclear envelope. Curr. Opin. Cell Biol. **7,** 301-309.

Harlow, E., and Lane, D. (1988). Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Hart, G.W. (1997). Dynamic O-linked glycolsylation of nuclear and cytoskeletal proteins. Annu. Rev. Biochem. **66,** 315-335.

Heese-Peck, A., Cole, R.N., Borkhsenious, O.N., Hart, G.H., and Raikhel, N.V. (1995). Plant nuclear pore complex proteins are modified by novel oligosaccharides with terminal N-acetylglucosamine. Plant Cell **7,** 1459-1471.

Hicks, G.R., and Raikhel, N.V. (1995). Protein import into the nucleus: An integrated view. Annu. Rev. Cell Dev. Biol. **11,** 155-188.

Hicks, G.R., Smith, H.M.S., Lobreaux, S., and Raikhel, N.V. (1996). Nuclear import in permeabilized protoplasts from higher plants has unique features. Plant Cell **8,** 1337-1352.

Hird, D.L., Worrall, D., Hodge, R., Smartt, S., Paul, W., and Scott, R. (1993) The anther-specific protein encoded by the *Brassica napus* and *Arabidopsis thaliana* A6 gene displays similarity to b-1,3-glucanases. Plant J. **4,** 1023-1033.

Jordan, E.G., Timmis, J.N., and Trewavas, A.J. (1980). The plant nucleus. In Tolbert, N.E. (ed.), The Biochemistry of Plants, Academic Press, Inc., Vol. 1, pp. 490-588.

Lane, B.G. (1994). Oxalate, germin, and the extracellular matrix of higher plants. FASEB J. **8,** 294-301.

Lane, W.S., Galat, A., Harding, M.W., and Schreiber, S.L. (1991). Complete amino acid sequence of the FK506 and rapamycin binding protein, FKBP, isolated from calf thymus. J. Prot. Chem. **10,** 151-160.

Lis, H., Joubert, F.J., Sharon, N. (1985). Isolation and properties of *N*-acetyllactosamine-specific lectins from nine *Erythrina* species. Phytochemistry **624,** 2803-2809.

Masuda, K., Xu, Z.-J., Takahashi, S., Ito, A., Ono, M., Nomura, K. and Inoue, M. (1997). Peripheral framework of carrot cell nucleus contains a novel protein predicted to exhibit a long alpha-helical domain. Exp. Cell Res. **232,** 173-181.

Merkle, T., Leclerc, D., Marshallsay, C., and Nagy, F. (1996). A plant *in vitro* system for the nuclear import of proteins. Plant J. **10,** 1177-1186.

Miller, M.W., and Hanover, J.A. (1994). Functional nuclear pores reconstituted with ß-1,4 galactose-modified *O*-linked *N*-acetylglucosamine glycoproteins. J. Biol. Chem. **269**, 9289-9297.

Ono, M., Sage-Ono, K., Inoue, H. and Harada, H. (1996). Transient increase in the level of mRNA for a germin-like protein in leaves of the short-day plant *Pharbitis nil* during the photoperiodic induction of flowering. Plant Cell Physiol. **37,** 855-861.

Panté, N., and Aebi, U. (1993). The nuclear pore complex. J. Cell Biol. **122**, 977-984.

Roquemore, E.P., Chou, T.-Y., and Hart, G.W. (1994). Detection of O-linked *N*-acetylglucosamine (O-GlcNAc) on cytoplasmic and nuclear proteins. Methods in Enzymology **230**, 443-460.

Willmitzer, L., and Wagner, K.G. (1981). The isolation of nuclei from tissue-cultured plant cells. Exp. Cell Res. **135**, 69-77.

Chapter 4

gp40, a tobacco glycoprotein modified with terminal *N*-acetylglucosamine and localized to the nuclear periphery, shows sequence similarity to aldose-1-epimerases

ABSTRACT

Macromolecules, such as proteins and different RNA species, move in and out of the nucleus through nuclear pore complexes (NPCs). Several glycoproteins are present at the NPC of tobacco suspension-cultured cells that are modified by O-linked oligosaccharides with terminal *N*-acetylglucosamine (GlcNAc). Some of these glycoproteins, referred to as tGlcNAc-protein, have been previously purified using *in vitro* GalTF assays and subsequent *Erythrina crystagalli* agglutinin affinity chromatography. Peptide sequence information was derived from one tGlcNAc-protein with an apparent molecular mass of ~40 kD, named gp40. In the current studies, we isolated the gene encoding gp40 and investigated the cellular localization of the protein produced. Interestingly, gp40 has 28 to 34% amino acid identity to aldose-1-epimerases from bacteria. No gene encoding an aldose-1-epimerase has been previously isolated from higher organisms. Polyclonal antibodies were generated against recombinant gp40. Consistent with its purification as a putative NPC protein, gp40 was associated with the nuclear periphery as shown by biochemical fractionation and immunofluorescent microscopy studies.

ABSTRACT

Macromolecules, such as proteins and different RNA species, move in and out of the nucleus through nuclear pore complexes (NPCs). Several glycoproteins are present at the NPC of tobacco suspension-cultured cells that are modified by O-linked oligosaccharides with terminal *N*-acetylglucosamine (GlcNAc). Some of these glycoproteins, referred to as tGlcNAc-protein, have been previously purified using *in vitro* GalTF assays and subsequent *Erythrina crystagalli* agglutinin affinity chromatography. Peptide sequence information was derived from one tGlcNAc-protein with an apparent molecular mass of ~40 kD, named gp40. In the current studies, we isolated the gene encoding gp40 and investigated the cellular localization of the protein produced. Interestingly, gp40 has 28 to 34% amino acid identity to aldose-1-epimerases from bacteria. No gene encoding an aldose-1-epimerase has been previously isolated from higher organisms. Polyclonal antibodies were generated against recombinant gp40. Consistent with its purification as a putative NPC protein, gp40 was associated with the nuclear periphery as shown by biochemical fractionation and immunofluorescent microscopy studies.

INTRODUCTION

The nuclear pore complex (NPC) is a proteinaceous supramolecular structure of ~ 125 MD. It controls the movement of macromolecules such as proteins and different RNA species in and out of the nucleus (Davis, 1995; Görlich and Mattaj, 1996). Overall, little is known about the structure and composition of plant NPCs, most of which has been obtained from microscopic and biochemical studies. The morphology of plant NPC appears to be similar as those described for amphibian NPCs (Chapter 1 and 2; for review, see Jordon *et al.*, 1980; Panté and Aebi, 1993; Goldberg and Allen, 1995). Further, using an antibody raised against the yeast NPC protein Nsp1p (Hurt, 1988), a 100-kD protein is recognized in enriched nuclear matrix preparation of carrot-suspension cultured cells and is located at the NPC (Scofield *et al.*, 1992). Studies by Hicks and Raikhel (1993; 1995a) show that proteins are present at the NE and NPCs from isolated monocot and dicot nuclei that can bind to nuclear localization signals (NLSs) *in vitro*. Consistent with these results, the plant homolog of the NLS-receptor complex protein importin α (Impα), is not only found in the nucleo- and cytoplasm, but also accumulates at the NE (Smith *et al.*, 1997).

Using wheat germ agglutinin (WGA), a lectin that binds specifically to *N*-acetylglucosamine (GlcNAc) as a probe, glycoproteins modified by GlcNAc are found at the plant nuclear periphery (Chapter 2; Heese-Peck *et al.*, 1995; Hicks *et al.*, 1996; Merkle *et al.*, 1996). As shown by EM and biochemical analyses, at least some of these glycoproteins are located at the NPC of tobacco suspension-cultured cells and are modified by terminal GlcNAc (Chapter 2; Heese-Peck *et al.*, 1995). Most glycans with

terminal GlcNAc are attached to the proteins via an O-linkage and are larger in size than five GlcNAc residues (Chapter 2; Heese-Peck *et al.*, 1995). Thus, this sugar modification is different from the single O-linked GlcNAc (O-GlcNAc) found on vertebrate NPC proteins (for review, see Forbes, 1992; Hicks and Raikhel, 1995b).

Using lectin affinity chromatography, we have previously purified several tGlcNAc-proteins and have obtained internal amino acid sequence information for three tGlcNAc-proteins, named gp33, gp40, and gp65 (Chapter 3). Here, we report the cloning of the gene encoding gp40 and the investigation of its subcellular localization. This tGlcNAc-protein showed significant sequence similarity to aldose-1-epimerases, and its possible function are discussed.

METHODS

Cloning and sequencing of gp40

Unless otherwise stated, all general molecular biology methods were performed as in Sambrook *et al.* (1989). A degenerate oligonucleotide primer (5'-**CCATCGAT**AAYAC NCCNTAYGAYTT-3'; with N representing A, C, G, T, and Y representing C and T), designated pt64s and based in part on the gp40-PT64 peptide sequence (NTPYDF), was synthesized containing a *Cla*I restriction site (shown in bold) at the 5'-end. The primers pt64s and a T7 promotor specific primer (Boehringer Mannheim), designated T7, were used for a polymerase chain reaction (PCR)-based amplification from a pcDNAII cDNA library (Invitrogen, San Diego) made from poly(A)$^+$ RNA from 4-h-elicitor-treated tobacco cells (Facchini and Chappel, 1992). Thermocycling was performed at 94°C for 1 min, 45°C for 1 min, 72°C for 2 min, for 30 cycles using 150 ng of cDNA template, 50 nM of pt64s, and 5 nM T7. One μL of the PCR reaction was then used as the template for a reamplification reaction under the same conditions described. The products were separated by agarose gel electrophoresis, recovered by electrophoresis onto NA-45 paper (Schleicher & Schüler), digested with *Cla*I and *Xba*I and ligated into pBluescript SK$^-$ (pBS-SK$^-$). Sequencing was performed according to manufacture's specification (Promega). Three of five independent clones, designated p64s/8, p64s/9, p64s/H, were identified that contained an ~ 600-bp fragment and encoded the entire peptide gp40-PT64.

Complementary DNA library screening and 5'-rapid amplification of cDNA ends

A *Cla*I-*Xba*I fragment isolated from p64s/8 was ^{32}P-labeled by the random primer method of Feinberg and Vogelstein (1985) and used as a probe to screen the pcDNAII cDNA library under high stringency conditions. Final wash conditions were done using 0.1 x SSC, 0.5% SDS at 65°C. Eight independent clones were isolated, partially sequenced, and sequence information of the clone pGP40-5e was used for further studies. To obtain the full length 5'end of the gene encoding gp40, 5'-rapid amplification of cDNA ends (RACE) was performed. Two non-degenerate oligonucleotide primers, designated G1a (5'-GGTGATGTGTGATCCAA-3') and G2a (5'-**CCATCGAT**CACCAATGTTCCAGTAA-3'; containing a *Cla*I site at the 5' end shown in bold), were synthesized based on the nucleotide sequence of pGP40-5e encoding the amino acid sequence 220-FGSHIT and 200-YWNIG, respectively. 5'-RACE was performed according to manufacture's specifications (GIBCO-BRL). First strand cDNA was synthesized from total RNA isolated from 3-d *N. tabacum* suspension-cultured cells (Puissant and Houdeline, 1990) using G1a (400 nM) and SuperScript^{TM}II (GIBCO-BRL). PCR amplification was performed using 5 μL of the dC-tailed cDNA, 200 μM dNTPs, 2 units *Pwo*I DNA Polymerase (Boehringer Mannheim), 0.5 units *Taq* DNA Polymerase (GIBCO-BRL), 400 nM nested primer G2a and 400 nM Abridged Anchor Primer (AAP; supplied by GIBCO-BRL). Thermocycling was done at 94°C for 30 sec, 43°C for 30 sec, 72°C for 2 min for 35 cycles followed by 72°C for 7 min. The products were ligated directly into the *Sma*I site of pBS-SK$^-$. Six independent clones, designated pRACE-3 to pRACE-8, were selected for sequencing.

Northern blot analysis

Total RNA was isolated from young leaves, old leaves, roots, and stems of 2 month-old *Nicotiana tabacum* cv Wisconsin 38 plants and from 2 to 8 d-old *Nicotina tabacum* suspension-cultured cells as described by Puissant and Houdeline (1990). For all RNA blots, 20 μg of total RNA was separated on a 1.2% agarose-formaldehyde gel. Ethidium bromide staining of the ribosomal bands was used to confirm equal loading of lanes. Gels were transferred to Hybond-N nylon membranes (Amersham, Il), and the RNA was crosslinked to the membrane using the UV Stratalinker 1800 (Stratagene) according to the manufacturer's specification. The 600-bp *Cla*I-*Xba*I fragment isolated from p64s/8 was ^{32}P-labeled by the random primer method (Feinberg and Vogelstein, 1985) and used as probe for Northern blot analysis. Final washes were done using 0.25 x SSC, 0.5% SDS at 60°C. Autoradiography was performed using Reflection NEF (DuPont) with amplification screens at -80°C.

Construction of the full length gp40 and gp40-GST fusion

Based on the sequence identity of the 5'RACE product present in pRACE-7 to the fragment in pGP40-5e in the overlapping region, the fragment of pRACE-7 was selected to be combined with the fragment of pGP40-5e to create pGP40-7/5e as follows. The approximately 900 bp *Xba*I and *Hind*III fragment was isolated from pGP40-5e and ligated into pBS-SK$^-$, designated p5e/SK$^-$. p5e/SK$^-$ and pRACE-7 were digested with *Xmn*I yielding each the following size fragments, 1.0 and 2.8 kb for p5e/SK$^-$, and 2.2 and 1.5 kb for pRACE-7. The 2.8 kb fragment of p5e/SK$^-$ and the 1.5 kb of pRACE-7 were isolated and ligated to give pGP40-7/5e containing the complete

ORF of gp40. Correct ligation of the two fragments was confirmed by sequencing across the *Xmn*I site of pGP40-7/5e. Further, both DNA strands of the combined insert were sequenced in the W.M. Keck facility (Yale University) using Taq FS DNA polymerase and fluorescent-dideoxy terminators in a cycle sequencing method. The resultant DNA fragments were electrophoresed and analyzed using an automated Applied Biosystems 373A Stretch DNA sequencer. PCR-based mutagenesis was used to introduce a *Xho*I site in frame upstream of the ATG (Met) of the open reading frame (ORF) of gp40-7/5e. Thermocycling was done for using 10 ng of template plasmid DNA (pGP40-7/5e), 400 nM dNTPs, 400 nM reverse primer (BMB), 400 nM 7pcrMs primer (5'-ATACA**CT*C*GAG**TTGACTATG-3'; with *Xho*I site shown in bold and mismatched nucleotide in italics), and 2.5 units *Pwo*I DNA polymerase (BMB) at 94°C for 45 sec, 45°C for 45 sec, 72°C for 3 min for 25 cycles followed by an extension reaction at 72°C for 7 min. A PCR product of 1.25 kb (encoding the full length gp40-7/5e) was isolated, digested with XhoI, and ligated into the XhoI site of pGEX-5X-1 (Pharmacia Biotech). In frame fusion of the gp40 ORF to glutathion-S-transferase (GST) was confirmed by sequence analysis. One clone, designated pGEX-D, was selected for overexpression of GST-gp40 fusion protein in *E. coli*.

Production of antibody against GST-gp40 fusion protein

E. coli strain BL21 (DE3) was transformed with pGEX-D (GST-gp40). Transformed cells were grown in 500 mL LB media to OD_{600} = 0.5-0.7 at 37°C, and proteins were induced with 0.2 mM isopropylthio-ß-D-galactose for 3 hrs. Cells were collected, lysed, and separated into soluble and insoluble fractions by centrifugation according to

Bar-Peled and Raikhel (1996). The insoluble fraction (containing GST-gp40 in inclusion bodies) was washed three times in 50 mM Hepes-KOH, pH 7.5, 10 mM $MgCl_2$, 25% sucrose, 1% Triton X-100, and by three washes in ddH_2O. Each wash step was followed by a centrifugation step at 15,000x*g* for 20 min at 4°C. After the final wash, the insoluble fraction containing the inclusion bodies was stored at -80°C. Proteins in inclusion bodies were separated by 10% SDS-polyacrylamide gels, stained in 0.05% Coomassie Blue-G in ddH_2O, and washed in ddH_2O. Horizontal gel strips containing GST-gp40 were excised, cut into small pieces, and stored at -80°C overnight. The fusion protein was eluted from the gel pieces by electroelution (Electro-Eluter Model 422, BioRad) using 25 mM Tris base, 192 mM glycine, 0.1% SDS at 10 mA/glass tube for 8 hr with one change of buffer. Solutions containing the eluted fusion protein were combined, lyophilized, and resuspended in less than 1 mL of ddH_2O. Protein concentration was approximated by comparison to a known amount of Broad Molecular Weight Standards (BioRad). Fusion protein (500-750 μg) was mixed with TiterMax (CytRx Corporation, Norcross, GA) and injected into a New Zealand White rabbit according to standard procedures (Harlow and Lane, 1988). Antibodies were made monospecific to gp40 by first depleting them of anti-GST antibodies as described by Bar-Peled and Raikhel (1996) and then by strip-purifying the depleted antibody against GST-gp40 fusion protein bound to Immobilon-P membranes (Harlow and Lane; 1988).

Nuclear Protein Isolation

Nicotiana tabacum suspension-cultured cells were maintained and subcultured, and nuclei were isolated in the presence of 0.6% Triton X-100 as in Chapter 3.

Protein Blot Analysis

Proteins were separated on SDS-polyacrylamide gels, transferred to Immobilon-P membranes (Amersham), blocked overnight in 3% BSA in Tris-buffered saline (TBS) plus 0.1% Tween 20 (TBST), and further processed according to standard procedures (Harlow and Lane, 1988). Where indicated, 1μL/mL *Erythrina cristagalli* agglutinin-alkaline phosphatase (ECA-AP; EY Laboratories, San Mateo, CA) was used as a probe. α-atIMPα (1:2,000 dilution; Smith *et al.*, 1997), α-GST (1:1000 dilution; M. Bar-Peled, unpublished), α-rat aldose-1-epimerase (1:500 dilution; Toyoda *et al.*, 1983), and α-gp40 (1:500 dilution; see below) were used as the primary antibodies. α-atIMPα and α-GST treated blots were incubated with goat-α-rabbit IgG-alkaline phosphatase (AP; 1:5000 dilution) and developed by standard procedures (Harlow and Lane, 1988). α-rat aldose-1-epimerase and α-gp40 treated blots were incubated with Protein A-horse radish peroxidase (HRP; Sigma; 1:7,500 dilution), and proteins were then visualized using a chemiluminescence kit according to the manufacturer's specification (Pierce; SuperSignal). Where specified, after detection by chemiluminescence, the protein blots were stripped as described in the manufacturer's protocol (Pierce) and reprobed with ECA-AP. The primary antibodies α-atIMPα were kindly provided by H.S.M. Smith, the α-rat aldose-1-epimerase by Dr. Y. Toyoda, and α-GST by Dr. M. Bar-Peled.

***Erythrina cristagalli* agglutinin (ECA) affinity chromatography**

Salt extracted nuclear proteins (S2; ~ 500 μg) were galactosylated *in vitro* with non-radioactive UDP-galactose, subjected to ECA affinity chromatography and protein blot analyses as described in Chapter 3. For analyses of ECA column fractions using α-rat

aldose-1-epimerase, twice the amount of proteins were loaded in the Lanes 1 (S3; column load) and 2 (unbound proteins).

Immunofluorescence Microscopy

N. tabacum cell suspension-cultured cells were protoplasted for 30 to 40 min, collected, spun onto lysine-coated slides (Shannon), and fixed in 3% paraformaldehyde (Smith *et al.*, 1997). About 2.5 x 10^6 protoplasts were used per slide to obtain relatively even distribution. For indirect immunofluorescent microscopy, labeling was performed according to Smith *et al.* (1997) with some modifications. The material was blocked in PBS, 0.05% Tween 20, 3% BSA for 30 min prior to the incubation with the primary antibody. Incubation of primary (α-gp40, 1:50 dilution; α-atIMPα; 1:12 dilution) and secondary antibodies (goat-α rabbit IgG-CY3; 1:20 dilution; Jackson) were done in the same blocking solution. After incubation with the secondary antibody, protoplasts were incubated with a DAPI/glycerol solution (Hicks *et al.*, 1996) and then washed extensively. Samples were examined with epifluorescent optics (Axiphot; Zeiss) using a 490-nm longpass filter for immunofluorescent labeling and a 365-nm bandpass filter for DAPI staining. Pictures were taken with p3200 TMAX film (Kodak). The same samples were then viewed by laser scanning microscopy as described by Smith *et al.* (1997). Images were digitized and prepared in Adobe Photoshop after removing scratches using the "dust and scratches" filter.

RESULTS

Cloning of the gene encoding gp40

The tGlcNAc-protein gp40 was purified by ECA affinity chromatography in large quantities for subsequent tryptic digestion (see Chapter 3). Internal amino acid sequence analysis yielded sequence information for three peptides, designated gp40-PT48, -PT64, and -PT72 (Chapter 3; Figure 3.3). These peptides were compared to sequences of known proteins and expressed sequence tags (ESTs) from eukaryotes and prokaryotes present in available databases; however, no sequence similarity was found. Therefore, polymerase chain reaction (PCR) amplification was employed to clone the gene encoding gp40 using degenerate oligonucleotide primers based on the information obtained from the gp40-peptides. Using a cDNA library made from poly(A)$^+$ RNA from *N. tabacum* suspension-cultured cells (Facchini and Chappell, 1992) as a template, a specific 600-bp PCR fragment was amplified using a pt64s primer (Fig. 4.1) and a T7 promotor specific primer present in the pcDNAII plasmid. After subcloning, three independent clones, designated p64s/8, p64s/9, and p64s/H, were isolated, and sequence analysis revealed that they were identical (data not shown). The PCR fragments also contained the nucleotide sequence encoding the entire peptide fragment gp40-PT64 confirming the identity of the PCR fragments (Figure 4.1); however, the remaining peptides gp40-PT48 and gp40-PT72 were not contained in this PCR fragment.

Screening of the tobacco cDNA library with the PCR fragment p64s/8 generated only partial clones with a size of about 0.75 to 0.95 kb (data not shown). Therefore, 5'-rapid amplification of cDNA ends (5'RACE) was employed to obtain the missing

gp40-PT64	N	T	P	Y	D	F	L	K	P	R
	AAY	ACN	CCN	TAY	GAY	TTY	YTN	AAR	CCN	MGN
	N	T	P	Y	D	F	L	K	P	R
p64s/8	AAT	ACC	CCA	TAC	GAC	TTC	TTG	AAA	CCC	CGT
p64s/9	AAT	ACC	CCG	TAC	GAC	TTC	TTG	AAA	CCC	CGT
p64s/h	AAC	ACC	CCG	TAT	GAC	TTC	TTG	AAA	CCC	CGT

Figure 4.1

The 5'end of the three specific PCR products, pt64s/8, pt64s/9, pt64s/H, were aligned with the degenerate nucleotide sequence of the peptide gp40-PT64 to confirm their identity. The nucleotide sequence of the primer pt64s used for PCR-based amplification is underlined. The nucleotide sequences are represented by the following codes: M = A or C; N = A, C, G, or T; R = A or G; Y = T or C.

5'end of the clone. Due to its proof-reading ability, *PwoI* DNA polymerase (supplemented with a low amount of *Taq* DNA polymerase) was used in the amplification reaction using nested oligonucleotide primers and total RNA isolated from 3-d old tobacco suspension-cultured cells as template. To confirm the sequence identity, nested primers were designed such that the 5'RACE product would have an 183-bp overlap with the 5'end of one of the truncated cDNA clones, GP40-5e (Figure 4.2; underlined; see Methods). A 750-bp fragment was amplified by 5'RACE, isolated, and subcloned for sequence analysis. The nucleotide and deduced amino acid sequences of three independent inserts (including RACE-7) were identical to the overlapping sequences of GP40-5e with the exception of a single nucleotide change (T for C at nt 169) that did not alter the deduced amino acid sequence (Figure 4.2). Significantly, the two peptides gp40-PT48 and gp40-PT72 were present in the deduced amino acid sequence of the 5'RACE product (Figure 4.2, marked in grey), confirming that the isolated gene indeed encodes gp40. The deduced amino acid sequence of the 5'RACE products differed from the peptide sequence of gp40-PT48 in one amino acid (I104T; see Figures 4.1 and 4.2), which may be due to an error during amino acid sequencing. It could not, however, be ruled out that gp40-PT48 was derived from another, very closely related protein with similar properties as gp40.

For further studies, the fragments RACE-7 and GP40-5e were combined using a *Xmn* I site to yield the full length clone GP40 (GenBank accession number: AF032386). The size of the combined fragment (1.25 kb) was consistent with the gp40 transcript size (see below; Figure 4.4).

```
AAAAAGTCTAAAATACACTAGAGTTGACTATGTCTTTAAAGATTAATTTGTTGGTCTGTTTGTTCATTTT  70
                              M  S  L  K  I  N  L  L  V  C  L  F  I  F

TCATCTGTTGGTTGCCGCTAGTGTTAGAGGGCACAAGATAGGGATCTATGAGATCAAGAAAGGAGATTTC 140
  H  L  L  V  A  A* S  V  R  G  H  K  I  G  I  Y  E  I  K  K  G  D  F

TCTGTTAAGATCACCAATTATGGTACCTCAATCATCTCTGTTCTTCTTCCTGATAAGCATGGAAAAATAG 210
 S  V  K  I  T  N  Y  G  T  S  I  I  S  V  L  L  P  D  K  H  G  K  ■

GTGATGTTGTTCTTGGATATGACACTATCGAGGAATACAAGAATGATACAAGTTATTTTGGAGCCACACT 280
G  D  V  V  L  G  Y  D  T  I  E  E  Y  K  N  D  T  S  Y  F  G  A  T  L

TGGAAGAGTTGCTAACCGGATTGGTGGTGCTCAATTTACTTTGAATGGAATCCATTATAAGCTTGTCCCC 350
  G  R  V  A  N  R  I  G  G  A  Q  F  T  L  N  G  I  H  Y  K  L  V  P

AATGAAGGCAAAAACATGTTGCATGGTGGCCCTAAAGGATTTAGCAAAGTTGTCTGGAAAGTAAGCAAGT 420
 N  E  G  K  N  M  L  H  G  G  P  K  G  F  S  K  V  V  W  K  V  S  K

ACGTGAAAGATGGTCCGTGTCCTTACATAACTCTAACCTACTACAGTGCTGATGGTGAAGAAGGATTTCC 490
Y  V  K  D  G  P  C  P  Y  I  T  L  T  Y  Y  S  A  D  G  E  E  G  F  P

TGGTGCTGTTCTTGCCTCTGTTACCTACACATTGAAAGATTCTTACAAACTTAGCGTGGTATTTAGGGCA 560
  G  A  V  L  A  S  V  T  Y  T  L  K  D  S  Y  K  L  S  V  V  F  R  A

AAAGCACTGAACAAGGCCACTCCAATTAACCTGTCTCACCACCCTTACTGGAACATTGGTGGTCACGACA 630
 K  A  L  N  K  A  T  P  I  N  L  S  H  H  P  Y  W  N  I  G  G  H  D

GTGGCGACGTCTTGTCCCAAGTTCTTCAAATTTTTGGATCACACATCACCCTAGTAGATAAACAGCTCAT 700
S  G  D  V  L  S  Q  V  L  Q  I  F  G  S  H  I  T  L  V  D  K  Q  L  I

TCCCACAGGCGAAATTGCCCCCATCAAGAACACGCCATATGATTTCTTGAAACCCCGTAAAGTTGGGAGC 770
  P  T  G  E  I  A  P  I  K  N  T  P  Y  D  F  L  K  P  R  K  V  G  S

AGGATCAATAAACTCAAAAATGGATATGACATCAACTATGTACTTGACAGCACTGAAAAAATGAAACCAG 840
 R  I  N  K  L  K  N  G  Y  D  I  N  Y  V  L  D  S  T  E  K  M  K  P

TGGGAATAGTGTATGATAAGAAATCAGGAAGAGTTATGGATGTACAAGCATCTTCACCTGGTGTCCAATT 910
V  G  I  V  Y  D  K  K  S  G  R  V  M  D  V  Q  A  S  S  P  G  V  Q  F

CTACACTGCAAATTTTGTTAATAATACAAAAGGGAAAGGTGGATTTGTGTATCAACCTCATTCAGCATTG 980
  Y  T  A  N  F  V  N  N  T  K  G  K  G  G  F  V  Y  Q  P  H  S  A  L

TCTTTGGAGACTCTAGTGTTCCCTGACGCTGTGAATCACCCTAATTTCCCGTCGACAATTGTAAATCCGG 1050
 S  L  E  T  L  V  F  P  D  A  V  N  H  P  N  F  P  S  T  I  V  N  P

GAGAGAAATACGTCCACTCTGTGTTGTATACGTTCTCCATAAAAAAAATAGAAGTACTAGAATGCTTATAT 1120
G  E  K  Y  V  H  S  V  L  Y  T  F  S  I  K  K  .

TTTTTTTATTTCCATTATCCTTGTGGTATTTGTATCTTCTGCATAAATTGATGTTTCTGTCAAGGTTGTA 1190

TGTTCTTCGTGACTAATAAAACACCTGACATTCAAAATAAAAAAAAAAAAAAAAAAAAA 1250
```

Figure 4.2

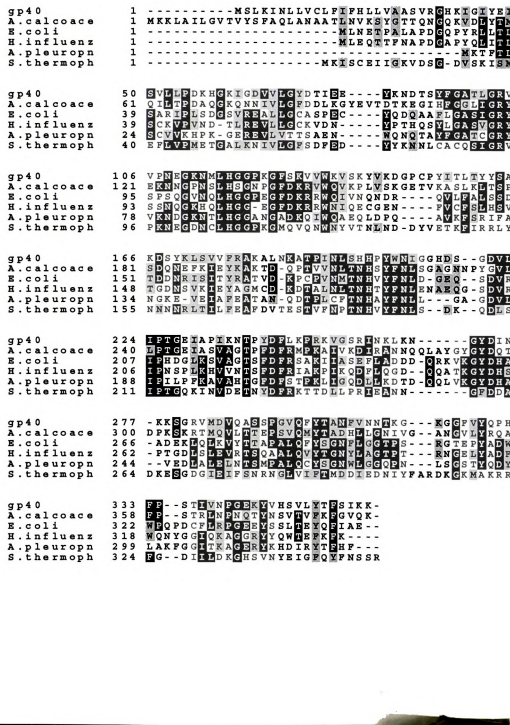
Nucleotide and deduced amino acid sequence of gp40 as derived from the 5'RACE product (RACE-7) and the truncated cDNA clone (GP40-5e; see Methods). Peptide sequences obtained from internal amino acid sequencing are marked in gray, and the primers used for 5'RACE are underlined. The stop codon is indicated by a period. The cleavage site of a putative signal sequence is marked by an astrix, and the putative NLS is marked in bold. The GenBank accession number of gp40 is AF032386.

The sequence of *GP40* contains an open reading frame encoding 357 amino acids whose predicted molecular weight of 39.3 kD was in good agreement with the apparent molecular mass of gp40. gp40 contains ~10% lysine residues, and consistent with it being glycosylated by O-linked oligosaccharides, gp40 is rich in serine and threonine (~13%). The sequence of gp40 contains at least one putative NLS (Figure 4.2; marked in bold; Raikhel, 1992), and a putative signal sequence with a cleavage site between amino acids 20 and 21 is found at the N-terminus (Figure 3C; see astrix; von Heijne, 1986). It is unclear whether the putative NLS or signal sequence are actually utilized for subcellular localization of gp40 (see Discussion).

Interestingly, the deduced amino acid sequence of gp40 showed 28-34% amino acid sequence identity and 45-51% amino acid sequence similarity over the whole protein to bacterial aldose-1-epimerases (Figure 4.3). These enzymes, also referred to as mutarotases, catalyze the interconversion of the anomeric forms (α and ß) of the carbon 1 of aldose sugars. Enzyme activity has been reported from bacteria, animals and plants (Bailey *et al.*, 1967; Mulhern *et al.*, 1973; Bouffard *et al.*, 1994 and references within). To date, no gene encoding an aldose-1-epimerase has been identified from higher organisms, although apparent *Arabidopsis* (Z25633), rice (D22462), and human (U11036) homologs exist in EST databases (data not shown).

Figure 4.3

Comparison of amino acid sequences of gp40 and bacterial aldose-1-epimerases. The deduced amino acid sequence of gp40 was aligned with the sequences of aldose-1-epimerases from *Acinetobacter calcoaceticus* (database accession number: X03893), *Escherichia coli* (U13636), *Haemophilus influenza* (U32764), *Actinobacillus pleuropneumoniae* (U63731), and *Streptococcus thermophilus* (M38175). Amino acids marked in black and grey boxes indicate sequence identity and similarity, respectively.


gp40 1 ----------MSLKINLLVCLFIFHLLVAASVRGHKIGIYEIKK-GDFSVKITNYGTSII
A.calcoace 1 MKKLAILGVTVYSFAQLANAATLNVKSYGTTTQNQKVDLYTMSNNNGWSVSFISFGGVIT
E.coli 1 ----------MLNETPALAPDQQPYRLITRNNANGMVVTLMDWGATLL
H.influenz 1 ----------MLEQTTFNAPDSAPYQLITQNNENGMKVQFMLWGAEWL
A.pleuropn 1 ----------MKTFLTSNLSFEKITLSDFGAAWL
S.thermoph 1 ----------MKISCEIIGKVDSG-DVSKISNENNNGVVISTLTTGAILQ
gp40 50 SVLLDKHGKIGDVVDGYDTISE----YKNDTSYFGATLGRVANRIGGAQFTTNGIHKKL
A.calcoace 61 QILTDDAQGKQNNIVLGFDDLKGYEVTDTKEGIHFGLLIGRYANRIGNAKFSDGKTTNL
E.coli 39 SARIPLSDGSVREALIGCASPSC----VQDQAAALGASIGRYANRIANSRYTFDGETVTL
H.influenz 39 SCKVEWNDTLREVLGCKVDN----NPTHQSLGASYGRYANRINANAPFENGKLTYL
A.pleuropn 24 SCVVKHPK-GEREVLVTTSAEN-----WQNQTAYFGATCGRYANRIANAEYQDNGKTYTL
S.thermoph 40 EPLVEMETGALKNIVLGFSDFSD----VYKNNLCACQSIGRVAGIGKASYTHNMVLKSL
gp40 106 VPNEGKNMLHGGPKGFSKVVWKVSKYVKDGPCPYITLTYYSADGEEGFPGAVLASVTYTL
A.calcoace 121 EKNNGFNSLHSGNPGFDKRVWQVKPLVSKGETVKASLKLTSPNGDQGFPGKLDVEVTYSL
E.coli 95 SFSQQVNQLHGGEGFDKRRWQIVNQNDK----QVLDAESSDGDQGFPGNLGANVQYEL
H.influenz 93 SSNQGKNQLHGGEGFDKESWNIQECGEN----TVGSSHESVDGDQGFPGNVDWSTWYTL
A.pleuropn 78 VKNDGKNTLHGGANGADKQIWQAEQLDPQ----AVKFSRIFADGEQGFPGEVVAVVTYRL
S.thermoph 96 PKNEGDNCLHGGPKGMQVQNWNYVTNLND-DYVETKIRRLYSVDGFPGDVTVSISYRL
gp40 166 KDSYKLSVVFRAKALNKATPFNLSHPYWNIGGHDS--GDVTSQVLQIFGSHITLVDKQL
A.calcoace 181 SDQNEFKTEYKAKTDLQTTVVNLTNHSYFNLSGAGNNPYGVLDHVVRLNAGRILVTDQNS
E.coli 151 TDENNIKSVYKAYGVDLKFCPVNNTNNVFDLADD-QRKVCGYDHAFLDQAKGDKLVVAPNWYS
H.influenz 148 TGDNSVKIEYAGMCDLKDTAELNLTNTYFNLENAEQG-SDVRETHLRLNADFELPVDAGG
A.pleuropn 134 NGKE-VEHAFEATAMLQDTPLCFNNLAYFNLL---GA-GDVLSHGLMINADEVLPVGAGG
S.thermoph 155 NNNNRLTLLFEAFDVTESVTNPTNEVYFNLS--DK--QDLSSHELQIVSDFNLDVKEVAQLH
gp40 224 IPTGEIAPIKNTPYDFLKPRKVGSRFNKLKN------GYDINYVFDSTEKMKPVGIVYD
A.calcoace 240 IPTGEIASVAGTPFDFRKPKAIVKDTRFNNQQLAYGVGYDQTWVLNQKSQGKLNLAAVVV
E.coli 207 KPEDGLKSVAGTPFDFSAPKPIASRIEQVNDD-QRKVCGYDHAFLDQAKGDKLVVAPNWYS
H.influenz 206 IPNSPLKHVVNTSFDFRIAKPIKQDLTQGD--QQATKGYDHSFIVNK--AWQKPCVLLTS
A.pleuropn 188 IPHLPFKAVAHTGFDFSTPKPLHGQDLKDTD-QQLVKGYDHAFKEVK--NSAKPTACLT-
S.thermoph 211 IPTGQKINVDETNYDFRKTTDLLPRHENNN------GFDDAFVVGGGTCDHVKEVAQLH
gp40 277 -KKSGRVMDVQAESFGVGFVTANPVVNNTKG---KGGFVVQPHSAFSLETLVFFDAVNHPN
A.calcoace 300 DPTSGRTMQVLFPSSVDWFKSADHFLGNTVG--ANGVLWQADAFALETKHFFDSPNQPT
E.coli 266 -ADEKLQFKVTTAPALVFYSGNPLSGTFS---NGTEPYADNQGDALESSFLFDSPNHPE
H.influenz 262 -PTGDLSSEVRTSQAALQVYTGNYLAGTFT---RNGELVADFSGIALETQCLFDPNHPE
A.pleuropn 244 --VEDLALELNTSMPALQCYSGNYLSGQFN---LSGSTVQDVAGVALEPFFFFDSPNQAS
S.thermoph 264 DKESGDGDIEFSNRNGLVIFHMDDIEDNIYFARDKGKMAKRREAIAMEAQTLSDAVNHKG
gp40 333 PE--STHVNPGSKVVHSVLYTFSIKK-
A.calcoace 358 PS--STRTMQNLQTFESVDWFKSGVQKC-
E.coli 322 WLQPDCFLRDGEEVSSLLEYQFIAE--
H.influenz 318 WQNYGGIQKAGKGRVYQWHEFKFK----
A.pleuropn 299 LAKFGGITKAGERVKHDIRYTFHF----
S.thermoph 324 TG--DILLDKGHSVNYEIGEQVFNSSR

As shown by RNA gel blot analysis, a ~1.3 kb transcript was detected in all examined tobacco tissue indicating that the gene encoding gp40 was expressed (Figure 4.4). The gp40 transcript was present in RNA isolated from 2 to 8 d-old tobacco suspension-cultured cells (Figure 4.4, Lanes 1-7) and from young leaves, old leaves, stems and roots of *N. tabacum* cv Wisconsin 38 plants (Figure 4.4, Lanes 8-11, respectively). (The isolation of RNA and the subsequent RNA gel blot analyses were done by Robyn Perrin during her rotation in Dr. Raikhel's laboratory).

gp40 is antigenically related to aldose-1-epimerases from rat

Sequence comparison indicated that gp40 has sequence similarity with bacterial aldose-1-epimerases. Even though aldose-1-epimerases have been purified from different animal tissues, no gene encoding an animal aldose-1-epimerase has been cloned. To investigate whether gp40 was related to the rat enzyme, we used polyclonal antibodies made against an aldose-1-epimerase purified from rat kidney (Toyoda *et al.*, 1983) for protein gel blot analysis. The α-rat epimerase antibodies were not able to detect any plant protein in tobacco protoplasts or enriched nuclear fractions such as the 0.5 M salt extracted fraction from purified nuclei (data not shown). For further enrichment of gp40, nuclear proteins extracted by 0.5 M salt were subjected to *in vitro* GalTF assays followed by ECA affinity chromatography. As previously described in Chapter 3, proteins that remained soluble during the *in vitro* GalTF assay were loaded onto ECA agarose columns. Column fractions were collected and analyzed by protein blot analysis using α-rat aldose-1-epimerase antibodies (Figure 4.5).

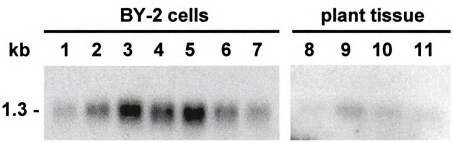


Figure 4.4

RNA gel blot analysis of gp40 mRNA levels from tobacco suspension-cultured cells and different plant tissues.
Total RNA was isolated from 2 to 8 d-old tobacco suspension-cultured cells (Lanes 1-7) and from young leaves, old leaves, stems, and roots of Wisconsin cv 38 tobacco plants (Lanes 8-11, respectively). Equal amounts of total RNA (20 μg) were separated on a formaldehyde agarose gel, immobilized to membrane, and hybridized with ^{32}P-labeled 64s/8 fragment.
Molecular size is indicated to the left in kb (kilobase).

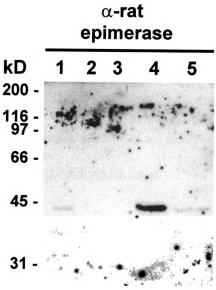


Figure 4.5

Detection of galactosylated gp40 after isolation on ECA affinity chromatography by α-rat aldose-1-epimerase antibodies.
Salt extracted nuclear proteins were subjected to nonradioactive *in vitro* GalTF assay in the presence of GalTF and separated into insoluble and soluble fractions by centrifugation. The soluble fraction (Lane 1) were loaded onto ECA agarose columns, and column fractions were collected as Lane 2, eluate; Lane 3, 0.1 M mannose wash; Lane 4, 0.05 M lactose elute; Lane 5, 1 M NaCl elute. All fractions were subjected to protein gel blot analysis using α-rat aldose-1-epimerase followed by incubation with Protein A-HRP.
Molecular mass standards are indicated at left in kilodaltons (kD).

A protein with a similar apparent molecular mass as the galactosylated gp40 was detected in the fraction loaded onto the ECA column. As found for gp40, this antigen bound to ECA and was eluted by the specific sugar lactose (Figure 4.5, Lanes 4), but not by the nonspecific sugar mannose (Figure 4.5, Lane 3). Further, it interacted strongly with ECA, because some antigen was present in the 1 M salt eluate (Figure 4.5, Lane 5), and thus showed very similar ECA binding properties as gp40 (Chapter 3, Figure 3.2, and see below). These results indicated that gp40 was at least antigenically related to rat aldose-1-epimerase. The α-rat epimerase antibodies were, however, not useful for more detailed biochemical analyses and intracellular localization of gp40, because they did not detect gp40 in more crude tobacco fractions (data not shown).

gp40 is a nuclear protein

To investigate the intracellular localization of gp40, polyclonal antibodies were prepared to recombinant gp40. A plasmid (pGEX-D) was constructed encoding the entire ORF of gp40 fused in frame to the C-terminus of glutathione S-transferase (GST). *Escherichia coli* cells were transformed with pGEX-D, and expression of GST-gp40 was induced by isopropylthio-ß-D-galactose. Proteins of uninduced and induced cells were collected, separated on SDS-polyacrylamide gels and visualized by Coomassie Brilliant Blue R 250 (Figure 4.6A). A protein with an apparent molecular mass of 64 kD was found in the induced (Figure 4.6A, Lane 2), but not in the uninduced cells (Figure 4.6A, Lane 1). The expressed protein had a lower apparent molecular mass than the expected $\sim$ 70 kD. Therefore, proteins were subjected to

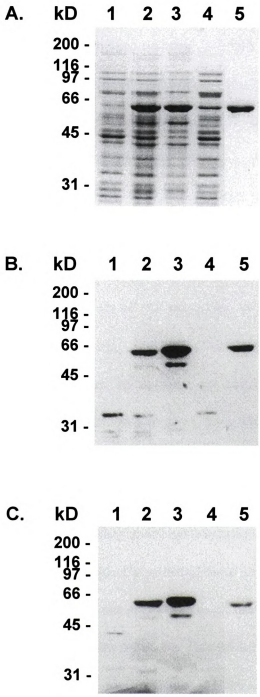
protein blot analysis using α-GST and α-rat aldose-1-epimerase antibodies (Figures 4.6B and C, respectively). Both antibodies detected a 64-kD protein in the induced (Lane 2), but not the uninduced cells (Lane 1), confirming the identity of the expressed fusion protein. After lysis of induced cells, soluble proteins were separated from the insoluble proteins by centrifugation. As shown by Coomassie Brilliant Blue R 250 and protein blot analyses, the GST-gp40 was expressed as an insoluble protein, most likely present in inclusion bodies (Figure 4.6, Lane 3). No fusion protein was detected in the soluble fraction (Figure 4.6, Lane 4).

Insoluble proteins were separated on SDS-polyacrylamide gels, and the expressed GST-gp40 was excised and eluted from the gel for antibody production. Isolation of GST-gp40 was confirmed by protein blot analysis (Figure 4.6B and C, Lane 5), and its quantity was approximated by comparing it to a known amount of BSA (Figure 4.6A, Lane 5; data not shown). Polyclonal antibodies made in rabbits were depleted of α-GST-antibodies as described by Bar-Peled and Raikhel (1996). To obtain monospecific antibodies against the recombinant gp40, the GST-depleted antibodies were further affinity purified against recombinant GST-gp40 that was immobilized onto membranes.

To examine the subcellular localization of gp40, protoplasts were isolated from tobacco suspension-cultured cells for subsequent nuclear isolation and biochemical fractionation. Equal amounts of protein were separated by SDS-polyacrylamide gel electrophoresis and visualized by Coomassie Brilliant Blue R 250 staining (Figure 4.7A) or analyzed by protein gel blot analysis (Figures 4.7B and 4.7C).

Figure 4.6

Immunodetection of recombinant glutathione S-transferase (GST)-gp40 fusion protein expressed in *E. coli*.
E. coli cells were transformed with pGEX-D (GST-gp40; uninduced, Lane 1), and expression of GST-gp40 was induced (induced, Lane 2). After lysis of induced cells, insoluble (containing inclusion bodies; Lanes 3) and soluble proteins (Lanes 4) were separated by centrifugation. GST-gp40 was isolated from the insoluble protein fraction by elution out of SDS-PAGE polyacrylamide gels (Lane 5). All lanes contained proteins from equal amount of cells ($7.5x10^7$). Protein were separated by SDS-polyacrylamide gel electrophoresis and visualized by Coomassie Brilliant Blue R 250 (A) or transferred to membrane for protein gel blot analysis. The GST-gp40 fusion proteins were detected using α-GST antibodies (B) or α-rat aldose-1-epimerase (C). Molecular mass standards are indicated at left in kilodaltons (kD).

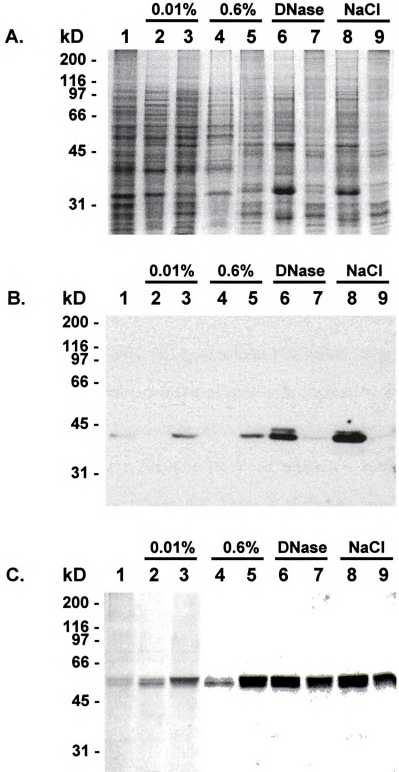

A.
kD
1 2 3 4 5
200 -
116 -
97 -
66 -
45 -
31 -
B.
kD
1 2 3 4 5
200 -
116 -
97 -
66 -
45 -
31 -
C.
kD
1 2 3 4 5
200 -
116 -
97 -
66 -
45 -
31 -

Using α-gp40 antibodies, a single protein with an apparent molecular mass similar to that of gp40, was detected in total protoplasts (Figure 4.7B, Lane 1). As expected for gp40, the antigen copurified with nuclei, as it was present in the crude nuclear (Figure 4.7B, Lane 3) and the purified nuclear fractions (Figure 4.7B, Lane 5). It was not detected in the soluble fraction of 0.01% Triton X-100 treated protoplasts (Figure 4.7B, Lane 2) indicating that it was not a cytosolic protein. Furthermore, it was neither associated with membranes nor present in the ER lumen, because it was not released upon exposure of crude nuclei to 0.6% Triton X-100 (Figure 4.7, Lane 4). The 40-kD protein was partially solubilized when purified nuclei were incubated with DNase I buffer (Figure 4.7, Lane 6). Furthermore, exposure of DNase I treated nuclei to 0.5 M salt led to the extraction of most of the remaining antigen (Figure 4.7, Lane 8). These fractionation studies indicated that the 40-kD protein was associated with peripheral structures of the nucleus. In addition, the results were in good correlation with our previous extraction studies, in which we have identified gp40 as a tGlcNAc-protein that was extracted from nuclei with 0.5 M salt (Chapter 2 and 3; Heese-Peck *et al.*, 1995).

Based on its apparent molecular mass and biochemical properties, we concluded that this antigen is most likely gp40. Upon enrichment in nuclear fractions, a second protein with a slightly larger apparent molecular mass than gp40 was also detected by the same antibody (Figure 4.7, Lanes 6 and 8). This protein may either represent gp40 with additional post translational modifications or an antigenically closely related nuclear protein with similar biochemical properties as gp40.

Figure 4.7

Cofractionation of gp40 with the nucleus.
Proteins were fractionated and extracted from tobacco suspension-cultured cells under the following conditions. Protoplasts (Lane 1) were broken in the presence of 0.01% Triton X-100 and separated into cytoplasm I (Lane 2) and crude nuclei (Lane 3) by centrifugation on a Percoll cushion. Crude nuclei were then incubated with 0.6% Triton X-100 and separated into cytosol II/released membrane proteins (Lane 4) and purified nuclei (Lane 5) by centrifugation on a Percoll gradient. Purified nuclei were treated with DNase I and separated into soluble (Lane 6) and insoluble (Lane 7) fractions by centrifugation. The insoluble fraction was then incubated with a 0.5 M NaCl buffer and separated into soluble (Lane 8) and insoluble (Lane 9) fractions by centrifugation. Equal amounts of proteins (20 μg) were loaded in each lane, separated by SDS-polyacrylamide gel electrophoresis and either stained by Coomassie blue (A) or transferred to Immobilon-P membranes for protein blot analysis (B, C). Antibody-reactive proteins were visualized using α-gp40 antibodies followed by incubation with Protein A-HRP (B) or α-IMPα antibodies followed by α-IgG-AP (C).
Molecular mass standards are indicated in kilodaltons (kD). 0.01, 0.01% Triton X-100; 0.6%, 0.6% Triton X-100; DNase, DNase I; NaCl, 0.5 M NaCl.


A.
kD
1
0.01%
2 3
0.6%
4 5
DNase
6 7
NaCl
8 9
200 -
116 -
97 -
66 -
45 -
31 -
B.
kD
1
0.01%
2 3
0.6%
4 5
DNase
6 7
NaCl
8 9
200 -
116 -
97 -
66 -
45 -
31 -
C.
kD
1
0.01%
2 3
0.6%
4 5
DNase
6 7
NaCl
8 9
200 -
116 -
97 -
66 -
45 -
31 -

In control experiments, it was observed that NT-IMPα, the protein that binds to NLSs (Smith *et al.*, 1997), possessed a different subcellular localization than gp40 (Figure 4.7C). In agreement with its proposed function as a nuclear/cytoplasmic shuttle protein and with immunofluorescence studies by Smith *et al.* (1997), NT-IMPα was found in cytoplasmic and nuclear fractions (Figure 4.7C, Lanes 1 to 5). The nuclear associated form of NT-IMPα behaved in part similar to gp40; it was partially extracted by DNase I and 0.5 M salt (Figure 4.7C, Lanes 6 and 8). Consistent with results described by Smith *et al.* (1997), NT-IMPα appeared to be much tighter associated with the nucleus than gp40, and a significant portion of the protein was still present in the insoluble 0.5 M salt fraction (Figure 4.7, Lane 6) .

To conclusively show that gp40 is indeed the 40-kD antigen detected by the α-gp40 antibodies, salt extracted nuclear proteins were subjected to nonradioactive *in vitro* GalTF labeling assays followed by ECA column affinity chromatography. Column fractions were collected as described above and analyzed by protein gel blot analysis using α-gp40 antibodies (Figure 4.8, Lanes 1 to 5). As expected, the antibodies recognized a protein with a slightly higher apparent molecular mass than that of the non-galactosylated 40-kD antigen due to the addition of Gal (Figure 4.8, Lane 1; data not shown). Most of the protein bound to the ECA agarose and was eluted with lactose and 1 M salt (Figure 4.8, Lanes 4 and 5), and it therefore displayed similar binding properties as gp40 (Figure 2A and B). When blots were stripped and reprobed with ECA-AP, this protein is also recognized by ECA-AP (Figure 4.8, Lane 7) further demonstrating that it was a tGlcNAc-protein.

In addition to the galactosylated gp40, two proteins with a slightly larger apparent molecular mass (43-45 kD) were detected in the lactose elute by the antibody (Figure 4.8, Lanes 4). These proteins are also tGlcNAc-proteins because they were detected by ECA (Figure 4.8, Lane 6), and as discussed above, these proteins may be closely related to gp40 or posttranslationally modified gp40.

Taken together, biochemical fractionation and ECA affinity chromatography showed that the protein detected by the α-gp40 antibodies was indeed gp40; it had the same apparent molecular mass, similar biochemical properties and elution profile as gp40, and it was also a tGlcNAc-protein. Further, these studies indicated that gp40 was a nuclear tGlcNAc-protein with similar biochemical properties as peripheral NPC proteins from vertebrates (Davis and Blobel, 1986; Kita *et al.*, 1993; Radu *et al.*, 1994).

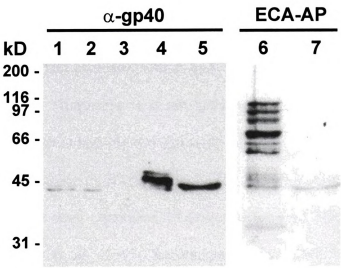


Figure 4.8

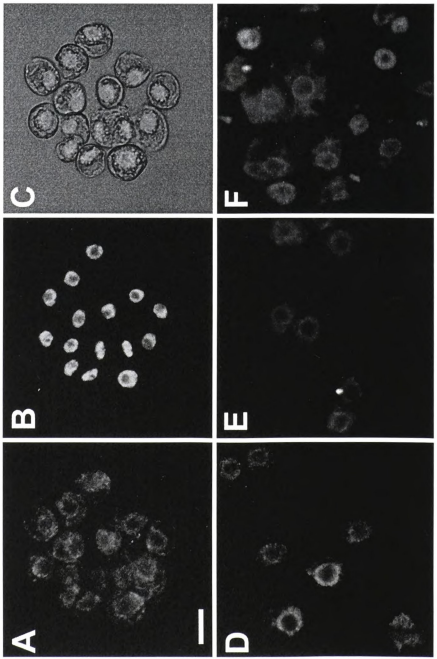
Purification of galactosylated gp40 by ECA affinity chromatography.
Salt extracted nuclear proteins were subjected to nonradioactive *in vitro* GalTF assay in the presence of GalTF and separated into insoluble and soluble fractions by centrifugation. The soluble fraction (Lane 1) was loaded onto ECA agarose columns, and column fractions were collected as Lane 2, eluate; Lane 3, 0.1 M mannose wash; Lane 4, 0.05 M lactose eluate; Lane 5, 1 M NaCl eluate. All fractions were subjected to protein gel blot analysis using α-gp40 antibodies followed by incubation with protein A-HRP (Lanes 1-5). The blot containing Lanes 4 and 5 was then stripped of the probes and reprobed with ECA-AP to visualize galactosylated tGlcNAc-proteins (Lanes 6 and 7, respectively). Molecular mass standards are indicated in kilodaltons (kD).

gp40 is a nuclear rim/NPC associated protein

Immunofluorescent microscopy was used to confirm the nuclear localization and to further investigate the subnuclear localization of gp40. In general, tobacco suspension-cultured cells were incubated with protoplasting enzymes for a short time (less than 30 min) to ensure cell preservation during the fixation and labeling procedures as confirmed by Nomarski optics (Figure 4.9C). The lightly protoplasted cells were then fixed and incubated with α-gp40 antibodies, followed by CY3-labeled secondary antibodies. Consistent with the biochemical fractionation studies (Figure 4.7B), immunofluorescent labeling was primarily detected around the nuclear rim using conventional light microscopy (Figure 4.9A).This labeling pattern was similar to that of NPC proteins from other organisms, in particular for yeast NPC proteins (Wimmer *et al.*, 1992; Grandi *et al.*, 1993; Siniossoglou *et al.*, 1996). Immunofluorescent labeling around the nuclear rim was further supported by staining protoplasts with diaminophenylindole (DAPI; Figure 4.9B). The dye DAPI binds to DNA and is commonly used to indicate the position and outline of the nucleus within a cell (Hicks *et al.*, 1996). We also observed in some cells low immunofluorescence labeling at the cell's periphery that was reminiscent of cell wall labeling (Figure 4.9A). At this point, we do not know whether this labeling may be significant or represent non-specific interaction of the antibody. If cells were protoplasted for a longer time (30 to 60 min), we could only detect immunofluorescent labeling at the nuclear rim (Figure 4.9D); these cells however, were not as well preserved.

Figure 4.9

gp40 is located at the nuclear periphery of tobacco protoplasts.
Tobacco suspension-cultured cells were protoplasted for less than 30 min (A-C, E, and F) or for 30-60 min (D). Protoplasted cells were fixed and incubated with α-gp40 (A-E) or α-IMPα (F) antibodies followed by CY3-labeled secondary antibodies. The cells were visualized by epifluorescent (A and D) or confocal laser scanning microscopy (E and F). Cells in (C) were visualized by Nomarski optics, and DAPI staining was visualized by epifluorescence (B).
Bar in (A) = 30 μM for (A)-(F).

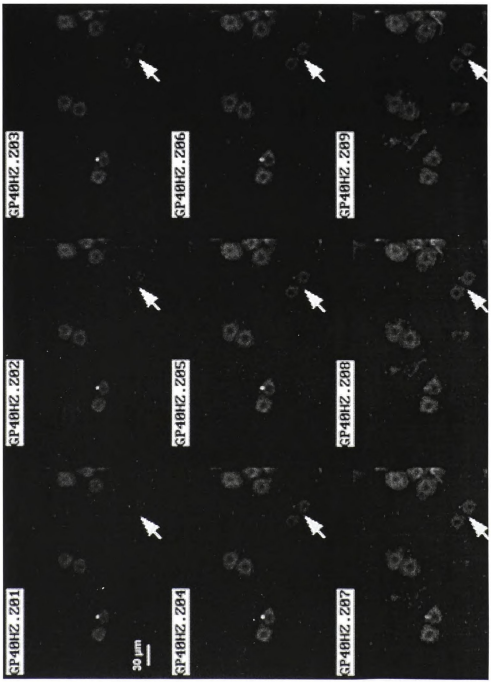

A
B
C
D
E
F

Labeling with CY3-labeled secondary antibody gave only low level of background labeling within the entire cell (data not shown) indicating that primary antibody labeling was specific. After locating gp40 by conventional light microscopy, immunofluorescently labeled protoplasts were viewed in optical sections by confocal laser scanning microscopy which confirmed that gp40 was primarily present at the nuclear rim with little labeling within the nucleus (Figure 4.9E). In control experiments, NT-IMPα was localized to the cytoplasm, the nucleus and the nuclear rim (Figure 4.9F). These results are consistent with those previously obtained by Smith *et al.* (1997) and with the biochemical fractionation studies (Figure 4.7C). Further, they confirmed that the protoplast used in these studies were properly fixed and processed for immunofluorescence.

To confirm that gp40 was localized at the nuclear rim, nine serial confocal sections were taken in 1μm intervals through a series of planes from below to near the top of particular nuclei (Figure 4.10, arrow). Importantly, the gp40-specific staining was found to be associated with the nuclear rim in all sections. The results obtained from the microscopy studies (Figures 4.9 and 4.10), taken together with the biochemical fractionation studies (Figures 4.7B and 4.8), clearly indicated that gp40 was a tGlcNAc-protein associated with the nuclear rim.

Figure 4.10

gp40 is localized at the nuclear rim as shown in serial optical sections obtained from laser scanning confocal microscopy.
Tobacco suspension-cultured cells were treated as in Figure 4.9E and observed in nine successive optical sections (Z01-Z09, 1μm intervals) using laser scanning confocal microscopy).
Bar in (Z01) = 30μm for (Z01)-(Z09).


GP40HZ.201
GP40HZ.202
GP40HZ.203
30 µm
GP40HZ.204
GP40HZ.205
GP40HZ.206
GP40HZ.207
GP40HZ.208
GP40HZ.209

DISCUSSION

In the present studies, a combination of PCR-based amplification, cDNA library screening, and 5'RACE were used to clone the gene encoding gp40. This tGlcNAc-protein showed significant amino acid sequence similarity to aldose-1-epimerases from bacteria. Using monospecific antibodies made against recombinant gp40, we confirmed that gp40 was a nuclear tGlcNAc-protein with similar biochemical properties as vertebrate NPC proteins. Consistent with its biochemical isolation, conventional light and laser scanning immunofluorescent microscopy indicated that gp40 was primarily found at the nuclear periphery of tobacco protoplasts.

gp40 is associated with structures at the nuclear periphery

As discussed in Chapter 3, the *in vitro* GalTF assay has served as an excellent tool to identify and purify nuclear tGlcNAc-proteins including gp40. The production of antibodies made against gp40 allowed a more detailed biochemical and microscopic analysis of gp40. In agreement with our initial purification scheme, fractionation studies showed that gp40 was a nuclear protein. Based on its extraction by 0.5 M salt, but not by detergent treatment, gp40 appeared to be peripherally associated with nuclear structures. In vertebrates, exposure of isolated nuclei to 0.5 M salt releases NPC proteins that are present at the periphery of this large proteinaceous complex (Davis and Blobel, 1986; Kita *et al.*, 1993; Radu *et al.*, 1994). Harsher conditions are usually required to extract proteins that are located to the NPC core structures as well as to the nuclear lamina/matrix, a filamentous structure found throughout the nucleus (Dwyer and Blobel, 1976; Davis and Blobel, 1986; Masuda *et al.*, 1997). At this point, we do not

fully understand the release of gp40 by DNase I treatment. Based on immunofluorescent microscopy, gp40 was found at the nuclear periphery, and only little labeling within the nucleus. Therefore, we do not believe that this low amount could account for the amount of gp40 released by DNAse I. It should be noted that the DNase I buffer used in the present studies contained Mg^{2+}, a divalent ion shown to be involved in the disassembly and reassembly of the nucleoplasmic basket in amphibian cells (Jarnick and Aebi, 1991). Our previous EM studies have demonstrated that proteins modified by GlcNAc are mainly located at or adjacent to tobacco NPC as shown by EM (Heese-Peck *et al.*, 1995). Consistent with these results, gp40 was primarily found at the nuclear periphery using immunofluorescent microscopy pointing towards a possible localization of gp40 to the NPC. Further, the labeling pattern was comparable to that of NPC proteins from other organisms, in particular for yeast NPC proteins (Wimmer *et al.*, 1992; Grandi *et al.*, 1993; Siniossoglou *et al.*, 1996).

As mentioned above, gp40 has a putative signal sequence as well as at least one putative NLS; However, we do not know whether they are used for gp40 localization. It has been shown for several growth hormone peptides and the rat prostatic protein probasin, that the presence of these two signals is not mutually exclusive, and both signal can be functional and localized for dual protein localization (Maher *et al.*, 1989; Spence *et al.*, 1989; Kimura, 1993; Kiefer *et al.*, 1994). If hte putative signal sequence is utilized for gp40 localization, it is possible that gp40 may associate with peripheral nuclear structures after translocation into the ER lumen. For example, the vertebrate NPC proteins, gp210 and Pom121, have functional signal sequences and appear to provide contact sites for NPC proteins in the cytoplasm as well as in the ER lumen

(Wozniak *et al.*, 1989; Greber *et al.*, 1990; Hallberg *et al.*, 1993). It should be, however, pointed out that gp40 is modified by O-linked oligosaccharides with terminal GlcNAc (Heese-Peck *et al.*, 1995; N.V. Raikhel, A. Heese-Peck, A. Bacic; unpublished). Although little is known about O-linked glycosylation in plants, it has not been reported to occur in the ER, but rather in the Golgi apparatus as found also for other higher organisms (Driouich *et al.*, 1993). Further in vertebrates, the O-GlcNAc is added in the cytoplasm by the O-linked GlcNAc transferase (for review, see Hart, 1997). The SPINDLY gene from *A. thaliana*, involved in gibberellin signal transduction (Jacobsen *et al.*, 1996), shows extensive sequence identity to genes encoding vertebrate transferase (Lubas *et al.*, 1997; Kreppel *et al.*, 1997) indicating that O-linked glycosylation may also occur in the cytoplasm in plants.

Sequence similarity of gp40 to aldose-1-epimerases

Interestingly, gp40 shares significant sequence identity (28-34%) to aldose-1-epimerases from several bacteria, with the highest to that of *A. calcoaceticus* (34% identity and 51% similarity). In addition, gp40 has a similar apparent molecular mass as the bacterial enzymes (Poolman *et al.*, 1990; Mollet and Pilloud, 1991; Maskell *et al.*, 1992; Bouffard *et al.*, 1994). Considering the evolutionary distance of bacteria to plants, it is noteworthy that bacterial aldose-1-epimerases have a similar percentage of sequence identity ($\sim$30%) with gp40 as with each other (Poolman *et al.*, 1990; Mollet and Pilloud, 1991; Maskell *et al.*, 1992; Bouffard *et al.*, 1994). Alignment of gp40 and bacterial sequences show several well conserved blocks suggesting structural and maybe even functional similarities. Based on early biochemical studies of purified aldose-1-

epimerases from different organisms, histidine, tyrosine, and tryptophan residues have been proposed to be components of the active site (Wallensfeld *et al.*, 1965; Hucho and Wallensfeld, 1971; Fishman *et al.*, 1973). In the sequence alignment, one tryptophan (W121), two histidine (H115, H192), and two tyrosine residues (Y163, Y194) are conserved between all bacterial aldose-1-epimerases as well as gp40. To date, no definitive amino acid residue has been assigned for aldose-1-epimerase function. Bacterial enzymes and gp40 also shared considerable, but lower sequence identity ($\sim$20%) to the carboxy-terminal portion of the fungal Gal10p from *S. cerevisiae* and *P. tannophilus*, but the fungal proteins have a predicted mass of about twice the size as gp40 and bacterial aldose-1-epimerases (Citron and Donelson; 1984; Skrzypek and Maleszka, 1994).

Even though aldose-1-epimerases have been purified and its enzyme activity characterized from different vertebrate tissues, no gene encoding an animal aldose-1-epimerase has been cloned. Interestingly, immunohistochemical studies indicate that aldose-1-epimerases in different rat tissues are nuclear proteins (Baba *et al.*, 1979; Toyoda *et al.*, 1983). However, no further biochemical fractionation and subnuclear localization have been reported. The polyclonal antibodies against rat kidney aldose-1-epimerase (Toyoda *et al.*, 1983) recognized gp40 that was isolated by ECA-affinity chromatography further supporting that gp40 was antigenically related to rat aldose-1-epimerase.

Possible functions of gp40

In bacteria, aldose-1-epimerase functions in carbohydrate metabolism (Bouffard *et al.*, 1994); however, its exact function is unknown in higher organisms. Bailey *et al.* (1967; 1970) argue against the involvement of the enzyme in metabolic sugar breakdown in higher organisms, because its distribution in mammalian tissue is not consistent with glycolytic capacity. In addition, a lack of any anomeric specificity for enzymes that phosphorylate or dephosphorylate glucose (including hexokinase, glucokinase, and pyrophosphate-glucose phosphotransferase) indicates a putative function of aldose-1-epimerases in sugar utilization rather than sugar metabolism. Based on inhibitor studies, the active site of aldose-1-epimerase has a similar structural organization as the glucose carrier from erythrocytes indicating a possible function in sugar binding and transport (Diedrich and Stringham, 1970a; 1970b). Due to its sequence similarity to bacterial aldose-1-epimerases, gp40 may also be functionally related to aldose-1-epimerases, but to date, we do not know whether gp40 possesses enzyme activity. To address possible activity, we expressed gp40 in fusion to GST (see above) or His-tag (data not shown) in *E. coli*. The recombinant fusion proteins were present in the insoluble fractions as inclusion bodies and, thus, could not be used for activity assays (data not shown). It should be considered that proper glycosylation may be required for enzyme activity. Expression in heterologous systems such as *E. coli* may not be desirable, because gp40 will not be properly glycosylated. Thus, purification of gp40 from plant tissue may be necessary that should exclude the use of the *in vitro* GalTF assay to avoid additional modification of gp40.

The antibodies made against gp40 may be helpful to design a new purification scheme and to directly isolate properly glycosylated gp40.

If gp40 proves to be an aldose-1-epimerase, gp40 may act on nucleosides and nucleotide-containing substrates including UDP-sugars, DNA and RNA. One should, however, keep in mind that the hydroxyl group at the C1 carbon of these molecules provide the link to the attached base. Thus, these molecules may not serve as substrate, because aldose-1-epimerases appear to require a free hydroxyl group at this particular carbon (Feingold, 1982).

gp40 is modified by O-linked oligosaccharides with terminal GlcNAc (Chapter 2; Heese-Peck *et al.*, 1995; N.V. Raikhel, A. Heese-Peck, T. Bacic, unpublished). Because in vertebrate O-GlcNAc proteins, the glycosylation is proposed to function in assembly and disassembly of multimeric protein complexes (for review, see Hart, 1997), the carbohydrate modification of gp40 may have a similar role in plants (see Chapter 2 for detailed discussion).

The localization of gp40 at the nuclear periphery allows new speculation about putative functions of aldose-1-epimerase-like proteins. A possible role of gp40 in nuclear import of glycoproteins comes from studies from Duverger *et al.* (1995). The authors report that glycosylated BSA is imported into the nucleus of permeabilized HeLa cells in a sugar-specific manner. The nuclear import of these glycoproteins shares some common features with the nuclear uptake of NLS-bearing proteins. Because import of glycoproteins, however, is not competed by NLS-bearing proteins, a different uptake mechanism has been proposed that may involve the presence of sugar-binding proteins at the NPC (Duverger *et al.*, 1995). It can be speculated that gp40 may serve

a similar role because it was not only located at the nuclear periphery, but also shared sequence similarity with aldose-1-epimerases, enzymes that need to bind sugars to accomplish their enzymatic function. The identification of gp40 may give new insights into the function of aldose-1-epimerases in eukaryotes, and future experiments will be designed to investigate possible roles of gp40 (Chapter 5).

References

Baba, S., Arimoto, Y., Yoshikawa, D., Toyoda, Y., Miwa, I., and Okuda, J. (1979). The localization of mutarotase in rat kidney. Experienta **35,** 1094-1097.

Bar-Peled, M., and Raikhel, N.V. (1996). A method for isolation and purification of specific antibodies to a protein fused to the GST. Anal. Biochem. **241,** 140-142.

Bailey, J.M., Fishman, P.H., and Pentchev, P.G. (1967). Studies on mutarotases. I. Purification and properties of a mutarotase from higher plants. J. Biol. Chem. **242,**-4263-4269.

Bailey, J.M., Fishman, P.H., and Pentchev, P.G. (1970). Studies on mutarotases. VI. Enzyme level and sugar reabsorption in developing rat kidney and intestine. J. Biol. Chem. **245,** 559-563.

Bouffard, G.G., Rudd, K.E., and Adhya, S.L. (1994). Dependence of lactose metabolism upon mutarotase encoded in the *gal* operon in *Escherichia coli*. J. Mol. Biol. **244,** 269-278.

Citron, B.A. and Donelson, J.E. (1984). Sequence of the *Saccharomyces* GAL region and its transcription *in vivo*. J. Bacteriol. **158**, 269-278.

Davis, L. (1995). The nuclear pore complex. Annu. Rev. Biochem. **64,** 865-896.

Davis, L.I., and Blobel, G. (1986). Identification and characterization of a nuclear pore complex protein. Cell **45**, 699-709.

Diedrich, D.F., and Stringham, C.H. (1970a). Mutarotase: Still candidate for the role of membrane carrier? Arch. Biochem. Biophys. **138,** 493-498.

Diedrich, D.F., and Stringham, C.H. (1970b). Active site comparison of mutarotase with the glucose carrier in human erythrocytes. Arch. Biochem. Biophys. **138,** 499-505.

Doye, V. and Hurt, E.C. (1997). From nucleoporins to nuclear pore complex. Curr. Opin. Cell Biol. **9**, 401-411.

Driouich, A., Faye, L. and Staehelin, L.A. (1993). The plant Golgi apparatus: a factory for complex polysaccharides and glycoproteins. Trends Biochem. Sci. **18,** 210-214.

Duverger, E., Pellerin-Mendes, C., Mayer, R., Roche, A.-C., and Monsigny, M. (1995). Nuclear import of glycoconjugates is distinct from the classical NLS pathway. J. Cell Sci. **108,** 1325-1332.

Dwyer, N. and Blobel, G. (1976). A modified procedure for the isolation of a pore complex-lamina fraction from rat liver nuclei. J. Cell Biol. **70,** 581-591.

Ebisu, S., Lonngren, J., and Goldstein, I.J. (1977). Interaction of pneumococcal S 14 polysaccharides with lectins from *Ricinus communis, Triticum vulgares*, and *Bandeiraea simplicifolia*. Carbohydrate Res. **58**, 187-191.

Facchini, P.J., and Chappell, J. (1992). Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. Proc. Natl. Acad. Sci. USA **89,** 11088-11092.

Feinberg, A.P., and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132,** 6-13.

Feingold, D.S. (1982). Aldo (and keto) hexoses and uronic acids. In Loewus, F.A. and Tanner, W. (ed.), Plant Carbohydrates I in Encyclopedia of Plant Physiology, Springer-Verlag, Vol. **13A**, 30-37.

Fishman, P.H., Kusiak, J.W., and Bailey, J.M. (1973). Studies on mutarotase: Photooxidation reactions and nature of the enzyme catalysis. Biochem. **12,** 2540-2544.

Forbes, D.J. (1992). Structure and function of the nuclear pore complex. Annu. Rev. Cell Biol. **8**, 495-527.

Gatz, C., Altschmied, J., Hillen, W. (1986). Cloning and expression of the *Acinetobacter calcoaceticus* mutarotase gene in *Escherichia coli*. J. Bacteriol. **168,** 31-39.

Goldberg, M.W., and Allen, T.D. (1995). Structural and functional organization of the nuclear envelope. Curr. Opin. Cell Biol. **7,** 301-309.

Görlich, D., Mattaj, I.W. (1996). Nucleocytoplasmic transport. Science **271,** 1513-1518.

Grandi, P., Doye, V., and Hurt, E.C. (1993). Purification of NSP1 reveals complex formation with 'GLFG' nucleoporins and a novel nucleoporin NIC96. EMBO J. **12**, 3061-3071.

Greber, U.F., Senior, A., and Gerace, L. (1990). A major glycoprotein of the nuclear pore complex is a membrane-spanning polypeptide with a large lumenal domain and a small cytoplasmic tail. EMBO J. **9,** 1495-1502.

Hallberg, E., Wozniak, R.W., and Blobel, G. (1993). An integral membrane protein of the pore membrane domain of the nuclear envelope contains a nucleoporin-like region. J. Cell Biol. **122**, 513-521.

Harlow, E., and Lane, D. (1988). Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Hart, G.W. (1997). Dynamic O-linked glycolsylation of nuclear and cytoskeletal proteins. Annu. Rev. Biochem. **66,** 315-335.

Heese-Peck, A., Cole, R.N., Borkhsenious, O.N., Hart, G.H., and Raikhel, N.V. (1995). Plant nuclear pore complex proteins are modified by novel oligosaccharides with terminal *N*-acetylglucosamine. Plant Cell **7,** 1459-1471.

Hicks, G.R., and Raikhel, N.V. (1993). Specific binding of nuclear localization sequences to plant nuclei. Plant Cell **5**, 983-994.

Hicks, G.R., and Raikhel, N.V. (1995a). Nuclear localization signal binding proteins in higher plant nuclei. Proc. Natl. Acad. Sci. USA **92**, 734-738.

Hicks, G.R., and Raikhel, N.V. (1995b). Protein import into the nucleus: An integrated view. Annu. Rev. Cell Dev. Biol. **11,** 155-188.

Hicks, G.R., Smith, H.M.S., Lobreaux, S., and Raikhel, N.V. (1996). Nuclear import in permeabilized protoplasts from higher plants has unique features. Plant Cell **8,** 1337-1352.

Hucho, F., and Wallenfels, K. (1971). The enzymatically catalyzed mutarotation: The mechanism of action of mutarotase (aldose-1-epimerase) from *Escherichia coli*. Eur. J. Biochem. **23,** 489-496.

Hurt, E.C. (1988). A novel nucleoskeletal-like protein located at the nuclear periphery is required for the life cycle of *Saccharomyces cerevisiae*. EMBO J. **7**, 323-4334.

Jacobsen, S.E., Binkowski, K.A., and Olszewski, N.E. (1996). SPINDLY, a tetratricopeptide repeat protein involved in gibberellin signal transduction in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **93,** 9292-9296.

Jarnik, M. and Aebi, U. (1991). Towards a more complete 3D structure of the nuclear pore complex. J. Struc. Biol. **107**, 291-308.

Jordan, E.G., Timmis, J.N., and Trewavas, A.J. (1980). The plant nucleus. In Tolbert,N.E. (ed.), The Biochemistry of Plants, Academic Press, Inc., Vol. 1, pp. 490-588.

Kiefer, P., Acland, P., Pappin, D., Peters, G., and 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**, 4126-4136.

Kimura, H. (1993). Schwannoma-derived growth factor must be transported into the nucleus to exert its mitogenic activity. Proc. Natl. Acad. Sci. **90**, 2165-2169.

Kita, K., Omata, S., and Horigome, T. (1993). Purification and characterization of a nuclear pore glycoprotein complex containing p62. J. Biochem. **113**, 377-382.

Kreppel, L.K., Blomberg, M.A., and Hart, G.W. (1997). Dynamic glycolsylation fo nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. J. Biol. Chem. **272,** 9308-9315.

Lane, W.S., Galat, A., Harding, M.W., and Schreiber, S.L. (1991). Complete amino acid sequence of the FK506 and rapamycin binding protein, FKBP, isolated from calf thymus. J. Prot. Chem. **10,** 151-160.

Lis, H., Joubert, F.J., Sharon, N. (1985). Isolation and properties of *N*-acetyllactosamine-specific lectins from nine *Erythrina* species. Phytochemistry **624,** 2803-2809.

Lubas, W.S., Frank, D.W., Krause, M., and Hanover, J.A. (1997). O-linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. J. Biol. Chem. **272,** 9316-9324.

Maskell, D.J., Szabo, M.J., Deadman, M.E., and Moxon, E.R. (1992). The *gal* locus from *Haemophilus influenzae*: cloning, sequencing and the use of *gal* mutants to study lipopolysaccharide. Mol. Microbiol. **6,** 3051-3063.

Maher, D.W., Lee, B.A, Donoghue, D.J. (1989). The alternatively spliced exon of the platelet-derived growth factor A chain encodes a nuclear targeting signal. Mol. Cell Biol. **9**, 2251-2253.

Masuda, K., Xu, Z.-J., Takahashi, S., Ito, A., Ono, M., Nomura, K. and Inoue, M. (1997). Peripheral framework of carrot cell nucleus contains a novel protein predicted to exhibit a long alpha-helical domain. Exp. Cell Res. **232,** 173-181.

Merkle, T., Leclerc, D., Marshallsay, C., and Nagy, F. (1996). A plant *in vitro* system for the nuclear import of proteins. Plant J. **10,** 1177-1186.

Miller, M.W., and Hanover, J.A. (1994). Functional nuclear pores reconstituted with ß-1,4 galactose-modified *O*-linked *N*-acetylglucosamine glycoproteins. J. Biol. Chem. **269**, 9289-9297.

Mollet, B., and Pilloud, N. (1991). Galactose utilization in *Lactobacillus helveticus*: isolation and characterization of the galactokinase (*galK*) and galactose-1-phosphate uridyl transferase (*galT*) genes. J. Bacteriol. **173,** 4464-4473.

Mulhern, S.A., Fishman, P.H., Kusiak, J.W., and Bailey, J.M. (1973). Physical characteristics and chemi-osmotic transformations of mutarotases from various species. J. Biol. Chem. **248,** 4163-4173.

Panté, N., and Aebi, U. (1993). The nuclear pore complex. J. Cell Biol. **122**, 977-984.

Poolman, B., Royer, T.J., Mainzer, S.E., and Schmidt, B.F. (1990). Carbohydrate utilization in *Streptococcus thermophilus*: characterization of the genes for aldose-1-epimerase (mutarotase) and UDPglucose 4-epimerase. J. Bacteriol. **172,** 4037-4047.

Puissant, C., Houdeline, L.-M. (1990). An improvement of the single-step method of RNA isolation by guanidinium thiocyanate-phenol-chloroform extraction. Biotechniques **9,** 148-149.

Radu, A., Blobel, G., and Wozniak, R.W. (1994). Nup107 is a novel nuclear pore complex protein that contains a leucine zipper. J. Biol. Chem. **269,** 17600-17605.

Raikhel, N.V. (1992). Nuclear targeting in plants. Plant Physiol. **100**, 1627-1632.

Roquemore, E.P., Chou, T.-Y., and Hart, G.W. (1994). Detection of O-linked *N*-acetylglucosamine (O-GlcNAc) on cytoplasmic and nuclear proteins. Methods in Enzymology **230**, 443-460.

Rout, M.P., and Wente, S.R. (1994). Pores for thought: nuclear pore complex proteins. TICB **4**, 357-365.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989). Molecular cloning: A laboratory manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Scofield, G.N., Beven, A.F., Shaw, P.J., and Doonan, J.H. (1992). Identification and localization of a nucleoporin-like protein component of the plant nuclear matrix. Planta **187**, 414-420.

Siniossoglou, S., Wimmer, C., Rieger, M., Doye, V., Tekotte, H., Wiese, C., Emig, S., Segref, A., and Hurt, E.C. (1996). A novel complex of nucleoporinswhich includes Sec13p and a Sec1p homologue is essential for normal nuclear pores. Cell **84,** 265-275.

Skrzypek, M. and Maleszka, R. (1994). A gene homologous to that encoding UDP-galactose-4-epimerase is inducible by xylose in the yeast *Pachysolen tannophilus*. Gene **140,** 127-129.

Smith, H.M.S., Hicks, G.R., and Raikhel, N.V. (1997). Importin α from *Arabidopsis thaliana* is a nuclear import receptor that recognizes three classes of import signals. Plant Physiol. **114,** 411-41.

Spence, A.M., Sheppard., P.C., Davie, J.R., Matuo, N., McKeehan, W.L., Dodd, J.G., and Matusik, R.J. (1989). Regulation of a bifunctional mRNA results in the synthesis of secreted and nuclear probasin. Proc. Natl. Acad. Sci. **86**, 7843-7847.

Toyoda, Y., Miwa, I., and Okuda, J. (1983). Multiple forms of mutarotases from the kidney, liver, and small intestine of rats: Purification, properties, subcellular localization and developmental changes. J. Biochem. **94,** 421-431.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Res **14,** 4683-4690.

Wimmer, C., Doye, V., Grandi, P., Nehrbass, U., and Hurt, E.C. (1993). A new subclass of nucleoporins that functionally interacts with nuclear pore protein NSP1. EMBO J. **11**, 5051-5061.

Wozniak, R.W., Bartnik, E., and Blobel, G. (1989). Primay structure analysis of an integral membrane glycoprotein of the nuclear pore. J. Cell Biol. **108,** 2083-2092.

Chapter 5

Conclusions and future directions

The NPC regulates communication between the nucleus and the cytoplasm by controlling the movement of macromolecules, such as proteins and RNA species, in and out of the nucleus. The identification of NPC proteins is crucial in elucidating the molecular mechanism that underlies the continuous transport of such diverse molecules.

When we started this project nearly five years ago, only seven genes encoding NPC proteins had been isolated from vertebrates and yeast (Chapter 1; Forbes, 1992). At that time and during the following years, no plant homologs to known NPC proteins were available in the EST databases. As determined by EM, antibodies against several vertebrate NPC proteins that were made available to us did not specifically crossreact with plant NPC proteins (A. Heese-Peck, O.N. Borkhsenious, N.V. Raikhel, unpublished). Thus, we had to consider other means to identify plant NPC proteins.

In vertebrates, some NPC proteins are posttranslationally modified by single O-GlcNAc residues (Davis, 1995; Hart, 1997). Because the lectin WGA binds specifically to GlcNAc, this sugar modification has served as an useful tool for identifying and purifying these glycoproteins (Forbes, 1992; Davis, 1995). As an initial step to identify plant NPC proteins, we demonstrated that NPC proteins from tobacco suspension-cultured cells were also modified by GlcNAc residues. As shown by EM using WGA as a probe, specific labeling was often found associated with or adjacent to NPCs (Chapter 2; Heese-Peck *et al.*, 1995). These glycoproteins possessed similar biochemical properties as described for vertebrate NPC proteins associated with peripheral NPC structures (Chapter 2).

Further sugar analysis showed that most of theses tobacco glycoproteins were modified by O-linked glycans with terminal GlcNAc residues. Interestingly, the plant carbohydrate modification appeared to be more complex than the single O-GlcNAc found on vertebrate NPC proteins. The plant glycans were composed of oligosaccharides larger in size than five GlcNAcs (Chapter 2; Heese-Peck *et al.*, 1995). It is noteworthy that yeast NPC proteins do not seem to contain any type of GlcNAc modification (Rout and Wente, 1994). Thus, the glycan complexity is different in the various kingdoms. Moreover, the function of any of the sugar modifications is not understood. The gene encoding O-GlcNAc transferase, the enzyme specific for the attachment of O-GlcNAc to proteins, has been recently isolated from vertebrates (Kreppel *et al.*, 1997; Lubas *et al.*, 1997). It shows extensive sequence similarity to the SPINDLY gene from *Arabidopsis* identified Dr. N. Olszewski laboratory (Univ. of Minnesota; Jacobsen *et al.*, 1996). The SPINDLY gene was originally isolated from a gibberellin responsive mutant indicating a possible involvement of this proteins in GA signal transduction. Further analysis of SPINDLY may help to gain a better understanding in the function of GlcNAc modifications in plants and animals. An initial step is the identification of target substrates, and the nuclear tGlcNAc proteins identified in Chapter 2, 3, and 4 may be candidates. In addition, Dr. Olszewski's laboratory is currently using the purification scheme developed in these studies (Chapter 3) to isolate other putative target proteins modified by terminal GlcNAc.

A more detailed characterization of the carbohydrate composition may further help to understand the possible functions of O-linked sugar modifications. These analyses are being pursued in collaboration with Dr. Bacic's laboratory (Univ. of

Melbourne, Australia). Although the lack of a purification scheme compatible with the methods used for sugar analysis hampered our initial attempts to elucidate the sugar composition of these unusual plant glycans (N.V. Raikhel, A. Heese-Peck, A. Bacic; unpublished), antibodies against gp40 (Chapter 4) may serve as a useful tool in designing a new protein purification scheme.

Galactosylation of tGlcNAc-proteins enabled us to purify these glycoproteins by ECA affinity chromatography (Chapter 3). This purification scheme was efficient in obtaining amino acid sequence information of three tGlcNAc-proteins, namely gp33, gp40, and gp65, with similar properties as described for vertebrate NPC proteins. The peptide sequence information obtained from gp40 led to the subsequent isolation of the gene encoding gp40 (Chapter 4). Consistent with its isolation as a putative NPC protein, gp40 was found at the periphery of the nucleus using immunofluorescent microscopy. The precise localization of gp40 will require EM immunocytochemistry. These experiments will be done with the help of Dr. V. Kovaleva, an EM specialist joining the Raikhel laboratory at the beginning of 1998.

As described in Chapter 4, gp40 may have a role in nuclear import. To address this, purified gp40 as well as α-gp40-antibodies may be used in the *in vitro* nuclear import system developed in our laboratory (Hicks *et al.*, 1996). Based on the selective permeabilization of the plasma membrane, this assay will provide an excellent means to study the function of gp40 in nuclear import of glycoproteins and NLS-containing proteins.

Sequence analysis indicated that gp40 was similar to aldose-1-epimerases. To date, we do not know whether gp40 also has aldose-1-epimerase activity. To address possible enzyme activity, we expressed gp40 in *E. coli*. The recombinant protein was insoluble and, thus, could not be used for activity assays (Chapter 4). It is possible that proper glycosylation of gp40 may be required for enzyme activity. Purification of gp40 from plant tissue may therefore be favored over expression in heterologous systems such as *E. coli* or yeast that do not provide proper glycosylation of gp40. Further, purification from plant tissue should exclude the use of *in vitro* GalTF assays, in which gp40 is modified by additional galactose. The antibodies made against gp40 (Chapter 4) may provide an useful tool to directly isolate properly modified gp40.

In general, the function of aldose-1-epimerases is not known in higher organisms. Thus, analysis of gp40 may provide new insight into the function of aldose-1-epimerases in eukaryotes.

References

Davis, L. (1995). The nuclear pore complex. Annu. Rev. Biochem. **64,** 865-896.

Forbes, D.J. (1992). Structure and function of the nuclear pore complex. Annu. Rev. Cell Biol. **8**, 495-527.

Hart, G.W. (1997). Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. Annu. Rev. Biochem. **66,** 315-335.

Heese-Peck, A., Cole, R.N., Borkhsenious, O.N., Hart, G.H., and Raikhel, N.V. (1995). Plant nuclear pore complex proteins are modified by novel oligosaccharides with terminal N-acetylglucosamine. Plant Cell **7,** 1459-1471.

Hicks, G.R., Smith, H.M.S., Lobreaux, S., and Raikhel, N.V. (1996). Nuclear import in permeabilized protoplasts from higher plants has unique features. Plant Cell **8,** 1337-1352.

Jacobsen, S.E., Binkowski, K.A., and Olszewski, N.E. (1996). SPINDLY, a tetratricopeptide repeat protein involved in gibberellin signal transduction in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **93,** 9292-9296.

Kreppel, L.K., Blomberg, M.A., and Hart, G.W. (1997). Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. J. Biol. Chem. **272,** 9308-9315.

Lubas, W.S., Frank, D.W., Krause, M., and Hanover, J.A. (1997). O-linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. J. Biol. Chem. **272,** 9316-9324.

Rout, M.P., and Wente, S.R. (1994). Pores for thought: nuclear pore complex proteins. TICB **4**, 357-365.