The Plant Cell, Vol. 10, 599–612, April 1998 © 1998 American Society of Plant Physiologists

A Glycoprotein Modified with Terminal *N*-Acetylglucosamine and Localized at the Nuclear Rim Shows Sequence Similarity to Aldose-1-Epimerases

Antje Heese-Peck[1] and Natasha V. Raikhel[2]

Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824-1312

Several glycoproteins that are present at the nuclear rim and at the nuclear pore complex of tobacco suspension-cultured cells are modified by O-linked oligosaccharides with terminal *N*-acetylglucosamine (GlcNAc). Here, we report on the purification of several of these glycoproteins, which are referred to as terminal GlcNAc (tGlcNAc) proteins. In vitro galactosylation of the tGlcNAc proteins generated glycoproteins with terminal galactosyl-β-1,4-GlcNAc and thus permitted their isolation by *Erythrina crystagalli* agglutinin affinity chromatography. Peptide sequence information derived from one tGlcNAc protein with an apparent molecular mass of 40 to 43 kD, designated gp40, made it possible to clone its gene. Interestingly, gp40 has 28 to 34% amino acid identity to aldose-1-epimerases from bacteria, and no gene encoding an aldose-1-epimerase has been isolated previously from higher organisms. Polyclonal antibodies were generated against recombinant gp40. Consistent with its purification as a putative nuclear pore complex protein, gp40 was localized to the nuclear rim, as shown by biochemical fractionation and immunofluorescence microscopy.

INTRODUCTION

In eukaryotes, the nucleus is separated from the cytoplasm by the nuclear envelope, and communication between these two compartments is essential for proper functioning of the cell. Molecules, such as proteins, different RNA species, T-DNA, and ions, move in and out of the nucleus through nuclear pore complexes (NPCs). The NPCs are enormous proteinaceous structures of ~125 MD that form aqueous channels across the nuclear envelope (Davis, 1995). The overall morphology of NPCs appears to be conserved among different organisms, and the NPC architecture has been extensively investigated in amphibian cells by using electron microscopy (Panté and Aebi, 1993; Goldberg and Allen, 1995).

Although much is known about NPC structure, little has been learned about the protein composition of the NPC. The identification of NPC proteins, however, is crucial in elucidating the molecular mechanism that underlies the continuous transport of diverse molecules in and out of the nucleus. In vertebrates and yeast, a combination of immunological, biochemical, and genetic approaches has led to the identification of one-third of the ~100 different NPC proteins (reviewed in Rout and Wente, 1994; Davis, 1995; Doye and Hurt, 1997). Most of our knowledge of plant NPC proteins has come from a few biochemical analyses, and no plant NPC protein has been identified and characterized molecularly. By using an antibody raised against the yeast NPC protein Nsp1p (Hurt, 1988), a 100-kD protein was recognized in an enriched nuclear matrix preparation of carrot suspension-cultured cells located at the NPC (Scofield et al., 1992). Studies by Hicks and Raikhel (1993, 1995a) have shown that proteins are present at the nuclear envelope, and NPCs from isolated monocot and dicot nuclei can bind to nuclear localization signals (NLSs) in vitro. Consistent with these results, the plant homolog of importin α (Impα), the NLS receptor protein, is found not only in the nucleus and cytoplasm but also accumulates at the nuclear envelope (Smith et al., 1997).

Using wheat germ agglutinin, which is a lectin that binds specifically to *N*-acetylglucosamine (GlcNAc), as a probe, we and others have found glycoproteins modified by GlcNAc at the plant nuclear periphery (Heese-Peck et al., 1995; Hicks et al., 1996; Merkle et al., 1996). As shown by electron microscopy and biochemical analyses, at least some of these glycoproteins are localized at the NPC of tobacco suspension-cultured cells and are modified by terminal GlcNAc (tGlcNAc; Heese-Peck et al., 1995). Most glycans with tGlcNAc are attached to the proteins via an O-linkage and are larger in size than five GlcNAc residues (Heese-Peck et al., 1995). Thus, this plant sugar modification is different

[1] Current address: Biozentrum of the University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.
[2] To whom correspondence should be addressed. E-mail nraikhel@pilot.msu.edu; fax 517-353-9168.

from the single O-linked GlcNAc (O-GlcNAc) found on vertebrate NPC proteins (reviewed in Forbes, 1992; Hicks and Raikhel, 1995b). In contrast to most other known sugar modifications of proteins, the addition of the vertebrate O-GlcNAc occurs in the cytoplasm (Hart, 1997). The gene encoding O-GlcNAc transferase, which is the enzyme specific for the attachment of O-GlcNAc to proteins, has been recently cloned from vertebrates (Kreppel et al., 1997; Lubas et al., 1997). It shows extensive sequence similarity to *SPINDLY*, a gene from Arabidopsis involved in gibberellin signal transduction (Jacobsen et al., 1996). Interestingly, O-GlcNAc modification has not been observed for any NPC protein from the yeast *Saccharomyces cerevisiae* (Rout and Wente, 1994), and no homolog of the O-GlcNAc transferase appears to be present in the yeast genome (Lubas et al., 1997).

Although the function of O-GlcNAc modification is not fully understood (Miller and Hanover, 1994), this glycosylation has been proven useful in purifying vertebrate NPC proteins for subsequent functional analysis (reviewed in Forbes, 1992; Davis, 1995; Hicks and Raikhel, 1995b). As an initial step in isolating genes encoding plant proteins at the nuclear rim and possibly at the NPC, we purified tobacco proteins modified by O-linked oligosaccharides with tGlcNAc, referred to as tGlcNAc proteins, by lectin affinity chromatography. Peptide sequence information from one tGlcNAc protein with an apparent molecular mass of 40 to 43 kD, designated gp40, made it possible to clone its gene and investigate its cellular localization. This tGlcNAc protein showed significant sequence similarity to aldose-1-epimerases, and its possible functions are discussed.

RESULTS

Isolation of Nuclear Glycoproteins with tGlcNAc

To begin our investigation of plant NPC proteins, we demonstrated previously that glycoproteins modified by GlcNAc are present at the nuclear rim and at the NPC of tobacco suspension-cultured cells. Most of these glycoproteins can be partially extracted with 0.5 M salt; thus, they display properties similar to those described for vertebrate NPC proteins. The majority of these glycoproteins are modified by O-linked oligosaccharides with tGlcNAc, and they can be detected by using wheat germ agglutinin, a lectin that binds to internal as well as tGlcNAc residues (Ebisu et al., 1977), and by in vitro galactosyltransferase (GalTF) assays (Heese-Peck et al., 1995). In in vitro GalTF assays, the enzyme GalTF adds galactose (Gal) specifically to terminal but not internal GlcNAc residues to generate glycoproteins with terminal galactosyl-β-1,4-*N*-acetylglucosamine (Gal-β-1,4-GlcNAc; Roquemore et al., 1994). Thus, this assay is designed specifically to detect proteins with tGlcNAc residues and was used to purify tGlcNAc proteins found at the nuclear rim and NPC of tobacco.

Nuclei were isolated from tobacco suspension-cultured cells in the presence of 0.6% Triton X-100 by using Percoll gradients. A 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 nuclear envelope are also removed; nevertheless, the nuclei remain intact because nuclear integrity is not dependent on the nuclear envelope (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 1, lanes 1 and 5), was incubated with radiolabeled UDP–^{3}H-Gal in the presence of GalTF (Figure 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 1, lane 6; Heese-Peck et al., 1995). The in vitro GalTF assay was performed at 37°C; at this temperature, some proteins precipitated out of solution. The precipitated proteins (Figure 1, lanes 3 and 7) were separated from the soluble proteins (Figure 1, lanes 4 and 8) by centrifugation, and almost all of the tGlcNAc proteins were found in the soluble fractions (Figure 1, lane 8). Thus, incubation at 37°C provided a convenient step to enrich for tGlcNAc proteins, as was observed when the insoluble (containing the precipitated proteins) and soluble protein patterns were compared using Coomassie blue staining (Figure 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 nonradioactive Gal, as previously described by Miller and Hanover (1994).

We then examined whether these galactosylated tGlcNAc proteins could be detected by *Erythrina crystagalli* agglutinin (ECA), a lectin shown to be specific for Gal-β-1,4-GlcNAc (Lis et al., 1985). When protein gel blots were incubated with ECA coupled to alkaline phosphatase (ECA-AP; Figure 1, lanes 9 to 12), proteins were detected by using the ECA probe, which had an apparent molecular mass similar to those being visualized by fluorography (Figure 1, cf. lanes 6 and 10 as well as 8 and 12, respectively). One additional glycoprotein with an apparent molecular mass of 35 kD was detected by using the ECA probe (Figure 1, lanes 9 to 12). Because this protein was also detected by the ECA probe before the in vitro GalTF assay (Figure 1, lane 9; see below), it was not considered to be a tGlcNAc protein. As defined above, tGlcNAc proteins require the addition of galactose in the in vitro GalTF assay for subsequent detection with the ECA probe. Taken together, these results indicate that labeling of tGlcNAc proteins with nonradioactive Gal and subsequent detection with ECA-AP provide an excellent alternative method for the detection of glycoproteins with tGlcNAc.

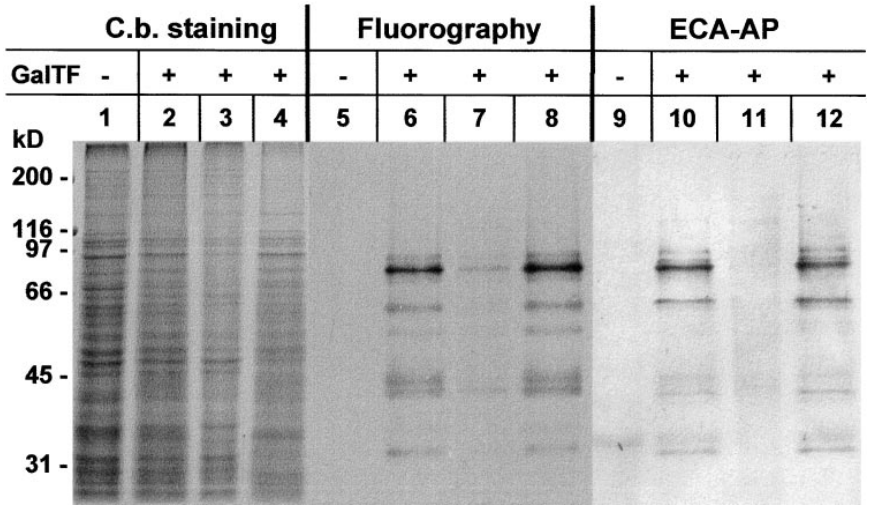

Figure 1. Detection of Galactosylated tGlcNAc Proteins by Using Fluorography and ECA Gel Blot Analysis.

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

The ability of ECA to bind to the galactosylated tGlcNAc proteins was used in the purification of these glycoproteins by lectin affinity chromatography. Salt-extracted proteins were galactosylated with nonradiolabeled Gal in the presence of GalTF, and soluble proteins were loaded onto ECA–agarose columns (Figure 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 2A). With the exception of a 35-kD glycoprotein (Figure 2A, asterisk; see below), the tGlcNAc proteins bound to the ECA–agarose column, and only small amounts were found in the flowthrough (Figure 2A, lane 2). The tGlcNAc proteins were eluted with the specific sugar lactose (Gal-β-1,4-Glu; Figure 2A, lane 4), a disaccharide closely related to Gal-β-1,4-GlcNAc and shown to be a competitor of Gal-β-1,4-GlcNAc binding (Lis et al., 1985). These glycoproteins were not eluted with the unrelated sugar mannose (Figure 2A, lane 3), demonstrating the specific interaction of the galactosylated tGlcNAc proteins with the ECA-agarose.

Tight interaction between glycoproteins and lectin was further disrupted by 1 M salt. One tGlcNAc protein with an apparent molecular mass of 40 to 43 kD, designated gp40, was present not only in the lactose but also in the 1 M salt eluate, indicating that gp40 was tightly bound to ECA-agarose (Figure 2A, lane 5). The same blot was subsequently stained with Coomassie blue to visualize all proteins in each column fraction (Figure 2B). These results confirmed that ECA affinity chromatography serves as an efficient means to purify tGlcNAc proteins (Figure 2B, lanes 4 and 5).

In control experiments, proteins were treated as described above, except that GalTF was omitted (Figures 2C and 2D). No proteins were present in the lactose- and 1 M salt-eluted fractions (Figures 2C and 2D, lanes 4 and 5, respectively), demonstrating 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 2C, lane 1 and asterisk), but it did not bind to the column (Figure 2C, lane 2). This glycoprotein is most likely the 35-kD protein that was detected by ECA-AP before in vitro GalTF assays (Figure 1, lane 9, and Figure 2A, lanes 1 and 2, and asterisk) and was not analyzed further.

To obtain sufficient material for internal amino acid sequence analysis, large-scale purification of tGlcNAc proteins was performed using ECA affinity column chromatography, as described above. Proteins eluted in the presence of lactose as well as 1 M salt were separated on 7.5 to 15% gradient SDS–polyacrylamide gels, transferred to membranes, and visualized by amido black staining. For tryptic digestion, sufficient material of gp40 present in the 1 M salt eluate was recovered. Internal amino acid sequence analysis of gp40

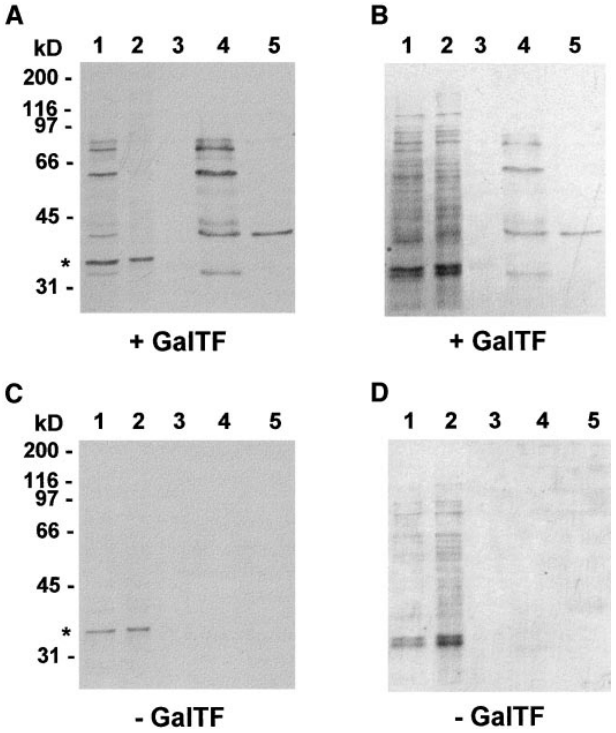

Figure 2. Purification of Galactosylated tGlcNAc Proteins by ECA Affinity Chromatography.

Salt-extracted nuclear proteins were subjected to a nonradioactive in vitro GalTF assay in the presence (+) or absence (−) of GalTF and separated into insoluble and soluble fractions. The soluble fractions (lanes 1) were loaded onto ECA–agarose columns, and column fractions were collected. Lanes 2 contain flowthrough; lanes 3, 0.1 M mannose wash; lanes 4, 0.05 M lactose eluate; and lanes 5, 1 M NaCl eluate. Molecular mass standards are indicated at left in **(A)** and **(C)** in kilodaltons. The asterisk indicates a nonspecific protein in **(A)** and **(C)**.

(A) Assay performed in the presence of GalTF. Fractions were subjected to protein gel blot analysis using ECA-AP.

(B) Assay performed in the presence of GalTF. Total proteins were visualized by immersing the blots in Coomassie blue.

(C) Assay performed in the absence of GalTF. Fractions were subjected to protein gel blot analysis using ECA-AP. Total proteins were visualized by immersing the blots in Coomassie blue.

(D) Assay performed in the absence of GalTF.

generated sequence information for three peptides, designated gp40-PT48, gp40-PT64, and gp40-PT72 (Figure 3A).

Cloning of the Gene Encoding gp40

The gp40 peptides were compared with sequences of known proteins and expressed sequence tags from eukaryotes and prokaryotes available in the databases; however, no sequence similarity was found. Therefore, polymerase chain reaction (PCR) amplification was employed to clone the gene encoding gp40. Degenerate oligonucleotide primers based on the information obtained from the gp40 peptides were used. A cDNA library made from poly(A)$^+$ RNA from tobacco suspension-cultured cells (Facchini and Chappell, 1992) was the template, from which a specific 600-bp PCR

A

```
gp40-PT48    I  G  G  A  Q  F  T  L  -  G  T  H [Y]
gp40-PT64    N  T  P  Y  D  F  L  K [P][R]
gp40-PT72    I  G  D  V  V  L  G  Y (D)[T]
```

B

```
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
```

C

```
  1  M S L K I N L L V C L F I F H L L V A A*S V R G
 25  H K I G I Y E I K K G D F S V K I T N Y G T S I
 50  I S V L L P D K H G K I G D V V L G Y D T I E E
 75  Y K N D T S Y F G A T L G R V A N R I G G A Q F
100  T L N G I H Y K L V P N E G K N M L H G G P K G
125  F S K V V W K V S K Y V K D G P C P Y I T L T Y
150  Y S A D G E E G F P G A V L A S V T Y T L K D S
175  Y K L S V V F R A K A L N K A T P I N L S H H P
200  Y W N I G G H D S G D V L S Q V L Q I F G S H I
225  T L V D K Q L I P T G E I A P I K N T P Y D F L
250  K P R K V G S R I N K L K N G Y D I N Y V L D S
275  T E K M K P V G I V Y D K K S G R V M D V Q A S
300  S P G V Q F Y T A N F V N N T K G K G G F V Y Q
325  P H S A L S L E T L V F P D A V N H P N F P S T
350  I V N P G E K Y V H S V L Y T F S I K K
```

Figure 3. Deduced Amino Acid Sequence of gp40.

(A) The peptide sequences of gp40-PT48, gp40-PT64, and gp40-PT72 were derived from an internal amino acid sequence analysis of gp40. The dash indicates that no amino acid was determined; residues within brackets indicate probable amino acids.

(B) The 5′ ends of the three specific PCR products—p64s/8, p64s/9, and p64s/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.

(C) The deduced amino acid sequence of gp40 was 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 shaded. The cleavage site of a putative signal sequence is marked by an asterisk, and the putative NLS is indicated in boldface. The GenBank accession number for the sequence is AF032386.

fragment was amplified using a pt64s primer and a T7 promotor–specific primer present in the pcDNAII plasmid. Three independent clones, designated p64s/8, p64s/9, and p64s/h, were isolated, and sequence analysis revealed that they are identical (data not shown). The PCR fragments contained the nucleotide sequence encoding the entire peptide fragment gp40-PT64, confirming the identity of the PCR fragments (Figure 3B); however, the remaining peptides—gp40-PT48 and gp40-PT72—were not present in the PCR fragments.

Screening of the tobacco cDNA library with PCR fragment p64s/8 resulted in the isolation of only partial clones (see Methods). Therefore, 5′ rapid amplification of cDNA ends (5′ RACE) was employed to obtain the missing 5′ end of the clone. Due to its proofreading ability, Pwol 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-day-old tobacco suspension-cultured cells as template. To confirm the sequence identity, nested primers (see Methods) were designed such that the 5′ RACE product would have a 183-bp overlap with the 5′ end of one of the truncated cDNA clones, *GP40-5e.* 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 (a T residue for a C residue at nucleotide 169) that did not alter the deduced amino acid sequence (Figure 3C). Significantly, gp40-PT48 and gp40-PT72 were present in the deduced amino acid sequence of the 5′ RACE product (Figure 3C, shaded), 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 3A and 3C), which may be due to an error during amino acid sequencing. However, it could not be ruled out that gp40-PT48 was derived from another very closely related protein with properties similar to gp40.

For further studies, the fragments RACE-7 and GP40-5e were combined using an XmnI site to yield the full-length clone *GP40* (GenBank accession number AF032386). The size of the combined fragments (1.25 kb) is consistent with the *GP40* transcript size of ~1.3 kb detected during RNA gel blot analysis (data not shown). The nucleotide sequence of *GP40* contains an open reading frame encoding 357 amino acids with a predicted molecular mass of 39.3 kD, which is in agreement with the apparent molecular mass of gp40. The deduced amino acid sequence of *GP40* contains ~10% lysine residues. Consistent with being glycosylated by O-linked oligosaccharides, gp40 is rich in serine and threonine (~13%). The sequence of gp40 encodes for at least one putative NLS (Figure 3C, marked in bold; Raikhel, 1992) and a putative signal sequence with a cleavage site between amino acids 20 and 21 (Figure 3C, see asterisk; von Heijne, 1986). At this point, it is not known whether the putative NLS or signal sequence is actually utilized for subcellular localization of gp40 (see Discussion).

Interestingly, the deduced amino acid sequence of gp40 shares 28 to 34% amino acid sequence identity and 45 to 51% amino acid sequence similarity with all of the aldose-1-epimerases shown in Figure 4. These enzymes, also referred to as mutarotases, catalyze the interconversion of the anomeric forms (α and β) at 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). To date, no gene encoding an aldose-1-epimerase has been identified from higher organisms, although apparent Arabidopsis (GenBank accession number Z25633), rice (D22462), and human (U11036) homologs exist in the expressed sequence tag databases (data not shown).

gp40 Is a Nuclear-Associated Protein

To investigate the intracellular localization of gp40, we prepared polyclonal antibodies to recombinant gp40. The entire open reading frame of gp40 was fused in frame to the C terminus of glutathione *S*-transferase (GST). The GST–gp40 fusion protein was expressed in *Escherichia coli* as an insoluble protein (data not shown) and isolated from SDS–polyacrylamide gels for antibody production in rabbits. Polyclonal antibodies were depleted of antibodies that recognized the GST portion of the fusion protein, as described by Bar-Peled and Raikhel (1996). To obtain monospecific antibodies generated against recombinant gp40, the GST-depleted antibodies were further affinity purified against recombinant GST–gp40 that was immobilized onto polyvinylidene fluoride (PVDF) membranes.

To examine the subcellular localization of gp40, we isolated protoplasts from tobacco suspension-cultured cells for subsequent nuclear isolation and biochemical fractionation. Equal amounts of protein were separated by SDS-PAGE and visualized by Coomassie blue staining (Figure 5A) or analyzed by protein gel blot analysis (Figures 5B and 5C). By using gp40 antibodies, a single protein was detected in total protoplasts that had an apparent molecular mass similar to gp40 (Figure 5B, lane 1). The antigen copurified with nuclei and was present in the crude nuclear (Figure 5B, lane 3) and the purified nuclear (Figure 5B, lane 5) fractions. It was not detected in the soluble fraction of 0.01% Triton X-100–treated protoplasts (Figure 5B, lane 2), indicating that it was not a cytosolic protein. In addition, it was not associated with any membrane and it was not present in the ER lumen, because it was not released when crude nuclei were exposed to 0.6% Triton X-100 (Figure 5B, lane 4). The antigen was partially solubilized when purified nuclei were incubated with DNase I buffer (Figure 5B, lane 6). Exposure of DNase I–treated nuclei to 0.5 M salt led to the extraction of most of the remaining antigen (Figure 5B, lane 8).

These fractionation studies indicated that the antigen was associated with peripheral structures of the nucleus. Moreover,

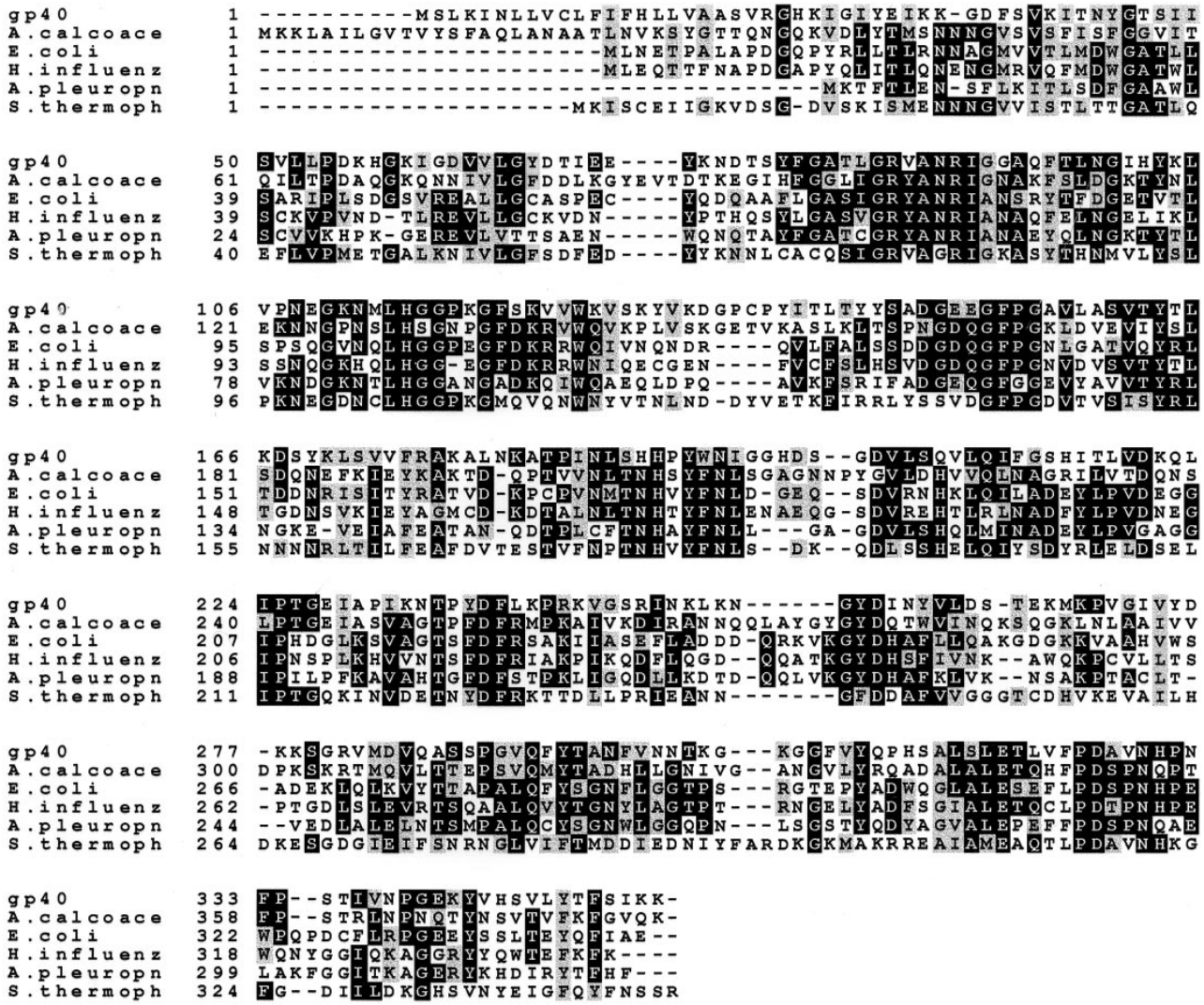

Figure 4. Comparison of Amino Acid Sequences of gp40 and Bacterial Aldose-1-Epimerases.

The deduced amino acid sequence of gp40 was aligned with the sequences of aldose-1-epimerases from *A. calcoaceticus* (A. calcoace; GenBank accession number X03893), *E. coli* (GenBank accession number U13636), *Haemophilus influenzae* (H. influenz; GenBank accession number U32764), *Actinobacillus pleuropneumoniae* (A. pleuropn; GenBank accession number U63731), and *Streptococcus thermophilus* (S. thermoph; GenBank accession number M38175). Amino acids marked in black and grey boxes indicate sequence identity and similarity, respectively. Gaps introduced for alignment of homologous regions are indicated by dashes.

the results correlated well with our previous extraction studies in which we identified gp40 as a tGlcNAc protein that was extracted from nuclei with 0.5 M salt (Figures 1 and 2; 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 higher apparent molecular mass than that of gp40 was also detected by using the gp40 antibody (Figure 5B, lanes 6 and 8). This protein may represent either gp40 with additional post-translational modifications or an antigenically closely related nuclear protein with biochemical properties similar to gp40. In control experiments, it was observed that NT-IMPα, an NLS receptor (Smith et al., 1997), possessed a different subcellular localization than that of gp40 (Figure 5C). In agreement with its proposed function as a nuclear/cytoplasmic shuttle protein and with studies by Smith et al. (1997), NT-IMPα was found in cytoplasmic and nuclear fractions (Figure

5C, lanes 1 to 5). The nuclear-associated form of NT-IMPα behaved in a manner similar to gp40 in part; it was partially extracted by using DNase I and 0.5 M salt (Figure 5C, lanes 6 and 8). Consistent with the results described by Smith et al. (1997), NT-IMPα appeared to be more tightly associated with the nucleus than was gp40, and a significant portion of the protein was still associated with the insoluble fraction after extraction with 0.5 M salt (Figure 5B, lane 6).

To show conclusively that gp40 was indeed the antigen detected by the gp40 antibodies, salt-extracted nuclear proteins were subjected to nonradioactive in vitro GalTF assays followed by ECA column affinity chromatography. Column fractions were collected as described above and analyzed by protein blotting, with the gp40 antibodies being used (Figure 6, lanes 1 to 5). As expected, the antibodies recognized a protein with a slightly higher apparent molecular mass than that of the nongalactosylated 40-kD antigen due to the addition of Gal (Figure 6, lane 1; data not shown). Most of the protein bound to the ECA–agarose and was eluted with lactose and 1 M salt (Figure 6, lanes 4 and 5); thus, it displayed binding properties similar to gp40 (Figures 2A and 2B). When blots were stripped and reprobed with ECA-AP, this protein was also recognized by ECA-AP (Figure 6, lane 7), demonstrating that it is a tGlcNAc protein. In addition to the galactosylated gp40, two proteins with slightly higher apparent masses (43 to 45 kD) were detected in the lactose eluate by the antibody (Figure 6, lane 4). These proteins were also tGlcNAc proteins because they were detected by ECA (Figure 6, lane 6). As discussed above, these proteins may be closely related to gp40 or post-translationally modified gp40.

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

gp40 Is Associated with the Nuclear Rim

Immunofluorescence microscopy was used to confirm the nuclear localization of gp40 and to investigate further its localization. Consistent with the biochemical fractionation studies (Figure 5B), immunofluorescence labeling was detected primarily around the nuclear rim by using conventional light microscopy (Figure 7A). A CY3-labeled secondary antibody gave only low-level 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 using confocal laser scanning microscopy, which confirmed that gp40 is present primarily

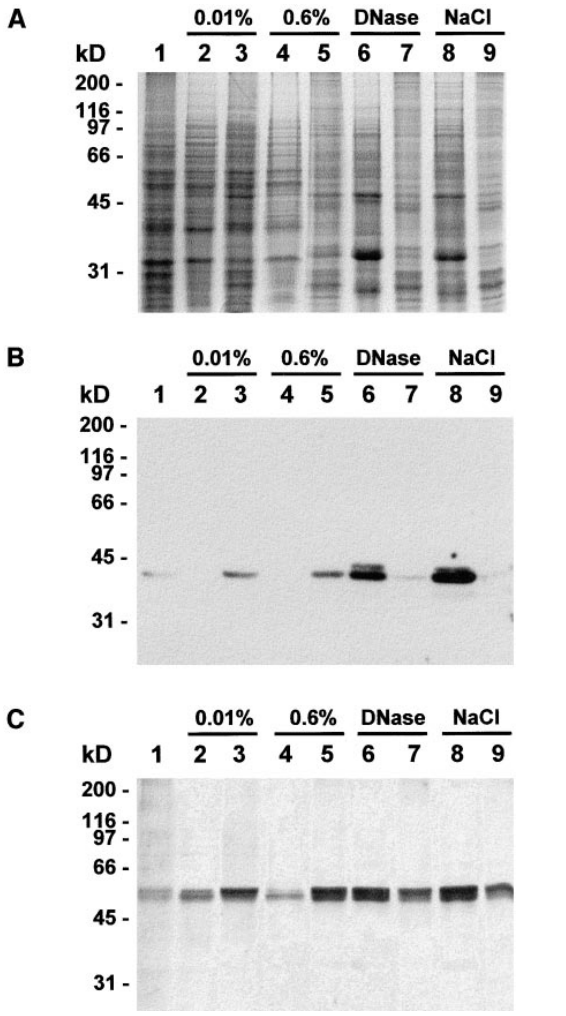

Figure 5. Cofractionation of gp40 with the Nucleus.

Proteins were fractionated and extracted from tobacco suspension-cultured cells under the following conditions. Protoplasts (lanes 1) were broken in the presence of 0.01% Triton X-100 and separated into cytoplasm I (lanes 2) and crude nuclei (lanes 3). Crude nuclei were then incubated with 0.6% Triton X-100 and separated into cytosol II/released membrane proteins (lanes 4) and purified nuclei (lanes 5). Purified nuclei were treated with DNase I and separated into soluble (lanes 6) and insoluble (lanes 7) fractions by centrifugation. The insoluble fraction was then incubated with a 0.5 M NaCl buffer and separated into soluble (lanes 8) and insoluble (lanes 9) fractions by centrifugation. Equal amounts of proteins (20 mg) were loaded in each lane and separated by SDS-PAGE. Molecular mass standards are indicated at left in kilodaltons. 0.01%, 0.01% Triton X-100; 0.6%, 0.6% Triton X-100; DNase, DNase I; NaCl, 0.5 M NaCl.
(A) Proteins stained with Coomassie blue.
(B) Proteins transferred to PVDF membranes. Antibody-reactive proteins were visualized using gp40 antibodies followed by incubation with protein A–horseradish peroxidase.
(C) Proteins transferred to PVDF membranes. Antibody-reactive proteins were visualized using IMPα antibodies followed by the addition of anti-IgG–alkaline phosphatase.

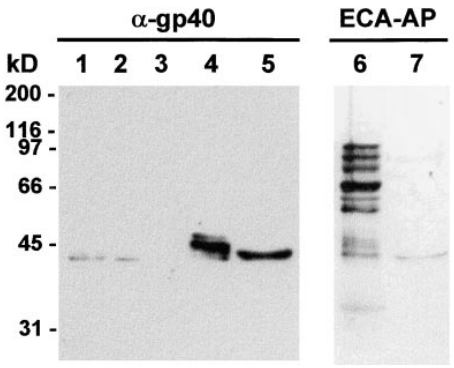

Figure 6. Purification of Galactosylated gp40 by ECA Affinity Chromatography.

Salt-extracted nuclear proteins were subjected to a 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. Lane 2 contains flowthrough; lane 3, 0.1 M mannose wash; lane 4, 0.05 M lactose eluate; and 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–horseradish peroxidase (lanes 1 to 5). The blot containing lanes 4 and 5 was then stripped of the antibodies and reprobed with ECA-AP to visualize galactosylated tGlcNAc proteins (lanes 6 and 7, respectively). Molecular mass standards are indicated at left in kilodaltons.

at the nuclear rim, with little labeling within the nucleus (Figure 7B). In control experiments using confocal laser scanning microscopy, NT-IMPα localization was detected in the cytoplasm, the nucleus, and the nuclear rim (Figure 7C). These results are consistent with those previously obtained by Smith et al. (1997) and with the biochemical fractionation studies (Figure 5C). The results obtained from microscopy studies (Figure 7), taken together with the biochemical fractionation studies (Figures 5B and 6), indicate that gp40 is a tGlcNAc protein associated with the nuclear rim.

DISCUSSION

In our study, we developed a purification scheme to isolate nuclear tGlcNAc proteins that previously were shown to have the characteristics of proteins present at the nuclear rim. Some of these proteins are localized to the NPC. Labeling of tGlcNAc proteins with Gal in vitro generated glycoproteins with a terminal Gal-β-1,4-GlcNAc modification. Due to the specific affinity of ECA for Gal-β-1,4-GlcNAc, lectin affinity chromatography provided an efficient means by which to purify the modified tGlcNAc proteins in large quantities for subsequent internal amino acid sequence analysis. Peptide sequence information was obtained from three isolated tGlcNAc proteins with apparent molecular masses of 33, 40 to 43, and 65 kD (named gp33, gp40, and gp65, respectively) (this study; A. Heese-Peck and N.V. Raikhel, unpublished data). Here, we have focused on the cloning of *GP40* and initial characterization of this protein. A combination of PCR-based amplification, cDNA library screening, and 5′ RACE was used to clone the gene encoding gp40. Its deduced amino acid sequence showed significant sequence similarity to aldose-1-epimerases from bacteria. By using monospecific antibodies raised against recombinant gp40, it was demonstrated that gp40 is a tGlcNAc protein associated with the nucleus. Consistent with its biochemical isolation as a putative NPC protein, conventional light and laser scanning immunofluorescence microscopy indicated that gp40 is localized primarily to the nuclear rim of tobacco protoplasts.

gp40 Is Localized at the Nuclear Rim

In our previous and current studies, the in vitro GalTF assay has served as an excellent tool for the identification and purification of nuclear tGlcNAc proteins, including gp40. The production of antibodies made against gp40 allowed a more detailed biochemical and microscopic analysis of this protein. In agreement with our initial purification scheme, fractionation studies showed that gp40 is a nuclear protein. Based on its extraction by 0.5 M salt but not by detergent treatment, gp40 appeared to be associated peripherally 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 within the NPC core structures or the nuclear lamina/matrix, which is a filamentous structure found throughout the nucleus (Dwyer and Blobel, 1976; Davis and Blobel, 1986; Masuda et al., 1997).

gp40 was also released by treatment with DNase I; however, based on immunofluorescence microscopy studies, gp40 was found at the nuclear periphery, and little labeling was observed within the nucleus. We do not believe that such a small amount could account for the amount of gp40 released by DNase I. It should be noted that the DNase I buffer used in our studies contained Mg^{2+}, a divalent ion shown to be involved in the disassembly and reassembly of the nucleoplasmic basket in amphibians (Jarnik and Aebi, 1991). However, because resolution of our micrographs was not very high, we were not able to detect whether gp40 was localized precisely to the NPC in tobacco nuclei.

As mentioned above, gp40 has a putative signal sequence as well as at least one putative NLS. To date, we do not know whether either of these two signals is 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 signals can be functional and utilized for dual pro-

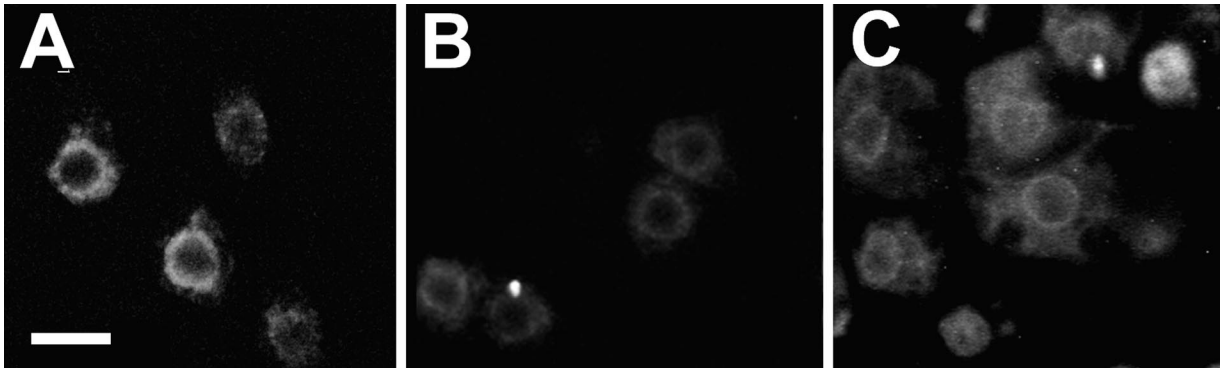

Figure 7. gp40 Is Localized at the Nuclear Rim of Tobacco Protoplasts.

Cells from which protoplasts were isolated were fixed and incubated with gp40 or IMPα antibodies followed by CY3-labeled secondary antibodies.
(A) gp40 antibodies detected an antigen at the nuclear rim. The cells were visualized by epifluorescence microscopy.
(B) gp40 antibodies detected an antigen at the nuclear rim. The cells were visualized by using confocal laser scanning microscopy.
(C) IMPα antibody detected an antigen at the cytoplasm, nuclear rim, and nucleus. The cells were visualized by using confocal laser scanning microscopy.
Bar in **(A)** = 30 μm for **(A)** to **(C)**.

tein localization (Maher et al., 1989; Spence et al., 1989; Kimura, 1993; Kiefer et al., 1994). If the putative signal sequence is utilized for gp40 localization, it is also 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 are localized to the pore membrane domain of the nuclear envelope. These NPC proteins appear to provide contact sites for proteins in the cytoplasm as well as in the ER lumen (Wozniak et al., 1989; Greber et al., 1990; Hallberg et al., 1993). However, it should be pointed out that gp40 is modified by O-linked oligosaccharides with tGlcNAc (Heese-Peck et al., 1995; N.V. Raikhel, A. Heese-Peck, and A. Bacic, unpublished data). 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 (Driouich et al., 1993). In addition, in vertebrates, O-GlcNAc is added in the cytoplasm by the O-linked GlcNAc transferase (reviewed in Hart, 1997). The *SPINDLY* gene from Arabidopsis (Jacobsen et al., 1996) shows extensive sequence identity to genes encoding the vertebrate O-linked GlcNAc transferase (Kreppel et al., 1997; Lubas 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 to 34%) with aldose-1-epimerases from several bacteria, with the highest being that of *Acinetobacter calcoaceticus* (34% identity and 51% similarity). In addition, gp40 has a molecular mass similar to bacterial enzymes (Gatz et al., 1986; Poolman et al., 1990; Mollet and Pilloud, 1991; Maskell et al., 1992; Bouffard et al., 1994). Considering the evolutionary distance between bacteria and plants, it is noteworthy that bacterial aldose-1-epimerases have a similar percentage of sequence identity (~30%) with gp40 as they do with each other (Gatz et al., 1986; Poolman et al., 1990; Mollet and Pilloud, 1991; Maskell et al., 1992; Bouffard et al., 1994).

Alignment of gp40 and bacterial sequences shows several well-conserved blocks, suggesting structural and maybe even functional similarities. Although no definitive amino acid residue has been assigned for aldose-1-epimerase function to date, biochemical studies of purified aldose-1-epimerases implicated histidine, tyrosine, and tryptophan residues as components of the active site (Hucho and Wallenfels, 1971; Fishman et al., 1973). In the sequencing alignment, one tryptophan (W-121), two histidine (H-115 and H-192), and two tyrosine (Y-163 and Y-194) residues are conserved between all bacterial aldose-1-epimerases and gp40. Bacterial enzymes and gp40 also share some sequence identity (~20%) with the C-terminal portion of the fungal protein Gal10p from yeast and *Pachysolen tannophilus.* However, the fungal proteins have a predicted mass of approximately twice the size of gp40 and bacterial aldose-1-epimerases (Citron and Donelson, 1984; Skrzypek and Maleszka, 1994).

Even though aldose-1-epimerases from different vertebrate tissues have been purified and their enzyme activity characterized, 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). No further biochemical fractionation and subnuclear localization have been reported. The polyclonal antibodies against rat kidney aldose-1-epimerase (Toyoda et al., 1983)

recognized the gp40 that was isolated by ECA affinity chromatography, further demonstrating that gp40 is antigenically related to rat aldose-1-epimerase (A. Heese-Peck and N.V. Raikhel, unpublished data). These antibodies, however, did not detect any protein in crude tobacco fraction and were not useful for analyses in plants.

Possible Functions of gp40

In bacteria, aldose-1-epimerase functions in carbohydrate metabolism (Bouffard et al., 1994); to date, its exact function in higher organisms is not known. Bailey et al. (1967, 1970) argue against the involvement of this enzyme in metabolic sugar breakdown in higher organisms, because its distribution in mammalian tissue is not consistent with glycolytic activity. 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 with mammalian cells, the active site of aldose-1-epimerase has a structural organization similar to 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 related functionally to aldose-1-epimerases, but to date, we do not know whether gp40 possesses aldose-1-epimerase activity.

Giving rise to new speculation about the putative functions of aldose-1-epimerase–like proteins is the localization of gp40 at the nuclear rim. Support for a possible role of gp40 in nuclear import of glycoproteins comes from studies by Duverger et al. (1995). These authors report that glycosylated BSA, which does not contain an NLS, is imported into the nucleus of permeabilized HeLa cells in a sugar-specific but NLS-independent manner. A different uptake mechanism has been proposed that may involve the presence of sugar binding proteins at the NPC (Duverger et al., 1995). One can speculate that gp40 may serve a similar role because it was not only located at the nuclear periphery but also shares sequence similarity with aldose-1-epimerases. These enzymes need to bind sugars to accomplish their enzymatic function.

To gain a better understanding of gp40's function, we will design future experiments to address possible aldose-1-epimerase activity and substrate specificity; however, these were beyond the scope of this study. For such studies, gp40 should be isolated from plant tissue by procedures that circumvent modifying the protein in GalTF assays, because proper glycosylation of gp40 may be required for enzyme activity. The antibodies will serve as a useful tool in designing a new protein purification scheme. To investigate a putative role for gp40 in the nuclear import of glycoproteins and NLS-containing proteins, purified gp40 as well as antibodies raised against gp40 may be used in the in vitro nuclear import system developed in our laboratory (Hicks et al., 1996) and in electron microscope immunocytochemistry to ascertain its precise localization within the nuclear rim. Analysis of gp40 as an aldose-1-epimerase–like protein may provide new insights into the function of aldose-1-epimerases in eukaryotes.

METHODS

Nuclear Protein Isolation

Nuclei were isolated in the presence of 0.6% Triton X-100 from tobacco suspension-cultured cells, as described by 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 1000*g*. The dense yellow-brown band on top of the Percoll cushion was designated the crude nuclear fraction, and the clear yellowish fraction above the band was referred to as the cytosol I fraction. Crude nuclei were incubated with NIB plus 0.6% Triton X-100 for 20 min and filtered twice through a 20-mm nylon mesh. The filtrate was then loaded onto Percoll gradients consisting of 1 mL of 50% Percoll in NIB and 7 mL of 10% Percoll in NIB. A centrifugation step (1000*g* for 10 min) yielded a yellow-gray band directly on top of the Percoll cushion (purified nuclear fraction) that was overlaid by a clear light yellow cytosolic fraction II. The purified nuclei were then processed as described previously (Heese-Peck et al., 1995).

For further nuclear protein isolation, purified nuclei were incubated with DNase I (Boehringer Mannheim) and separated into soluble (S1) and insoluble (P1) fractions. The P1 fraction was then incubated in 0.5 M NaCl and 50 mM Hepes, pH 6.8 (2 μL per 10^6 nuclei), followed by a centrifugation step to yield a soluble (S2) fraction and an insoluble (P2) fraction (Heese-Peck et al., 1995). In general, 3 to 3.5 μg of protein per μL was present in S2. Protein concentrations were determined using Bradford protein assay (Bio-Rad), and proteinase inhibitors were used during all steps (Heese-Peck et al., 1995).

Protein Gel Blot Analysis

Proteins were separated on SDS–polyacrylamide gels, transferred to a polyvinylidene fluoride (PVDF) membrane (Millipore, Bedford, MA), blocked overnight in 3% BSA in Tris-buffered saline plus 0.1% Tween 20, and further processed according to standard procedures (Harlow and Lane, 1988). Where indicated, 1 mg/mL *Erythrina cristagalli* agglutinin–alkaline phosphatase (ECA-AP; EY Laboratories, San Mateo, CA) was used as a probe, and antibodies raised against importin α (AtIMPα; Smith et al., 1997) or gp40 (see below) were used as primary antibodies, followed by incubation with goat anti–rabbit IgG–AP or protein A–horseradish peroxidase (Sigma), respectively. Proteins detected by protein A–horseradish peroxidase were visualized using a chemiluminescence kit, according to the manufacturer's specifications (SuperSignal; Pierce, Rockford, IL). Where specified, after detection by chemiluminescence, the protein blots were stripped, as described in the manufacturer's protocol (Pierce), and reprobed with ECA-AP.

Galactosyltransferase Labeling Assays

In vitro galactosyltransferase (GalTF) labeling of salt-extracted proteins (S2) was performed essentially as described by Roquemore et al. (1994), using 100 milliunits of GalTF (Sigma or FLUKA, Ronkonkoma, NY) and either 4 μCi UDP–^{3}H-Gal (1.0 mCi/mL; 17.3 Ci/mmol; Amersham, Arlington Heights, IL) or 500 μM UDP–Gal (Sigma) in a 100-μL reaction. After incubation at 37°C for ~3 hr, samples were cleared from precipitated proteins by a centrifugation step at 12,000*g* for 5 min at room temperature into a soluble (S3) fraction and an insoluble (P3) fraction. Proteins were separated on 9% SDS–polyacrylamide gels, stained with Coomassie Brilliant Blue R 250, and transferred to a PVDF membrane, according to standard methods (Harlow and Lane, 1988; see above). Blots containing ^{3}H-Gal–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 nonradioactive galactose were probed with ECA-AP and processed as described above.

ECA Affinity Chromatography

Salt-extracted nuclear proteins (S2; ~500 mg) were galactosylated in vitro in the presence of nonradioactive UDP–Gal and GalTF, as described above. In a control reaction, proteins were incubated with nonradioactive UDP–Gal but in the absence of the GalTF. ECA affinity chromatography was based on Miller and Hanover (1994), with some modification. Before loading onto the lectin columns, cleared supernatants were diluted 1:3 with 20 mM Hepes–KOH, pH 7.3, 5 mM $MgCl_2$, and 2 mM β-mercaptoethanol to lower the salt concentration in the samples. All column steps were performed at 4°C, and all buffers contained a proteinase inhibitor mix, as was described for nuclear protein isolation.

Diluted protein samples were batch bound for 1 or 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, and 2 mM β-mercaptoethanol). After unbound proteins were collected, the columns were washed extensively with CWB (15 to 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, and 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 gel blotting, as described above. In some cases, protein blots were incubated with diluted Coomassie blue.

Peptide Sequence Analysis

For peptide sequencing, proteins were extracted from 3.5×10^9 nuclei and galactosylated in vitro in the presence of nonradioactive UDP–Gal. 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 to 15% SDS–polyacrylamide gels and transferred to PVDF membranes (Millipore) in 25 mM Tris base, 192 mM glycine, 0.01% SDS, and 10% methanol by stepwise electroblotting (0.6 mA/cm^2, 1 mA/cm^2, and 2 mA/cm^2 for 15 min each) and using a semidry protein electroblotting apparatus (model 6000; E & K Scientific Products, Saratoga, CA). Blots were washed extensively in double-distilled H_2O, and proteins were detected in 0.1% amido black (Sigma), 45% methanol, and 2.5% acetic acid. Horizontal strips corresponding to gp33, gp40, and gp65 were excised and washed extensively in double-distilled H_2O.

The following procedures were performed by W.S. Lane and personnel at the Harvard Microchemistry Facility for HPLC, mass spectrometry, and peptide sequencing (Harvard University, Cambridge, MA). Strips containing proteins were subjected to in situ digestion with trypsin (Fernandez et al., 1994). The resulting peptide mixture was separated by reverse-phase column on a diode array detector (model 1090 HPLC/1040; Hewlett-Packard, Palo Alto, CA). Optimum fractions from the chromatogram were chosen based on differential UV absorbance at 205, 277, 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 on a Finnigan Lasermat 2000 (Hemel, UK). Selected fractions were subjected to automated Edman degradation on an Applied Biosystems 494A, 477A (Foster City, CA), or Hewlett-Packard G1005A. Details of strategies for the selection of peptide fractions and their microsequencing have been described previously (Lane et al., 1991).

Cloning and Sequencing of gp40

Unless stated otherwise, all general molecular biology methods were done as given in Sambrook et al. (1989). A degenerate oligonucleotide primer (5′-**CCATCGAT**AAYACNCCNTAYGAYTT-3′; with N representing A, C, G, and T and Y representing C and T), designated pt64s and based in part on the gp40-PT64 peptide sequence (NTPYDF), was synthesized containing a ClaI restriction site (shown in boldface) at the 5′ end. The pt64s primer 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, CA) made from poly(A)$^+$ RNA from 4-hr elicitor-treated tobacco cells (Facchini and Chappell, 1992). Thermocycling was performed at 94°C for 1 min, 45°C for 1 min, and 72°C for 2 min for 30 cycles using 150 ng of cDNA template, 50 nM of pt64s, and 5 nM of T7. One milliliter 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 and Schuell), digested with ClaI and XbaI, and ligated into pBluescript SK− (Stratagene, La Jolla, CA). Sequencing was performed according to the manufacturer's specifications (Promega, Madison, WI). Three of five independent clones, designated p64s/8, p64s/9, and p64s/h, that were identified contained an ~600-bp fragment and encoded the entire peptide gp40-PT64.

cDNA Library Screening and 5′ Rapid Amplification of cDNA Ends

A ClaI-XbaI 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 × SSC (1 × SSC is 0.15 M NaCl and 0.015 M sodium citrate) and 0.5% SDS at 65°C. Eight independent clones were isolated and partially sequenced, and sequence information for 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 (5′ RACE) was performed. Two nondegenerate oligonucleotide primers, designated G1a (5′-GGTGATGTGTGATCCAA-3′) and G2a (5′-**CCATCGAT**CACCAATGTTCCAGTAA-3′; containing a ClaI site at the 5′ end shown in boldface), were synthesized based on the nucleotide sequence of pGP40-5e encoding the amino acid sequence 220-FGSHIT and 200-YWNIG, respectively (Figure 3C).

5′ RACE was performed according to the manufacturer's specifications (Gibco BRL). First-strand cDNA was synthesized from total RNA isolated from 3-day-old tobacco suspension-cultured cells (Puissant and Houdeline, 1990) by using G1a (400 nM) and SuperScriptTMII (Gibco BRL). PCR amplification was performed using 5 mL of the poly(dC)-tailed cDNA, 200 mM deoxynucleotide triphosphates, 2 units of PwoI DNA polymerase (Boehringer Mannheim), 0.5 units of Taq DNA polymerase (Gibco BRL), 400 nM nested primer G2a, and 400 nM Abridged Anchor Primer (supplied by Gibco BRL). Thermocycling was done at 94°C for 30 sec, 43°C for 30 sec, and 72°C for 2 min for 35 cycles, followed by 72°C for 7 min. The products were ligated directly into the SmaI site of pBluescript SK−. Six independent clones, designated pRACE-3 to pRACE-8, were selected for sequencing.

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 900-bp XbaI and HindIII fragment isolated from pGP40-5e and ligated into pBluescript SK− was designated p5e/SK−. p5e/SK− and pRACE-7 were digested with XmnI, yielding 1.0- and 2.8-kb fragments for p5e/SK− and 2.2- and 1.5-kb fragments for pRACE-7. The 2.8-kb fragment of p5e/SK− and the 1.5-kb fragment of pRACE-7 were isolated and ligated to give pGP40-7/5e containing the complete open reading frame of gp40.

Correct ligation of the two fragments was confirmed by sequencing across the XmnI site of pGP40-7/5e. This fusion was expected to give a 1.25-kb RNA transcript, which is consistent with the ~1.3-kb RNA transcript observed in RNA gel blot analysis (data not shown). Furthermore, both DNA strands of the combined insert were sequenced in the W.M. Keck facility (Yale University, New Haven, CT) by using Taq FS DNA polymerase and fluorescent dideoxy terminators in a cycle sequencing method. The resulting DNA fragments were electrophoresed and analyzed using an automated Applied Biosystems 373A Stretch DNA sequencer.

PCR-based mutagenesis was used to introduce an XhoI site in frame upstream of the ATG (Met) of the open reading frame of gp40-7/5e. Thermocycling was performed using 10 ng of template plasmid DNA (pGP40-7/5e), 400 nM deoxynucleotide triphosphates, 400 nM reverse primer (Boehringer Mannheim), 400 nM 7pcrMs primer (5′-ATACA**CTCGAG**TTGACTATG-3′; with the XhoI site shown in boldface and the mismatched nucleotide in italics), and 2.5 units of PwoI DNA polymerase (Boehringer Mannheim) at 94°C for 45 sec, 45°C for 45 sec, and 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 Biotechnology, Piscataway, NJ). In-frame fusion of the gp40 open reading frame to glutathione *S*-transferase (GST) was confirmed by sequence analysis. One clone, designated pGEX-D, was selected for overexpression of the GST–gp40 fusion protein in *Escherichia coli.*

Production of Antibody against the GST–gp40 Fusion Protein

E. coli strain BL21 (DE3) was transformed with pGEX-D (GST-gp40) or pGEX-5X-1 (GST). Transformed cells were grown in 500 mL of Luria-Bertani media to an OD_{600} of 0.5 to 0.7 at 37°C; proteins were induced and cells were lysed 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, and 1% Triton X-100 and three times in double-distilled H_2O. Each wash step was followed by a centrifugation step at 15,000*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 on 10% SDS–polyacrylamide gels, stained in 0.05% Coomassie blue in double-distilled H_2O, and washed in double-distilled H_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; Bio-Rad), using 25 mM Tris base, 192 mM glycine, and 0.1% SDS at 10 mA per glass tube for 8 hr with one change of buffer.

Solutions containing the eluted fusion protein were combined, lyophilized, and resuspended in <1 mL of double-distilled H_2O. Protein concentration was approximated by comparison to a known amount of broad molecular weight standards (Bio-Rad). Fusion protein (500 to 750 mg) 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 PVDF membranes (Harlow and Lane, 1988).

Immunofluorescence Microscopy

Protoplasts were isolated from tobacco cell suspension-cultured cells for 30 to 40 min, collected, spun onto lysine-coated slides (Shandon Lipshaw, Pittsburgh, PA), and fixed in 3% paraformaldehyde (Smith et al., 1997). Approximately 2.5×10^6 protoplasts were used per slide to obtain relatively even distribution. For indirect immunofluorescent microscopy, labeling was performed according to Smith et al. (1997), with the following modifications. The material was blocked in PBS containing 0.05% Tween 20 and 3% BSA for 30 min before incubation with the primary antibody. Incubation of primary (anti-gp40, 1:50 dilution; anti-AtIMPα, 1:12 dilution) and secondary antibodies (goat anti–rabbit IgG-CY3, 1:20 dilution; Molecular Probes, Inc., Eugene, OR) were done in the presence of the same blocking solution. After incubation with the secondary antibody, protoplasts were incubated with a 4′,6-diamidine-2-phenylindole/glycerol solution (Hicks et al., 1996) and then washed extensively.

Samples were examined with epifluorescence optics (Axiophot; Carl Zeiss, Thornwood, NY) by using a 490-nm longpass filter for immunofluorescent labeling. Pictures were taken with p3200 TMAX film (Kodak) and were processed commercially. 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.

ACKNOWLEDGMENTS

We thank the members of the Raikhel laboratory, Drs. Glenn Hicks and Scott Peck, for many stimulating discussions, Harley Smith and Dr. S. Owens for help with the immunofluorescence microscopy, and Drs. Anton Sanderfoot and Diane Bassham for critical reading of the manuscript. We are grateful to Dr. Joseph Chappell for the tobacco pcDNAII cDNA library, Dr. Y. Toyoda for the anti–rat aldose-1-epimerase antibodies, and H. Smith for the anti–AtIMPα antibodies. We also acknowledge Kurt Stepnitz for photography and Marlene Cameron for art work. This work was supported by a grant from the U.S. Department of Energy (No. DE-FG02-91ER20021).

Received November 10, 1997; accepted February 5, 1998.

REFERENCES

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

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.

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.

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 S14 polysaccharides with lectins from *Ricinus communis*, *Triticum vulgares*, and *Bandeiraea simplicifolia.* Carbohydr. 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. (1985). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132,** 6–13.

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

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

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, NY: Cold Spring Harbor Laboratory Press).

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., 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,** 4323–4334.

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. Struct. Biol. **107,** 291–308.

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. USA **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 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.

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. Protein Chem. **10,** 151–160.

Lis, H., Joubert, F.J., and 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.

Maher, D.W., Lee, B.A., and 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.

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.

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 α-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 β-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 UDP glucose 4-epimerase. J. Bacteriol. **172,** 4037–4047.

Puissant, C., and 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 Enzymol. **230,** 443–460.

Rout, M.P., and Wente, S.R. (1994). Pores for thought: Nuclear pore complex proteins. Trends Cell Biol. **4,** 357–365.

Sambrook, J., Fritsch, E.F., and 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.

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,** 411–417.

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. USA **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.

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

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