3120–3129 *Nucleic Acids Research, 1999, Vol. 27, No. 15*

Isolation and characterisation of the cDNA encoding a glycosylated accessory protein of pea chloroplast DNA polymerase

Amos Gaikwad, Krishna K. Tewari[1], Dhirendra Kumar, Weiliang Chen and Sunil Kumar Mukherjee*

International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Marg, New Delhi 110 067, India and [1]Department of Molecular Biology and Biochemistry, University of California, Irvine, CA 92717, USA

Received April 7, 1999; Revised May 26, 1999; Accepted June 22, 1999 DDBJ/EMBL/GenBank accession no. Y11824

ABSTRACT

The cDNA encoding p43, a DNA binding protein from pea chloroplasts (ct) that binds to cognate DNA polymerase and stimulates the polymerase activity, has been cloned and characterised. The characteristic sequence motifs of hydroxyproline-rich glycoproteins (HRGP) are present in the cDNA corresponding to the N-terminal domain of the mature p43. The protein was found to be highly O-arabinosylated. Chemically deglycosylated p43 (i.e. p29) retains its binding to both DNA and pea ct-DNA polymerase but fails to stimulate the DNA polymerase activity. The mature p43 is synthesised as a pre-p43 protein containing a 59 amino acid long transit peptide which undergoes stromal cleavage as evidenced from the post-translational *in vitro* import of the precursor protein into the isolated intact pea chloroplasts. Surprisingly, p43 is found only in pea chloroplasts. The unique features present in the cloned cDNA indicate that p43 is a novel member of the HRGP family of proteins. Besides p43, no other DNA-polymerase accessory protein with O-glycosylation has been reported yet.

INTRODUCTION

Replicative DNA polymerases are generally assisted by a group of accessory proteins for processive, faithful and rapid DNA synthesis. The three well characterised accessory proteins of eukaryotic DNA polymerases are the proliferating cell nuclear antigen (PCNA), replication factor C (RFC) and replication protein A (RPA) (1). The PCNA and RP-A analogues of prokaryotic DNA polymerases are also well known (2). Apart from these three types, other proteins with the DNA-polymerase accessory activities have also been reported from various sources. The adenovirus DNA-binding protein, Ad-DBP (3), and the UL42 protein from Herpes Simplex virus (4) are two examples. Organellar DNA polymerases also require accessory proteins but only a few have been identified and characterised (5,6). Very few similar proteins from the plant system are known (7). Since the mechanism of DNA replication, especially chromosomal DNA replication, in higher plants is poorly understood (8,9), DNA polymerase accessory proteins of plants have remained largely unknown.

Biochemical activity of a host of proteins is regulated through the post-translational modifications, the most common being glycosylation (10,11). Although the precise role of glycosylation of accessory proteins of the DNA polymerases as mentioned above remains to be elucidated, glycosylation of many proteins contributing to cell-proliferation, cellular rigidity, etc., has been well documented (12). In plants, there is a superfamily of proteins termed the hydroxyproline-rich glycoproteins (HRGP) which are generally involved in cell wall formation, building resistance towards pathogens and various stages of cellular development (13,14). The HRGPs contain oligoarabinosides and/or heteropolysaccharides as the principal carbohydrates which are O-linked to the hydroxyproline (HyP) and/or serine (and threonine) residues. The degree of arabinosylation at the hydroxyproline residues depends on the extent of clustering of HyP amino acids, which are very often preceded by a serine (S) residue (15,16). HRGPs are found abundantly either in extracellular matrices, cell walls or as integral membrane proteins (12,17), but their intracellular localisation, if any, is yet to be demonstrated.

In our continuing efforts to understand the molecular mechanism of pea chloroplast DNA replication (18–20), we have reported earlier the identification of a 43 kDa DNA polymerase accessory activity [p43, a DNA binding protein from pea chloroplasts (ct)] (6). To gain insight into its functional significance, we have isolated and characterised the cDNA encoding this accessory protein. Analysis of the cDNA and partial amino acid sequencing of the p43 protein revealed that the N-terminal region of p43 is very rich in contiguous HyP residues. Thus, p43 was predicted to be O-glycosylated (especially O-arabinosylated) to a great extent, a fact which was confirmed by biochemical and physical evidence. p43 can be classified as a new member of the HRGP family because of its interesting biochemical and physiological features. Chemically deglycosylated p43 (i.e. p29) bound to DNA and the cognate DNA polymerase, but failed to activate the pea ct-DNA polymerase. Since p43 is targeted to pea chloroplasts, evidence is provided,

*To whom correspondence should be addressed. Tel: +91 11 6181242; Fax: +91 11 6162316; Email: sunilm@icgeb.res.in

for the first time, that a member of the HRGP family could also be located intracellularly and its glycosylation might be important for the DNA-polymerase accessory activity.

MATERIALS AND METHODS

Screening of cDNA library

A cDNA library was constructed using 3 μg of poly(A)$^+$ RNA of leaf tissues of pea seedlings at the *Eco*RI/*Xho*I cloning sites of the λUni-ZAP XR vector and employing the cDNA synthesis kit and protocol of the manufacturer (Stratagene). A total of 1.2×10^6 plaques were immunoscreened using anti-p43 antibodies that were adsorbed with bacterial antigens (21). Very few (13 only) positive clones were identified by the immunotechnique. Of these only two clones were confirmed by hybridisation with the 5'-labelled degenerate oligonucleotide probes derived from microsequencing of p43 (Table 1). These cDNAs were sequenced. Since the latter clones contained only 0.9 kb insert (Fig. 1e, clone C1) which was not sufficient to represent the entire cDNA, the phage library was rescreened using the 0.9 kb excised cDNA. The positive plaques were identified and purified. Subsequently, the phagemid clones were prepared following the excision methods and those containing the 1.3 kb insert DNA were chosen for DNA sequencing.

Table 1. Nucleotide sequence of oligonucleotide primers

Primers	Sequence (5'-3') (coordinate number in cDNA)
LG1	TAYTTYGARGARACIYTIAAYGTITAYGAY (680) (709)
S1	CATATTATTYTBGATATTYTBGARAAR (713) (739)
L1	ATTGATAAYYTBGATTAYYTBTAYGAR (836) (862)
DK20-N3	CGCGGATCCATGACATGGCCACCAACATACTGC (392) (412)
DK20-C1	CGCGGATCCTTATTAGAAATACACTTC (958) (941)
DK20-R5	CTTCTCTTACTTTCCTCTAGAGG (216) (194)

I, inosine; Y = C + T; R = A + G; B = G + T + C. The underlined region is the added restriction site.

N-terminal and internal amino acid sequencing

An 18 amino acid long N-terminal sequence and sequences of five other trypsin fragments (Fig. 1a) of p43 were obtained using the ABI model 492A precise sequence analyser machine. The sequence information was made available by the Worcester Foundation for Biomedical Research (Worcester, MA). Mass spectral analysis (MALDI-TOF) of these peptides was carried out to confirm the identity. Selected portions of three of these peptides namely YFEETLNVYD, HIILDILEK and IDNLDYLYL, were used to design the oligonucleotide probes LG1, S1 and L1, respectively, as shown in Table 1.

Lectin blots

Assays with the lectin blots were carried out using the DIG Glycan differentiation kit and application protocols of the manufacturer (Boehringer Mannheim Biochemical). Positive control glycoproteins, supplied with the kit, were tested first to examine the efficacy of the probing lectins. Deglycosylated p43 (p29) and lysozyme were used as negative controls.

Deglycosylation of p43 by trifluoromethanesulfonic (TFMS) acid

About 100 μg of lyophilised and Tris-free p43 along with 2 mg of lysozyme (Sigma) was incubated with a mixture of 330 μl of anisole and 670 μl of TFMS at 4°C for 2 h. Deglycosylation was carried out as mentioned (22). Post-dialysis buffer exchange of deglycosylated p43 was finally carried out by filtration using the centricon system (Amicon, 3 kDa cut off). To remove excess lysozyme from this preparation, four rapid cycles of dialysis with higher pore size membranes were performed. The best preparation i.e. deglycosylated p43(-L) contained lysozyme and deglycosylated p43 in the molar ratio of 1:2.

Monosaccharide determination

About 30 μg of p43 along with inositol as an internal standard and monosaccharide standards was subjected to methanolysis (1 M methanolic HCl, 16 h, 80°C). Released monosaccharides were derivatised with pyridine/chloromethylsilane/*N*,*O*-bis (trimethylsilyl) acetamides (3:2:5, v/v). Following trimethylsilyation, the TMS-O-methylglycosides were analysed by gas chromatography and mass spectroscopy as described by Moody *et al.* (23).

In vitro translation of p43 specific cDNA

The F1 or F2 cDNA (Fig. 1a) was cloned as an *Eco*RI–*Xho*I fragment in the mammalian shuttle vector pSGI (modified, Stratagene) and the recombinant plasmid was used as a template. *In vitro* translation was carried out with either the TNT coupled rabbit reticulocyte lysate (RRL) or the TNT coupled wheat germ extract (WGL) systems (Promega), and ^{35}S-cysteine (specific activity > 1000 Ci/mmol, NEN-Dupont) following the manufacturer's protocols (Promega). Immunoprecipitation of the *in vitro* translated proteins was, if necessary, carried out in RIPA buffer employing standard protocols (21).

Transport of the *in vitro* synthesised protein into intact chloroplasts

Intact chloroplasts from 7–8 day old pea leaves were isolated as described (24). Import reactions were performed following the published procedure (25). The reaction mixture was incubated in the presence of sunlight (2 mmol photons/m^2/s) at 22°C for various intervals of time. Post-import intact chloroplasts were re-isolated (26) and the ct-proteins were analysed by SDS–PAGE and autoradiography.

PCR amplification

PCR amplifications were performed in a Perkin-Elmer DNA Thermal Cycler 480 (27,28). Many of the primers carried restriction sites at their 5' ends. Either 100 ng of total DNA from plasmid-library or 10 ng of specific DNA was used as a template.

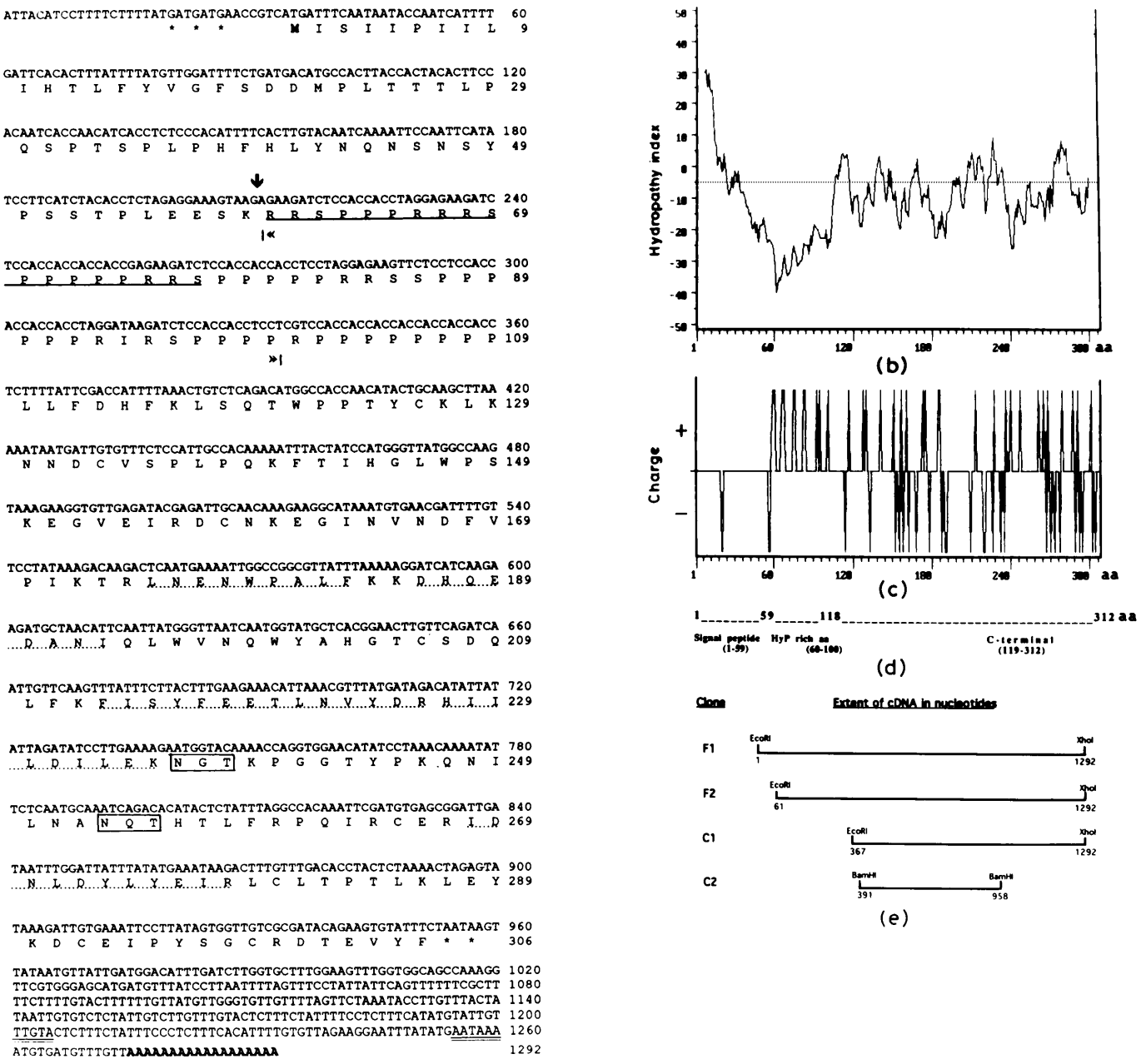

Figure 1. Primary sequence and organisation of F1 cDNA. (**a**) The sequence shown begins at or very near the 5' end of the mRNA. The in-frame stop codons both at the 5' and 3' ends of the ORF are shown by asterisks. The putative initiation methionine is shown in bold letters. The cleavage site of the transit peptide is shown by the vertical bold arrow pointing downwards. The putative O-glycosylation region is shown as two pairs of arrows that are associated with the vertical bars. The prolines of this region are perhaps post-translationally modified to hydroxyprolines. The sequences of N-terminal and internal peptides of p43 as obtained experimentally are shown by bold and dotted underlines, respectively. The putative N-glycosylation sites are boxed. The probable poly(A) processing sites are shown by double-underlines. (**b**) Hydropathy plot of the deduced amino acid sequence of p43. The hydropathy values of each amino acid were determined using an interval of 9 amino acids. Values above and below the dashed line indicate hydrophobic and hydrophilic regions respectively. (**c**) Plot showing charge distribution across the entire ORF. (**d**) Various modules of p43. (**e**) Terminal restriction sites and the nucleotide coordinates of the important clones used.

Northern blots

Total cellular RNA was extracted from the leaves of various plants and tissues of the pea plant. About 25 µg of total RNA was electrophoresed in a 1% agarose/formamide gel at 1.5 V/cm for 17 h. The fractionated RNA was then blotted onto Nylon membrane (Gene Screen). For hybridisation reaction, the blotted membrane was incubated with 2×10^7 c.p.m. of ^{32}P-labelled F1 cDNA (specific activity ~10^8 c.p.m./µg) at 65°C for 16 h. The

membranes were washed and exposed for autoradiography. For reuse of the blot, the bound radiolabel was stripped by incubating the blot in 0.1× SSC and 0.5% SDS at 95°C for 30 min.

Primer extension

The procedure was carried out essentially as described by Maniatis *et al.* (21). RNA was prepared independently from three sources: (i) pea leaves for obtaining total cellular RNA; (ii) *in vitro* system using the recombinant plasmid pSG1-F1 as the template and the TNT coupled RRL system (Promega) in the presence of 20 μg/ml cycloheximide to block translation; (iii) defined *in vitro* system using T7 RNA polymerase (USB) along with the pSG1-F1 template.

An appropriate amount of RNA sample from each source was annealed with ~5 ng (5×10^5 c.p.m.) of 5' labelled DK20R5 primer. Following annealing, extension of the primer was carried out using MMLV reverse transcriptase (USB) for 2 h at 37°C. The extension products were analysed in a 6% polyacrylamide urea sequencing gel and autoradiographed. For the control reaction, the linearised plasmid pSG-F1 was used as the template.

Other methods

DNA-polymerase assays, co-immunoprecipitation of DNA polymerase along with p43 using anti-p43 antibodies, South-western blotting, etc. were carried out as described earlier (6). Other methods including gel electrophoresis, 5', 3' and uniform labelling of DNA, Southern blotting, western blotting, DNA sequencing, etc., were carried out according to the published procedures (21).

RESULTS

Isolation of cDNA clones for p43

The proteins required for the replication of pea ct-DNA are possibly nuclear encoded (29). Hence the pea cDNA library was immunoscreened and 13 clones were selected. Of these, only two hybridised with the degenerate oligonucleotide probes, namely, LG1, S1 and L1 (Table 1). The cDNA inserts of these clones were used to rescreen the entire phage library. About 30 positive clones of various insert sizes were obtained on re-screening ~10 000 plaques. Only four clones harbouring ~1.3 kb cDNA inserts were chosen for DNA sequencing. Three of the inserts were identical (clone F1, Fig. 1e) in sequence and the fourth one (F2) lacked only ~60 bases from the 5' end of F1. The DNA sequence of the clone F1 is shown in Figure 1a.

Features of F1 cDNA

The 5' and 3' untranslated regions along with an open reading frame (ORF) are shown in Figure 1a. There are three tandem in-frame stop codons upstream of the ORF coupled with one at the 3' end, signifying the completeness of the ORF within the cloned F1 DNA. Only two methionine (met) residues are present throughout the entire ORF at the amino acid positions 1 and 22, respectively. Assuming that the most favourable context of initiation codon recognition can be defined as CC A/G CC (ATG) G A/C T (30), the methionine at position 1 has better fit than the methionine at position 22 as the initiation codon. Moreover, (as shown later) the clone F2, lacking the first methionine but retaining the second one, failed to generate any specific *in vitro* translatable product. On the other hand, a protein of the expected size was synthesised using the F1 clone under similar conditions (Fig. 4). These observations point to the methionine at amino acid position 1 as the start codon.

The ORF begins with a transit peptide of 59 amino acids which probably determines the chloroplast localisation of p43. The stromal cleavage site is shown by a bold vertical arrow (Fig. 1a). Since the N-terminal amino acid sequence of p43 was obtained (bold underlining in Fig. 1a), it was easy to assign the cleavage point. The transit peptide is rich in serine (15%) and threonine (10%) residues but poor in acidic residue content (aspartic acid 3%, glutamic acid 3%). There is also a consensus serine residue located 8 amino acids upstream of the cleavage point at lysine. Seventy-three percent of the residues of this transit peptide assume the random coil conformation according to the secondary structure prediction by the GGBSM method (31). All these features agree well with those found in other chloroplast transit peptides (32,33). Later, we have also shown the cell-free translated product of the clone F1 translocated in the purified intact pea chloroplasts (Fig. 4) using the transit peptide mechanism.

Judging from the nucleotide sequences alone, the expected size of the mature protein would be ~29 kDa (amino acids 60–306) instead of 43 kDa (6). This discrepancy was accounted for by the glycoprotein character of p43 (Fig. 2). N-terminal amino acid sequencing of p43 (amino acids 60–78 of ORF) revealed that all the prolines (P) of the underlined region of Figure 1a are indeed hydroxyprolines (HyP). Amino acid composition analysis also showed (data not shown) that each molecule of p43 contained about 25 and 16 residues of HyP and proline, respectively. Moreover, considering that the motif S-HyP_{3-6} is a major target for O-arabinosylation (15,16), it appears that all the contiguous prolines following a serine residue (amino acids 60–100 in Fig. 1a) are post-translationally modified to hydroxyprolines. Consequently, this N-terminal region of p43 could be highly arabinosylated. Besides O-arabinosylation, high O-glycosylation is also predicted (34) at each of the five serine residues of the same region (amino acids 69–96). There are also two potential N-glycosylation sites at amino acid positions 236–238 and 253–255, respectively, as indicated by boxes in Figure 1a. The presence of most of these carbohydrates in mature p43 protein had also been experimentally demonstrated (Fig. 2, Tables 2 and 3). Deglycosylation studies with p43 revealed that the mass of protein moiety was only 29 kDa.

The secondary structure analysis revealed that the protein p43 is highly hydrophilic (Fig. 1b) and can assume a rod like extended conformation. It is predicted to contain very few α-helices and does not appear to have any transmembrane domain. Many HRGPs contain the motifs of crosslinking that are essential for the cell wall proteins (35,36). But p43 is devoid of such motifs. This agrees well with the fact that p43 is a stromally located protein (6). The predicted amino acid sequence of p43 contains eight cysteines and six histidines but no Zn-finger motif. Although many potential phosphorylation and N-myristoylation sites are present within p43, the biochemical implications of these have not been studied here.

Further downstream of the termination codon of the ORF, a distinct poly(A) processing signal, namely AATAAA (nt 1255–1260) is present. In plants, unlike in animal systems, a single poly(A) signal may not be sufficient and often a multitude of signals act together for formation of the 3' end of

Table 2. Moiety of sugars identified from the lectin blots

Lanes in Fig. 2A	Probing lectins	Sugars recognised	Abundance of sugars
1	*Datura stramonium* agglutinin (DSA)	Galactose β(1–4) *N*-acetyl galactosamine	+
2	*Maackia amurensis* agglutinin (MAA)	Sialic acid linked α(2–3) to galactose	+++
3	Peanut agglutinin (PNA)	Core disaccharide galactose β(1–3) *N*-acetyl galactosamine	+++
4	*Galanthus nivalis* agglutinin (GNA)	Terminal mannose α(1–3), α(1–6) or α(1–2) linked to mannose	+++
5	*Sambucus nigra* agglutinin (SNA)	Sialic acid linked α(2–6) to galactose	+

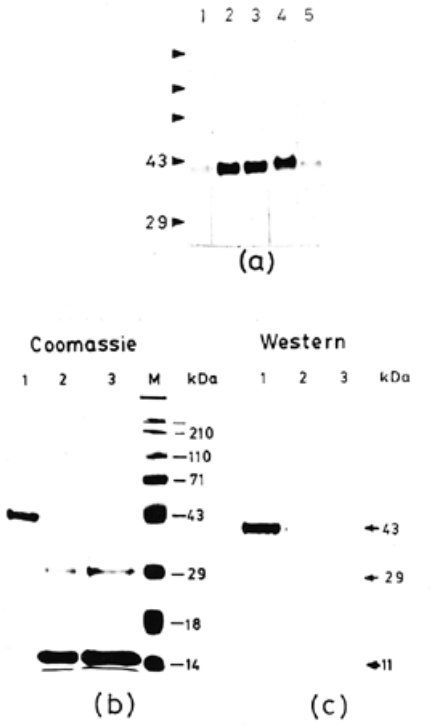

Figure 2. p43 is glycosylated. (**a**) Identification of various sugars by lectin blot: 2 µg of p43 were loaded in each lane and blotted on nitrocellulose membrane following electrophoresis. Sugar of each blot was identified by probing with specific lectins as mentioned in Table 2. (**b**) Molecular weight of the deglycosylated p43: 2 µg of p43 (lane 1) and 6 and 12 µg of deglycosylated preparation of p43 (lanes 2 and 3, respectively) were analysed by SDS–12% PAGE. 6 and 12 µg of deglycosylated preparation of p43 contained 400 and 800 ng of deglycosylated p43 (or p29), respectively. The molecular weight markers are indicated on the right. (**c**) Western blotting of the glycosylated and deglycosylated proteins. 500 ng of p43 in lane 1, 36 and 18 µg of deglycosylated preparation of p43 in lanes 2 and 3, respectively, were probed immunologically by anti-p43 antiserum. Dilution of the antisera used was 1:2000. Lysozyme did not crossreact with anti-p43 antisera.

the mRNAs. In the F1 clone, at least two such additional signals namely TTGTA (nt 1201–1205) and TTGTG (nt 1233–1237) are present. A 30 bp GT-rich region (nt 1092–1122) can also serve as another putative polyadenylation signal.

Glycosylation of p43

Theoretical analysis revealed that p43 would be O-arabinosylated at the HyP residues (37–39). For determination of sugars other than arabinose, a commercially available DIG Glycan differentiation kit (BMB Biochemical) was used, employing various lectins to identify the specific carbohydrate moieties. Figure 2a and Table 2 show that galactose and mannose were present in p43. The presence of others namely sialic acid, etc. was ruled out by the solvolysis technique as mentioned below.

Table 3. Determination of monosaccharides by solvolysis of p43

Monosaccharides	%
Arabinose	87
Galactose	4
Xylose	2
Mannose	1
Glucose	8
Sialic acid	–
Hexosamine	–
Uronic acid	–

Other monosaccharides were detected by solvolysis of p43 in methanolic HCl as mentioned in Materials and Methods. The gas chromatographic and mass spectral results revealed that p43 is composed of arabinose (87%), galactose, xylose, mannose and glucose but no sialic acid or uronic acid (GlcA/GalA) as shown in Table 3. Since glucose is a very common contaminant in this kind of sugar analysis and was not detected by lectin-blot investigation, the presence of glucose in p43 cannot be taken seriously. It is thus apparent that p43 is glycosylated and arabinose constitutes the major component of glycosylation.

In order to study the extent of glycosylation of p43, chemical deglycosylation studies were undertaken. The protein was deglycosylated by TFMS acid under mild conditions keeping lysozyme in the reaction mixture as an internal control and protective reagent as described in Materials and Methods. Figure 2b shows that the SDS-molecular weight of the maximally deglycosylated species of p43 was only 29 kDa, consistent with the prediction from the ORF. Western blot analysis (Fig. 2c) with the deglycosylated protein revealed that no polypeptide fragment smaller than 29 kDa was detectable, thereby suggesting that the deglycosylation treatment did not result in random hydrolysis of the peptide backbone (22). However, deglycosylation affected antibody recognition since a higher amount of the deglycosylated protein was required to

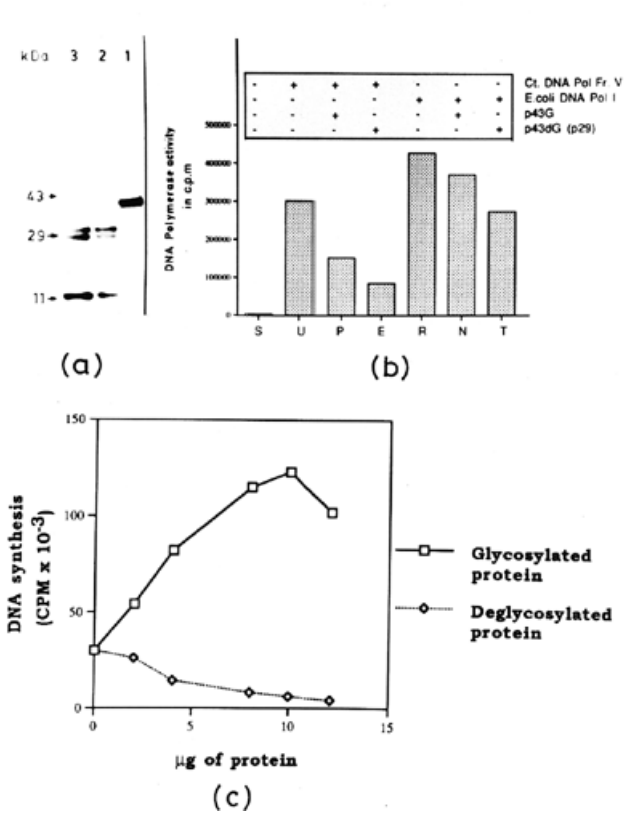

Figure 3. Biochemical properties of deglycosylated p43. (**a**) Autoradiogram of South-western blot containing p43 and its deglycosylated form. 1 µg of p43 in lane 1 and 2.5 and 5 µg of deglycosylated preparation of p43 in lanes 2 and 3, respectively, were examined for DNA binding activity by South-western technique. Deglycosylated preparation of 2.5 and 5 µg contained ~150 and 300 ng of deglycosylated protein (p29) respectively. (**b**) Co-immunoprecipitation of DNA-polymerase and p43 using anti-p43 antibodies. About 1.5 U fraction V ct-DNA polymerase (6), 0.5 U *E.coli* Pol 1 (USB), 7 µg p43 and 8 µg deglycosylated (-L) p43 (or p29) were used whenever necessary along with the 50 µg rabbit anti-p43 antibodies. DNA synthesis was quantitated by measuring TCA insoluble [^{3}H]TMP incorporation with activated calf thymus DNA as template (6). G and dG stand for glycosylated and deglycosylated forms of p43, respectively. (**c**) DNA synthesis in the presence of glycosylated and deglycosylated proteins. About 0.2 U of fr. V ct-DNA polymerase was used in the standard assay for DNA synthesis (6). Deglycosylated (-L) p43 and native p43 were used as the source for deglycosylated and glycosylated proteins.

generate a comparable signal on the western blot (compare lane 2 of Fig. 2c with lane 3 of Fig. 2b).

Biochemical properties of deglycosylated p43 (p29)

The native p43 possesses three distinct biochemical functions, namely, non-specific DNA binding, cognate DNA polymerase binding and the stimulation of the cognate DNA polymerase activity (6). Since the mild treatment with TFMS acid is known to retain the functional properties of the treated protein (40), we examined whether the deglycosylated p43 protein also maintained the native characteristics.

Results of the South-western blot (Fig. 3a) showed that the deglycosylated protein retained the DNA-binding ability. The signal observed at 14 kDa was due to the DNA-binding activity of lysozyme that was present during the deglycosylation treatment. Excess lysozyme was absent in the deglycosylated p43(-L) preparation.

Deglycosylated p43(-L) preparation was used to evaluate its DNA polymerase binding and activation properties. The DNA polymerase binding activity was assayed by co-immunoprecipitation as described earlier (6). In a reaction mixture containing the DNA polymerase, p43 and anti-p43 antibodies, the polymerase was co-immunoprecipitated with p43 using ProteinG-Sepharose beads. The polymerase-depleted supernatant was then assayed for DNA polymerase activity. The depletion of polymerase activity was reflective of the binding of p43 to the DNA polymerase (6). It is evident from Figure 3b that deglycosylated p43 (i.e. p29) could bind to the ct-DNA polymerase as efficiently as the native p43, if not better. Binding of deglycosylated p43 was also observed with *Escherichia coli* DNA Pol 1. Binding to the DNA polymerase remained unaffected even in the presence of excess amount (40 µg) of carrier proteins like BSA or lysozyme. Thus glycosylation of p43 is not required for its binding to DNA or the cognate DNA polymerase.

To investigate the effect of deglycosylation on the activation of ct-DNA polymerase, DNA synthesis was performed (6) in the presence of either native p43 or deglycosylated p43 using activated calf thymus DNA as a template. Figure 3c clearly shows that the deglycosylated protein inhibited the ct-DNA polymerase instead of activating it. A similar inhibition by deglycosylated p43 was also observed when *E.coli* DNA Pol 1 was used for DNA-synthesis, whereas the optimal activation of *E.coli* Pol 1 by native p43 was only 1.5-fold under similar reaction conditions (data not shown). The presence of carrier proteins like BSA or lysozyme did not significantly affect the DNA synthesis. Assuming that the chemical treatment for deglycosylation had left the peptide backbone and the amino acid residues of the activation domain of p43 more or less unchanged, glycosylation of p43 might have a role in stimulating the ct-DNA polymerase activity.

Role of the transit peptide in import

The functionality of the 59 amino acid long transit peptide (41) was studied in a post-translational *in vitro* transport assay of the precursor protein into intact pea chloroplasts. The ^{35}S-labelled precursor proteins were synthesised *in vitro* and were resolved by SDS–PAGE (lanes 1 and 7 of Fig. 4). *In vitro* synthesised proteins were also immunoprecipitated using antibodies to p43 to identify the specific precursor protein of interest (not shown). The major band in each lane from 1 to 6 represents the *in vitro* translated pre-p43. Since the molecular mass of the pre-p43 was only ~36 kDa (instead of 50 kDa, which should have been the expected size of glycosylated pre-p43), it appears that the *in vitro* synthesised protein was not significantly glycosylated. The size of the precursor protein did not change even when the WGL based TNT system was used. Incidentally, no specific protein could be synthesised using the DNA of the isolate F2 (data not shown). A pre-33 kDa chloroplast-localised RNA binding protein was also synthesised (lane 7) similarly and was used for the comparative analysis.

In vitro synthesised proteins were incubated with purified intact chloroplasts for various time intervals and their transport within the chloroplasts was monitored as mentioned (26). Lanes 4 and 5 of Figure 4 show that only a small fraction of pre-p43 was imported within the chloroplasts after an extended

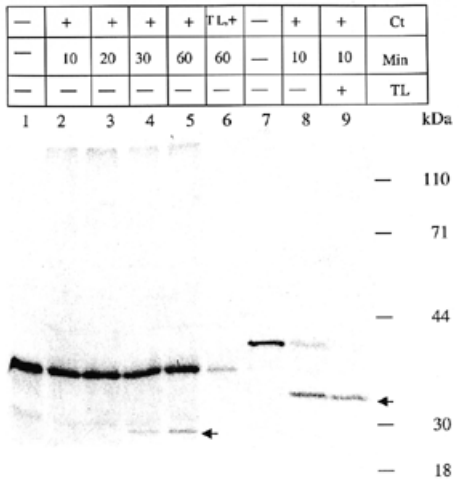

Figure 4. Import of proteins into intact chloroplasts. Translocation of the pre-p43 and pre-33 kDa proteins into intact chloroplasts (equivalent to 50 µg of chlorophyll) was carried out. Lane 1, *in vitro* translated, radiolabelled (4×10^6 c.p.m.) pre-p43 protein. Lanes 2–5, radiolabelled (2×10^7 c.p.m., each lane input) pre-p43 protein that was imported into intact chloroplasts for 10, 20, 30 and 60 min respectively. Lane 6, import of the radiolabelled pre-p43 (2×10^7 c.p.m.) protein into chloroplasts that were pre-treated with thermolysin. Lane 7, *in vitro* translated radiolabelled (0.6×10^6 c.p.m.) pre-33 kDa protein. Lane 8, radiolabelled (3×10^6 c.p.m.) 33 kDa protein that was imported into intact chloroplasts for 10 min. Lane 9, thermolysin treatment (30 min on ice) of post-import chloroplasts used for experiments shown in lane 8. Molecular weight marker positions are shown on the right hand side. Arrows indicate the positions of the imported and thermolysin resistant mature protein. Some of the reaction conditions are indicated on the top. + and ─ indicate presence and absence, respectively.

incubation period and the size of the imported protein was ~30 kDa. When immunoprecipitation of the imported protein was carried out, the distinct presence of a 30 kDa protein was observed but the signal was very weak due to the low efficiency of transport (data not shown). The reduction of the precursor polypeptide size by ~6 kDa is consistent with the proposal that a 59 amino acid long transit peptide would be cleaved off by a stromal protease. It is worth noting that a pre-33 kDa RNA binding ct-protein transported in the intact chloroplasts with usual high efficiency under similar transport conditions (lanes 7–9). Chloroplasts pre-treated with thermolysin were also used for transport studies. No binding of pre-p43 to the pre-treated chloroplasts was expected in view of the proteolytic digestion of the chloroplast membrane bound translocation machinery. Instead some residual binding was observed under the present experimental conditions as shown in lane 6. This binding reflects either the incomplete proteolysis of the translocation apparatus (which may be unlikely) or tight binding of pre-p43 to the membrane lipids of the chloroplasts.

Analysis of p43 protein-specific transcripts

In previous studies (6, unpublished observations) with the pea plant, it was demonstrated that the p43 protein is expressed in a species and leaf-tissue specific manner. To correlate the absence of the protein with either the physical absence or non-translatability of the transcripts, a northern blot analysis of total cellular RNAs from different plant species and different tissues of the pea plant was carried out using radiolabelled F1 DNA as the probe.

Figure 5a shows that an ~1.4 kb long transcript was present in the leaves of pea only and not in any other tested monocots or dicots. Similarly, it is also apparent from Figure 5b that the p43 specific 1.4 kb transcript could be found only in the leaves and not in other tissues of pea.

Figure 5b also shows that the mobility of denatured cDNA was slower than that of the p43-specific mRNA. Assuming no breakage of 5' end of mRNA during isolation, the faster mobility could be attributed to the higher electronegativity and more structured conformation of RNA than that of single-stranded cDNA. The ratio of intensity of the bands present in lanes of pea L and ss cDNA in Figure 5b was 1:3. Taking into account the differences of autoradiographic exposure time for the above two lanes and assuming that the nuclear mRNA content is only ~4% of total cellular RNA (42), the amount of 1.4 kb long transcript would be ~1.2% (by weight) of nuclear messages. Hence, p43 specific transcripts are highly abundant inside the leaf-tissues.

We also wanted to isolate the clone with the largest 5' untranslated region from the pBluescript SK(–) based phagemid library. A 5' RACE technique was adopted using the upstream vector-specific T3 20mer and downstream gene specific primer, DK20R5 (Table 1). The results of Figure 5c show that the maximum size of the amplifiable products was only 320 bp with two independently isolated library preparations. An amplified fragment of the same size was also obtained using the template DK20 (i.e. clone F1 in pBluescript vector). Lane 2 of Figure 5c shows that the PCR amplification was specific for the pair of primers since no amplification was possible with any single primer (namely DK20R5). Thus, it appears that the maximal 5' untranslated region (i.e. the region upstream of the primer DK20R5) was also present within the clone F1.

A primer extension analysis was used to strengthen the above conclusion. The template RNA, derived from either pea leaves (lane 2 of Fig. 5d) or F1-specific template (lanes 3 and 4), was annealed to the primer DK20R5 and reverse transcribed. The nascent DNA products are shown in Figure 5d. The start points of p43 specific messages transcribed *in vivo* (lane 2) and *in vitro* (either lane 3 or 4) were located at 192 and 235 nt, respectively, upstream of the primer DK20R5. On the other hand, the extended product was also 225 nt long when the pSG-F1 plasmid linearised with *Eco*RI enzyme (lane 1) was used as the source of control DNA. These data thus reconfirm the result shown in Figure 5c, i.e. the F1 clone contained the largest and probably the complete 5' untranslated region.

DISCUSSION

In the context of pea ct-DNA replication, p43 seems to be an important protein because of its biochemical properties. In order to understand its functional role, the encoding cDNA was isolated and characterised. Pea ct-DNA failed to hybridise with either the isolated intact cDNA or the selected portions of it (namely the LG1, L1 and S1 oligonucleotides shown in Table 1), thereby confirming that p43 is encoded only by nuclear DNA.

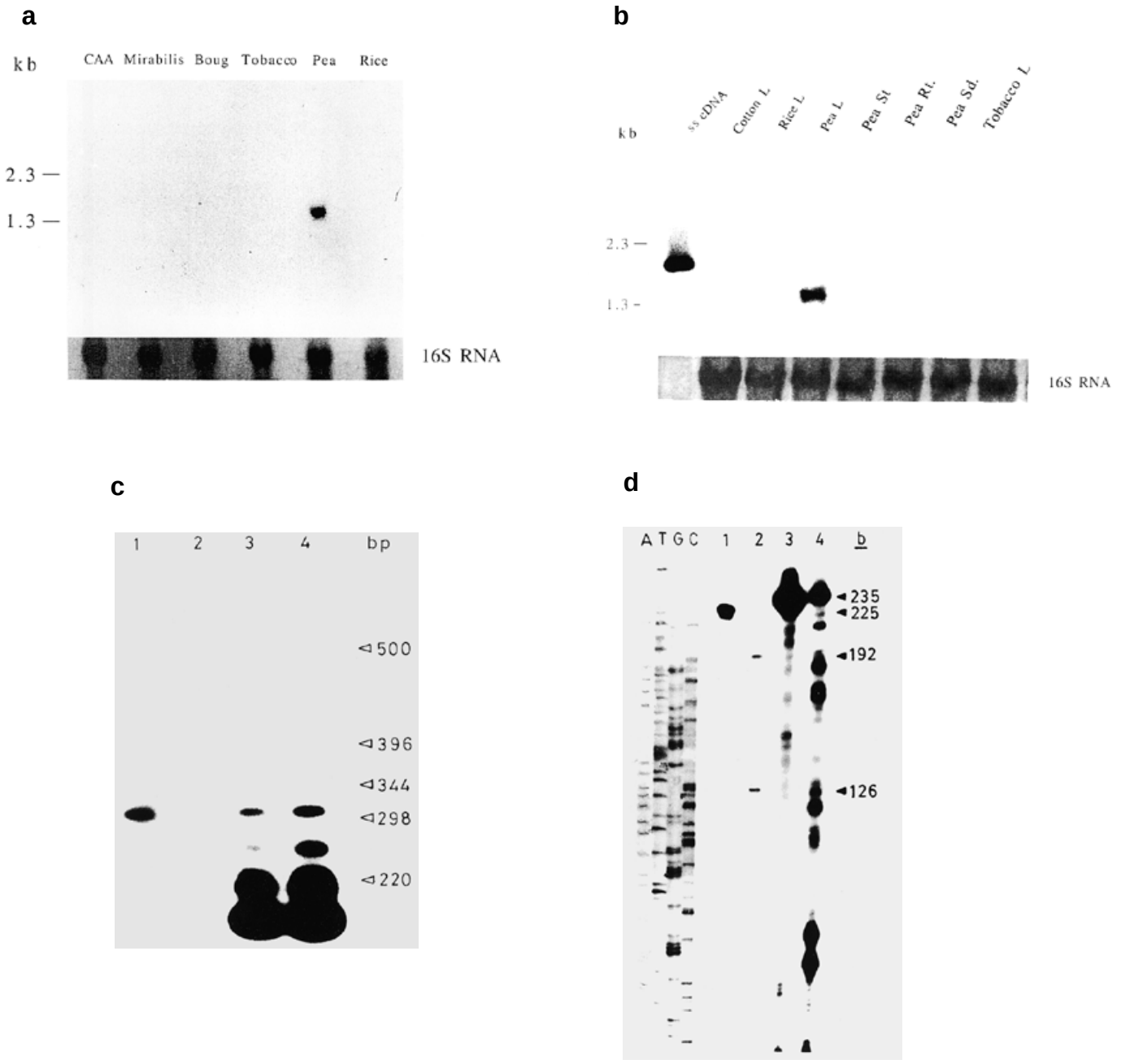

Figure 5. The integrity and completeness of the 5' region of the isolated F1 cDNA. (**a**) Northern blot of transcripts from the leaves of various plants. About 25 µg of total cellular RNA from the leaves of each of the plants like *Clerodendrum aculeatum* (CAA), *Mirabilis*, *Bougainvillea*, tobacco, pea and rice, along with 10 µg of standard RNA markers (Gibco-BRL) were size-separated in 1% agarose-formamide gel. RNA was blotted and hybridised with uniformly labelled F1 cDNA. The same blot was also probed with 16S DNA for loading controls as shown in the lower panel. (**b**) Northern blot of transcripts from various tissues of pea plant. About 25 µg of total RNA from each of the tissues like leaves (L), stem (St), Root (Rt), imbibed seed (Sd) of pea and similar amounts from other sources were electrophoresed followed by blotting, hybridisation and autoradiography. About 200 ng of single-stranded F1 DNA with *Eco*RI/*Xho*I ends (ss cDNA) was also loaded in the same formamide agarose gel. The autoradiographic exposure time for lanes marked as ss cDNA and cotton L was only 2 h. while that of other lanes of the blot was 12 h. The same blot was also hybridised with 16S DNA for loading control as shown in the lower panel. (**c**) 5' RACE. The template for the PCR-amplified product shown in lane 1 was 5 ng of DK20 whereas the same for lanes 3 and 4 were 100 ng of each of two different preparations of phagemid based cDNA library. Lane 2 shows the control PCR reaction where only the F1 cDNA specific primer DK20R5 was present but T3 20mer was omitted. A standard 1 kb DNA ladder was also electrophoresed along with the PCR products to estimate the size of the specific bands. (**d**) Primer extension. 5'-end labelled DK20R5 primer was annealed to various templates and extended by using either *Taq* DNA polymerase (lane 1) or MMLV reverse transcriptase (lanes 2–4). Lane 1 shows the primer extended product with 2 µg of *Eco*RI digested pSG1-F1 plasmid as the template DNA. For lane 2, ~300 µg of total RNA from the pea leaves was employed as the source of template. About 25 µl of cell free transcription products was also used when synthesised with either the TNT coupled RRL system (lane 3) or T7 RNA polymerase (lane 4). For accurate molecular weight determination, primer extension products were electrophoresed along with the dideoxynucleotide terminated (A, T, G or C) sequencing reaction products. Sequencing reaction was carried out using a recombinant template DNA of predetermined nucleotide sequence.

There are many lines of evidences to establish the authenticity of the isolated cDNA F1. First, the approach adopted for isolation itself was stringent enough to explore the correct cDNA. Both the immunoprobes and the oligonucleotide probes, derived from internal amino acid sequences of p43, were employed simultaneously to minimise the presence of any false positive clones. Second, the *in vitro* translated product of the F1 cDNA template immunoprecipitated with anti-p43 antibodies and the immunoprecipitated product was of the right size (36 kDa) as predicted from the ORF in F1 cDNA. In addition, the translated product acted as a precursor protein and translocated *in vitro* through the purified intact chloroplasts using the transit peptide as predicted from the cDNA sequences. The molecular size of the imported protein was 30 kDa, equivalent to the size of the deglycosylated p43 (Fig. 4). Third, the size of p43 specific transcripts found *in vivo* was almost equivalent to the size of the cDNA F1 (Fig. 5). Northern blot data showing tissue-specific abundance of these transcripts also supported the previously reported (6) findings of western blot experiments. Finally, the N-terminal and few internal regions of p43 that were sequenced also matched perfectly well with the predicted amino acid sequences of the corresponding regions from the F1 cDNA (Fig. 1). Taking all the evidence together, it is clear that cDNA F1 is specific for p43.

The p43 specific transcripts were abundant in the leaf-tissue. Hence, during re-screening of the phage library, the frequency of positive clones was reasonably high (30 out of 10 000). In comparison, the frequency of the immunopositive clones was very low (13 out of a million). The low frequency is probably due to two reasons, namely (i) low or no bacterial expression of the F1 cDNA; (ii) poor recognition of bacterially expressed and perhaps non-glycosylated protein by the antibodies raised against native (i.e. glycosylated) p43. In our laboratory it has been demonstrated that only the C-terminal part of F1 cDNA was expressible in bacteria (A.Gaikwad and S.K.Mukherjee, unpublished). So the small population of the cDNA library containing only the expressible C-terminal part of the appropriate cDNA might had been recognised by the immunoprobe. It was also found that the antibodies to p43 reacted with deglycosylated p43 poorly (Fig. 2).

Many unique properties of the F1 cDNA became apparent from our study. The N-terminal part of p43 bears partial homology to other HRGPs because of the repeated presence of S-$Hyp_{n\geq 3}$ as the O-glycosylation motif. However, apart from this, p43 does not show any resemblance to other HRGPs. The cDNA of many HRGPs contains signal sequences for extracellular localisation whereas the F1-cDNA contains the transit peptide for intracellular compartmentalisation. Many of the HRGPs contain motifs for intermolecular crosslinking, self-assembly nucleation or increased molecular rigidity and hydrophobicity (12); some others are arabinogalactans, commonly found in the roots of leguminous plants, possessing the transmembrane helical domains at the C-terminal ends (16). Such features are not present within the protein encoded by the F1 cDNA. From the cDNA analysis, p43 seems more like a hybrid or fusion protein with the HRGP fusion domain at the N-terminal end. The solaneceous lectins also exhibit similar fusion characteristics or structural patterns. Therefore, we compared the properties of these lectins with p43. The solaneceous lectins from potato tuber agglutinated red blood cells but failed to bind DNA, whereas p43 failed to agglutinate red blood cells but bound strongly to DNA (unpublished). Thus p43 does not appear to belong to the family of solaneceous lectins and probably is a unique member of the HRGP family.

Homology search did not reveal any significant homology of p43 with any of the leguminous lectins (43,44). However, a stretch of 8 amino acids namely FTIHGLWP (amino acids 141–148) has been found common to both p43 and stylar glycoproteins located at the self-incompatibility alleles (S) of *Solanum tuberosum* and *Nicotiana alata* (45). These extracellular proteins are also referred to as S-like RNases and function as RNA hydrolytic enzymes to play their roles in the reproductive barrier (46). But, p43 bound to RNA very poorly (6) and did not show RNase activity either in solution or in activity-gels (unpublished). In this view, p43 is also distinct from the S-like RNases with regard to cellular localisation and functional properties.

Though p43 is biochemically distinct from its structural counterparts, its chloroplastic localisation is evident from many viewpoints. First, p43 was derived from intact pea chloroplasts using chromatographic separation techniques. Second, anti-p43 antibodies recognised p43 specifically from the proteins of pea chloroplasts that were treated with thermolysin, a protease known to remove surface protein contaminants of chloroplasts (6). Third, Figure 4 (this paper) shows that pre-p43 could be targeted to the intact chloroplasts with the help of a 59 amino acid long transit peptide. Since the pre-p43 would be glycosylated *in vivo*, chloroplast targeting of p43 would perhaps be easier within the leaf tissues. Fourth, immunoelectron microscopic studies recently carried out in our laboratory aptly pointed out that anti-p29 (deglycosylated p43) antibodies recognised only the chloroplasts and not other sections, namely the nucleus, cytosol, cell wall, etc. of the pea leaf tissues (A.Gaikwad and S.K.Mukherjee, unpublished). Hence p43 ought to be a chloroplast localised protein.

All the above facts put together help conclude that p43 is a novel protein. It is not only a novel variant of HRGPs but also is the first glycoprotein being reported with the cognate DNA-polymerase accessory activity. In plants, proteins analogous to p43 are not yet known.

It has been suggested earlier that the domains of p43 for binding to DNA and ct-DNA polymerase are probably different (6). Since the chemically deglycosylated p43 bound to the pea ct-DNA polymerase but failed to activate the polymerase (Fig. 3), it appears that all the three distinct biochemical functions of p43, i.e., DNA-binding, DNA-polymerase binding and activation of DNA polymerase, probably reside in spatially separate domains. Currently the domain analysis of p43 is being actively pursued.

The inefficient transport of *in vitro* synthesised pre-p43 within the pea chloroplast could be attributed to a variety of factors including the glycosylation of p43. First, within the leaf tissues, the chloroplast localised p43 is glycosylated. The golgi compartment of the dicotyledonous plants contains enzymes that catalyse O-glycosylation of hydroxyprolines (47). Hence the transport of p43 to the chloroplasts could be routed through the golgi compartments within the leaf tissues. But the *in vitro* translated protein obtained from the F1 cDNA template was not glycosylated and the putative signal for the golgi compartment, if any, was also not removed from the precursor unlike what might generally happen *in vivo*. Second, the predicted

amino acid sequences suggest that p43 is mostly hydrophilic in nature. The presence of sugars might help bind p43 better on the outer lipid bilayer membrane of the chloroplast prior to the initiation of the import process. The transit peptide guiding the chloroplast transport process probably could act at a stage subsequent to the step of its binding (48). The other possibilities for inefficient translocation like participation of cytosolic factor(s) or other modifications of p43, etc. cannot also be ruled out. Thus the results of Figures 3 and 4 indicate that glycosylation might at least partially control two important biological functions of p43, namely the activation of DNA polymerase and the transport within chloroplasts.

ACKNOWLEDGEMENTS

We are indebted to Prof. A. Bacic, School of Botany, University of Melbourne, Australia for carrying out the monosugar compositional analysis of p43 (Table 3) in his laboratory. Dr C. Saha provided her help for the lectin blot work. We are thankful to Prof. S. Sopory and Dr Y. Vaishnav for many helpful discussions. Thanks are due to Ms R. Radha and Ms G. Srinivasan for their excellent secretarial assistance.

REFERENCES

1. Hubscher,U., Maga G. and Podust V.N. (1996) In DePamphilis,M.L. (ed.), *DNA Replication in Eukaryotic Cells.* Cold Spring Harbor Laboratory Press, New York, pp. 525–543.
2. Onrust,R., Finkelstein,J., Naktinis,V., Tumer,J., Fang,L. and O'Donnell,M. (1995) *J. Biol. Chem.*, **270**, 13348–13391.
3. Dekker,J., Kanellopoulos,P.N., Loonstra,A.K., Van Oosterhout,J.A.W., Leonard,K., Tucker,P.A. and Uander Vliet,P.C. (1997) *EMBO J.*, **16**, 1455–1463.
4. Hernandez,T.R. and Lehman,I.R. (1990) *J. Biol. Chem.*, **265**, 11227–11232.
5. Williams,A.J. and Kaguni,L.S. (1995) *J. Biol. Chem.*, **270**, 860–865.
6. Weiliang,C., Gaikwad,A., Mukherjee,S.K., Roychoudhury,N.R., Kumar,D. and Tewari,K.K. (1996) *Nucleic Acids Res.*, **24**, 3953–3961.
7. Burton,S.K., Van't Ho,J. and Bryant,J.A. (1997) *Plant J.*, **12**, 357–365
8. Hernandez,P., Martin-Parras,L., Martinez-Robles,M.L. and Schvartzman,J.B. (1993) *EMBO J.*, **12**, 1475–1485.
9. Van't Hof,J. (1996) In DePamphilis,M.L. (ed.), *DNA Replication in Eukaryotic Cells.* Cold Spring Harbor Laboratory Press, New York, pp. 1005–1027.
10. Kobata,A. (1992) *Eur. J. Biochem.*, **209**, 483–501.
11. Lis,H. and Sharon,N. (1993) *Eur. J. Biochem.*, **218**, 1–27.
12. Showalter,A.M. and Varner,J.E. (1989) *The Biochemistry of Plants.* Academic Press, New York, pp. 485–520.
13. Kieliszewski,M.J. and Lamport,D.T. (1994) *Plant J.*, **5**, 157–172.
14. Kreuger,M. and van Holst,G.-J. (1996) *Plant Mol. Biol.*, **30**, 1077–1086.
15. Kieliszewski,M.J., Showalter,A.M. and Leykam,J.F. (1994) *Plant J.*, **5**, 849–861.
16. Kieliszewski,M.J., O'Neill,M., Leykam,J. and Orlando,R. (1995) *J. Biol. Chem.*, **270**, 2541–2549.
17. Chen,C.-G., Pu Z.-Y., Moritz,R.L., Simpson,R.J., Bacic,A., Clarke,A.E. and Mau,S.L., (1994) *Proc. Natl Acad. Sci. USA*, **91**, 10305–10309.
18. Reddy,M.K., Roy-Choudhuri,N., Kumar,D., Mukherjee,S.K. and Tewari,K.K. (1994) *Eur. J. Biochem.*, **220**, 933–941.
19. Mukherjee,S.K., Reddy,M.K., Kumar,D. and Tewari,K.K. (1994) *J. Biol. Chem.*, **269**, 3793–3801.
20. Kumar,D., Mukherjee,S.K., Reddy,M.K. and Tewari,K.K. (1995) *J. Exp. Bot.*, **46**, 767–776.
21. Maniatis,T., Fristch,E.F. and Sambrook,J. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, New York.
22. Edge,A.S.B., Faltynek,C.R., Liselotte,H., Reichert,L.E. and Weber,P. (1981) *Anal. Biochem.*, **118**, 131–137.
23. Moody,S.F., Handman,E., MacConville,M.J. and Bacic,A. (1993) *J. Biol. Chem.*, **268**, 8457–18466.
24. Bartlett,S.G., Grossman,A.R. and Chua,N.H. (1982) In Edelman,M. (ed.), *Methods in Chloroplast Molecular Biology*. Elsevier Biomedical Press, Amsterdam, pp. 1081–1091.
25. Chen,Q., Osteryoung,K. and Vierling,E. (1994) *J. Biol. Chem.*, **269**, 13216–13223.
26. Tranel,P.J., Froehlich,J., Goyal,A. and Keegstra,K. (1995) *EMBO J.*, **14**, 2436–2446.
27. Lin,Q., Ma,L., Burkhart,W. and Spremulli,L.L. (1994) *J. Biol. Chem.*, **269**, 9436–9444.
28. Wu,D.Y., Ugozzoli,L., Pal,B.K., Qian,J. and Wallace,B. (1991) *DNA Cell Biol.*, **10**, 233–238.
29. Bogorad,L. (1991) In Bogorad,L and Vasil,I.K. (eds), *The Molecular Biology of Plastids.* Academic Press, New York, pp. 93–117.
30. Grunert,S. and Jackson,R.J. (1994) *EMBO J.*, **13**, 3618–3630.
31. Gascuel,O. and Golmard,J.L. (1988) *Comput. Appl. Biosci.*, **4**, 357–365.
32. Von Heijne,G., Steppuhn,J. and Herman,G. (1989) *Eur. J. Biochem.*, **180**, 535–545.
33. Von Heijne,G. and Nishikawa,K. (1991) *FEBS Lett.*, **278**, 1–3.
34. Hansen,J.E., Lund,O., Engelbrecht,J., Bohr,H., Nielsen,J.O., Hansen,J.E.S. and Brunak,S. (1995) *Biochem. J.*, **308**, 801–813.
35. Hong,J.C., Cheong,Y.H., Nagao,R.T., Bahk,J.D., Cho,M.J. and Key,J.L. (1994) *Plant Physiol.*, **104**, 793–796.
36. Parmentier,Y., Durr,A., Marbach,J., Hirsinger,C., Criqui,M.C., Fleck,J. and Jamet,E. (1995) *Plant Mol. Biol.*, **29**, 279–292.
37. Rubinstein,A.L., Marquez,J., Svarez-Cervera,M. and Bedinger,P.A. (1995) *Plant Cell*, **7**, 2211–2225.
38. Brownleader,M.D. and Dey,P.M. (1993) *Planta*, **191**, 457–469.
39. Sheng,J., Jeong,J. and Mehdy,M.C. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 828–832.
40. Desai,N.N., Allen,A.K. and Neuberger (1983) *Biochem. J.*, **211**, 273–276.
41. Broeck,G.V.D., Timko,M.P., Kausch,A.P., Cashmore,A.R., Mantagu,M.V. and Estrella,L.H. (1985) *Nature*, **313**, 358–363.
42. Lessard,P., Decroocq,V. and Thomas,M. (1997) In Clark,M.S. (ed.), *Plant Molecular Biology.* Springer-Verlag, Berlin, Heidelberg, pp. 154–200.
43. Kardailsky,I.V., Sherrier,D.J. and Brewin,N.J. (1996) *Plant Physiol.*, **111**, 49–60.
44. Rini,J.M. (1995) *Annu. Rev. Biophys. Biomol. Struct.*, **24**, 551–577.
45. Kaufmann,H., Salanini,F. and Thompson,R.D. (1991) *Mol. Gen. Genet.*, **226**, 457–466.
46. Dodds,P.N., Clarke,A.E. and Newbigin,E. (1996) *Plant Mol. Biol.*, **31**, 227–238.
47. Sticher,L., Hofsteenge,J., Milani,A., Neuhaus,J.M. and Meins,F. (1992) *Science*, **257**, 655–657.
48. Kouranov,A. and Schnell,D.J. (1996) *J. Biol. Chem.*, **271**, 31009–31012.