*Articles*

# Synthesis and Structure Determination by NMR of a Putative Vacuolar Targeting Peptide and Model of a Proteinase Inhibitor from *Nicotiana alata*[†,||]

Katherine J. Nielsen,[‡] Justine M. Hill,[‡] Marilyn A. Anderson,[§] and David J. Craik*[,‡]

*Centre for Drug Design and Development, University of Queensland, St. Lucia, Queensland 4072, Australia, and School of Biochemistry, La Trobe University, Bundoora, Victoria 3083, Australia*

*Received September 18, 1995; Revised Manuscript Received October 31, 1995*[⊗]

ABSTRACT: NA-proPI is a 40.3-kDa multidomain precursor protein found in the stigma of the ornamental tobacco *Nicotiana alata.* It is selectively targeted to the vacuole and, as the plant matures, is processed to produce a series of five 6-kDa proteinase inhibitors (one chymotrypsin and four trypsin reactive sites) which are thought to play a vital role in plant protection against insect pests. A putative sixth domain with a chymotrypsin reactive site is likely to be formed by three disulfide bridges linking the N- and C-terminal fragments of NA-proPI. This domain contains two distinct structural elements: a 54-residue sequence with high identity to each of the five repeated PI domains, and a nonrepeated 25-residue sequence at the C-terminus which is proposed to contain a vacuolar targeting peptide. The structure of the putative sixth domain was predicted using a combination of secondary structure prediction and homology modeling based on the known structure of one of the intact domains. A 26-residue peptide corresponding to the nonrepeated C-terminal sequence and encompassing the putative vacuolar targeting sequence was synthesized and its structure determined using $^1$H NMR spectroscopy and simulated annealing calculations. The peptide was found to adopt an amphipathic helical structure. The calculations based on NOE data suggested that the helix is curved, with a hydrophobic concave face and a hydrophilic convex face. This curvature is consistent with an observed periodicity in backbone NH chemical shifts. The structure of the entire sixth domain was modeled by combining the experimentally determined structure of the putative vacuolar targeting peptide with the homology model of the PI domain. In this model the α-helix of the putative targeting peptide protrudes from the otherwise compact PI domain. This observation is consistent with the requirement for targeting sequences to be relatively exposed for recognition by the sorting apparatus. As there is no consensus on the structure of vacuolar targeting sequences, this study provides a valuable insight into their potential mechanism of interaction with the cellular sorting apparatus.

Serine proteinase inhibitors (PIs)[1] are found in a wide variety of plants and have been implicated in the protection of leaves, storage, and reproductive organs against insects and potential pathogens, including bacteria and fungi (see Ryan, 1990, and Richardson, 1991, for reviews). In a recent report (Atkinson et al., 1993), a series of PIs was identified in the stigma of the ornamental tobacco, *Nicotiana alata*. These PIs, which accumulate in the vacuole (Atkinson, 1992), have a molecular mass of approximately 6 kDa and are derived *in vivo* from the post-translational processing of a 40.3-kDa precursor protein.

The precursor protein, henceforth referred to as NA-proPI, comprises five domains of high sequence identity, each containing a potential PI reactive site. This is somewhat unusual, as most serine PIs in plants have only one or two kinetically independent reactive sites. The production of several PI domains on a single polypeptide chain may facilitate a more rapid accumulation of PI protein in plant tissues at times of stress by comparison to single- and double-headed inhibitors. In NA-proPI, the five domains are separated by the linker amino acid sequence EEKKND and are flanked by N-terminal (Nter) and C-terminal (Cter) fragments, as illustrated in Figure 1. Cleavage in the linker sequence and subsequent trimming by endogenous peptidases produce the five 6-kDa PIs, referred to as C1 and T1−T4 because of their activity as chymotrypsin and trypsin inhibitors, respectively (Heath et al., 1995). NA-proPI is present in the stigma at the bud stages of flower development, but as the flower matures, less NA-proPI is detected, and there is a concomitant increase in the levels of the 6-kDa PIs (Atkinson et al., 1993). C1 and T1−T4 have been isolated in sufficient quantities (Heath et al., 1995) to be studied using $^1$H NMR techniques (Nielsen et al., 1994,

[†] This work was supported by a grant from the Australian Research Council.

[||] Coordinates of the structure of the putative vacuolar targeting peptide have been deposited in the Brookhaven Protein Data Bank under the file name 1VTP.

* Author to whom correspondence should be addressed.

[‡] University of Queensland.

[§] La Trobe University.

[⊗] Abstract published in *Advance ACS Abstracts,* December 15, 1995.

[1] Abbreviations: PI, proteinase inhibitor; NA-proPI, *Nicotiana alata* precursor protein; Nter, N-terminal fragment; Cter, C-terminal fragment; NMR, nuclear magnetic resonance; DQF-COSY, double quantum filtered *J*-correlated spectroscopy; TOCSY, total correlated spectroscopy; NOESY, two-dimensional nuclear Overhauser spectroscopy; NOE, nuclear Overhauser effect; 2D, two-dimensional; 3D, three-dimensional; DSS, 4,4-dimethyl-4-silapentane-1-sufonate; SA, simulated annealing; RMSD, root mean square deviation.

0006-2960/96/0435-369$12.00/0 

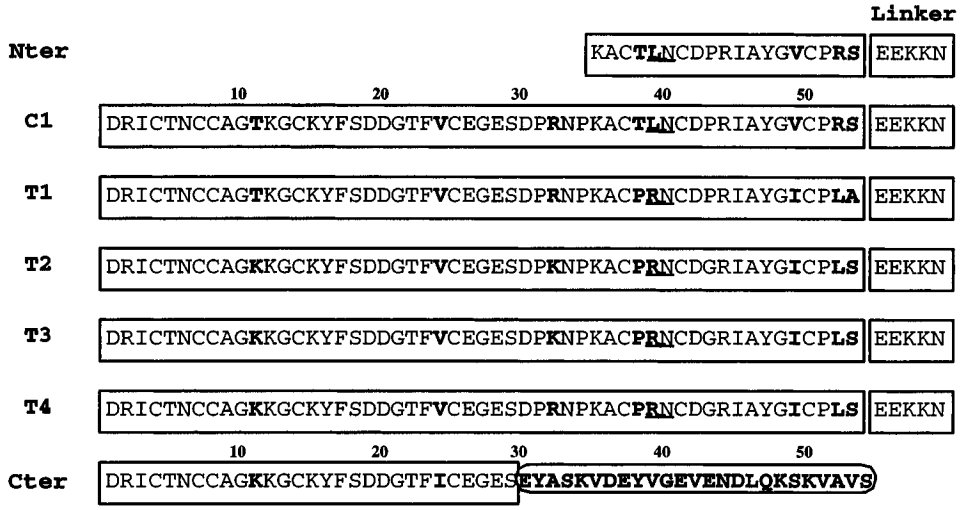


FIGURE 1: Amino acid sequence of NA-proPI showing the individual PI domains (C1 and T1−T4), the linker sequences, the N- and C-terminal sequences (Nter and Cter), and the putative vacuolar targeting sequence (oval). Underlined residues indicate the PI reactive site, and residues shown in bold indicate sequence differences. The first residue (D) in each PI domain may form the C-terminal part of the linker sequence, depending upon the location of the cleavage site.

1995). Nter and Cter have not been isolated, but are predicted to exist on the basis of the cDNA clone, NA-PI-II (Atkinson et al., 1993).

The 3D structure and disulfide connectivities of C1 have been determined using $^1$H NMR spectroscopy (Nielsen et al., 1994). C1 is a disk-shaped molecule comprised of a triple-stranded $\beta$-sheet, a short section of $3_{10}$ helix, several turns, and a loop region containing the putative reactive site. The N- and C-termini are exposed, consistent with the fact that this protein is produced via proteolytic cleavage from NA-proPI. The structure of C1 is similar to that of PCI-1, a chymotrypsin inhibitor isolated from potato tubers (Greenblatt et al., 1989). This is of significant interest, since PCI-1 is most likely derived from proteolytic cleavage of the double-headed potato type II proteinase inhibitor (PI-II) which has no linker regions comparable to the EEKKND sequence present in NA-proPI. Therefore, while the two precursor proteins have different tertiary arrangements of their subunits, their individual PI domains adopt similar 3D structures.

The five PIs (C1 and T1−T4) share between 85% and 100% sequence identity and have similar 3D structures (Nielsen et al., 1995). Part of Cter (residues 1−29) and all of Nter in NA-proPI have high sequence identity to residues 1−29 and 35−53, respectively, in C1. It has been proposed that, although well separated in the primary sequence, these may fold together, linked by three disulfide bridges, to form a sixth PI domain similar to the other five PIs (Atkinson et al., 1993; Nielsen et al., 1994). To date, there has been no evidence to support this hypothesis. In this study, secondary structure prediction and computer modeling based on the experimentally determined structure of C1 are used to predict the tertiary structure of the putative sixth PI. In addition, a peptide corresponding to the last 26 residues of Cter, Cter(29−54), has been synthesized and its structure determined using NMR spectroscopy. Figure 1 shows that this peptide sequence has no homology to the remainder of NA-proPI. Linkage of Nter and Cter to form a sixth PI domain will incorporate, as an appendage, 25 residues of this nonrepeated sequence. It is proposed here that this unique appendage at the C-terminus may contain the vacuolar targeting sequence.

Many plant proteins are synthesized with N- or C-terminal propeptides that are removed before or after transport to the vacuole. Experiments with transgenic plants have shown that deletion of these propeptide sequences results in secretion of the protein (Bednarek et al., 1990) and, conversely, addition of these elements to proteins that are normally secreted results in their redirection to the vacuole (Bednarek & Raikhel, 1991). Several vacuolar targeting motifs have been identified (see Nakamura & Matsuoka, 1993, and Chrispeels & Raikhel, 1992, for recent reviews), but these display little sequence identity, and there have been no reports of their 3D structures. X-ray diffraction studies of barley lectin and its precursor form are in progress, and analysis of their structures should reveal the folding motif of the C-terminal propeptide (Wright et al., 1993).

The lack of consensus on the sequence of vacuolar signal peptides has prompted the suggestion that the sorting apparatus either recognizes a common, and as yet unidentified, structural feature, or that there are several different receptors (Chrispeels & Raikhel, 1992). The presence of more than one receptor has been demonstrated in yeast using a series of mutants deficient in vacuolar targeting (Paravicini et al., 1992). The gene encoding one of the yeast receptors has been cloned (Marcusson et al., 1994), and there is preliminary evidence for a vacuolar sorting receptor in plants (Kirsh et al., 1994). Bednarek et al. (1990) emphasized the potential importance of structural features in vacuolar targeting when they predicted that the 15-residue C-terminal propeptide of barley lectin adopts an amphipathic $\alpha$-helix. Further analysis, however, indicated that short amino acid segments (FAEAI and LVAE) within the propeptide were sufficient (although less efficient) for targeting, suggesting that the core of the sorting signal is composed of several hydrophobic residues followed by an acidic residue (Bednarek & Raikhel, 1992). Given the current interest and uncertainty in the mechanisms of vacuolar targeting, it is of importance to determine the structure of the putative vacuolar targeting sequence in NA-proPI and to compare this with predicted structures of known C-terminal targeting sequences.

## MATERIALS AND METHODS

*Materials.* Boc-L-amino acids were from the Peptide Institute (Osaka, Japan) and Bachem Bioscience (Philadelphia). *t*-Boc-Ser(Bzl)-$OCH_2$-Pam-(S-DVB)-resin was obtained from Applied Biosystems (Foster City, CA.). HBTU was obtained from Richelieu Biotechnologies (Quebec, Canada). Trifluoroacetic acid (TFA), *N*,*N*-diisopropylethylamine (DIEA), and *N*,*N*-dimethylformamide (DMF), all peptide synthesis grade, were purchased from Auspep (Melbourne, Australia). Acetonitrile (HPLC grade) was obtained from BDH (Poole, England). HF was purchased from Mallinckrodt. Deuterated solvents were obtained from Cambridge Isotope Laboratories (Woburn, MA). Other reagents and solvents were analytical reagent grade.

*Peptide Synthesis.* Cter(29−54) (SEYASKVDEYVGEVENDLQKSKVAVS) was synthesized by manual stepwise solid-phase peptide synthesis using an *in situ* neutralization protocol for Boc chemistry (Schnölzer et al., 1992). The following side chain protection was used: Asn (Xan), Asp (OcHex), Glu (OcHex), Lys (Cl-Z), Ser (Bzl), and Tyr (BrZ). The peptide was cleaved from the resin and simultaneously deprotected by treatment with anhydrous HF/*p*-cresol/*p*-thiocresol (18:1:1 v/v; 0 °C, 1 h). Crude peptide product was precipitated with diethyl ether, dissolved in 50% aqueous acetic acid, diluted with water, and lyophilized. The peptide was purified to homogeneity by semipreparative reverse-phase HPLC (Delta Pak $C_{18}$ column, 7.8 × 300 mm, Waters) using a linear gradient of 0−60% acetonitrile in water and 0.1% trifluoroacetic acid over 60 min. The overall yield of purified peptide was 26%. Analytical HPLC (Vydac $C_{18}$ column, 4.6 × 150 mm) and ionspray mass spectrometry confirmed the purity of the peptide.

*Secondary Structure Predictions*. The conformational preferences of the putative sixth PI domain, Cter(29−54), and several C-terminal propeptides required for vacuolar targeting of plant proteins were simulated using the program ALB (Ptitsyn & Finkelstein, 1983). This program predicts elements of secondary structure on the basis of the physical properties of polypeptides when they are placed in aqueous, partially hydrophobic, and fully hydrophobic environments.

*Homology-Based Modeling*. The structure of the putative sixth PI domain, C2, was based on the PDB coordinates obtained for C1 (Nielsen et al., 1994) and the structure determined for Cter(29−54) in this study. An α-helix incorporating residues 29−54 was appended to the putative sixth PI domain using the Insight II (Biosym Technologies) molecular modeling system.

*NMR Spectroscopy.* Samples were prepared to a final concentration of 2 mM in the following solvent systems: 90% $H_2O$/10% $D_2O$ (pH 3.6 and 4.7, 280 K) or 33% TFE/67% $H_2O$ (pH 3.1, 280, and 300 K). All solvent percentages are by volume.

$^1$H NMR spectra were recorded on a Bruker ARX-500 spectrometer equipped with a shielded gradient unit. Low-temperature studies employed a temperature-controlled stream of cooled air using a Bruker BCU refrigeration unit and B-VT 2000 control unit. 2D NMR spectra were recorded in phase-sensitive mode using time-proportional phase incrementation for quadrature detection in the $t_1$ dimension (Marion & Wüthrich, 1983). The 2D experiments included DQF-COSY (Rance et al., 1983), TOCSY (Braünschweiler & Ernst, 1983) using a MLEV-17 spin-lock sequence (Bax & Davis, 1985) with a mixing time of 80 ms, and NOESY (Jeener et al., 1979) with mixing times of 200 and 300 ms. For DQF-COSY experiments, solvent suppression was achieved using selective low-power irradiation of the water resonance during a relaxation delay of 1.8 s. Solvent suppression for NOESY and TOCSY experiments was achieved using a modified WATERGATE sequence (Piotto et al., 1992) in which two gradient pulses of 2 ms duration and 6 G $cm^{-1}$ strength were applied either side of a binomial pulse. Spectra were routinely acquired with 4096 complex data points in $F_2$ and 512 increments in the $F_1$ dimension, with 32 scans per increment (64 for NOESY).

Slowly exchanging NH protons were detected by acquiring a series of 1D and TOCSY spectra of the peptide immediately following dissolution in 100% $D_2O$ or 33% TFE/67% $D_2O$. The $^3J_{NH\text{-}H\alpha}$ coupling constants were measured from DQF-COSY spectra, strip-transformed to 8K × 1K. 1D slices from this spectrum were analyzed using a simple peak simulation routine to obtain approximate coupling constants.

Spectra were processed on a Silicon Graphics Indigo workstation using UXNMR (Bruker) and FELIX (Biosym Technologies) software. The $t_1$ dimension was zero-filled to 2048 real data points, and 60° or 90° phase-shifted sine bell window functions were applied prior to Fourier transformation. Spectra were referenced to residual solvent peaks, calibrated externally using 4,4-dimethyl-4-silapentane-1-sulfonate (DSS).

Peak volumes in the 300 ms 33% TFE/67% $H_2O$ NOESY spectrum were classified as strong, medium, and weak corresponding to upper bound interproton distance restraints of 2.7, 3.5, and 5.0 Å, respectively (Clore et al., 1986). Appropriate pseudoatom corrections were applied to non-stereospecifically assigned methylene and methyl protons (Wüthrich et al., 1983). In addition, 1.5 and 2.0 Å was added to the upper limits of distances involving methyl protons and phenyl rings, respectively.

*Structure Calculations.* Three-dimensional structures were calculated using a simulated annealing (SA) and energy minimization protocol within the program X-PLOR (version 3.1; Brünger et al., 1986; Brünger, 1992). An *ab initio* SA protocol (Nilges et al., 1988) was used, starting from template structures with randomized $\phi$ and $\psi$ angles and extended side chains, to generate a set of 50 structures. The SA protocol consisted of a total of 30 ps of molecular dynamics (800 K) followed by cooling (300 K). Refinement under the influence of the CHARMM forcefield (Brooks et al., 1983) was achieved using the conjugate gradient Powell algorithm (Clore et al., 1986) with 1000 cycles of energy minimization.

## RESULTS

*Secondary Structure Prediction*. The sequence alignment in Figure 1 illustrates the high sequence identity of Nter with residues 35−53 in each of C1 and T1−T4. Similarly, the first 29 residues of Cter have high sequence identity with the corresponding residues in T1−T4. Thus, the combination of Nter with Cter(1−29) can, in principle, form a complete domain similar to the PIs C1 and T1−T4. Such a domain would require the formation of three interchain disulfide bridges between Nter and Cter and one intrachain disulfide

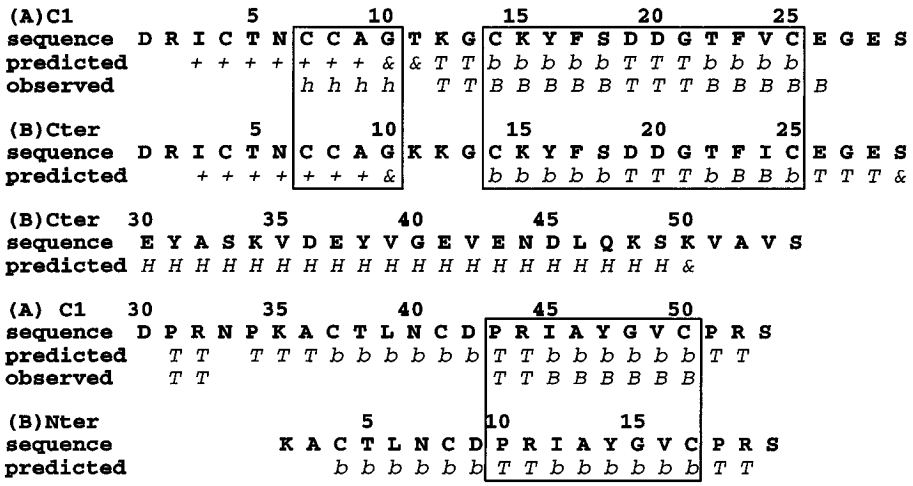


FIGURE 2: A comparison of the predicted and observed secondary structure for (A) C1 aligned with the predicted secondary structure for (B) the C- and N-terminal sequences (Nter and Cter). Boxed regions are those regions in Cter (or Nter) that are predicted and observed in C1. *B*: $\beta$-sheet definite; *b*: $\beta$-sheet probable; *: $\beta$-sheet possible, but not predicted; *H*: $\alpha$-helix definite; *h*: $\alpha$-helix probable; &: $\alpha$-helix possible, but not predicted; +: both $\alpha$-helix and $\beta$-sheet possible; *T*: turn. Different forms of helix (i.e., $\alpha$-helix and $3_{10}$ helix) are not distinguishable using this algorithm.

bridge within Cter to give the same disulfide connectivities observed for C1 and T1−T4.

To model the structure of this domain, it was first necessary to determine the intrinsic propensity of the individual fragments to adopt defined secondary structure. Figure 2 shows the results of secondary structure predictions for Nter and Cter and a comparison of the predicted and observed secondary structure for C1. It is of interest to note that most of the secondary structure elements which were experimentally identified in C1 (Nielsen et al., 1994) are correctly predicted by the ALB algorithm (Ptitsyn & Finkelstein, 1983). This includes the $\beta$-hairpin incorporating residues 14−25 (residue 26 forms part of the $\beta$-hairpin, but this is not predicted), the third strand of $\beta$-sheet (residues 45−50), two of the three remaining turns (i.e., those centered on residues 31−32 and 43−44), and to a lesser degree, the short stretch of helix (residues 7−10), where either helix or $\beta$-sheet was predicted to be possible. The major difference between the predicted and observed structure is at the reactive site loop (residues 37−41) which is relatively extended and not well-defined in the solution structures (Nielsen et al., 1994) but predicted to form $\beta$-sheet. As expected, Nter and Cter(1−29) are predicted to have similar secondary structure to C1. The fact that the isolated peptide fragments are predicted to have a high intrinsic tendency to form most elements of secondary structure independently provides support for the existence of the putative sixth PI inhibitor, C2, formed by the linked Nter and Cter(1−29) fragments, and suggests that its structure is similar to those experimentally determined for C1 (Nielsen et al., 1994) and T1−T4 (Nielsen et al., 1995).

Before the structure of the sixth PI was modeled, it was also necessary to obtain structural information for the remainder of the C-terminus. In particular, it was of interest to predict the structure of the 25-residue sequence corresponding to the C-terminus of NA-proPI. This represents the only major difference between a module comprising the linked Nter and Cter and the other PI molecules which have been isolated from the plant stigma (Heath et al., 1995). Application of the structure prediction protocol to this sequence shows that Cter(29−54) is likely to consist entirely of $\alpha$-helix (Figure 2).

*$^1$H NMR Spectroscopy*. To confirm the above prediction, Cter(29−54) was synthesized by Boc chemistry solid-phase techniques and its structure determined using $^1$H NMR spectroscopy. Spectra were recorded both in $H_2O$ and in 33% TFE/$H_2O$.

*(A) Chemical Shift Assignment*. The $^1$H NMR spectra of Cter(29−54) in aqueous solution are relatively overlapped; however, complete resonance assignment was achieved using well-established techniques (Wüthrich, 1986). Ambiguities due to peak overlap were resolved by comparison of NOESY and TOCSY spectra recorded at different pH values (pH 3.6 and 4.7). Addition of 33% TFE to the aqueous solution resulted in greater chemical shift dispersion, making resonance assignment relatively straightforward. Figure 3 shows NOESY spectra illustrating the sequential assignment of Cter(29−54) in $H_2O$ and 33% TFE. The complete resonance assignments in $H_2O$ are available as supporting information.

*(B) Secondary Structure*. The deviation of $^1$H chemical shifts from their "random coil" values has been shown to have a strong relationship to protein secondary structure (Wüthrich, 1986; Pastore & Saudek, 1990). The secondary shifts for the H$\alpha$ (Wishart et al., 1991) and NH (Wüthrich, 1986) protons of Cter(29−54) in $H_2O$ and TFE are presented in Figure 4. In aqueous solution, a general upfield trend of H$\alpha$ secondary shifts suggests that the peptide has a preference for helical conformation. On addition of TFE, which has recently been shown to have negligible effects on H$\alpha$ shifts in random coil peptides (Merutka et al., 1995), this upfield trend is accentuated, particularly over residues 32−51. From residues 32 to 51 the H$\alpha$ shifts are on average 0.25 ppm upfield of random coil values in TFE, compared with 0.08 ppm in $H_2O$. This indicates that the secondary structure is the same in both solvents but more stable in 33% TFE. These results are consistent with previous findings that TFE is conducive to helix stabilization in peptides with inherent propensities for helix formation (Dyson et al., 1992; Sönnichsen et al., 1992).

The NH chemical shifts change periodically along the peptide sequence, with three maximum positive $\Delta\delta$ values at positions 39, 43, and 46 and five maximum negative $\Delta\delta$ values at positions 35, 37, 41, 44, and 51 (Figure 4B). This is consistent with previous observations that NH chemical

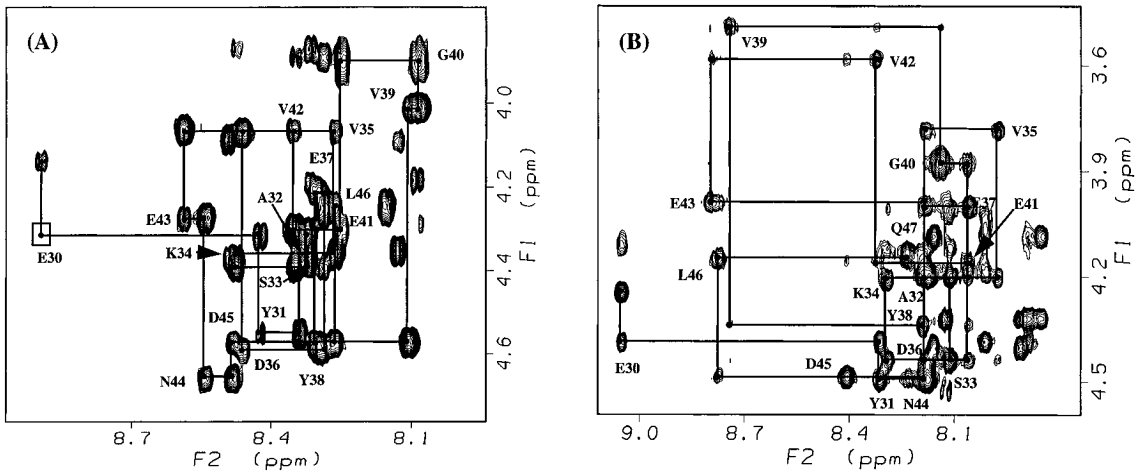


FIGURE 3: NH−Hα regions of NOESY spectra for Cter(29−54) in (A) 90% $H_2O$/10% $D_2O$ at pH 4.7 and 280 K and (B) 33% TFE/63% $H_2O$ at pH 3.1 and 280 K. Mixing times are 300 and 200 ms, respectively.

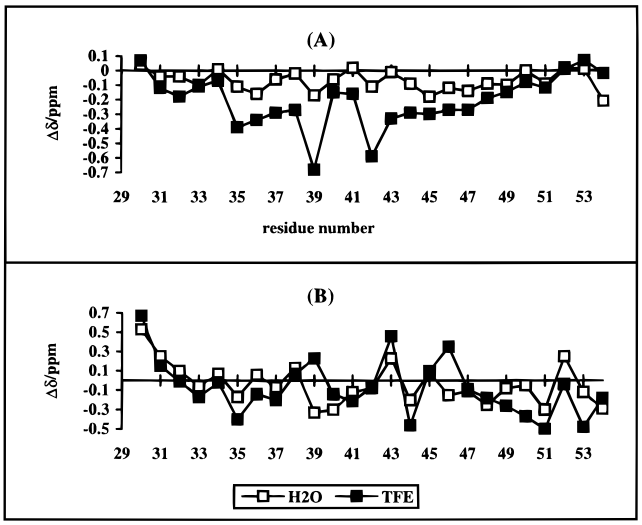


FIGURE 4: Secondary chemical shifts for (A) Hα and (B) NH protons of Cter(29−54) in $H_2O$ (280 K) and TFE (300 K).

shifts in many amphipathic and/or curved α-helices show a 3−4 residue periodicity (Kuntz et al., 1991; Jiménez et al., 1992; Zhou et al., 1992).

Summaries of the sequential and medium-range NOEs, $^3J_{NH\text{-}H\alpha}$ coupling constants, and slowly exchanging amide protons for Cter(29−54) in $H_2O$ and TFE are given in Figure 5, panels A and B, respectively. In $H_2O$, the observation of strong Hα−$NH_{i+1}$ and corresponding weaker NH−$NH_{i+1}$ NOEs is not indicative of a high population of helical conformations, suggesting a significant proportion of partially unfolded and/or extended conformations. This is supported by the absence of medium-range NOEs. However, the presence of a transient α-helix in $H_2O$ is supported by the measurement of reasonably low $^3J_{NH\text{-}H\alpha}$ coupling constants (5−7 Hz) and 11 slowly exchanging amide protons, which are present more than 2 h after dissolution in $D_2O$. Addition of TFE to the aqueous solution produced a significant change in a range of spectral parameters, highly characteristic of a stable helical conformation. The observation of strong NH−$NH_{i+1}$ and corresponding weak Hα−$NH_{i+1}$ connectivities, together with the presence of several weak Hα−$NH_{i+3}$ and medium to strong Hα−$H\beta_{i+3}$ connectivities, suggests the presence of α-helix from residues 32 to 51. The lack of Hα−$NH_{i+2}$ NOEs and the presence of Hα−$NH_{i+4}$ NOEs over this region indicates that the α-helix is regular and

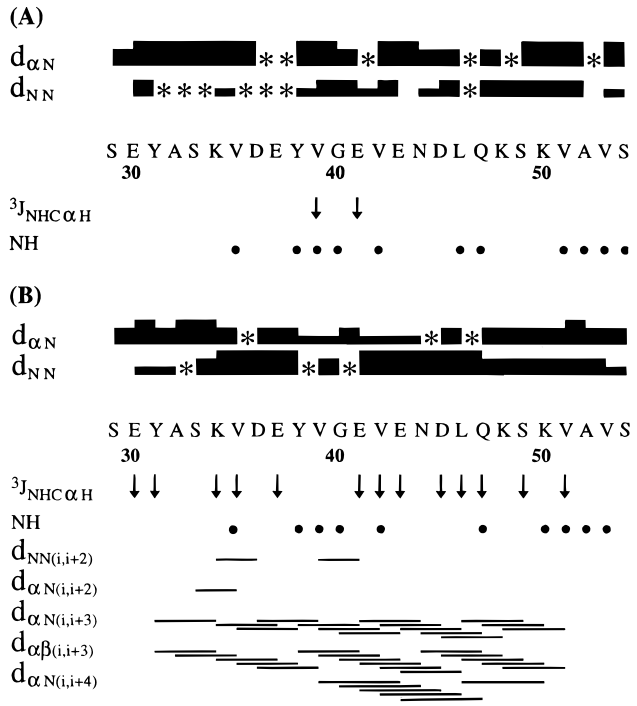


FIGURE 5: Summary of the sequential and medium-range NOE information, $^3J_{NH\text{-}H\alpha}$ coupling constant and slow exchange NH data observed for Cter(29−54) in (A) $H_2O$, pH 3.6, and (B) 33% TFE. Relative NOE intensities are represented by the heights of the bars. Overlapping NOEs are indicated by an asterisk (*). Down arrows indicate $^3J_{NH\text{-}H\alpha}$ coupling constants ≤6 Hz. Filled circles indicate NH resonances present with significant intensity more than 2 h after dissolution of the peptide in $D_2O$ or 33% TFE/$D_2O$.

stable. The presence of helical structures is supported by the observation of 10 slowly exchanging NH protons, which were identical to those observed in $D_2O$, and low (<6 Hz) $^3J_{NH\text{-}H\alpha}$ coupling constants (Figure 5B). However, the relative intensities of the NH−$NH_{i+1}$ and Hα−$NH_{i+1}$ connectivities indicate that the first five N-terminal residues differ in conformation from the rest of the peptide. This is supported by the observation of a Hα−$NH_{i+2}$ connectivity near the N-terminus, indicating fraying or a turn at this end of the α-helix.

The $^1H$ NMR studies have shown unambiguously that Cter(29−54) adopts at least transient or "nascent" helical conformations in aqueous solution as an isolated peptide fragment. In TFE, with increased hydrophobicity, the helical

Table 1: Structural Statistics for Cter(29−54)

| structural parameter | 20 converged structures |
|---|---|
| energy statistics (kcal mol$^{-1}$)[a] | |
| $F_{NOE}$ | 2.09 ± 0.23 |
| $E_{L\text{-}J}$ | −116 ± 4 |
| RMS deviations from NOE restraints (Å) | 0.020 ± 0.001 |
| RMS deviations from idealized geometry | |
| bonds (Å) | 0.008 ± 0.001 |
| angles (deg) | 2.06 ± 0.03 |
| impropers (deg) | 0.148 ± 0.017 |

[a] $F_{NOE}$ is the energy related to NOE violations and $E_{L\text{-}J}$ is the Lennard-Jones van der Waals energy calculated with the CHARMM (Brooks et al., 1983) empirical energy functions.

structure is stabilized over most of the peptide. The helix formed in this environment stretches from approximately residues 32 to 51, with the N- and C-terminal regions showing less definition than the core region. Due to the lack of information on the structure of vacuolar targeting peptides and the preliminary evidence from NH chemical shift periodicity of some curvature of the helical structure, it was considered important to calculate a set of structures for Cter(29−54). Identification of structural features common to vacuolar targeting peptides will provide an insight into the sorting receptors that function in vacuolar targeting in plants.

*(C) Three-Dimensional Structure*. A total of 268 distance restraints derived from 131 intraresidual, 78 sequential, and 59 medium-range NOEs was used in the calculation of 50 simulated annealing (SA) structures. Thirteen restraints on the $\phi$ dihedral angle were also included. All SA runs converged to produce structures with a common fold which were in good agreement with the experimental restraints and had low total energies. The 20 structures with the lowest energies were chosen to represent the solution structure of Cter(29−54). These structures satisfied the experimental restraints, with an average of only two interproton distance violations greater than 0.1 Å per structure and no torsion angle violations greater than 3°. Structural statistics are shown in Table 1.

A superimposition of the backbone atoms of residues 32−51 for the 20 refined structures of Cter(29−54) (Figure 6) shows that the helix is well-defined over most residues, apart from some dynamic fraying at the N- and C-termini. The helix is substantially curved in all of the structures, consistent with the periodicity of the NH chemical shifts. The precision of the final set of structures is documented in Figure 7, which displays the RMSD to the average structure for each residue and the angular order parameter, *S*, as a measure of the spread of $\phi$, $\psi$, and $\chi_1$ dihedral angles within the set of structures (Hyberts et al., 1992; Pallaghy et al., 1993). The RMSD by residue for the backbone atoms shows that the structure is best defined from residues 32−51, and least defined at the helix termini. The backbone pairwise RMSD for residues 32−51 is 1.63 Å. The high values of *S* (mostly ≥0.99) for the $\phi$ and $\psi$ dihedral angles (Figure 7B,C) of residues 32−51 and their consistent location in the α-region of the Ramachandran plot show that the helical structure in this region of the peptide is well-defined. The side chain dihedral angles, $\chi_1$, are not as well-defined (Figure 7D); however, several *S* values approach unity, indicating good definition of the side chains in these regions.

Analysis of the final structures for possible H-bond partners of the observed slowly exchanging NH protons was conducted using SSTRUC [Smith, 1991 (unpublished program)]. SSTRUC identifies H-bonds on the basis of NH···C═O distance and relative orientation using the algorithm of Kabsch and Sander (1983). The results show that the C═O oxygen atoms of residues 38−46 favor the *i*,*i*+4 H-bond observed in regular α-helices over the *i*,*i*+3 H-bonds observed in $3_{10}$ helices. This indicates that, over the well-defined region of the structure, regular α-helix rather than $3_{10}$ helix is the dominant folded form in the structural equilibrium.

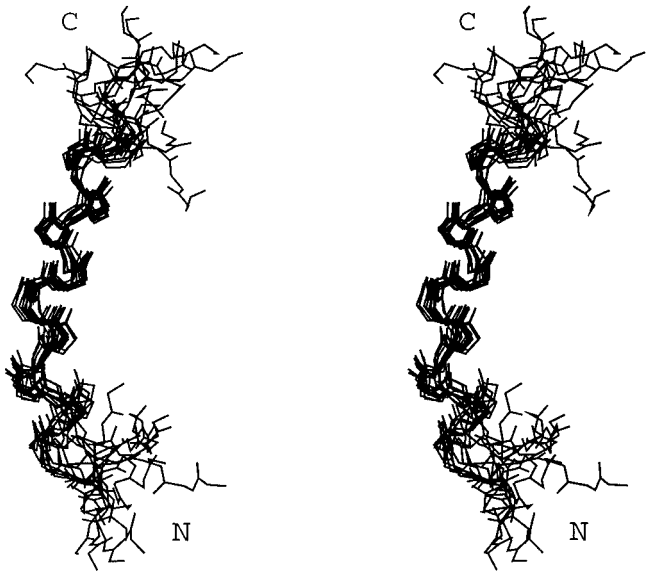


FIGURE 6: Stereoview of the superimposition of the backbone atoms of residues 32−51 for the 20 refined structures of Cter(29−54) derived from the experimental NOE data in 33% TFE.

*Modeling of the Linked N- and C-Terminal Fragments.* In NA-proPI, Nter and Cter are likely to be linked by three disulfide bridges, producing an overall cyclic structure with six distinct PI domains (Figure 8A). The alternative hypothesis in which Nter and Cter are unconnected would result in a linear protein with only five PI domains and the N- and C-terminal fragments, both with at least one unpaired cysteine residue (Figure 8B). Given the high sequence identity of Nter and Cter(1−29) with the five known PIs (C1 and T1−T4), the identical cysteine positioning, and the similarity in the secondary structure predictions, it is likely that the "cyclic" structure exists in the precursor protein.

Based on the results described above, it is possible to build a model of the putative sixth PI using coordinates obtained from the 3D structure calculations for C1 (Nielsen et al., 1994) with appropriate residue substitutions for Nter and Cter(1−29) and addition of the helical region for Cter(29−54). Figure 9 illustrates the two participating peptide fragments based on this homology modeling. Nter contains the reactive site, one $\beta$-turn, and a peripheral $\beta$-strand, while Cter contains the short stretch of $3_{10}$ helix, a $\beta$-turn, and the $\beta$-hairpin followed by the α-helix. A regular and "linear" α-helix was used in this model in preference to one of the calculated structures to emphasize that (i) while curvature of the helix is indicated in the isolated peptide, it will not necessarily occur in the intact domain (i.e., solvent effects may play a role) and (ii) fraying of the termini in the calculated structures makes it difficult to determine the relative orientation of the helix and the PI domain.

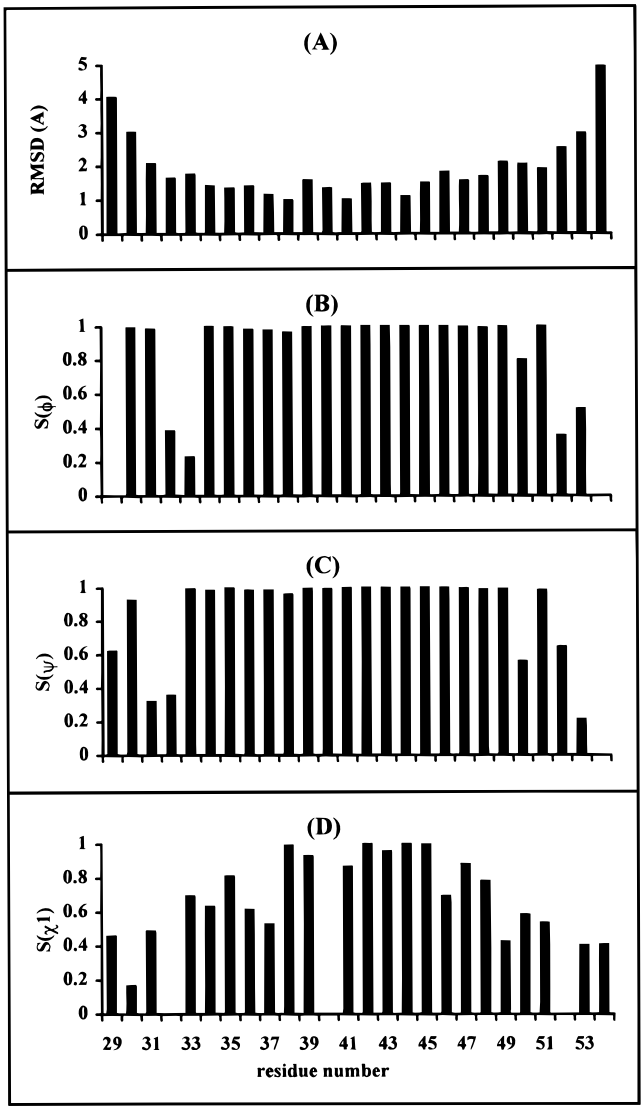


FIGURE 7: Plots as a function of residue number of (A) the RMSDs for the backbone atoms, and (B) to (D) the angular order parameters for the backbone ($\phi$ and $\psi$) and first side chain ($\chi_1$) dihedral angles. All parameters were calculated from the final 20 refined structures.

## DISCUSSION

A model for the 3D structure of the sixth PI domain in NA-proPI, henceforth referred to as C2, has been proposed based on the high sequence identity of this region to the five isolated PIs and the results of secondary structure prediction. An analogous arrangement of PI domains has been reported for another member of this class of inhibitors (potato type II) which has two PI domains with substantial sequence identity to the NA-proPI domains. In this case, the chymotrypsin domain (PCI-1) is in the middle of the amino acid sequence of the precursor protein, and the trypsin reactive domain is believed to arise from the linkage of the flanking N- and C-termini by three disulfide bridges (Greenblatt et al., 1989). Like NA-proPI, the structure of this precursor protein, PI-II, has not been determined experimentally. The trypsin reactive domain reported to be present in PI-II has not been isolated as a separate domain; however, the chymotrypsin reactive domain has been isolated and its X-ray structure determined (Greenblatt et al., 1989). The high sequence identity between C1 and PCI-1 (71%) and their identical disulfide connectivity pattern and similar 3D structure suggest that these PIs belong to the same class even though the number of domains in their respective precursors differ greatly and the peptide linkage between the domains found in NA-proPI is absent in PI-II.

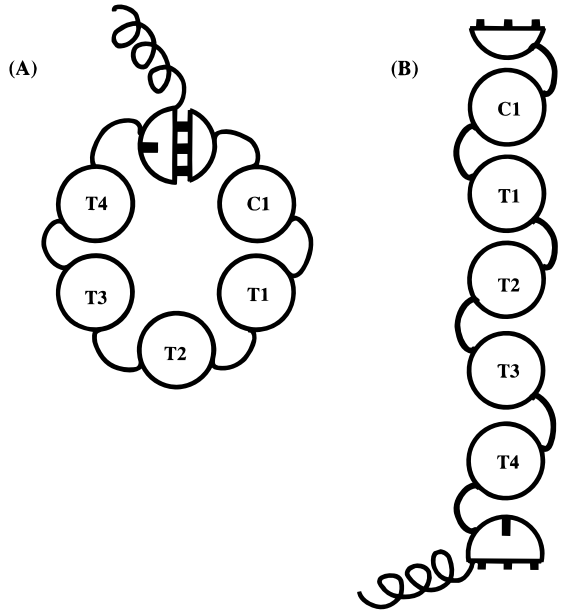


FIGURE 8: Possible models for the precursor protein, NA-proPI, indicating the possible arrangements of PI domains. (A) "Cyclic" model in which Nter and Cter are linked by three interchain disulfide bridges (shown as long rods). (B) "Linear" model in which Nter and Cter are unconnected, resulting in a linear protein with only five PI domains and the N- and C-terminal fragments with one or more unpaired cysteine residues (shown as short rods).

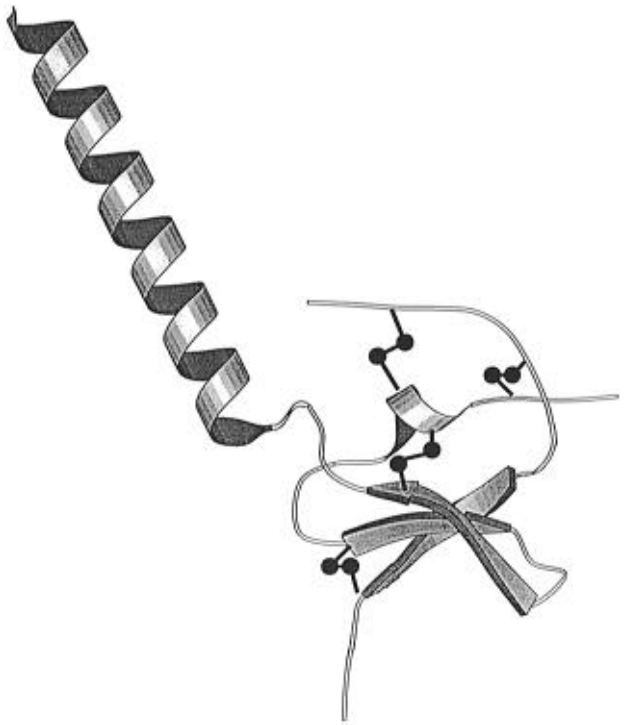

FIGURE 9: A Molscript view (Kraulis, 1991) of Cter and Nter linked by the three interchain disulfide bridges to produce C2. As the purpose of the diagram is to illustrate the relative sizes of the PI domain and the targeting signal, the helix shown is a regular α-helix, rather than one of the calculated structures which display a degree of curvature.

The $^1$H NMR results indicate that the putative vacuolar targeting peptide, Cter(29−54), adopts at least transient helical conformations in aqueous solution as an isolated

Table 2: Predicted Secondary Structure for Several C-Terminal Vacuolar Targeting Peptides in Different Solvent Environments[a,b]

| Peptide/environment | Primary sequence/secondary structure prediction |
|---|---|
| Wheat Germ Agglutinin A[c] | **V F A E I T A N S T L L Q E** |
| aqueous | * * * * * * * * |
| lipid | + h h h h h h & * * * |
| Barley Lectin[d] | **V F A E A I A A N S T L V A E** |
| aqueous | * * * * * * * * * * * * |
| lipid | + h h h H H h h & & + * * * |
| Rice Lectin[e] | **D G M A A I L A N N Q S V S F E G I I E S V A E L V** |
| aqueous | h h h h h h & * + + h H H H H H H H h h h |
| lipid | & H H H H H H & & T T & & H H H H H H H H H H & |
| *N. tabacum* chitinase[f] | **N G L L V D T M** |
| aqueous | * * * * * |
| lipid | * + + + |
| *N. tabacum* β-1,3-glucanase[g] | **V S G G V W D S S V E T N A T A S L V S E M** |
| aqueous | T T T + + + & & & & + + + |
| lipid | T T T & & h h h h H h h h h |
| *N. plumbaginifolia* β-1,3-glucanase[g] | **F S D R Y W D I S A E N N A T A A S L I S E M** |
| aqueous | * * * * * * & & & & + + + & |
| lipid | & & + h h H H H H H h h h |

[a] Predictions carried out using the program ALB (Ptitsyn & Finkelstein, 1983). [b] *T* = turn predicted; *H* = helix predicted; *h* = helix probable; & = helix possible but not predicted; * = β-strand possible but not predicted; + = helix or β-strand possible but not predicted. [c] Raikhel & Wilkins, 1987. [d] Bednarek et al., 1990. [e] Wilkins & Raikhel, 1989. [f] Neuhaus et al., 1991. [g] Bol et al., 1990.

peptide fragment. In the more hydrophobic environment provided by TFE, the helical structure is stabilized over most of the peptide. The structure of Cter(29−54) is best defined over residues 32−51, with some fraying at the helix termini. Interestingly, the structure calculations suggest that the helix is not linear, but curved. Several factors support the existence of this helix curvature, as opposed to it being an artifact of 3D structure calculation. First, no long-range NOEs were observed that could falsely distort the helical structure. Second, the empirical NH chemical shifts show a 3−4 repeat periodicity consistent with many amphipathic and/or curved α-helices (Kuntz et al., 1991; Jiménez et al., 1992; Zhou et al., 1992). The observation of NH chemical shift variation can be explained by the change of H-bond length in an amphipathic α-helix (Wagner et al., 1983; Kuntz et al., 1991; Zhou et al., 1992), where short H-bonds are correlated with low-field shifts and long H-bonds with shifts to higher field. The curvature of helices is caused by these differences in the H-bonding on opposite faces of the helix, with shorter H-bonds in the center of the hydrophobic face and longer H-bonds in the center of the hydrophilic face. In support of this, the slow exchange NH protons in $D_2O$ and TFE solutions are found predominantly along the hydrophobic face of the helix which, in the structures, is concave, as is the case for the curved helices described by Barlow and Thornton (1988). The relevance of the helix curvature to the function of the peptide is not known. The curvature may, in fact, be enhanced by solvent effects as was found for the long helix bridging the two domains of T4 lysozyme, which was found to be more curved in solution (McLeish et al., 1994) than in the crystal state (Weaver & Matthews, 1987).

The positioning of the α-helical region is arbitrary with respect to the remainder of the putative sixth PI domain in the model shown in Figure 9. However, its significant longitudinal size, relative to the individual PIs, and its location close to the surface exposed reactive site indicate that it is likely to be exposed in the precursor protein. For this reason, and because the α-helical region represents the only major difference between this domain and the remaining five PIs, it is logical to assume that Cter(29−54) fulfils a unique function in NA-proPI. One possible function is that of the vacuolar targeting peptide, which must exist to direct NA-proPI through the Golgi apparatus to the vacuole for processing and export. In support of this, it is generally thought that vacuolar targeting peptides must be exposed on the surface of the protein to be recognized by the sorting machinery (Chrispeels, 1991; Nakamura & Matsuoka, 1993).

Three classes of vacuolar targeting peptides are found in plant proteins: C-terminal propeptides composed of several hydrophobic residues and with an overall negative charge, N-terminal propeptides with a short region of hydrophobic amino acids interrupted by a hydrophilic residue, and, in some cases, exposed regions within the amino acid sequence. The lack of sequence conservation between these signals suggests that multiple receptors are involved in vacuolar sorting in plants (Chrispeels & Raikhel, 1992; Nakamura & Matsuoka, 1993). Cter(29−54) is a C-terminal peptide enriched in negative charge which supports its proposed role in vacuolar targeting. Further support is provided by predicted secondary structures of other C-terminal plant vacuolar targeting peptides. A summary of the secondary structure predictions done in the current study for a range of established targeting sequences (Table 2) shows that many

of these are likely to adopt α-helical conformations, at least under hydrophobic conditions. Wilkins and Raikhel (1989) have also predicted the C-terminal domains of wheat germ agglutinin A, rice lectin, and *β*-1,3-glucanase to adopt amphipathic α-helical structures. However, an exception is the short chitinase sequence which has an equal preference for α-helix and *β*-sheet.

Bednarek et al. (1990) originally proposed that an amphipathic α-helix was required for vacuolar sorting. Subsequent deletion experiments have indicated that small peptide fragments of four amino acid residues are sufficient for targeting, but efficiency is increased when more residues are present (Bednarek & Raikhel, 1992; Schroeder et al., 1993). Conversely, recent experiments have indicated that vacuolar sorting is dependent not only on the information contained within the targeting signal, but also on its correct 3D presentation for efficient recognition by the sorting apparatus. From the recently published X-ray structure of a chitinase from barley seeds (Hart et al., 1993) it can be inferred that a C-terminal extension would stick out from the compact catalytic domain and thus be accessible to the vacuolar sorting system. Preliminary X-ray diffraction studies of barley lectin and its precursor form have also shown that the C-terminal extension is likely to be folded into a domain spatially distinct from the mature four-domain polypeptide. Since the presence of the targeting peptide alone is not sufficient to redirect fusion proteins to the vacuole with 100% efficiency, it is postulated that both the correct 3D fold as well as its assembly in relation to the four-domain module are important factors for efficient recognition in the sorting process (Wright et al., 1993).

Some features seem to occur frequently in C-terminal propeptides and could play a role in vacuolar targeting. These include stretches of hydrophobic amino acids and a preference for negative charges. Results to date indicate that a general accessibility of the peptide to the sorting system could be among the most important properties of the C-terminal vacuolar targeting peptides. It is of interest, however, that many of these peptides are predicted to be α-helical in a hydrophobic environment. The α-helical structure provides an efficient way of presenting a series of amino acid side chains to be recognized by the sorting apparatus. In this case, the α-helical region of C2 protrudes as a "tail" from the more compact, globular, PI domain despite its relatively smaller size (i.e., shorter sequence). Other types of targeting peptides have also been found to adopt α-helical conformations, including mitochondrial (Verner & Schatz, 1988) and periplasmic (Rizo et al., 1993) targeting peptides. The results of this study indicate that C-terminal vacuolar targeting sequences have a preference for α-helical conformation; however, there is a need for further studies in this area before a consensus on the structure of these peptides can be made.

## SUPPORTING INFORMATION AVAILABLE

A table of $^1$H NMR assignments for Cter(29−54) in 90% $H_2O$/10% $D_2O$ at pH 4.7 and 280 K (2 pages). Ordering information is given on any current masthead page.

## REFERENCES


Atkinson, A. H. (1992) Ph.D. Thesis, University of Melbourne.

Atkinson, A. H., Heath, R. L., Simpson, R. J. Clarke, A. E., & Anderson, M. A. (1993) *Plant Cell 5*, 203−213.

Barlow, D. J., & Thornton, J. M. (1988) *J. Mol. Biol. 201*, 601−619.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355−360.

Bednarek, S. Y., & Raikhel, N. V. (1991) *Plant Cell 3*, 1195−1206.

Bednarek, S. Y., & Raikhel, N. V. (1992) *Plant Mol. Biol. 20*, 133−150.

Bednarek, S. Y., Wilkins, T. A., Dombrowski, J. E., & Raikhel, N. V. (1990) *Plant Cell 2*, 1145−1155.

Bol, J. F., Linthorst, H. J. M., & Cornelisson, B. J. C. (1990) *Annu. Rev. Phytopathol. 28*, 113−138.

Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D. States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187−217.

Brünger, A. T. (1992) *X-PLOR Manual Version 3.1*, Yale University, New Haven, CT.

Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4130−4134.

Chrispeels, M. J. (1991) *Annu. Rev. Plant Physiol. Plant Mol. Biol. 42*, 21−53.

Chrispeels, M. J., & Raikhel, N. V. (1992) *Cell 68*, 613−616.

Clore, G. M., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986) *J. Mol. Biol. 191*, 523−551.

Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., & Wright, P. E. (1992) *J. Mol. Biol. 226*, 795−817.

Greenblatt, H. M., Ryan, C. A., & James, M. N. G. (1989) *J. Mol. Biol. 205*, 201−228.

Hart, P. J., Monzingo, A. F., Ready, M. P., Ernst, S. R., & Robertus, J. D. (1993) *J. Mol. Biol. 229*, 189−193.

Heath, R. L., Barton, P. A., Simpson, R. J., Reid, G. E., Lim, G., & Anderson, M. A. (1995) *Eur. J. Biochem. 230*, 250−257.

Hyberts, S. G., Goldberg, M. S., Havel, T. F., & Wagner, G. (1992) *Protein Sci. 1*, 736−751.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst R. R. (1979) *J. Chem. Phys. 71*, 4546−4553.

Jiménez, M. A., Blanco, F. J., Rico, M., Santoro, J., Herranz, J., & Nieto, J. L. (1992) *Eur. J. Biochem. 207*, 39−49.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577−2637.

Kirsh, T., Paris, N., Butler, J. M., Beevers, L., & Rogers, J. C. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3403−3407.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Kuntz, I. D., Kosen, P. A., & Craig, E. C. (1991) *J. Am. Chem. Soc. 113*, 1406−1408.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.

Marcusson, E. G., Horazdovsky, B. F., Ceregfino, J. L., Gharakhanian, E., & Emr, S. D. (1994) *Cell 77*, 579−586.

McLeish, M. J., Nielsen, K. J., Najbar, L. V., Wade, J. D., Lin, F., Doughty, M. B., & Craik., D. J. (1994) *Biochemistry 33*, 11174−11183.

Merutka, G., Dyson, H. J., & Wright, P. E. (1995) *J. Biolmol. NMR 5*, 14−24.

Nakamura, K., & Matsuoka, K. (1993) *Plant Physiol. 101*, 1−5.

Neuhaus, J.-M., Sticher, L., Meins, F., & Boller, T. (1991) *Proc. Natl. Acad. Sci. USA 88*, 10362−10366.

Nielsen, K. J., Heath, R. L., Anderson, M. A., & Craik, D. J. (1994) *J. Mol. Biol. 242*, 231−242.

Nielsen, K. J., Heath, R. L., Anderson, M. A., & Craik, D. J. (1995) *Biochemistry 34*, 14304−14311.

Nilges, M., Gronenborn, A. M., Brünger, A. T., & Clore, G. M. (1988) *Protein Eng. 2*, 27−38.

Pallaghy, P. K., Duggan, B. M., Pennington., M. W., & Norton, R. S. (1993) *J. Mol. Biol. 234*, 405−420.

Paravicini, G., Horazdovsky, B. F., & Emr, S. D. (1992) *Mol. Biol. Cell 3*, 415−427.

Pastore, A., & Saudek, V. (1990) *J. Magn. Reson. 90*, 165−176.

Piotto, M., Saudek, V., & Sklenar, V. (1992) *J. Biomol. NMR 2*, 661−665.

Ptitsyn, O. B., & Finkelstein, A. V. (1983) *Biopolymers 22*, 15−25.

Raikhel, N. V., & Wilkins, T. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 6745−6749.

Rance, M., Sørenson, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 177*, 479−495.
Richardson, M. (1991) *Methods Plant Biochem. 5*, 259−305.
Rizo, J., Blanco, F. J., Kobe, B., Bruch, B., Bruch, M. D., & Gierasch, L. M. (1993) *Biochemistry 32*, 4881−4894.
Ryan, C. A. (1990) *Annu. Rev. Phytopathol. 28*, 425−429.
Schnölzer, M., Alewood, P., Jones, A., Alewood, D., & Kent, S. B. H. (1992) *Int. J. Peptide Protein Res. 40*, 180−193.
Schroeder, M. R., Dombrowski, J. E., Bednarek, S. Y., Borkfisenious, O. N., & Raikhel, N. V. (1993) *J. Exp. Bot. 44* (Suppl.), 315−319.
Sönnichsen, F. D., van Eyk, J. E., Hodges, R. S., & Sykes, B. D. (1992) *Biochemistry 31*, 8790−8798.
Verner, R., & Schatz, G. (1988) *Science 241*, 1307−1313.
Wagner, G., Pardi, A., & Wüthrich, K. (1983) *J. Am. Chem. Soc. 105*, 5948−5949.
Weaver, L. H., & Matthews, B. W. (1987) *J. Mol. Biol. 193*, 189−199.
Wilkins, T. A., & Raikhel, N. V. (1989) *Plant Cell 1*, 541− 549.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991) *J. Mol. Biol. 222*, 311−333.
Wright, C. S., Schroeder, M. R., Raikhel, N. V. (1993) *J. Mol. Biol. 233*, 322−324.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949−971.
Zhou, N. E., Zhu, B.-Y., Sykes, B. D., & Hodges, R. S. (1992) *J. Am. Chem. Soc. 114*, 4320−4326.

BI952228I