Crystal Structure of Cardosin A, a Glycosylated and Arg-Gly-Asp-containing Aspartic Proteinase from the Flowers of *Cynara cardunculus* L.*

(Received for publication, April 15, 1999)

Carlos Frazão‡, Isabel Bento‡, Júlia Costa‡, Cláudio M. Soares‡, Paula Veríssimo§, Carlos Faro§, Euclides Pires§, Jon Cooper¶, and Maria A. Carrondo**

From the ‡Instituto de Tecnologia Química e Biomédica, Universidade Nova de Lisboa, Apartado 127, 2780-Oeiras, the ‡Departamento de Bioquímica, Universidade de Coimbra, Apartado 3126, 3000-Coimbra, Portugal, and the ¶Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Southampton SO16 7PX, United Kingdom

Aspartic proteinases (AP) have been widely studied within the living world, but so far no plant AP have been structurally characterized. The refined cardosin A crystallographic structure includes two molecules, built up by two glycosylated peptide chains (31 and 15 kDa each). The fold of cardosin A is typical within the AP family. The glycosyl content is described by 19 sugar rings attached to Asn-67 and Asn-257. They are localized on the molecular surface away from the conserved active site and show a new glycan of the plant complex type. A hydrogen bond between Gln-126 and Man4 renders the monosaccharide oxygen O-2 sterically inaccessible to accept a xylosyl residue, therefore explaining the new type of the identified plant glycan. The Arg-Gly-Asp sequence, which has been shown to be involved in recognition of a putative cardosin A receptor, was found in a loop between two β-strands on the molecular surface opposite the active site cleft. Based on the crystal structure, a possible mechanism whereby cardosin A might be orientated at the cell surface of the style to interact with its putative receptor from pollen is proposed. The biological implications of these findings are also discussed.

Aspartic proteinases (AP) are a class of enzymes (EC 3.4.23) involved in a number of physiological and pathological processes such as blood pressure homeostasis (renin), retroviral infection (human immunodeficiency virus proteinase), hemoglobin degradation in malaria (plasmepsin), intracellular proteolysis (cathepsin D), and digestion (pepsin) (see Ref. 1 for a recent review). Additionally, AP play an important role in the food industry, e.g., the cheese industry (chymosin) or in soya and cocoa processing. Inhibitors of AP enzymes have significant therapeutic potential for treatment of hypertension, AIDS, tumor invasiveness, and peptic ulcer disease. Despite their distribution in the living world, occurring from retrovirus to mammals, aspartic proteinases share significant similarities in primary and tertiary structures. Members of this class display two Asp-Thr/Ser-Gly motifs within their sequences and are specifically inhibited by pepstatin, a peptide produced by *Streptomyces*. Several high resolution x-ray structures of mammalian, fungal, and retroviral aspartic proteinases are available. The overall three-dimensional structure consists of two domains of similar secondary structure, dominated by orthogonally packed sheets with several small helical segments. In eukarotic aspartic proteinases each domain contributes one of the two catalytic aspartate residues to form the active site center located at a long and deep cleft between the two domains. Conversely, in retroviral aspartic proteinases, which are dimeric proteins consisting of two identical subunits, each subunit contributes one catalytic aspartate. It is thought therefore that eukarotic aspartic proteinases have evolved divergently from a primitive dimeric enzyme resembling retroviral proteinases by gene duplication and fusion.

Plant AP have been detected, extracted, and characterized from seeds, leaves, and flowers of a broad variety of plant species, being probably involved in specific physiological roles (2). They are also found in the digestive pitcher fluid of insecitrophic plants (3). Although known plant AP sequences show similarities to those of other AP, most of them contain in their cDNAs an insertion coding for a polypeptide segment of about 100 residues, known as PSI (plant-specific insert). The PSI bears no similarities to animal or microbial AP but shows significant sequence similarities with cDNA saposin precursor sequences (4). This saposin-like domain has been shown to be removed entirely or partially during maturation and processing of plant aspartic proteinase precursors (5, 6), rendering two-chain mature enzymes with a domain organization similar to that of the mammalian or microbial AP.

Cardosins are AP from the flowers of *Cynara cardunculus* L. (7, 8), whose milk-clotting activity has been exploited in Portugal in the manufacture of traditional cheeses since the Roman era. The molecular and enzymatic properties of cardosin A, the most abundant within cardosins, have been studied in detail (7, 9, 10). The enzyme is accumulated in protein storage vacuoles of the stigmatic papillae, the female receptive organ, being also present, although less abundantly, in vacuoles of the epidermic cells of the style (11). A unique feature of cardosin A among plant aspartic proteinases is the presence of an Arg-Gly-Asp (RGD) sequence, a well known cell attachment motif characteristic of integrin-binding proteins (12). This sequence, which has also been identified in the cardosin A-like proteinase from the stigmas of *Cynara humilis*, is apparently involved in...
the interaction between the proteinase and a 100-kDa protein from pollen.2 Together with its putative receptor, cardosin A is thought to participate in adhesion-mediated proteolytic mechanisms that operate in pollen-pistil interaction.

The present paper presents the mature cardosin A crystallographic structure with extensive description of its glycans and localization of the putative RGD adhesion site.

**EXPERIMENTAL PROCEDURES**

**Purification, Crystallization, and Data Collection—**Cardosin A was isolated and purified according to Verissimo et al. (13). Crystals of the glycosylated enzyme were obtained by the vapor diffusion method using PEG 4k as precipitating agent and were optimized by macro seeding (14). They belong to the monoclinic space group C2, with cell dimensions shown in Table I, and contain two molecules in the asymmetric unit. Diffraction data from three different crystals were successively obtained. The first crystal did not diffract beyond 2.85 Å resolution at room temperature and was prone to radiation damage when using synchrotron radiation at DESY, EMBL outstation in Hamburg, Germany. The second crystal measured with in-house equipment, led to similar data quality if under cryogenic conditions. Finally, the third crystal, under cryogenic conditions and with synchrotron radiation, diffracted to 1.72 Å (14). The first data set was used for structure determination and for the starting refinement cycles. The late stages of refinement were calculated with data composed by the highest resolution set completed with data from the in-house collection, in particular for the strongest, low resolution intensities (see Table I for final data statistics). The diffraction images were integrated with DENZO, and intensities were scaled and merged with SCALEPACK (15).

**Structure Solution and Refinement—**The structure of cardosin A was solved by AMoRE (16) using the structure of human cathepsin D (17) as search model. The molecular replacement solution found for the two molecules, using 15–3.5 Å data, showed a correlation coefficient 47% (33% for the first noise peak) and an R factor of 45.5% (57% for the first noise pick). In early stages of refinement a target sequence was used, composed partially of cardosin A sequence (that had not been totally determined by then) and completed with the cyprosin sequence (18). A model based on the cathepsin D structure was constructed, where alanines replaced all residues that differed between the target and the cathepsin D sequences. These were mutated to the expected sequence at a later stage, along with the refinement progress. Refinement was initially carried out with data to 2.85 Å and X-PLOR, using the simulated annealing/slow cooling protocol with strict non-crystallographic symmetry (19). Electron density maps with sigma A coefficients (20) were examined using TURBO (21), and the model was gradually completed and corrected, alternating with X-PLOR refinement cycles. When the complete cardosin A sequence was obtained and the final diffraction data to a resolution of 1.72 Å was available, the refinement proceeded using SHELXL with restrained (1–4 distances) non-crystallographic symmetry positional refinement and restrained non-crystallographic symmetry atomic displacement factor refinement (22). Solvent molecules were gradually introduced; some side chains were modeled with alternative conformations; crystal anisotropic correction was applied, and riding model hydrogen atoms were also refined (see Table I for final refinement statistics). Coordinates have been deposited in the Brookhaven Protein Data Bank under accession code 1bsf.

**Energy Minimization of the Substrate Model in the Active Site—**The structure of molecule A of cardosin without sugar residues or water molecules, with the exception of the catalytic water, was used in the energy minimization studies. The peptide Leu-Ser-Phe-Met-Ala-Leu was built in the active site with capped terminal ends (an acetyl group for the N-terminal and an N-methyl for the C-terminal) to avoid end charge effects. Hydrogen atoms were positioned using GROMOS (23), and their position was optimized using 1000 steps of steepest descents energy minimization. The initially docked structure was energy-minimized for 500 steps of steepest descents method while keeping all cardosin atoms fixed by the use of positional restraints. Finally, the system was subjected to a further 10,000 cycles of steepest descents energy minimization. Positional restraints were used for C-α atoms outside of the active site zone and adjacent flaps to ensure the preservation of the main fold of the protein under the pseudo-vacuum conditions. These calculations were performed using the GROMOS force field with polar and aromatic hydrogen atoms (23, 24) and, since solvent was not included, with a modified version of the program PROEM implementing the distance-dependent dielectric function of Mehler and Solmajer (25).

**RESULTS AND DISCUSSION**

**Crystal Structure—**In order to compare cardosin A with other known AP structures, the pepsin sequence numbering (26) was used to name the cardosin A residues throughout this paper.

The refined cardosin A crystallographic structure includes two independent, glycosylated, aspartic protease molecules, composed of two (30 and 15 kDa) peptide chains, in a total of 649 amino acids in the asymmetric unit (a.u.). The glycosyl content is described by 19 sugar rings attached to the protein moieties. The sugars are localized on the molecular surface, distributed between two glycosylation sites in each of the molecules, i.e. one site per polypeptide chain. Whereas the N termini of both molecules are not visible in the electron density maps, the two C termini could be fully modeled. The cardosin A sequence begins with the hydrophilic segment (5) Asp–2–Ser–1, Gly4, and it is therefore conceivable that the N terminus may be disordered in the solvent. In particular, molecule 1 was modeled beginning on Gly5 and molecule 2 from Ser1 onward. For the C terminus, on the contrary, one finds a mainly hydro-

---

2 C. Faro, M. Ramalho-Santos, M. Vieira, A. Mendes, R. Andrade, P. Verissimo, X.-L. Lin, J. Tang, and E. Pires, submitted for publication.
phobic sequence, Phe\textsuperscript{322}, Ala\textsuperscript{323}, Glu\textsuperscript{324}, Ala\textsuperscript{325}, and Ala\textsuperscript{326}. This sequence, essentially conserved among plant AP that also contain the PSI, has been proposed by Faro and co-workers (5) to play a role in vacuolar signaling, as it is known to be sufficient for the vacuolar targeting of barley lectin (27). It is noteworthy that this C-terminal sequence is well defined on both cardosin A molecules, with an average temperature factor (B) of 17.9 \text A\textsuperscript{2}. Data from primary sequence analysis (5) have shown that the PSI excision occurs between residues 238 and 240, with a small heterogeneity occurring at the N terminus of the 15-kDa chain. The electron density allows positioning of Asn\textsuperscript{238}, the C terminus of the first chain, but fades away afterward and appears back only by Glu\textsuperscript{243}, which is in accordance with the absence of PSI in mature cardosin A. This exci-

![Figure 1](image-url)

**Figure 1.** A, cartoon representation (49, 50) of the two cardosin A molecules in the a.u. They face each other through an extensive area, although the actual molecule to molecule contacts are relatively few. The two N-linked glycans are represented as ball-and-sticks with side chains of linking Asn\textsuperscript{67} and Asn\textsuperscript{257}. The active site aspartate side chains, as well as those from a putative molecular adhesion RGD motif (12) (Arg\textsuperscript{176}, Gly\textsuperscript{177}, and Asp\textsuperscript{178}) are also depicted as ball-and-stick representation. The missing PSI domain is indicated near its chain termini. B, accessible surface representation (51) of the contact regions between the two cardosin molecules in the a.u. Molecule 1 (left) and 2 (right) facing surfaces are represented after a 180\degree rotation around a vertical axis of one of the molecules. The contacts between the two molecules produce a decrease of the local solvent-accessible area represented in blue with the rest of the surface in white. The contacts are highly delocalized over the intermolecular surfaces and are spread over a wide region.
The AP structure is located where AP structures usually have a solvent-exposed loop, at the end of a conserved anti-parallel \(\beta\)-sheet, covering the crest of the C domain, at one side of the active site canyon. Solvent molecules, in a total of 528, were modeled as waters, and a total of 18 residues (including two of the sugar units) were modeled using two alternate conformations.

As cardosin A is essentially formed by the duplication of a motif of four anti-parallel \(\beta\)-strands and a helix, which is repeated twice in each domain, a characteristic of the AP family (28, 29), the Ramachandran plot shows a clustering on the \(\beta\) domain (45.3% of \(\beta\)-strands and 13.5% of helical motifs).

A stereochemical check of the molecules with PROCHECK (30) indicated the presence of 89.9% amino acids in the most favored region, 9.5% in the additional allowed region, 0.2% in the generously allowed region, and finally 0.4% (2 amino acids) in the disallowed region. This last pair corresponds to Leu295 residues of both molecules, modeled with very clear electron density maps, and located on the tip of a type IV \(\beta\)-turn, lying on the molecular surface but with relevant hydrophobic interactions with inner protein moiety. The distribution of isotropic displacement parameters is reasonable with an average \(B\) of 23.0 Å\(^2\) for the buried atoms and with higher values where no secondary structure was assigned. There is a clear correlation between high \(B\) factors and those residues that do not superimpose upon fitting the two molecules. They show equivalent folds, deviating significantly from one another (more than twice the overall main chain r.m.s. of 0.474 Å) only for residues 45–48 and 253–254 which are involved in crystal packing contacts, or on the poorly defined loops (see below). The final maps are in general very clear, as expected from 1.72-Å resolution data. However, there were three solvent-exposed loop regions that could not be modeled using the electron density maps.

---

**TABLE II**

Distance matrix among AP

|                  | Cardosin A (1b5f) | Cathepsin D (1lyb) | Yeast protease A (2jxr) | Chymosin B (4cms) | Mouse renin (1smr) | Cod pepsin (1am5) | Plasmepsin II (1smr) | Rhizomucor miehei (2asi) |
|------------------|-------------------|-------------------|------------------------|------------------|-------------------|-------------------|----------------------|-------------------------|
| Cardosin A A      | 46.3              | 45.5              | 45.2                    | 43.6             | 40.0              | 33.2              | 27.3                 |
| Human cathepsin D| 77.0              | 45.6              | 47.0                    | 49.4             | 48.9              | 34.9              | 24.6                 |
| Yeast protease A  | 78.8              | 78.6              | 37.2                    | 39.7             | 40.0              | 32.6              | 28.1                 |
| Bovine chymosin B| 79.4              | 75.5              | 98.8                    | 93.5             | 41.2              | 34.1              | 28.3                 |
| Mouse renin       | 83.0              | 70.5              | 92.4                    | 67.1             | 88.7              | 29.0              | 32.2                 |
| Cod pepsin        | 91.6              | 71.5              | 91.6                    | 126.7            | 107.7             | 123.7             | 20.5                 |
| Plasmepsin II     | 110.2             | 105.3             | 112.1                   | 126.7            | 126.4             | 113.5             | 158.5                |
| **Rhizomucor miehei** | 129.3             | 140.2             | 126.9                   | 128.7            |                   |                   |                      |

---

**FIG. 3.** Sub-site specificity mapping of cardosin A. Schematic representation (21) of a \(\kappa\)-casein chain fragment, with peptide scissile bond Phe\(^{105}\)–Met\(^{106}\), where milk clotting is initiated in cheese production. Cardosin A residues within 4.0 Å of the docked \(\kappa\)-casein fragment are listed and grouped at their sub-sites (\(S_n\) and \(S_{n+}\)).
Aspartic Proteinase from C. cardunculus

alone. They were at the end of the “flap,” between residues 75 and 79, and the variable loops 46 to 47 and 159 to 160.

Three disulfide bridges were found in the cardosin A mature protein, two within the first peptide chain (Cys<sup>45</sup>/Cys<sup>56</sup> and Cys<sup>206</sup>/Cys<sup>210</sup>) and the third within the second peptide chain (Cys<sup>249</sup>/Cys<sup>282</sup>), at positions known to form inter-cysteinyl bonds in the AP family (31). These three covalent bridges do not link the two peptide chains, which are therefore held together only by hydrophobic interactions and hydrogen bonds arising with the AP fold.

One cis-peptide bond was found between Thr<sup>22</sup> and Pro<sup>23</sup>. This cis-Pro is a conserved feature of the AP family, on the tip of a conserved VIb β-turn.

**Dimer Arrangement and Crystal Packing**—The two cardosin A molecules in the a.u. (Fig. 1A) are related by a pseudo-crystallographic 2-fold axis (32) and present an intriguing intermolecular surface. The inter-molecular contacts require only 4.4% of each monomer’s accessible surface area, but they are spread over a substantial area (Fig. 1B). A total of 38 atomic contacts between the two molecules are closer than 4.0 Å, including 3 salt bridges. Only two water molecules were found bridging the two molecules. For comparison purposes, the crystal packing contacts between pairs of molecules involve in some cases a larger number of interactions (up to 68), but they are spread over a significantly smaller accessible surface.

**Overall Three-dimensional Comparison of Cardosin A and Other AP**—The secondary structure consists essentially of β-strands and some helical motifs and follows the pepsin-like single chain folding topology (33). The molecule is bilobal with two domains separated by a large cleft where the active site is located. As cardosin A is the first AP from the plant kingdom whose structure has been determined, a run of COMPARER (34), using a representative set of AP crystallographic structures, was performed to compare in detail the cardosin A model against other members of the AP family. The three-dimensional comparison also took into account local environmental variables such as types of hydrogen bonding, amino acid solvent accessibility, Ramachandran parameters, cis-peptide, and disulfide bonds. The comparison was carried out against the models of vacuolar yeast proteinase A (35), of lysosomal human cathepsin D (17), of the milk-clotting secreted enzymes bovine chymosin (36), and the AP from Rhizomucor miehei (37), of the cod fish pepsin, of the subglandulary mouse renin (38), and of the malaria parasitic protozoan plaspepsin II (39). All models were retrieved from Protein Data Bank (40). As a result, a distance matrix (34) was obtained (Table II), and a phyletic dendrogram was generated (Fig. 2), where cardosin A clusters together with the other vacuolar AP from yeast. The two secreted enzymes chymosin and the AP from R. miehei, which like cardosin A are used in milk clotting, are more distantly related.

**Substrate Binding Cleft and Specificity Sub-site Mapping**—The identification of the residues involved in substrate specificity for individual AP has been pursued in order to understand the specificity determinants for each new member of the family and is an important source of motivation in site-directed mutation studies. It is the base for rationalization of the enzyme activity versus selectivity relationships. The physiological cardosin A substrate has not yet been definitely established, but the most important human application of cardosins is manufacture of exquisite cheese. In order to identify the residues involved in substrate binding, it was necessary to obtain a model of cardosin A-substrate transition state complex as no structure of cardosin A complexed to a peptide inhibitor has yet been obtained. However, the substrate-binding pockets can be identified by analogy with other AP, for which structures of enzyme-inhibitor complexes have been determined. The structure of an inhibitor complexed with renin (41) was fitted to the cardosin A coordinates. This was the first guide to model a fragment of κ-casein, corresponding to residues 102–108, containing the specific Phe<sup>105</sup>–Met<sup>106</sup> bond that is cleaved in the proteolysis of milk micelle proteins in the production of cheese (9). The docked structure was energy-minimized keeping cardosin Cα-atoms outside the active site zone and adjacent flap restrained to their initial positions to ensure the preservation of the main fold of the protein under the pseudo-vacuum conditions of the minimization. The optimization of the enzyme-substrate complex led to only minor changes in the structure of the free enzyme (data not shown), an indication of the specificity of the enzyme toward this particular substrate sequence. Residues on each of the specificity sub-sites were defined as those having atoms within 4.0 Å of residues flanking the scissile Phe–Met bond (Fig. 3), corresponding to 127 atomic contacts including five putative hydrogen bonds. These involve almost exclusively main chain atoms, with the exception for the cardosin Thr<sup>218</sup> OG1, κ-casein Phe<sup>105</sup> N contact.

**Active Site**—In common with the other AP, the active site of cardosin A is located between the two lobes of the molecule at the bottom of a large cleft. The base of the active site cleft is made of β-strands forming the typical, two abuttingψ-like structures that contain the two catalytic aspartates (Asp<sup>32</sup> and Asp<sup>215</sup>). The active site is one of the most conserved regions among the AP family and has been used to screen among new protein sequences for potential AP enzymes. In plant AP the DTG triad of the N-domain contributes to the active site as in other known eukaryotic AP, but the C-domain triad is mutated into DSG. This mutation, however, does not disturb the usual “fireman’s grip” three-dimensional hydrogen bond network arrangement surrounding the active site. The side chains of the aspartates are held coplanar and within hydrogen bonding involving main chain and conserved side chain groups. A water molecule is bound to both aspartate carboxyls by hydrogen bonds. This water molecule has been implicated in catalysis since it may become partially displaced upon substrate binding and polarized by one of the aspartate carboxyls (42). The water may then nucleophilically attack the peptidic scissile bond to form a tetrahedral intermediate, which is bound non-covalently to the enzyme. It has been proposed that the tetrahedral inter-
mediate is stabilized by hydrogen bonds to the negatively charged carboxyl of aspartate 32. Fission of the scissile main chain C—N bond is accompanied by transfer of a proton to the leaving amino group either from Asp$^{219}$ or from bulk solvent.

PSI—It has been proposed that the PSI domain may be involved in the association of plant aspartic proteinase precursors to the membrane during its intracellular transport and may possibly contribute to the targeting to the vacuole (4, 5). PSI sequence alignments from available plant AP, and their comparison with other protein sequences, revealed a striking resemblance with prosaposin sequences (4, 43). There is a particular match through the regions involving the C-half of a saposin, followed by a short inter-saposin linker and the N-half of the next saposin, within the tandem sequence of saposin precursors. This alignment involves the formation of three disulfide bridges and a consensual glycosylation site. However, in PSI the connection between the now circularly permuted two halves of saposin modules is established by a 20–30-residue long linker, which does not have a homologous counterpart in prosaposin sequences. A secondary structure prediction of this linker using simultaneously several available protocols (44) did not lead to a definitive consensual fold (data not shown). The rest of the PSI three-dimensional domain should resemble the saposin fold (45), tightly held together through their three inter-helical disulfide bridges. This domain is attached to the protease moiety by two extended linking segments, protruding from the crest of the C-lobe at one side the enzymatic canyon and is susceptible to protease attack. Processing of procardosin A has been shown to occur at these sites (5). Cleavage seems to occur first between the 31-kDa chain and PSI and afterward at the border of PSI and the 15-kDa chain.

Glycosylation—The glycan primary structures of the two glycosylation sites of cardosin A have been found to be of the plant complex type (10). In the first glycosylation site at Asn$^{67}$ five or six monosaccharide residues are visible in the electron density maps of molecules 1 and 2, respectively (Fig. 4, A and B). The N-linked carbohydrate is held tightly to the polypeptide backbone of the parent protein molecule due to extensive van der Waals contacts and five (or six for molecule 2) hydrogen bonds to main and side chain residues of the polypeptide. In the case of molecule 2, additional interactions (three hydrogen bonds) were found due to crystal packing contacts (Fig. 4B). From the primary sequence determination of the oligosaccharide, it has been found that although both glycosylation sites contain complex type glycans, the first is rather unusual in that it does not bear Xyl$\beta$1,2 linked to the branching Man$\beta$4. The three-dimensional structure shows that a hydrogen bond between Glu$^{136}$ (ND2) of the parent protein and that Man$\beta$4 (O-2) renders the monosaccharides oxygen sterically inaccessible to accept a xylosyl residue, transferred by xylosyltransferase. Several lines of evidence have shown that the transfer of xylose to the plant N-linked oligosaccharides occurs in the Golgi before the transfer of Fuc$\alpha$1,3 to the inner core GlcNAc (46). The configuration of this new complex glycan suggests that the plant xylosyltransferase does not require the presence of xylose in the acceptor motif.

Cathepsin D and yeast proteinase A are also glycosylated at Asn$^{67}$ (Fig. 4A), and the conserved presence of a well defined glycosylation site in a similar region of these three aspartic proteinases might indicate a common biological role. It is possible that they protect these hydroxides against accelerated proteolytic cleavage and therefore play a role in the stability of the enzymes, similar to what was found for recombinant renin expressed in COS cells (47) or for Mucor aspartic proteinase expressed in yeast (48).

In the second glycosylation site, Asn$^{257}$, only four monosaccharide rings were detected (Fig. 4C). The N-linked GlcNAc contacts only Asn$^{257}$ and Thr$^{270}$, with no hydrogen bonds with the parent molecule, which explains why fewer sugar rings were detected in the electron density maps. However, the two molecules in the crystal interact through a hydrogen bond via Fuc$\alpha$3(0-3) of molecule 1 and a symmetry mate of Fuc$\alpha$3(0-5) of molecule 2 (Fig. 4C). These stabilize the glycan conformation, allowing the visualization of four monosaccharide residues, in contrast to what happens for the second glycosylation site of cathepsin D (Asn$^{183}$) and yeast proteinase A (Asn$^{266}$), where only the single, inner core GlcNAc was observed. Due to the localization of the oligosaccharide on the protein surface, away from the active site cleft of cardosin A molecule, they should not have any effect on the activity or specificity of the enzyme.

Conclusions—Cardosin A is an aspartic proteinase found to accumulate to high levels in protein storage vacuoles of the stigmatic papillae of C. cardunculus L. and has been suggested to be involved in pollen-pistil interaction. It is synthesized as a single chain precursor but, by removal of the plant-specific insert domain (PSI), is converted into the mature active two-chain protease. The excised domain, whose structure has been predicted to be saposin-like, should be located above the active site, opposite the “flap” that covers that proteolytic site. The two glycans, of the plant complex type, attached to the polypeptide backbone are extensively visible in the present crystal structure. The unusual absence of a xylosyl residue in one of the glycans is explained by steric hindrance, due to hydrogen bonding between the branching Man$\beta$4 and a peptide side chain. The glycans are likely to be important for the stability and/or correct processing rather than for activity, in view of their localization away from the active site. The unique feature of cardosin A among plant aspartic proteinases is, however, the presence of the RGD cell-attachment motif. This sequence is located at the base of the molecule, opposite to the active site, and projects itself out of the molecular surface, thus explaining why it can be recognized by the 100-kDa protein from pollen, previously identified as a putative cardosin A receptor. The interaction between these two proteins is apparently mediated by the RGD sequence. As a possible mechanism, cardosin A might be orientated at the cell surface so that its RGD sequence could be recognized by the receptor from pollen. In this mechanism cardosin A is transported along the secretory pathway associated with the membrane via the PSI domain of its precursor. Following fusion of the transport vesicles with the plasma membrane, the RGD sequence is facing outward and in this way can be easily recognized by the 100-kDa RGD-binding protein. In the first stage, before removal of the saposin-like domain, the proteinase would be anchored to the membrane and would then be released upon processing at the PSI cleavage sites. The reported findings also suggest that cardosin A from the papillar pool may be stored in the vacuole in a dimeric/oligomeric form due to its high concentration. Whether dimerization is required for activation remains to be investigated, but it is likely to be an important determinant in the regulation of the enzyme activity and specificity.

Acknowledgment—We thank Dr. H. A. Nagarajaram for helpful discussions on the COMPARER analysis and PSI structure prediction studies.

REFERENCES
1. James, M. N. G. (ed) (1998) Aspartic proteinases, Plenum Press, New York
2. Kervinen, J., Tormakangas, K., Runenberg-Roos, P., Guruprasad, K., Blundell, T. L., and Teeri, T. H. (1995) Adv. Exp. Med. Biol. 362, 241–254
3. Tokei, Z. A., Woon, W. C., and Chambers, S. M. (1974) Planta 119, 39–46
4. Guruprasad, K., Tormakangas, K., Kervinen, J., and Blundell, T. L. (1994) FEBS Lett. 352, 131–136
5. Ramalho-Santos, M., Verissimo, P., Cortes, L., Samyn, B., van Beeumen, J., Pires, E., and Faro, C. (1998) Eur. J. Biochem. 255, 133–138
