

# Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting

Jukka Kervinen<sup>1,2</sup>, Gregory J. Tobin<sup>3</sup>,  
Júlia Costa<sup>4</sup>, David S. Waugh<sup>5</sup>,  
Alexander Wlodawer<sup>1</sup> and  
Alexander Zdanov<sup>1,6</sup>

<sup>1</sup>Protein Structure Section and <sup>5</sup>Protein Engineering Group, Macromolecular Structure Laboratory, ABL-Basic Research Program and <sup>3</sup>Laboratory of Cell and Molecular Structure- SAIC, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702, USA and <sup>4</sup>Instituto de Biologia Experimental e Tecnológica/Instituto de Tecnologia Química e Biológica, 2780 Oeiras, Portugal

<sup>2</sup>Present address: Fox Chase Cancer Center, Institute for Cancer Research, 7701 Burholme Avenue, Philadelphia, PA 19111, USA

<sup>6</sup>Corresponding author  
e-mail: zdanov@ncifcrf.gov

**We determined at 2.3 Å resolution the crystal structure of prophytepsin, a zymogen of a barley vacuolar aspartic proteinase. In addition to the classical pepsin-like bilobal main body of phytepsin, we also traced most of the propeptide, as well as an independent plant-specific domain, never before described in structural terms. The structure revealed that, in addition to the propeptide, 13 N-terminal residues of the mature phytepsin are essential for inactivation of the enzyme. Comparison of the plant-specific domain with NK-lysin indicates that these two saposin-like structures are closely related, suggesting that all saposins and saposin-like domains share a common topology. Structural analysis of prophytepsin led to the identification of a putative membrane receptor-binding site involved in Golgi-mediated transport to vacuoles.**

**Keywords:** aspartic proteinases/phytepsin/saposin-like domain/transport to vacuoles/zymogen structure

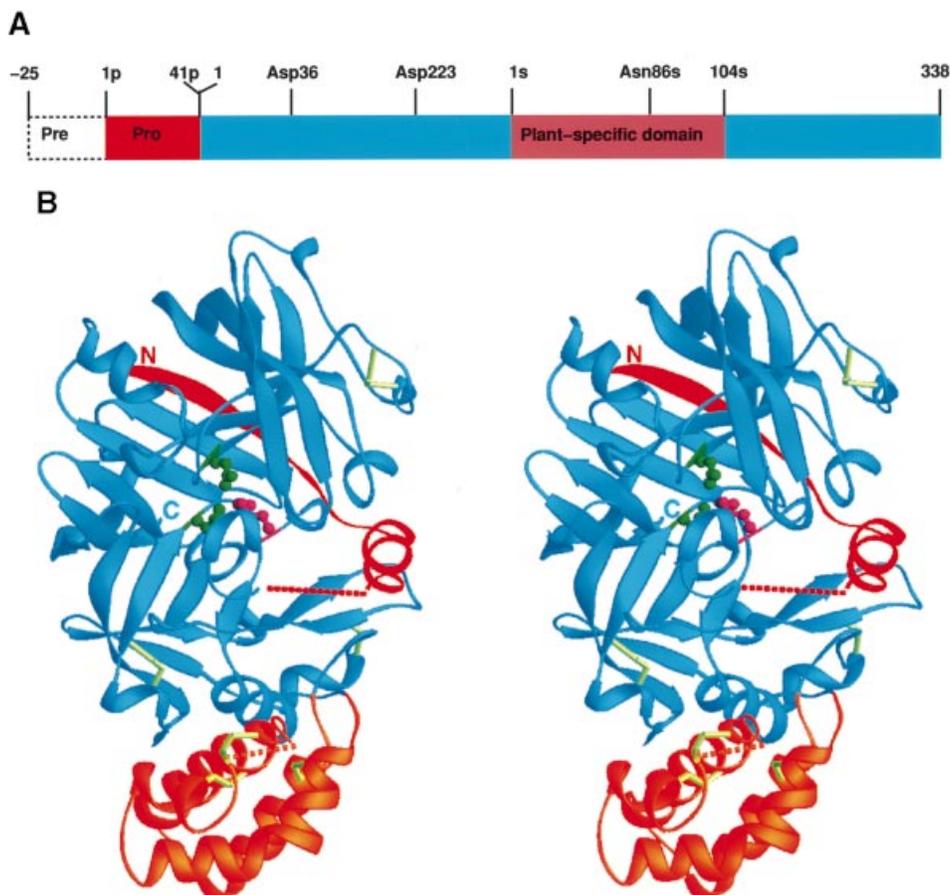
## Introduction

Aspartic proteinases (APs) (EC 3.4.23), found in animals, plants, fungi, yeast, some bacteria and viruses, constitute one of the four distinct superfamilies of proteolytic enzymes (Davies, 1990; Rawlings and Barrett, 1995). These well known and extensively characterized enzymes include pepsins, chymosin, renin, cathepsin D and human immunodeficiency virus type 1 (HIV-1) proteinase, among others. Common features of all APs are an active site cleft that divides two  $\beta$ -barrel domains and contains two catalytic aspartates (numbered 32 and 215 in pepsin), acidic pH optima for enzymatic activity, inhibition by pepstatin and preferential cleavage specificity for peptide bonds between bulky hydrophobic side chains. However, substantial differences exist in the catalytic properties, the cellular localization and, consequently, in their biological

function (Davies, 1990; Rawlings and Barrett, 1995; Blundell, 1998).

All non-viral APs are synthesized as inactive precursors (zymogens), in which the N-terminal propeptide is bound to the active site cleft, thus preventing undesirable degradation during intracellular transport and secretion (Dunn, 1997; Khan and James, 1998). In addition to their inhibitory function, prosegments play an important role in the correct folding, stability and intracellular sorting of many zymogens (Koelsch *et al.*, 1994). Whereas the structures of the active APs from many animals and microbes are known (Rawlings and Barrett, 1995; Blundell, 1998), the only zymogen structures that have been determined thus far are for porcine (James and Sielecki, 1986; Sielecki *et al.*, 1991) and human pepsinogens (Bateman *et al.*, 1998), human progastricsin (Moore *et al.*, 1995; Khan *et al.*, 1997) and proplasmepsin II from the human malaria parasite *Plasmodium falciparum* (Bernstein *et al.*, 1999). It has been shown that the prosegment coils around the mature enzyme and either extends into the active site cleft (Khan and James, 1998), where it provides a steric block for the approaching substrates, or bumps into the C-terminal domain thereby destroying the active site (Bernstein *et al.*, 1999). The zymogen structure is stabilized by several hydrogen bonds and salt bridges between the positively charged propeptide and the negatively charged mature enzyme regions, including interactions between Lys36p (pepsin numbering) of the propeptide and the active site aspartates.

Phytepsin (EC 3.4.23.40, previously abbreviated to HvAP) resides in barley grains, roots, stems, leaves and flowers (Tormakangas *et al.*, 1994). Based on its primary structure (Runeberg-Roos *et al.*, 1991), hydrolytic specificity (Kervinen *et al.*, 1993), sensitivity to AP inhibitors (Sarkkinen *et al.*, 1992) and vacuolar localization (Runeberg-Roos *et al.*, 1994; Marttila *et al.*, 1995), phytepsin is considered to be a plant homolog of mammalian lysosomal cathepsin D and yeast vacuolar proteinase A. Although the exact function of phytepsin and other plant APs is still unclear, phytepsin may participate in metabolic turnover and in protein processing events occurring in barley tissues (Runeberg-Roos *et al.*, 1994; Tormakangas *et al.*, 1994). In addition, phytepsin is highly expressed in several plant tissues undergoing apoptosis (Runeberg-Roos and Saarma, 1998). The sequence similarity of plant APs to their animal counterparts (Runeberg-Roos *et al.*, 1991; Cordeiro *et al.*, 1994; Hiraiwa *et al.*, 1997), as well as the comparative modeling of phytepsin (Guruprasad *et al.*, 1994), suggests that plant APs also possess a topology typical for all APs. However, plant APs from diverse sources (Runeberg-Roos *et al.*, 1991; Cordeiro *et al.*, 1994; Verissimo *et al.*, 1996; Hiraiwa *et al.*, 1997) contain an internal region consisting of ~100 residues not present in animal or microbial APs. This region is thus



**Fig. 1.** Structure of prophytepsin. The propeptide is shown in red, the mature protein is cyan and the plant-specific domain is orange. **(A)** Schematic representation of the open reading frame of pre-prophytepsin. Residue numbering and approximate locations of catalytic residues Asp36/223 and a glycosylation site (Asn86s) are indicated. **(B)** Stereo ribbon representation of the crystal structure of prophytepsin. Catalytic Asp residues are green, Lys11 is pink, and six disulfide bridges are yellow. Disordered parts of the propeptide and plant-specific domain are marked by dashed lines. (Figures 1B, 2A and 4A were prepared with the program RIBBONS; Carson, 1991.)

called a plant-specific insert. The evolutionary or biological importance of this unique region is obscure. The sequence, however, is highly similar to that of saposins, which are lysosomal sphingolipid-activating proteins in mammalian cells, as well as that of saposin-like proteins such as NK-lysin and sulfated glycoprotein-1, and saposin-like areas in acyloxyacyl hydrolase and acid sphingomyelinase (Munford *et al.*, 1995; Liepinsh *et al.*, 1997).

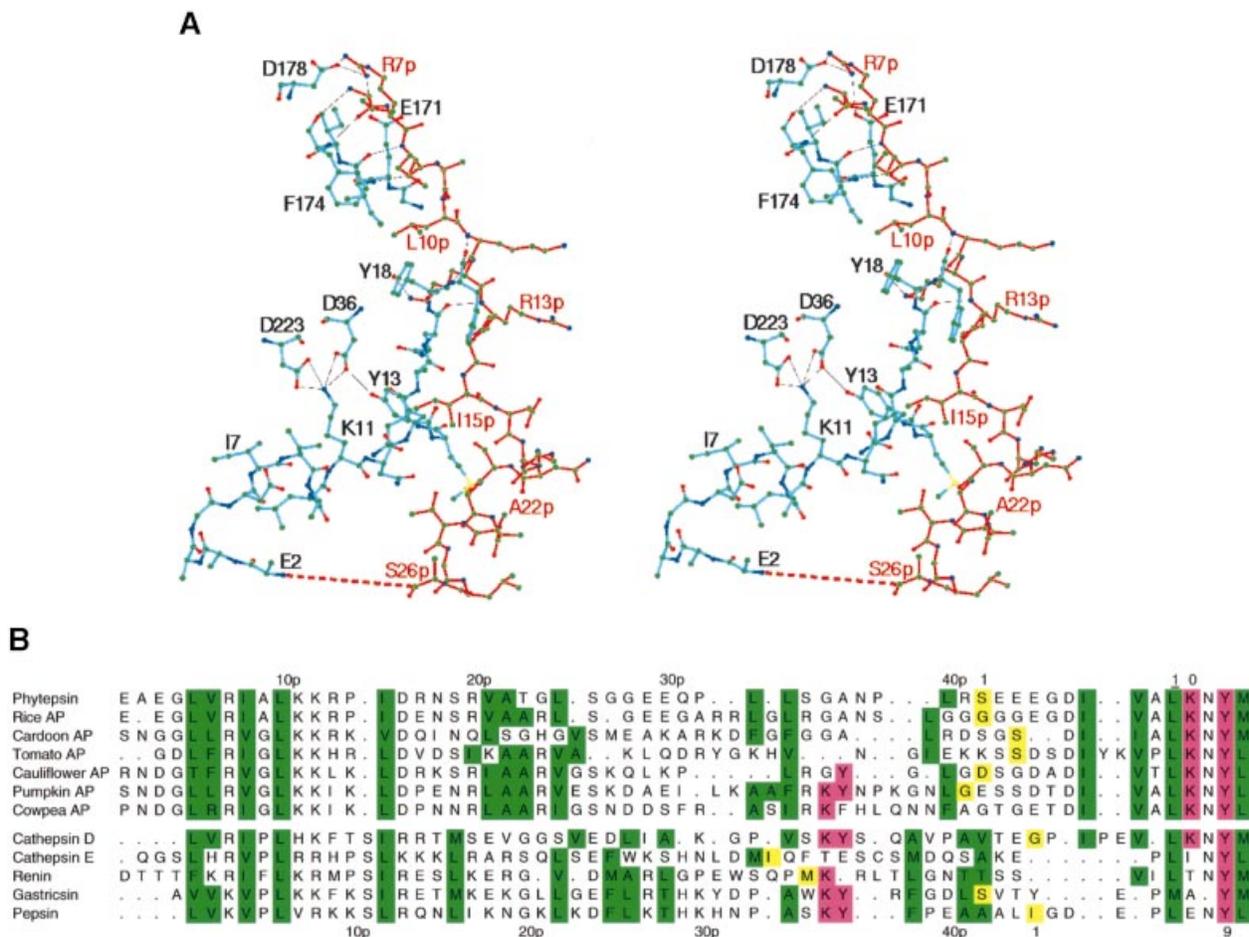
Plant APs have been difficult to crystallize because, when purified from original sources, the preparations usually contain several enzyme forms or processing intermediates (Sarkkinen *et al.*, 1992; Verissimo *et al.*, 1996). We recently developed a method to express the zymogen of barley vacuolar AP, prophytepsin, in baculovirus-infected insect cells and were able to obtain a correctly folded and highly homogeneous protein preparation for crystallographic and other studies (Glathe *et al.*, 1998). Here, we describe the first zymogen structure of an intracellular AP that also includes the saposin-like domain. The putative role of this domain in vacuolar targeting of the proenzyme is also discussed.

## Results and discussion

### Overall structure

The primary translation product of phytepsin in baculovirus-infected insect cells consists of the pre-proenzyme

(Figure 1A). The endoplasmic reticulum (ER) signal sequence (pre-part) was removed, presumably by endogenous enzyme(s) present in the cells. The resulting single-chain polypeptide corresponding to 53 kDa glycosylated prophytepsin was purified (Glathe *et al.*, 1998) and crystallized. The fold of prophytepsin (Figure 1B) can be divided into three main elements: a propeptide consisting of 41 amino acid residues; 338 residues corresponding to a two-domain mature protein; and an independent plant-specific domain of 104 residues, which is inserted into the C-terminal domain of the enzyme. The mature phytepsin is formed by two polypeptide chains, 1–247 and 248–338, and its fold is very typical for other APs. It consists of two topologically similar  $\beta$ -barrel-like domains with the active site aspartic residues Asp36 and Asp223 located in the interdomain cleft. A six-stranded  $\beta$ -sheet covers the hydrophobic core of the molecule at the bottom of the cleft, with three conserved disulfide bridges stabilizing the structure (Figure 1B). The propeptide wraps around the interdomain cleft in such a way that its N-terminal  $\beta$ -sheet is involved in the formation of the six-stranded  $\beta$ -sheet, while the helical part of the propeptide approaches the active site from the opposite side and partially covers it. We were able to trace only residues 6p–26p of the propeptide, out of 41 residues present in its sequence. The rest of the propeptide and the first residue of the mature



**Fig. 2.** (A) Stereo diagram of the interactions of the propeptide with the mature enzyme moiety. The propeptide is red, the mature protein is cyan, carbon atoms are green, nitrogens are blue, oxygens are red and sulfurs are yellow. (B) Sequence alignment of propeptides of plant and animal APs. Numbering for prophytepsin is given at the top of the sequences and that for pig pepsinogen below. Dots indicate gaps with no homology in the sequences. The putative or known N-terminus of the mature enzyme is yellow, the conserved hydrophobic regions are green and conserved Lys/Tyr residues are pink. The DDBJ/EMBL/GenBank accession Nos of the sequences are as follows: phylypsin, X56136; rice AP, D32144; cardoon AP, X81984; tomato AP, L46681; cauliflower AP, X80067; pumpkin AP, AB002695; cowpea AP, U61396; human cathepsin D, M63138; human cathepsin E, M84424; human renin, L00064; human gastricisin, U75272; pig pepsin, J04601. All names refer to the active enzyme forms.

protein were not visible in the electron density map, indicating their disorder. Interestingly, the relative orientation of the helical part of the propeptide differs between the three crystallographically independent molecules, even if we take into account the interdomain shifts (discussed below). However, the side chain interactions between the propeptide and the rest of the protein remain similar.

The plant-specific insertion, 1s–104s, forms an independent domain that is attached to the C-terminal domain of prophytepsin by two connecting strands (245–3s and 103s–253). Structurally, the plant-specific domain comprises five amphipathic  $\alpha$ -helices, which are folded into a single compact entity and are linked with each other by three disulfide bridges. The sole glycosylation site in prophytepsin (Costa *et al.*, 1997), Asn86s, is located on a short solvent-exposed loop between helices four and five of the plant-specific domain. Although we observed some shapeless, bulky electron density in the vicinity of Asn86s in molecules I and II, unambiguous interpretation of this density was not possible. Therefore, we did not model any oligosaccharide and did not include it in either the crystallographic refinement or in the final model. It was also not possible to model a 27 residue polypeptide

(38s–64s) located between helices three and four of the plant-specific domain, since no electron density corresponding to it was present.

#### **Analysis of the propeptide and a mechanism of inactivation**

As discussed above, the propeptide wraps around the mature phylypsin and, together with the N-terminal strand and the flap of the mature enzyme, sterically blocks the active site, leading to its total inactivation (Figure 1B). The N-terminal strand 6p–12p is involved in the formation of a six-stranded  $\beta$ -sheet, making several main chain hydrogen bond interactions with the neighboring strand 169–174, while the side chain of Arg7p forms a salt bridge with Glu171 and Asp178 (Figure 2A). The hydrophobic side chains of Val6p, Ile8p, Leu10p and Ile15p, as well as the hydrophobic part of the side chains of Lys11p, Lys12p and Arg13p, are involved in the formation of an internal hydrophobic core. After a short loop (13p–15p), the propeptide forms an amphipathic helix (16p–26p) (Figure 2A), which is oriented antiparallel to the helix 114–120 of the mature enzyme, and interacts with it through its hydrophobic side. The rest of the propeptide

and the first residue of the mature protein (Gly27p–Arg41p and Ser1) are disordered. The propeptide and the visible part of the N-terminus of the mature protein are anchored in the active site cleft by ionic interaction of Lys11 with the catalytic aspartates 36 and 223. In addition, the hydroxyl group of Tyr13 forms a hydrogen bond to Asp36, while its aromatic ring locates in the S1 hydrophobic pocket of the enzyme; main chain 9–13 makes hydrogen bonds to the flap, and Ile7 is positioned in the S1' binding pocket. Taken together, the propeptide, the 13 residues of the N-terminus of the mature enzyme and the  $\beta$ -hairpin containing residues 75–88 and known as the 'flap' cover the active site completely. Such a position and conformation of the first 54 residues of the prophytepsin relative to the rest of the protein are locked by two ionic interactions, a number of hydrogen bonds and by hydrophobic contacts, including binding into the S1 and S1' binding pockets.

Sequence alignment of the prosegments from several plant and animal APs (Figure 2B) shows some diversity in their length and considerable variations in the residue types. In general, it is clear that there is some similarity among the first 10 residues of the propeptide, as well as among the first 14 residues of the mature enzyme. However, the similarity within each group of the proteins is much higher, especially among the animal proenzymes. In addition, zymogens of most APs from higher organisms, such as pepsinogen, progastricsin, lysosomal procathepsin D and yeast proteinase A (not shown), contain the invariant residues Lys36p–Tyr37p in a prosequence and a conserved Tyr9 in the mature enzyme region (pepsin numbering). In pepsinogen (James and Sielecki, 1986; Sielecki *et al.*, 1991; Bateman *et al.*, 1998) and progastricsin (Moore *et al.*, 1995), these residues interact directly with the catalytic aspartates and thus stabilize the inactive zymogen structure in a manner similar to Ile7...Lys11...Tyr13 of prophytepsin. In other words, the role of an internal anchor to active site aspartates, assumed in animal APs by Lys36p of the propeptide, is played by Lys11 belonging to the mature protein part of prophytepsin. Based on the sequence homology (Figure 2B), it is very likely that this is the case for plant APs in general, although two sequences of isoenzymes of rice (DDBJ/EMBL/GenBank accession No. D12777) and cardoon (Verissimo *et al.*, 1996) APs were found which do not have a Lys11 homolog. In the few examples where this key lysine is present both in the propeptide region and in the mature protein, such as in pumpkin AP and cathepsin D, it is conceivable that both residues could play analogous roles, depending on particular circumstances. It has been shown, for example, that at higher or neutral pH, the mature N-terminus of cathepsin D moves back to the active site cleft and blocks it in a manner similar to that observed in the prophytepsin (Lee *et al.*, 1998). In addition, Lys11 of prophytepsin makes ionic interactions with the active site aspartates very similar to those of the analogous residue of cathepsin D (Figure 1B).

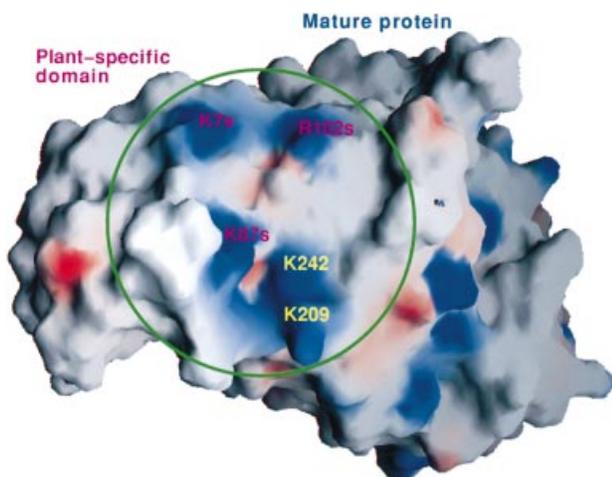
### **The plant-specific domain is homologous to the saposin-like protein family**

The plant-specific domain forms an independent subunit in the prophytepsin structure (Figure 1B). The domain comprises five amphipathic  $\alpha$ -helices forming a helical bundle or 'helical cage', where the internal surface is

covered exclusively with hydrophobic side chains. The five helices (H1–H5) include residues 4s–10s, 14s–22s, 27s–35s, 65s–85s and 89s–103s, respectively, and thus the domain displays 59%  $\alpha$ -helical secondary structure. The helices are oriented in an up–down–up–down manner (if we consider helices H1 and H2 which are separated only by a kink as one helix), and interact with each other mainly through hydrophobic contacts and three interhelical disulfide bridges, Cys6s–Cys100s, Cys31s–Cys72s and Cys37s–Cys69s. The plant-specific domain is attached to the main body of prophytepsin by two flexible polypeptides, which permit its slightly different orientation ( $\sim 1.5$ – $2.0^\circ$ ) relative to the main body of the protein in the three crystallographically independent molecules. Nevertheless, the interface between the plant-specific domain and the mature protein is preserved. Although the outer surface of the domain is mainly hydrophilic, it also contains a hydrophobic patch contacting the mature protein. This patch is a continuation of the internal hydrophobic core and is formed by the side chains of Met74s, Val77s, Trp78s, Leu99s and Leu103s. On the side of the mature enzyme, it interacts with Phe213, Ala234 and the hydrophobic parts of Thr233, Thr237, Glu238, Glu241 and Lys242. These side chains form a local hydrophobic core in the interface between the plant-specific domain and the mature phytepsin, surrounded by hydrogen bonds and ionic interactions.

We found that six out of nine positively charged residues of the plant-specific domain proved either to be completely disordered or to have very flexible side chains. However, it is very unlikely that any of these six residues form positively charged clusters involving more than two residues on the surface of the domain, since Lys28s–Lys29s are too far from Arg64s. Additionally, although we do not know where Arg59s is located, some weak, scattered electron density indicates that Arg43s and Arg50s are located on the loop, which is extended away from the helical bundle of the plant-specific domain. Instead, Lys7s, Lys87s and Arg102s of the plant-specific domain along with the Lys209 and Lys242 of the mature enzyme clearly constitute a positively charged ring framing the hydrophobic region, in which the side chains of Leu91s and Tyr95s, as well as of Leu99s (which is a part of the interface hydrophobic core), play a major role. This region might be a targeting receptor-binding site (see below). It is composed of residues belonging both to the plant-specific domain and to mature phytepsin and is located on the outer surface of the V-shaped interdomain interface (Figure 3).

The plant-specific domain of phytepsin is the first member of the family of saposin-like proteins for which the three-dimensional structure has been determined by X-ray crystallography (Figure 4A). However, the structure of porcine NK-lysin has been elucidated recently by NMR spectroscopy (Liepinsh *et al.*, 1997). NK-lysin is a basic polypeptide consisting of 78 amino acid residues with an antibacterial activity and the capability of lysing tumor cell lines (Andersson *et al.*, 1995). Comparison of the saposin-like domain of prophytepsin with NK-lysin shows remarkable similarity between these proteins (Figure 4B), including the approximate length and position of the helices, disulfide bridges, and a kink between helices H1 and H2 in phytepsin and H3 and H4 in NK-lysin. As seen



**Fig. 3.** Putative membrane/receptor-binding site. The site is shown inside the green circle. Positively charged regions are blue, and negatively charged regions are red. This figure was created with the program GRASP (Nicholls, 1992).

in the structures (Figure 4B) and in the sequence alignment (Figure 4C), the N- and C-terminal parts of the saposin-like domain are swapped in prophytepsin. Hence, as it has been pointed out previously (Munford *et al.*, 1995; Ponting and Russell, 1995), the plant-specific insert is equivalent to interchanged N- and C-terminal domains in the saposin-like family of proteins. However, the swapping does not impact either the orientation or the direction of the helices, since the peptides linking the domain to the main body of phytepsin are spatially close to each other, whereas there is a short loop between helices H2 and H3 in NK-lysin. Conversely, there is a loop between helices H3 and H4 in the structure of prophytepsin instead of the N- and C-termini in NK-lysin. Structural data and sequence homology between different members of the family of saposin-like proteins strongly suggest that saposins, saposin-like proteins and saposin-like structural domains in other proteins all share a closely related, disulfide-bonded compact structure with a five-helix topology.

#### **Packing of the prophytepsin molecules in the crystal unit cell**

There are three molecules (I–III) in the asymmetric unit. Molecules I and II form layers that are orthogonal to the crystallographic 2-fold screw axis, whereas molecule III is located in between the layers. Molecules I and II, as well as II and III, are related to each other by pseudo 2-fold rotation axes (178.1 and 176.7°, respectively), whereas molecules I and III are related by a rotation of 132.8°. The conformation of all three molecules is similar for the majority of the residues, particularly in the areas corresponding to the mature enzyme. However, the relative positions of the propeptide helices and of the plant-specific domains are quite different. Notably, at least some of these variations may be attributed to different relative orientations of the N- and C-terminal domains corresponding to the mature phytepsin. The relative movement of these domains has been reported previously for other APs (Sielecki *et al.*, 1991; Sali *et al.*, 1992; Lee *et al.*, 1998; Bernstein *et al.*, 1999). The root-mean-square (r.m.s.) deviation for the superposition of C $\alpha$  atoms of molecules

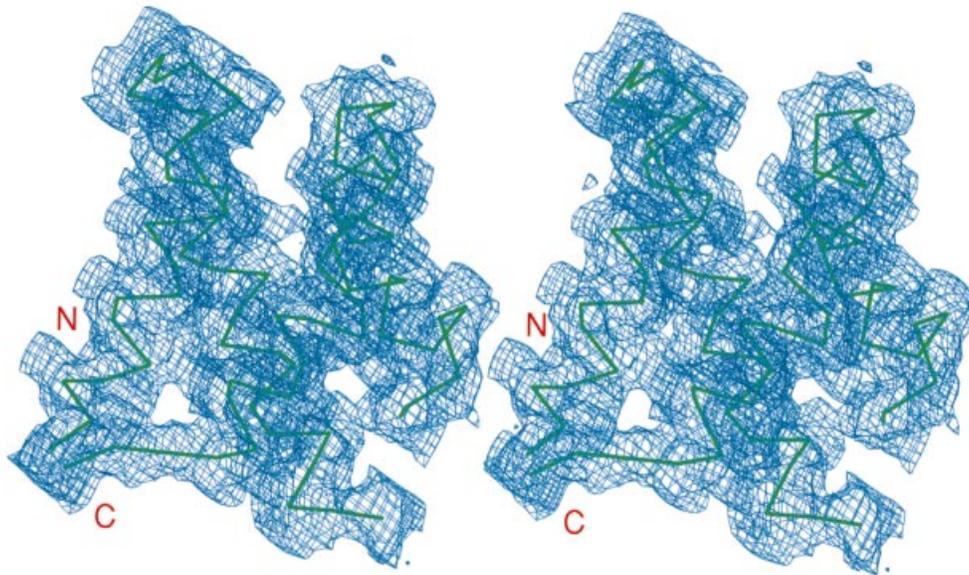
I and II, I and III, and II and III are 1.24, 1.15 and 0.74 Å, respectively. However, when the propeptides and plant-specific domains were excluded, and the C- and N-terminal domains including the central motif (Sali *et al.*, 1992) were superimposed pairwise, these deviations became 0.4–0.6 Å. To achieve such good superposition, one has to rotate and translate the C- and N-terminal domains independently. We found that such shifts between the domains are 0.4° and 1.6 Å for molecules I and II, 3.7° and 0.2 Å for molecules I and III, and 1.1° and 0.5 Å for molecules II and III.

#### **Unique structural features of inactivation/activation of plant APs**

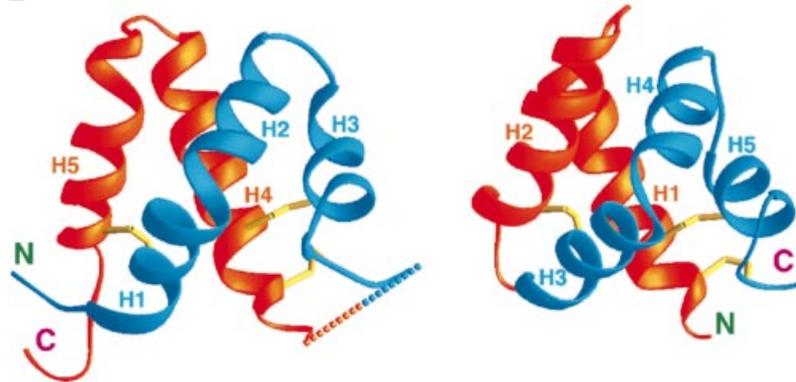
There are two quite different modes of inactivation of the mature APs in their zymogens. In the mammalian proenzymes, the active site is pre-formed (the distance between Asp32 and Asp215 is ~3 Å) and the propeptide covers the active site. This conformation is anchored by the ionic interactions of Lys36p (pepsinogen numbering) with the active site aspartates, together with Tyr9 and Tyr37p which occupy the S1 and S1' binding pockets, while their hydroxyl groups make hydrogen bonds to Asp32 and Asp215, respectively (Sielecki *et al.*, 1991; Moore *et al.*, 1995). The N-terminal 13 residues of the mature enzyme participate in covering the active site, keeping the flap wide open. A distinct second mode was found in the structure of proplasmepsin II from *P.falciparum* (Bernstein *et al.*, 1999). The position of the propeptide is such that it bumps into the C-terminal domain, initiating a relative rotation of the domains of ~14°, separating the active site aspartates by >6 Å and thus distorting the active site. In prophytepsin, the propeptide only partially blocks the pre-formed active site, with additional steric hindrance provided by the N-terminal 13 residues of the mature enzyme and by the flap, which assumes a completely closed conformation.

Although no structure of the mature phytepsin has been reported, a high resolution structure of another mature plant AP, cardosin A (Bento *et al.*, 1998), is now available (PDB code 1b5f). Because the amino acid sequence identity between phytepsin and cardosin exceeds 70%, their structures must be very similar, and we assume that the structure of cardosin A can be used as a surrogate of mature phytepsin. Figure 5 shows the superposition of the mature parts of prophytepsin and cardosin, with an r.m.s. deviation of 0.63 Å. We conclude that only minor changes occur upon maturation of plant APs, and the flaps assume almost the same positions both in the mature enzymes and in the zymogens. As previously mentioned, the propeptide and the N-terminus of the mature part are anchored by the interaction of Lys11 with the active site aspartates, whereas Tyr13 and Ile7 occupy the S1 and S1' binding sites. In addition, the hydroxyl groups of Tyr13 and Ser39 form hydrogen bonds to the catalytic Asp36. Tyr80, which is an analog of Tyr75 in pepsinogen (Sielecki *et al.*, 1991; Moore *et al.*, 1995), forms a hydrogen bond to Ne1 of Trp43 (Trp39 in pepsin numbering). This bond is a common feature of all mature APs, whereas sometimes it is absent in the proenzymes (Moore *et al.*, 1995). Thus, a unique feature of the proenzymes of plant APs is that not only are the active site aspartic acids in the active conformation, but also the rest of the enzyme appears

**A**



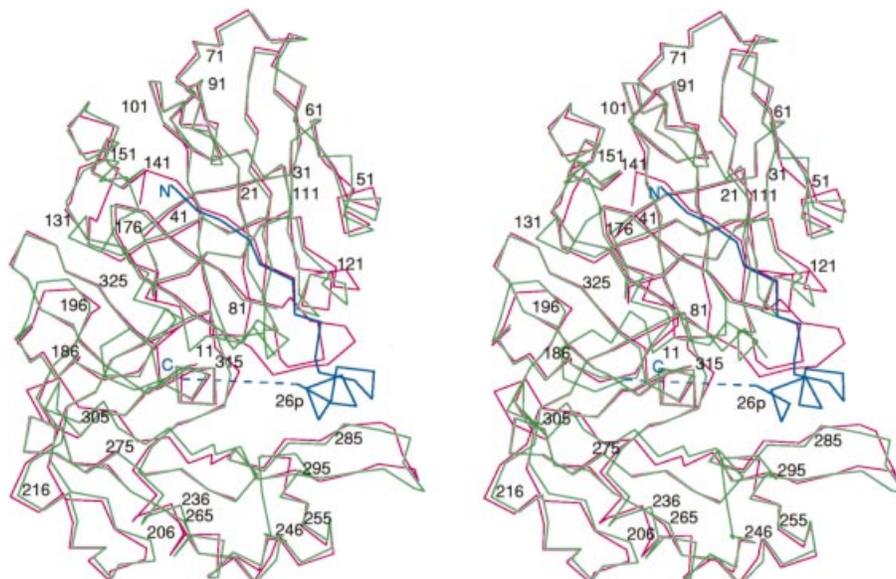
**B**



**C**

	helix 1		helix 2		helix 3					
	1 s	10 s	20 s	30 s	40 s	50 s				
Phyatepsin	V V S Q E C K T	V S Q Y G Q Q I L D	L L L A E T Q	P K K I C S Q V G L C T F	D G T R G V S A G	I R S V V				
Cardoon AP	V M S Q Q C K T	L V S Q Y G K T M I E	M L L S E A Q	P D K I C S Q M K L C T F	D G A R D A S S I	I E S V V				
Tomato AP	I V S M E C K T	I V S Q Y G E M I W D	L L V S G I R	P D Q V C S Q A G L C F L	D G S Q H V S S N	I R T V V				
Pumpkin AP	V V S Q Q C K A	V V A Q Y G Q T I M D	L L L S E A D	P K K I C S Q I N L C T F	D G T R G V S M G	I E S V V				
Saposin C	S L S E E C Q E	V V D T Y G S S	I L S I L L E E V S	P E L V C S M L H L C	. . S G T	. . . . .				
AOAH	F L K T T C Y L V I	D K F G S D I I K L L S A	D M N A D V	V C H T L E F C K Q N	. . . . .	. . . . .				
NK-lysin	I L R G L C K K I	M R S F L R R I	S W D I L T G K K	P Q A I C V D I K I C K E	. . . . .	. . . . .				
	4 0	I	5 0	6 0	7 8					
					II	III				
	helix 4		helix 5							
	6 0 s		7 0 s		8 0 s		9 0 s		10 0 s	
Phyatepsin	D D E P V K S N G L R A	D P M C S A C E M A V V W M	Q N Q L A Q . .	N K T Q D L I L D Y V N Q	L C N R L P					
Cardoon AP	D E N N G K S S S G V	H D E M C T F C E M A V V W M	Q N Q I K R . .	N E T E D N I I N Y V N E	L C D R L P					
Tomato AP	E R E T E G S S V G E A	P L C T A C E M A V V W M	Q N Q L K Q . .	E Q T K E K V L E Y V N Q L	L C E K I P					
Pumpkin AP	D E N A G K S S D S L	H D G M C S V C E M T V V W M	Q N Q L R Q . .	N Q T K E R I I N Y N E	L C D R M P					
Saposin C	. . . . .	S D V Y C E V C E F L V K E	V T K L I D N . .	N K T E K E I L D A F D K	M C S K L P K					
AOAH	. . . . .	N G H T C V G C V L V V S V I	. E Q L A Q V H	N S T . . . V Q A S M E R	L C S Y L P					
NK-lysin	. . . . .	G Y F C S C R K I Q K L	. E D M V G P	Q P N E D T V T Q A A S Q V	C D K L K					
		1	III	II	1 0	2 0	3 0	I		

**Fig. 4.** (A) Electron density map ( $2F_o - F_c$ ) for the plant-specific domain, contoured at the  $1.0\sigma$  level. (B) Comparison of saposin-like motifs of prophytepsin and NK-lysin. The plant-specific domain of prophytepsin is on the left (N-terminal domain, cyan; C-terminal domain, orange). NK-lysin is on the right (N-terminal domain, orange; C-terminal domain, cyan). (C) Sequence alignment of the plant-specific domains with human saposin C (DDBJ EMBL/GenBank, accession No. M32221), human acylglycerol hydrolase (M62840) and NK-lysin (Andersson *et al.*, 1995). DDBJ/EMBL/GenBank accession Nos for plant AP sequences are given in the legend to Figure 2B. Numbering for prophytepsin is given at the top of the sequences and that for NK-lysin is below. The conserved hydrophobic regions are green, conserved Pro and Phe/Tyr residues are pink, the N-glycosylation site is blue and conserved Cys residues are yellow. The order of disulfide connections between the Cys residues is marked by roman numerals. The positions of  $\alpha$ -helices are denoted for the prophytepsin structure. (Figures 4A and 5 were created with the program CHAIN; Sack, 1988.)



**Fig. 5.** Superposition of the  $C_{\alpha}$  tracing of the mature part of prophytepsin and cardosin A. Prophytepsin is shown in green with the propeptide in blue (the dashed line shows the disordered part of it), and cardosin is in violet.

to require only minor conformational changes for its activation.

From the studies of gastric APs, it is known that the proenzyme remains inactive at neutral pH. Upon secretion into the acidic environment in the stomach, the low pH initiates a disruption of the salt bridges connecting the propeptide to the mature enzyme, and the subsequent conformational changes lead to autocatalytic proteolysis of the prosegment. Removal of the propeptide is followed by conformational rearrangement of ~10 residues of the N-terminus that form the sixth strand in the  $\beta$ -barrel structure of the molecule (Dunn, 1997; Khan and James, 1998). Prophytepsin is capable of pH-regulated autocatalytic processing which results in an active enzyme residing in vacuoles (Runeberg-Roos *et al.*, 1994; Glathe *et al.*, 1998). Based on the structure of prophytepsin, it seems likely that conformational changes for the processing and activation follow a similar scheme. At low pH, the active site residues Asp36 and Asp223, as well as Glu171 and Asp178, become protonated and the ionic interactions disappear. Both the propeptide and the future N-terminus of the mature enzyme can now dissociate from the active site cleft, thus allowing autolysis to take place. However, in contrast to gastric APs, which only need autocatalytic steps for the activation process (Dunn, 1997; Khan and James, 1998), phytepsin apparently also needs additional vacuolar proteases for the completion of the maturation (Glathe *et al.*, 1998).

It has been shown recently (Lee *et al.*, 1998) that cathepsin D undergoes major structural rearrangements between the acidic and neutral pH. Although the network of interactions in the active site of the high pH cathepsin D is very similar to its counterpart in prophytepsin, they are not identical. For instance, in order to accommodate the inactive conformation of the N-terminus, cathepsin D requires an opening of the flap up to 8.0 Å at high pH (Lee *et al.*, 1998), while maximum deviation of only 2.6 Å along the active site cleft occurs at the tip of the flap (Gly81) in the structure of prophytepsin. By analogy, it is

possible that the N-terminus of the mature phytepsin undergoes similar conformational changes, depending on the prevailing pH in vacuoles. However, this suggested inactivation mechanism needs further investigation, since the exact pH for the conformational change is not known, and thus it remains to be elucidated if the normal pH fluctuation in vacuoles is capable of triggering the conformational changes.

#### **Role of saposin-like domain in vacuolar targeting**

Proteins residing in vacuoles are intracellularly transported to the *trans*-Golgi network and, analogous to animal cells, plant proteins also seem to interact with membrane-associated receptor proteins during the Golgi-mediated intracellular transport to the vacuoles (Neuhaus and Rogers, 1998; Sanderfoot *et al.*, 1998). The prophytepsin structure presented here provides evidence that the plant-specific domain may have a role in the vacuolar targeting. The association of saposins and saposin-like proteins with membrane lipids and membrane-bound proteins is well demonstrated (Misasi *et al.*, 1998; Wilkening *et al.*, 1998; Vaccaro *et al.*, 1999). The prophytepsin structure also revealed a putative membrane-binding region in the plant-specific domain, which includes an adjacent area in the mature enzyme (Figure 3). Thus, it is likely that the saposin-like domain brings prophytepsin into contact with membranes and possibly also with a membrane-bound receptor in the Golgi apparatus. The resulting complex is then packed into vesicles and prophytepsin is carried into vacuoles. The strong binding of the saposins to certain glycolipids (Misasi *et al.*, 1998) also indicates a possible role for protein–lipid interactions in protein sorting to the vacuole in plant cells. The saposin-like domain would mediate accumulation of phytepsin in the *trans*-Golgi network membrane microdomains, which would be sorted into vesicles targeted to the vacuole. This sorting would follow a mechanism similar to that described in mammalian cells, where sphingolipid–cholesterol microdomains in membranes mediate the targeting of proteins to the

apical membrane of polarized cells (Harder and Simons, 1997). Upon arrival in the vacuole, prophytepsin is activated by proteolytic cleavage and the saposin-like domain is removed, breaking the interaction with the putative membrane receptor or the membrane itself. Further supporting this hypothesis, a recent study showed that during the final maturation steps of prophytepsin (Glathe *et al.*, 1998) and cardosin (Ramalho-Santos *et al.*, 1998), most of the plant-specific domain is cleaved, including a putative membrane-binding region. Thus, most probably, this domain is removed since it becomes unnecessary after the arrival of the proenzyme into vacuoles.

Some direct evidence also exists on the role of the plant-specific domain in vacuolar transport. A recent study carried out with tobacco protoplasts showed that deletion of the plant-specific domain led to the secretion of truncated phytepsin, whereas a specific mutation in the propeptide or deletion of the *N*-glycosylation site did not affect vacuolar targeting (Tormakangas, 1997). Further corroborating the hypothesis, tobacco chloroplasts have recently been shown to contain an AP with a fully conserved active site region, but the cDNA encoding the protein does not include a plant-specific domain or a proregion typical of plant APs (Nakano *et al.*, 1997).

The structure we present here reveals a mechanism for the self-inactivation of prophytepsin during its route to vacuoles. Based on the sequence homology, this mechanism seems to be common for all related plant APs. In addition, the saposin-like domain in plant APs may have an essential role in vacuolar targeting by bringing these enzymes into contact with membranes and, possibly, also with membrane-bound receptor proteins. However, it must be noted that plant cells contain at least two functionally distinct vacuolar compartments, and phytepsin has been shown to reside in both types of vacuoles (Paris *et al.*, 1996). Although the saposin-like domain seems to represent one kind of vacuolar targeting signal, other targeting signals may be needed for the correct sorting of prophytepsin and other plant APs to several kinds of vacuoles.

## Materials and methods

### Crystallization and data collection

Phytepsin, originally isolated (Sarkkinen *et al.*, 1992) and sequenced (Runeberg-Roos *et al.*, 1991) from barley (*Hordeum vulgare* L.) grains, was expressed in baculovirus-infected insect cells and purified in its 53 kDa proform (Glathe *et al.*, 1998). Crystallizations were carried out using the hanging drop vapor diffusion technique at room temperature by mixing prophytepsin (8 mg/ml in 20 mM Tris-HCl pH 8.0, 0.1 M NaCl) with an equal volume of the reservoir buffer containing 15% polyethylene glycol 6000, 50 mM magnesium acetate, 0.1 M sodium cacodylate, pH 6.7, 0.02% sodium azide. Although obtaining small crystals of prophytepsin was relatively easy, attempts to obtain large crystals with good diffracting properties proved to be very difficult. Regardless of the modifications of the crystallization procedures, crystals remained small or formed clusters or stacks of thin plates. However, on one occasion, two single crystals (~1.0×1.0×0.2 mm) were obtained. These plates were cleaved and the pieces were used for X-ray analyses. Crystals belong to the monoclinic system, space group P2<sub>1</sub>, unit cell parameters *a* = 66.0 Å, *b* = 160.9 Å, *c* = 81.4 Å,  $\beta$  = 109.6° with three molecules per asymmetric unit. A mother liquor containing 15% glycerol was used as a cryoprotectant. X-ray diffraction data were collected from one crystal at 100 K using synchrotron radiation (CHESS, beam station F2, ADSC 1K CCD detector). The data set consisted of 170 frames, each corresponding to 1.5° oscillation and exposed for 50 s. Diffraction intensities were processed with the HKL suite of programs (Otwinowski and Minor, 1997).

**Table I.** Structure refinement

Resolution range (Å)	10.0–2.3
Unique reflections [ $I \geq 3\sigma(I)$ ]	49 301
Completeness	69.5%
$R_{\text{merge}}$	5.9%
$R_{\text{cryst}}$	22.4%
No. of non-hydrogen atoms	9341
R.m.s. deviations from ideality:	
Bond lengths (Å)	0.012
Bond angles (°)	1.8
Dihedral angles (°)	26.7
Improper angles (°)	1.7

### Structure determination and refinement

The structure was solved by the molecular replacement technique with the AmoRe program package (Navaza, 1994), using coordinates of cathepsin D (PDB code 1lya) as a starting model. Initially, only two of the three independent molecules in the asymmetric unit were located (molecules I and II); nevertheless, a few cycles of refinement with the program X-PLOR (Brünger, 1992) were carried out with this partial model. The resulting structure was superimposed on porcine pepsinogen (Sielecki *et al.*, 1991) (PDB code 2psg) and human progastriecin (Moore *et al.*, 1995) (PDB code 1htr), and a new model, including only areas common to both proenzymes, was generated. A new run of AmoRe (with a new search model) gave a solution for molecules I and III. This solution was confirmed independently by electron density maps calculated on the basis of molecules I and II only. Bricogne's modification of Sim's weights, implemented in the program package PHASES (Furey and Swaminathan, 1995), was used to construct phase probability distributions, later followed by the DM procedure (Collaborative Computational Project, Number 4, 1994). The map that was calculated with the resulting phases clearly contained electron density not only for molecules I and II, which were included in the phase calculation, but also for molecule III, which was found in the second run of AmoRe as well. After all three molecules were properly placed, routine crystallographic refinement was initiated, and the parts of the structure that were absent in the initial model were added gradually.

The final model (Table I) includes residues 6p–26p of the propeptide (out of 1p–41p), and residues 2–164, 168–37s and 64s–338 (residues forming plant-specific domain are 1s–104s). All comparisons of the atomic coordinates were performed with the program ALIGN (Cohen, 1997). The coordinates and structure factors have been deposited in the Protein Data Bank with the PDB code 1QDM for immediate release.

## Acknowledgements

Kirsi Tormakangas, Pia Runeberg-Roos, Teemu H. Teeri and Mart Saarna (University of Helsinki, Finland) are warmly thanked for their advice in the initial steps of this study. We also thank Ms Grace H. Li for her skilled technical assistance and Anne Arthur for her editorial assistance. We acknowledge CHESS for allowing us access to the data collection facility. A.W. would like to thank the Master and Fellows of the Sidney Sussex College and the Department of Biochemistry, University of Cambridge, for a Visiting Fellowship. Research sponsored in part by the National Cancer Institute, DHHS, under contract with ABL. J.K. was also supported by the Academy of Finland. The contents of this publication do not necessarily reflect the views or policies of the Department of Human Services, nor does mention of trade names, commercial product or organizations imply endorsement by the US Government.

## References

- Andersson, M. *et al.* (1995) NK-lysin, a novel effector peptide of cytotoxic T and NK cells. Structure and cDNA cloning of the porcine form, induction by interleukin 2, antibacterial and antitumour activity. *EMBO J.*, **14**, 1615–1625.
- Bateman, K.S., Chernaia, M.M., Tarasova, N.I. and James, M.N.G. (1998) Crystal structure of human pepsinogen A. In James, M.N.G. (ed.), *The Aspartic Proteinases*. Plenum Press, New York, NY, pp. 259–263.

- Bento, I. *et al.* (1998) Crystallization, structure solution and initial refinement of plant cardosin-A. In James, M.N.G. (ed.), *The Aspartic Proteinases*. Plenum Press, New York, NY, pp. 445–452.
- Bernstein, N.K., Cherney, M.M., Loetscher, H., Ridley, R.G. and James, M.N.G. (1999) Crystal structure of the novel aspartic proteinase zymogen proplasmepsin II from *Plasmodium falciparum*. *Nature Struct. Biol.*, **6**, 32–37.
- Blundell, T.L. (1998) The aspartic proteinases—an historical overview. In James, M.N.G. (ed.), *The Aspartic Proteinases*. Plenum Press, New York, NY, pp. 1–13.
- Brünger, A. (1992) *X-PLOR Version 3.1: A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
- Carson, M. (1991) RIBBONS 4.0. *J. Appl. Crystallogr.*, **24**, 958–961.
- Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
- Cohen, G.E. (1997) ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J. Appl. Crystallogr.*, **30**, 1160–1161.
- Cordeiro, M.C., Xue, Z.-T., Pietrzak, M., Pais, M.S. and Brodelius, P.E. (1994) Isolation and characterization of a cDNA from flowers of *Cynara cardunculus* encoding cyprosin (an aspartic proteinase) and its use to study the organ-specific expression of cyprosin. *Plant Mol. Biol.*, **24**, 733–741.
- Costa, J., Esteves, C.L., Faro, C.J., Kervinen, J., Pires, E., Verissimo, P., Wlodawer, A. and Carrondo, M.A. (1997) The glycosylation of the aspartic proteinases from barley (*Hordeum vulgare* L.) and cardoon (*Cynara cardunculus* L.). *Eur. J. Biochem.*, **243**, 695–700.
- Davies, D.R. (1990) The structure and function of the aspartic proteinases. *Annu. Rev. Biophys. Chem.*, **19**, 189–215.
- Dunn, B. (1997) Splitting image. *Nature Struct. Biol.*, **4**, 969–972.
- Furey, W. and Swaminathan, S. (1995) Phases-95: a program package for the processing and analysis of diffraction data from macromolecules. *Methods Enzymol.*, **276**, 546–552.
- Glathe, S., Kervinen, J., Nimtz, M., Li, G.H., Tobin, G.J., Copeland, T.D., Ashford, D.A., Wlodawer, A. and Costa, J. (1998) Transport and activation of the vacuolar aspartic proteinase phytepsin in barley (*Hordeum vulgare* L.). *J. Biol. Chem.*, **273**, 31230–31236.
- Guruprasad, K., Tormakangas, K., Kervinen, J. and Blundell, T.L. (1994) Comparative modeling of barley-grain aspartic proteinase: a structural rationale for observed hydrolytic specificity. *FEBS Lett.*, **352**, 131–136.
- Harder, T. and Simons, K. (1997) Caveolae, DIGs and the dynamics of sphingolipid-cholesterol microdomains. *Curr. Opin. Cell Biol.*, **9**, 534–542.
- Hiraiwa, N., Kondo, M., Nishimura, M. and Hara-Nishimura, I. (1997) An aspartic endopeptidase is involved in the breakdown of propeptides of storage proteins in protein-storage vacuoles of plants. *Eur. J. Biochem.*, **246**, 133–141.
- James, M.N.G. and Sielecki, A.R. (1986) Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature*, **319**, 33–38.
- Kervinen, J., Sarkkinen, P., Kalkkinen, N., Mikola, L. and Saarma, M. (1993) Hydrolytic specificity of the barley grain aspartic proteinase. *Phytochemistry*, **32**, 799–803.
- Khan, A.R. and James, M.N.G. (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.*, **7**, 815–836.
- Khan, A.R., Cherney, M.M., Tarasova, N.I. and James, M.N.G. (1997) Structural characterization of activation 'intermediate 2' on the pathway to human gastricsin. *Nature Struct. Biol.*, **4**, 1010–1015.
- Koelsch, G., Mares, M., Metcalf, P. and Fusek, M. (1994) Multiple functions of pro-parts of aspartic proteinase zymogens. *FEBS Lett.*, **343**, 6–10.
- Lee, A.Y., Gulnik, S.V. and Erickson, J.W. (1998) Conformational switching in an aspartic proteinase. *Nature Struct. Biol.*, **5**, 866–871.
- Liepinsh, E., Andersson, M., Ruyschaert, J.M. and Otting, G. (1997) Saposin fold revealed by the NMR structure of NK-lysin. *Nature Struct. Biol.*, **4**, 793–795.
- Marttila, S., Jones, B.L. and Mikkonen, A. (1995) Differential localization of two acid proteinases in germinating barley (*Hordeum vulgare*) seed. *Physiol. Plant*, **93**, 317–327.
- Misasi, R. *et al.* (1998) Colocalization and complex formation between prosaposin and monosialoganglioside GM3 in neural cells. *J. Neurochem.*, **71**, 2313–2321.
- Moore, S.A., Sielecki, A.R., Chernaia, M.M., Tarasova, N.I. and James, M.N.G. (1995) Crystal and molecular structures of human progastricsin at 1.62 Å resolution. *J. Mol. Biol.*, **247**, 466–485.
- Munford, R.S., Sheppard, P.O. and O'Hara, P.J. (1995) Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure. *J. Lipid Res.*, **36**, 1653–1663.
- Nakano, T., Murakami, S., Shoji, T., Yoshida, S., Yamada, Y. and Sato, F. (1997) A novel protein with DNA binding activity from tobacco chloroplast nucleoids. *Plant Cell*, **9**, 1673–1682.
- Navaza, J. (1994) An automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.
- Neuhaus, J.M. and Rogers, J.C. (1998) Sorting of proteins to vacuoles in plant cells. *Plant Mol. Biol.*, **38**, 127–144.
- Nicholls, A. (1992) GRASP: graphical representation and analysis of surface properties. Columbia University, New York, NY.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Paris, N., Stanley, C.M., Jones, R.L. and Rogers, J.C. (1996) Plant cells contain two functionally distinct vacuolar compartments. *Cell*, **85**, 563–572.
- Ponting, C.P. and Russell, R.B. (1995) Swaposins: circular permutations within genes encoding saposin homologues. *Trends Biochem. Sci.*, **20**, 179–180.
- Ramalho-Santos, M., Verissimo, P., Cortes, L., Samyn, B., Van Beeumen, J., Pires, E. and Faro, C. (1998) Identification and proteolytic processing of procardosin A. *Eur. J. Biochem.*, **255**, 133–138.
- Rawlings, N.D. and Barrett, A.J. (1995) Families of aspartic peptidases and those of unknown catalytic mechanism. *Methods Enzymol.*, **248**, 105–120.
- Runeberg-Roos, P. and Saarma, M. (1998) Phytepsin, a barley vacuolar aspartic proteinase, is highly expressed during autolysis of developing tracheary elements and sieve cells. *Plant J.*, **15**, 139–145.
- Runeberg-Roos, P., Tormakangas, K. and Ostman, A. (1991) Primary structure of a barley-grain aspartic proteinase. A plant aspartic proteinase resembling mammalian cathepsin D. *Eur. J. Biochem.*, **202**, 1021–1027.
- Runeberg-Roos, P., Kervinen, J., Kovaleva, V., Raikhel, N.V. and Gal, S. (1994) The aspartic proteinase of barley is a vacuolar enzyme that processes probarley lectin *in vitro*. *Plant Physiol.*, **105**, 321–329.
- Sack, J.S. (1988) CHAIN—a crystallographic modeling program. *J. Mol. Graph.*, **6**, 244–245.
- Sali, A., Veerapandian, B., Cooper, J.B., Moss, D.S., Hofmann, T. and Blundell, T.L. (1992) Domain flexibility in aspartic proteinases. *Proteins*, **12**, 158–170.
- Sanderfoot, A.A., Ahmed, S.U., Marty-Mazars, D., Rapoport, I., Kirchhausen, T., Marty, F. and Raikhel, N.V. (1998) A putative vacuolar cargo receptor partially colocalizes with AtPEP12p on a prevacuolar compartment in *Arabidopsis* roots. *Proc. Natl Acad. Sci. USA*, **95**, 9920–9925.
- Sarkkinen, P., Kalkkinen, N., Tilgmann, C., Siuro, J., Kervinen, J. and Mikola, L. (1992) Aspartic proteinase from barley grains is related to mammalian lysosomal cathepsin D. *Planta*, **186**, 317–323.
- Sielecki, A.R., Fujinaga, M., Read, R.J. and James, M.N. (1991) Refined structure of porcine pepsinogen at 1.8 Å resolution. *J. Mol. Biol.*, **219**, 671–692.
- Tormakangas, K. (1997) Structure, expression and intracellular targeting of barley aspartic proteinase. PhD Thesis. Institute of Biotechnology and Department of Biosciences, Division of Genetics, University of Helsinki, Finland.
- Tormakangas, K., Kervinen, J., Ostman, A. and Teeri, T.H. (1994) Tissue-specific localization of aspartic proteinase in developing and germinating barley grains. *Planta*, **195**, 116–125.
- Vaccaro, A.M., Salvioli, R., Tatti, M. and Ciaffoni, F. (1999) Saposins and their interaction with lipids. *Neurochem. Res.*, **24**, 307–314.
- Verissimo, P., Faro, C., Moir, A.J., Lin, Y., Tang, J. and Pires, E. (1996) Purification, characterization and partial amino acid sequencing of two new aspartic proteinases from fresh flowers of *Cynara cardunculus* L. *Eur. J. Biochem.*, **235**, 762–768.
- Wilkening, G., Linke, T. and Sandhoff, K. (1998) Lysosomal degradation on vesicular membrane surfaces. *J. Biol. Chem.*, **273**, 30271–30278.

Received April 26, 1999; revised and accepted May 26, 1999