Processing, Activity, and Inhibition of Recombinant Cyprosin, an Aspartic Proteinase from Cardoon (Cynara cardunculus)*

(Received for publication, January 29, 1999)

Paul C. White‡‡, Maria C. Cordeiro§, Daniele Arnold‡, Peter E. Brodelius‡, and John Kay‡***

From the ‡School of Biosciences, Cardiff University, P. O. Box 911, Cardiff CF1 3US, Wales, United Kingdom, the §Department of Natural Sciences, Kalmar University, P. O. Box 905, S-39129 Kalmar, Sweden, and the ¶Department of Immunology, Institute for Cell Biology, Eberhard Karls University, D-72076 Tubingen, Germany

The cDNA encoding the precursor of an aspartic proteinase from the flowers of the cardoon, Cynara cardunculus, was expressed in Pichia pastoris, and the recombinant, mature cyprosin that accumulated in the culture medium was purified and characterized. The resultant mixture of microheterogeneous forms was shown to consist of glycosylated heavy chains (34 or 32 kDa) plus associated light chains with molecular weights in the region of 14,000–18,000, resulting from excision of most, but not all, of the 104 residues contributed by the unique region known as the plant specific insert. SDS-polyacrylamide gel electrophoresis under non-reducing conditions indicated that disulfide bonding held the heavy and light chains together in the heterodimeric enzyme forms. In contrast, when a construct was expressed in which the nucleotides encoding the 104 residues of the plant specific insert were deleted, the inactive, unpurified precursor form (procyprosin) accumulated, indicating that the plant-specific insert has a role in ensuring that the nascent polypeptide is folded properly and rendered capable of being activated to generate mature, active proteinase. Kinetic parameters were derived for the hydrolysis of a synthetic peptide substrate by wild-type, recombinant cyprosin at a variety of pH and temperature values and the subsite requirements of the enzyme were mapped using a systematic series of synthetic inhibitors. The significance of the susceptibility of cyprosin to inhibitors of human immunodeficiency virus proteinase and particularly of renin, some of which were found to have subnanomolar potencies against the plant enzyme.

Sequences have been elucidated recently for genes encoding aspartic proteinases of plant origin, e.g. barley (1), rice (2), tomato (3), and oilseed rape and Arabidopsis (4). All of these sequences predict that, by comparison with the well studied aspartic proteinases from mammals, fungi, yeasts like Saccharomyces (5) and Candida (6), parasites like Plasmodium falciparum (7), and viruses such as HIV (8), approximately 100 extra amino acids are introduced into the C-terminal domain of the newly synthesized plant polypeptides. The function of this region or plant-specific insert (9) is currently unknown, but, as it does share considerable sequence identity with a group of mammalian sphingolipid activator proteins known as saposins (10, 11), it has been postulated to bind selectively to certain lipids and thus direct the precursor form of the aspartic proteinase into the appropriate cytological compartment in the plant cell (10).

Alignment of the sequences of the inserts predicted by the plant aspartic proteinase genes with those of saposins correctly positions all six cysteine residues and a glycosylation site and also maintains the pattern of hydrophobic residues described previously for all known saposins (12). However, each plant-specific insert does not consist of a single saposin domain but appears to correspond to the C-terminal portion of one saposin domain linked to the N-terminal portion of a second saposin domain. On this basis, the plant-specific inserts have been described as swaposins (12), indicating that they are likely to have a structure similar to that of the saposins, but with N- and C-terminal halves interchanged in sequence order. Relatively little information has been reported to date on aspartic proteinases originating from plant tissues, but, in the few enzymes that have been isolated, this swaposin domain is not present and appears to have been excised by post-translational processing. Nothing is known of the enzyme(s) responsible in plants, but the excision of each insert is imprecise, resulting in the generation of a complex mixture of heterogeneous, mature aspartic proteinases within the tissues of each of the plants that has been studied, e.g. from seeds of barley (13), pumpkin (14), and Arabidopsis thaliana (15), from rice (16) and from flowers of the cardoon, Cynara cardunculus (9, 17, 18). Commonly, the enzymes thus generated are heterodimers with molecular weights in the region of 40,000–45,000.

This complexity of natural isoforms is compounded even further by the expression of several genes in the plant tissues, each encoding closely related enzymes so that physicochemical and enzymatic characterization of naturally occurring aspartic proteinases isolated directly from plants has been made rather difficult. A recombinant approach was thus employed in attempts to gain an initial insight into the significance of the plant-specific insert and to examine the activity and specificity of an aspartic proteinase from the plant kingdom. We have expressed in the methylotrophic yeast, Pichia pastoris, the cDNA encoding the precursor of one aspartic proteinase from the flowers of the cardoon, C. cardunculus (19). Extracts of these flowers have been documented previously to contain a number of isoforms of aspartic proteinases called cyprosins and cardosins (9, 17, 18) and have been used for centuries as coagulants in traditional cheese-making in regions of southern Europe, particularly the Iberian peninsula (20).
MATERIALS AND METHODS

Gene Cloning and Mutagenesis—A cDNA clone encompassing the full-length precursor of a cyropsin was isolated by courtesy of Dr. M. Pietrzak (Basel, Switzerland) by rescanning of a cDNA library prepared from flower buds of C. cardunculus, as described previously (19, 21). This was amplified by PCR using Vent DNA polymerase (New England Biolabs) with appropriate forward and reverse primers. The 1192-bp amplicon thus generated was digested with HindIII and subcloned into the pUC18 vector. In turn, this recombinant pUC18 was used as template DNA for overlapping PCR mutagenesis reactions, as described previously (22). Mutations were introduced at appropriate locations by two initial and one subsequent PCR reaction. To confirm the sequence connecting the proposed region to the mature cyropsin enzyme, a 241-bp fragment spanning from the pUC18 vector to the desired, mutated cleavage junction was amplified using appropriate forward and reverse primers (5'-GTT GGA TCC TCA AGC TGC TTC TGC AAA-3' and 5'-CCT CAG AAA AGC AGC GAA GCC-3', respectively). The second PCR amplified a 478-bp fragment using a forward primer (5'-GG AAT TCC TTC TCA CCT ACT GCA TTT TCG GTC-3') and a reverse primer (5'-GGA TCC TCA CCT ACT GCA TTT TCG GTC-3') in the region of the spectrum and thus interferes with detection of the product of peptide substrate digestion.

Samples of conditioned medium harvested after 6 days were adjusted to pH 4.0 by the addition of 1 M sodium formate buffer, pH 3.0, and dialyzed at 4 °C against 10 mM sodium formate buffer, pH 4.0, containing 50 mM sodium chloride with five changes of buffer. Dialyates were applied to a Hi-Trap SP column fitted into a fast protein liquid chromatography instrument (Amersham Pharmacia Biotech), and elution was continued with the same pH 4.0 buffer. Under these conditions, the yellow Pichia-derived pigment was not retained by the column. After extensive washing, recombinant cyropsin was eluted using a linear gradient of 50–500 mM NaCl in the sodium formate buffer at pH 4.0. Fractions were monitored for activity and combined as appropriate.

For N-terminal sequencing were subjected to SDS-PAGE under reducing or non-reducing conditions and blotted onto polyvinylidene difluoride membranes, and appropriate bands were subjected to automated Edman degradation, as described previously (23). Attempts to derive C-terminal sequence on relevant bands were carried out (24), using a Hewlett-Packard G1009A C-terminal sequencing system. Deglycosylation reactions were carried out with N-glycosidase F (Roche Molecular Biochemicals, Mannheim, Germany) in 250 mM Tris-HCl buffer, pH 8.8, at 37 °C for 1 h.

A naturally occurring preparation of isoform 3 of cyropsin was purified to homogeneity from the dried flowers of C. cardunculus as described previously (17).

Kinetic parameters for hydrolysis of the chromogenic substrate Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu were derived spectrophotometrically as described previously (23). Values of $K_m$ were derived from the equation $v = \frac{\sqrt{S_{max}}}{K_m} \times [E_i]$, where the concentration of active enzyme $E_i$ was determined by active site titrations using preparations of recombinant cyropsin of precisely defined concentration (22). Inhibition constants were derived at pH 5.0, and the estimated error on all measurements was always less than ±15%. However, it was necessary to use final concentrations of cyropsin in the assay cuvettes of approximately 5 nM, and so $K_i$ values for tight-binding inhibitors (where $K_i < 0.1 \times E_i$) are given as best estimates, indicated by the $< $ symbol. The synthetic inhibitors of carboxypeptidase Y (H-Y-WL), carboxypeptidase A, and aminopeptidase N were the respective gifts of Drs. J. Brzin and B. Strukelj (J. Stefan Institute, Ljubljana, Slovenia), Dr. R. J. Penkasky (formerly of the University of South Dakota, Vermillion, SD), and Dr. L. H. Phylip (School of Biosciences, Cardiff University, Cardiff, Wales).

RESULTS AND DISCUSSION

The nucleotide sequence of the procyropsin clone isolated by rescanning of the cDNA library from flower buds of C. cardunculus (19, 21) has been deposited in the EMBL/GenBank data bases under the accession number X81984. The amino acid sequence predicted by this clone is aligned with that of human procathepsin D in Fig. 1. From this, it is apparent that (i) the plant gene encodes an insert of 104 residues within the C-terminal domain that is not present in the mammalian enzyme, and (ii) both proteins are predicted to have one common glycosylation motif (at Asn83-Gly-Thr95-P6-P5-P4-P3-P2-P1*P1) and the other known site of carbohydrate attachment in cathepsin D (at Asn183-Val184-Thr185). Gly-Thr95-P6-P5-P4-P3-P2-P1*P1 is not present in the cyropsin sequence, although an additional glycosylation motif (at Asn234-Glu-Thr235-P2-P1*P1) is

The nomenclature system –$P_2-P_3-P_4-P_5-P_6$– is used to depict amino acids adjacent to the residues in the $P_1$ and $P_3$ positions that contribute the asparagine peptide bond (indicated by an asterisk).

In keeping with IUB convention, putative plant-specific insert res-
present in the plant-specific insert of cyprosin (Fig. 1 and see Introduction). Excluding the plant-specific insert, the sequence of the mature enzyme region of this cyprosin shares 52% identity with that of cathepsin D; in contrast, the two prosegments have very little similarity to one another.

As described under “Materials and Methods,” the cDNA encoding this wild-type form of procyprosin was introduced into pET-3a. Expression in *E. coli* strain BL21 (DE3) pLysS resulted in the accumulation of recombinant protein at a high level (estimated to be 10–20 mg/liter of culture) but this was insoluble and misfolded. Despite extensive efforts, suitable conditions to prepare significant amounts of properly folded protein could not be established. Consequently a different expression system was required and the methylotrophic yeast, *P. pastoris*, was selected since this has been used previously to produce an aspartic proteinase (precursor) in a soluble, properly folded form (25). The procyprosin construct was subcloned into the pPICZαC plasmid, and appropriate recombinants of the *P. pastoris* cells (KM71 strain) were selected as described under “Materials and Methods.”

Expression of the procyprosin gene resulted in the appearance in the culture medium of a cluster of immunoreactive bands in the 32–34-kDa region, together with a second cluster in the region between 14 and 18 kDa (Fig. 2A). The relative proportions of each band in each of the clusters varied from batch-to-batch of induced culture medium so the clusters are referred to as heavy and light chains, respectively. The heavy chain (32–34 kDa) cluster was apparent as early as the second day of induction, while the light chains (14–18 kDa) became visible after about day 5. The recombinant protein was purified batch-to-batch of induced culture medium so the clusters are referred to as heavy and light chains, respectively. The heavy chain (32–34 kDa) cluster was apparent as early as the second day of induction, while the light chains (14–18 kDa) became visible after about day 5. The recombinant protein was purified

---

**Fig. 1.** Alignment of the deduced amino acid sequence of procyprosin (CYP) with that of human procathepsin D (HCD). The numbering system used is that of pig pepsin. The absence of a residue is indicated by a dash. The positions of the active site Asp residues and potential N-glycosylation sites are indicated by ▲ and ●, respectively. Known disulfide bonds in cathepsin D and other eukaryotic aspartic proteinases are between cysteine residues (indicated by ●) 45–50, 206–210, and 250–283. The N-terminal sequences of heavy and light chains determined by Edman degradation are denoted by arrows.

---

**Fig. 2.** Accumulation of recombinant cyprosin. A selected strain of *P. pastoris* in which the wild-type procyprosin cDNA had recombined into the host chromosome was induced by the addition of methanol (final concentration = 0.5%) to the BMMY culture medium every 24 h. A, aliquots were removed at the indicated times and centrifuged, and each medium sample was analyzed by SDS-PAGE under reducing conditions, followed by staining with Coomassie Blue. Marker proteins of Mr approximately 43,000, 29,000, 18,000, and 14,000 migrated as indicated. B, aliquots of medium harvested from *Pichia* cells grown in the absence (–) and presence (+) of isovaleryl-pepstatin (IVP; 400 nM) were analyzed by SDS-PAGE, under reducing conditions.
Recombinant Cyprosin

The 34- and 32-kDa heavy chains were resolved from one another by SDS-PAGE under reducing conditions and blotted onto polyvinylidene difluoride membrane. N-terminal sequencing by Edman degradation revealed that both were identical and contained overlapping sequences (in the ratio 40:60).

Determined: ALRDSGSDII
Predicted from cDNA: ALRDSGIDI1AL
(Fig. 1)

SEQUENCES 1 AND 2

The distinction in the sizes of these heavy chains must therefore be accounted for by differences at their C termini. The N termini identified for the recombinant heavy chains are displaced further upstream by three and one residue, respectively, from the Asp residue (Fig. 1) that was suggested to be at the N terminus of the heavy chain of a cyprosin isoform isolated from flowers of C. cardunculus (21). Processing of the recombinant procyprosin precursor had thus taken place to remove the prosegment by cleavage at two adjacent sites in the sequence, resulting in microheterogeneity at the N termini of the resultant (34/32 kDa) heavy chains of mature cyprosin. After SDS-PAGE under reducing conditions and blotting onto polyvinylidene difluoride, attempts were also made at N-terminal sequencing of the 14–18-kDa light chains. In all cases, however, multiple residues were detected in every cycle, making assignment difficult, but indicating that the array of light chains observed was likely to have arisen from microheterogeneity at their N termini. However, when samples of recombinant cyprosin were subjected to SDS-PAGE under non-reducing conditions, a different effect was observed. A typical example is depicted in Fig. 3B (compare lanes 3 and 5). This preparation of recombinant cyprosin consisted almost exclusively of 32-kDa heavy chain under reducing conditions but in the absence of reducing agent, the dominant band detected was at ~46 kDa. N-terminal analysis of this 46-kDa band gave the same microheterogeneous pair of sequences, as was detected for the 34- and 32-kDa bands from the reduced SDS gel. In addition, however, a third sequence was detected in the non-reduced sample.

Edman degradation: VNELXDLRP
Predicted from cDNA: VNELCDLRLP
(Fig. 1)

SEQUENCE 3

This sequence is likely to have been derived from a light chain that had remained attached to the heavy chains under the non-reducing conditions, thereby indicating that the recombinant 34/32-kDa heavy chains and the light chains may be linked by disulfide bonds formed between sequences of the plant-specific insert that remain attached at the C- and N termini of the heavy and light chains, respectively, after excision of the bulk of the plant-specific insert by the Pichia cells. The presence of Cys3·Thr in the N-terminal sequence of this light chain indicated its availability for disulfide bonding.

No data were obtained from attempts at C-terminal sequencing of the 34-kDa heavy chain, but, for the 32-kDa heavy chain, leucine was identified as the C-terminal residue, with threonine occupying the penultimate position. This ~Thr-Leu combination occurs only once in the sequence Cys3·Lys4·Thr5·Leu6 near the predicted N terminus of the plant-specific insert region (Fig. 1). Proteolytic processing after this leucine residue would thus result in Cys3·Thr remaining within the heavy chain, which would be predicted to have an M_r of 29,000 on SDS-

from medium harvested after 6 days, as described under “Materials and Methods.” Acidification and dialysis of the medium at pH 4.0, followed by chromatography on a HiTrap SP column, successfully removed the yellow pigment that is a persistent contaminant released into the culture medium of induced Pichia cells. The recombinant protein that was eluted by the salt gradient from the HiTrap SP column did not emerge in a sharp peak but rather was a broad smear of material that was absorbed at 280 nm and that reacted with the cyprosin antisera (data not shown). Early fractions of material contained a mixture of 34- and 32-kDa heavy chains in which the 34-kDa band was predominant (Fig. 3A, lane 1), while in the later fractions the 32-kDa band was most abundant (Fig. 3A, lane 2). All of the fractions also contained light chains migrating in the 14–18-kDa range, which stained only weakly with Coomassie Blue. The 34- and 32-kDa heavy chains and the light chains were all immunoreactive with anti-cyprosin antiserum (data not shown). Thus, it was apparent that all of the contaminating proteins had been removed and that the only remaining proteins were derived from recombinant (pro)cyprosin. The yield of (total) purified protein was approximately 1 mg/liter of culture medium. However, microheterogeneity was clearly evident.

Fig. 3. Analysis by SDS-PAGE of recombinant proteins and naturally occurring cyprosin. A, early (lane 1) and late (lane 2) fractions eluted from chromatography of recombinant cyprosin on a HiTrap SP column, were resolved under reducing conditions. B, a late fraction aliquot of recombinant wild-type cyprosin (lanes 3 and 5) and a sample of naturally occurring cyprosin (isoform 3; lanes 4 and 6) were subjected to electrophoresis under reducing (lanes 3 and 4) and non-reducing (lanes 5 and 6) conditions. C, aliquots of recombinant wild-type cyprosin before (lane 7) and after (lane 8) treatment with N-glycosidase F were analyzed under reducing conditions. Lane 9 depicts a mixture of 34- and 32-kDa heavy chains in which the 34-kDa chain indicated its availability for disulfide bonding. Neither the 34- nor 32-kDa heavy chain, but, for the 32-kDa heavy chain, leucine was identified as the C-terminal residue, with threonine occupying the penultimate position. This ~Thr-Leu combination occurs only once in the sequence Cys3·Lys4·Thr5·Leu6 near the predicted N terminus of the plant-specific insert region (Fig. 1). Proteolytic processing after this leucine residue would thus result in Cys3·Thr remaining within the heavy chain, which would be predicted to have an M_r of 29,000 on SDS-PAGE.

The presence of Cys97, a late type cyprosin before (~Thr-Leu-Ala-Phe~. Lanes 1–6 and 9 were stained with Coomassie Blue after electrophoresis, whereas lanes 7 and 8 were revealed by Western blotting with an anti-cyprosin antisera, demonstrating that the antibodies are capable of recognizing protein epitopes. Markers of M_r approximately 43,000, 29,000, 18,000, and 14,000 migrated as indicated. Since the cyprosin light chain bands were difficult to visualize (panel A), only the heavy chain regions of the gels are depicted in lanes 3–9.
that had been separated as described previously (17) from the mixture of naturally occurring enzyme forms extracted from the flowers of the plant. This sequence begins at residue 103I of the plant-specific insert (Fig. 1) and is located downstream from Cys97I. Processing in planta to remove the plant-specific insert must therefore have taken place at a slightly different location from those observed in P. pastoris, i.e. downstream from Cys97I, which serves therefore to indicate the C-terminal boundary of the plant-specific insert. No C-terminal sequence has been obtained for the heavy chain of any naturally occurring cyprosin, but the homologous cardosin A heavy chain has been reported recently (9) to have microheterogeneity at its C terminus, corresponding to cleavage at the N-terminal end of the plant-specific insert, after the residues (equivalent to) Lys238I, Val240I, and Met241I (Fig. 1). This also suggests that processing in planta may take place at adjacent site(s) to those observed in Pichia, i.e. upstream from the Cys32I-Lys4I-Thr4I-Leu4I sequence at the beginning of the plant-specific insert.

Processing in Pichia was only marginally affected by inclusion (at 400 nm final concentration) of the aspartic proteinase inhibitor, pepstatin, in the culture medium of the Pichia cells. This caused only a small increase in size of the 34- and 32-kDa bands to 36 and 33 kDa, respectively (Fig. 2B), but did not result in the accumulation of a protein band consistent in size with that of the unprocessed procyprosin precursor. As this report was nearing completion, Glathe et al. (26) reported the expression of the cDNA encoding the precursor of the aspartic proteinase, phytepsin, from barley (Hordeum vulgare L.) using baculovirus. In this heterologous system, the intact precursor accumulated in the medium, enabling it to be purified and its ability to be processed under defined conditions in vitro to be determined. It was shown to be capable of undergoing autophosphorylation to generate initially bands of 36 and 17 kDa, as monitored by SDS-PAGE under reducing conditions, followed by further processing to produce finally subunits of approximately 28 and 11 kDa, respectively. Microheterogeneity was detected at the N terminus of the 11-kDa subunit but the residues identified were (equivalent to) Tyr92I and Val93I (of cyprosin; Fig. 1); thus, autoprocessing had taken place finally at an identical location to that observed with recombinant procyprosin in the present study. In the phytepsin case, no indication was given of the migration pattern of the subunits on SDS-PAGE under non-reducing conditions. However, these autophosphorylation reactions of phytepsin only occurred when the pH was below 4.5 (26). In our case, the medium of the Pichia cells was buffered well above pH 4.5 (pH 6.0), yet it was mature cyprosin and not the intact precursor that accumulated in the medium. Since the two plant proteinases are 78% identical, unless the initial translation product of procyprosin encounters pH values below 4.5 during its translocation through the lumen of the secretory apparatus in the Pichia cells, it would seem unlikely that the (multiple forms) of mature cyprosin that accumulated in the medium had been produced as a consequence of autocatalytic reaction of the recombinant precursor. A more likely explanation is that processing took place through the action of host cell proteinases at exposed sites that were susceptible not only to heterocatalytic attack under the conditions likely to be encountered in living cells but also to autocatalytic cleavage under the somewhat more extreme conditions that can be employed in vitro. Just as in the present case with recombinant and naturally occurring forms of cyprosin, the subunits of recombinant phytepsin differed slightly from those of the enzyme extracted from barley seeds (26).

Some precursors of mammalian/fungal aspartic proteinases require the action of extrinsic proteinases to generate the mature enzyme, e.g. prorenin, whereas others such as pepsinogen

---

4 M. C. Cordeiro and P. E. Brodelius, unpublished observation.
undergo autocatalytic activation (27). In the case of procathepsin D, which is the mammalian enzyme most closely related to cyprosin and phytpepsin, cysteine proteinases are required for activation to occur within the lysosomes of the cell although, in vitro, limited autoprocessing occurs under acidic conditions to generate an activation intermediate (known as pseudo-cathepsin D) in which only 26 of the 44 residues of the prosegment have been removed (28). In order to achieve complete autocatalytic removal of all 44 prosegment residues in vitro, it was necessary to introduce an autoactivation sequence at the junction between the prosegment and mature enzyme region of human procathepsin D (28). A comparable alteration was introduced into the procyanosin sequence by changing the five residues immediately preceding the plant insert by the corresponding residues linking the prosegment to the N terminus of the (mature chain of the) mature cyprosin (Fig. 1) from −Phe40-Gly-Gly-Ala-Leu-Arg45P−Asp1− to −Phe40F-Ala-Ala-Phe-Leu-Arg45P−Asp1−. Expression of this construct in Pichia resulted in the appearance of active, mature enzyme that was indistinguishable in its features and behavior from wild-type cyprosin (data not shown). In order to gain some insight into the significance of the plant specific insert, a further adaptation was introduced into this mutant construct by deleting the nucleotides encoding the 104-residue plant-specific insert of procyprosin and replacing the five residues (−Lys238−Gly-Val-Met-Seq242−; Fig. 1) immediately preceding the plant insert by the corresponding residues (−Val238−Pro-Leu-Ile-Gln242−) from human cathepsin D, which are known from the crystal structure to form a surface loop in this region of the human enzyme (29). Expression of this insert-deleted construct in Pichia resulted in the production of lower amounts of recombinant protein. The protein that was detected (Fig. 3C, lane 9) had a molecular weight (46,000) which corresponded to that of unactivated (insert-deleted) procyprosin and it was inactive against the synthetic peptide substrate. Thus, it seems that, at least in Pichia cells where the processes do approximate those likely to be operative in plant cells, the residues of the plant-specific insert would appear to be essential to ensure that the nascent polypeptide is folded properly and rendered capable of being activated to generate mature enzyme, albeit in slightly altered forms from the multiple isoforms that are generated within the flowers of the plant (9, 17, 26).

Kinetic parameters (K_m, k_cat) were determined for the hydrolysis of a synthetic chromogenic peptide substrate by a purified preparation of wild-type, recombinant cyprosin in which the heavy chain consisted predominantly of the 32-kDa form (Table I, A). Consistent values for K_m (varying by less than 2-fold) were observed across the pH range 3.0–6.0. Similarly, the values for k_cat varied by only 2-fold in absolute magnitude but did increase progressively between pH 3.0 and 4.5–5.0, decreasing again at yet higher pH values, as has been observed previously for aspartic proteinases from other species (30). Values were also measured for the hydrolysis of the same substrate at pH 5.0 by one of the naturally occurring isoforms (isozyme 3) of cyprosin, purified from the flowers of the cardoon plant as described previously (17). The K_m value obtained (25 ± 5 μM) was little different from that measured for the recombinant form of cyprosin produced in Pichia (Table I, A). The k_cat value determined for the natural isoform(29 ± 3 s⁻¹) was somewhat higher (by 3-fold) than the corresponding value derived at pH 5.0 for the recombinant form of the enzyme. A homology-based model for plant aspartic proteinases constructed on the basis of their similarity to mammalian/fungal aspartic proteinases for which structures have been solved by x-ray crystallography, indicated that the plant-specific insert residues are likely to be located adjacent to the active site cleft (10). Since the specificity constants (k_cat/K_m) measured for isoform 3 and recombinant cyprosin at pH 5.0 were 1.2 and 0.18 μM⁻¹ s⁻¹, respectively (Table I, A), it may be that the residual residues remaining after excision of (most of) the plant-specific insert by the Pichia cells, have a minor influence on k_cat (and hence on k_cat/K_m) but have relatively little effect on substrate interaction (as reflected in K_m). Comparable situations have been reported for pseudocathepsin D relative to cathepsin D (29) and from a comparison of recombinant and naturally occurring forms of plasmapesin I, an aspartic proteinase from the malaria parasite Plasmodium falciparum.5

The temperature dependence of the kinetic parameters for chromogenic substrate hydrolysis by the recombinant cyprosin was also determined. The specificity constant (k_cat/K_m) increased progressively in magnitude, reaching a maximum value at 55 °C but activity was still readily detectable at 65 °C (Table I, B). The stability of the recombinant proteinase at 55 °C was measured at pH 5.0, 6.0, and 7.5, and half-lives of 40, 40, and <2 min were determined, respectively. Recombinant cyprosin thus displays a remarkable stability at temperatures up to 55 °C and at pH values as high as 6.0.

The recombinant plant proteinase was not affected (IC_{50} > 2,000 nM) by a plant derived protein isolated from potatoes, which is a potent inhibitor (K_m = 4 nM) of cathepsin D (32). Similarly, the naturally occurring protein inhibitors of pepsin/cathepsin E and yeast proteinase A from the parasitic worm A. lumbricoides (33) and S. cerevisiae (34), respectively, did not inhibit the recombinant cyprosin to any significant extent (IC_{50} > 2,000 nM).

In contrast, the naturally occurring peptide isovaleryl-pepstatin, which contains a central statine moiety as the transition state analogue occupying the P_i-P_1' positions in the acylated pentapeptide (35), showed subnondial molar potency as an inhibitor of recombinant cyprosin (inhibitor 1 in Table II). Whereas this response might perhaps have been expected as typical of an aspartic proteinase (30), a completely distinct acylated pentapeptide (inhibitor 2; Table II) was almost as effective as an inhibitor of cyprosin, despite having only the central statine residue in common with the sequence of isovaleryl pepstatin. Replacement of statine (which has a leucine side chain in P_i) by its cyclolhexyl alanine analogue (ACHPA, inhibitor 3) resulted

5 L. Tyas, R. P. Moon, K. Rupp, J. Westling, R. G. Ridley, J. Kay, D. E. Goldberg, and C. Berry, submitted for publication.
in a 5-fold reduction in potency (compare inhibitors 3 and 2 in Table II). In longer inhibitors spanning nine subunits (Table III), replacement of statine by ACHPA also resulted in a reduction of potency (by 2.5-fold; compare inhibitors 8 and 6 in Table III); replacement by the variant containing a phenylalanine substituent in P1, (AHPPA, inhibitor 7; Table III) produced a compound with comparable inhibitory potency to that of the statine-containing inhibitor 6. Consequently, statine was retained as the centerpiece, occupying the P1-P1' positions of inhibitors, and the effect of substitutions in other positions was examined systematically.

In the P3 position, truncation of the (CH3)2CH-CH2-CO- (isovaleryl) substituent by removal of a methylene carbon (generating (CH3)2CH-CO- = isobutyl) resulted in an ~3-fold improvement in Kᵢ value toward recombinant cyprosin (cf. inhibitors 9 and 6 in Table III). Replacement of the methylene carbon atom with an oxygen (in (CH3)2CH-O-CO- = t-butoxycarbonyl) caused an additional 6-fold improvement in potency, resulting in an inhibitor with a potency of 1 nM (inhibitors 10 and 6 in Table III). It may be that the oxygen atom in the -O-CO- arrangement of inhibitor 10 can enter into a hydrogen bonding arrangement with an H-bond donor in the enzyme that is not possible with the -CH2-CO- equivalent in inhibitor 6.

At the P3 position, the replacement of phenylalanine by homophenylalanine (which has an additional -CH3- in its side chain) caused a significant improvement (10-fold) in Kᵢ (cf. inhibitors 12 and 6 in Table III). Replacement of Phe with 2-naphthylalanine (inhibitor 13) had only a minor (3-fold) effect on potency, whereas a 1-naphthylalanine substituent in the P3 position resulted in an inhibitor (number 14) that showed almost equivalent potency toward cyprosin as that measured for pepstatin (Table II). In contrast to this increase in size upon replacing the benzene ring of phenylalanine with the 1-naphthylalanine substituent, removal of the benzene ring altogether, resulting in an alanine substituent in P3 (compare 11 and 10 in Table III), caused a reduction in potency of more than 3 orders of magnitude.

In the P2' position, substitution of His by the more hydrophobic isomer, Ile, resulted in a 5-fold improvement in inhibitor potency (cf. 18 and 6 in Table III), whereas replacement by the longer, straight-chain analogue, δ-N-acetylornithine (inhibitor 19) had only a minimal effect. At the P2' position, replacement of the hydrophobic Phe residue by the (charged), hydrophilic imidazole side chain of His resulted in a reduction in potency of 4-fold (cf. compounds 20 and 18; Table III) toward recombinant cyprosin.

Thus, on the basis of the data presented in Tables II and III, it may be concluded that the compound ethoxycarbonyl-1-naphthylalanine-o-amino- n-butyric acid-statine-Ile-Phe-NH₂ would be likely to be an extremely potent, if somewhat water-insoluble inhibitor of recombinant cyprosin. It will be apparent then that the active site cleft of the plant enzyme is very hydrophobic.

In this regard, however, it has long been held that, among the aspartic proteinases, cathepsin D has one of the most hydrophobic active site clefts in its specificity requirements (36). For the sake of comparison, then, the Kᵢ values that we have reported previously (37) for the interaction of this human enzyme with the same set of inhibitors, are included in Tables II and III. From these data, it is apparent that each inhibitor binds substantially tighter (varying from ~5- to 800-fold) to recombinant cyprosin than to human cathepsin D, with the exception of pepstatin and inhibitor 11, which contains an Ala in the P3 position. The dislike of cyprosin for a small hydrophobic substituent in P3 reflects the situation reported previously (37) for human renin, gastricsin, and cathepsins D and E. In contrast, human pepsin exhibits a striking preference for Ala rather than Phe as a P₃ substituent (37). The inhibitors described in Tables II and III were synthesized originally as potential inhibitors of human renin, which is one of, if not the, most specific proteinase so far described. It acts solely on one protein substrate angiotensinogen, to generate angiotensin I, and so there have been substantial efforts made to design specific renin inhibitors, such as those in Tables II and III, as potential anti-hypertensive agents. It seems all the more remarkable then that inhibitors designed on the basis of the sequence of residues in plasma angiotensinogen recognized by this highly specific enzyme involved in the mammalian cardiovascular system, should prove to be so effective against a proteinase from the flowers of the cardoon plant.

Following this rationale, with the advent of AIDS, several synthetic compounds have also been developed recently as inhibitors of the aspartic proteinase from human immunodeficiency virus (8). These have been shown to have differing potencies toward the human aspartic proteinases (38–41). Consequently, these were also examined for their effectiveness toward recombinant cyprosin. HBY-793 (compound 21), which is a symmetrical compound containing a central dihydroxethylene moiety (38), was found to act as an inhibitor of cyprosin (Table IV). It is, however, much less effective toward the plant enzyme than HIV protease or some of the human enzymes for which subnanomolar Kᵢ values were obtained. Ritonavir (compound 22) contains a dihydroxethylene transition state analogue and has been shown to have Kᵢ values of 20 and 8 nM, respectively, for human cathepsin D and cathepsin E (39), so is not totally specific for HIV protease either. This trend was reflected with recombinant cyprosin against which Ritonavir was a weak (Kᵢ = 110 nM) inhibitor. In total contrast, Saquinavir (inhibitor 23 in Table IV) has been determined previously (40) to be completely specific for the aspartic proteinases from HIV-1, HIV-2, and SIV and to have no measurable effect on any other aspartic proteinase, including those from other retroviruses (41). This dihydroxethylamine-containing inhibitor was found to inhibit recombinant cyprosin, albeit weakly (Table IV) at 140 nM. In contrast, Indinavir (compound 24 in Table IV),

| Inhibitor no. | Structure | Kᵢ (nM) | 
|--------------|-----------|---------| 
| 1 | Iva -Val-Val-Sta -Ala-Sta | 0.5 | <0.1 | 
| 2 | Boc -Phe-His-Sta -Leu-Phe-NH₂ | 2 | 35 | 
| 3 | Boc -Phe-His-ACHPA-Leu-Phe-NH₂ | 10 | 140 | 
| 4 | Ac -Phe-His-ACHPA-Leu-Phe-NH₂ | 9 | 550 | 
| 5 | Etoc-Phe-His-ACHPA-Leu-Phe-NH₂ | 2 | 210 |
which also contains a hydroxyethylamine transition state analogue, did not inhibit cyprosin to any significant extent.

From these data, it is evident that the active site cleft of cyprosin can readily accommodate chemically synthesized inhibitors of mammalian and retroviral aspartic proteinases although the plant enzyme is not susceptible to protein inhibitors including one, itself of plant origin (potato). Thus, it would seem that the most distinctive feature of the plant aspartic proteinase is the plant-specific insert. This conserved feature has been hypothesized to direct the newly translated plant polypeptides into appropriate cytomorphological compartments within the cell. However, in the present study, recombinant cyprosin was secreted by the yeast cells into the medium so that it did not appear to contain appropriate intracellular targeting signals that were functional in the Pichia cells. Thus, in order to gain further insight into the significance of the plant-specific insert, it will be necessary to express mutants encoding diminishing lengths of the insert, within plant cells.

Nevertheless, from the present study, the plant-specific insert appears to be necessary for the production of the recombinant precursor in such a form that it can be processed to produce the mature enzyme.

Acknowledgments—We gratefully appreciate the respective contributions made by M. Pietrzak (Basel, Switzerland), Edy Segura (Protein Core Facility, University of Florida), and Wieland Keilholz (AnalyTeck, Tubingen, Germany) in clone isolation, N-terminal sequencing, and C-terminal sequencing. We also acknowledge the generosity of J. Brzin, B. Strukelj, R. J. Peanasky, L. H. Phylip, J. A. Martin, D. Kempf, J. Knolle, and D. F. Veber, for their kind provision of the inhibitors used.

REFERENCES

1. Runeberg-Roos, P., Tirmakangas, K., and Ostman, A. (1991) Eur. J. Biochem. 202, 1015–1027
2. Asakura, T., Watanabe, H., Abe, K., and Arai, S. (1995) Eur. J. Biochem. 232, 77–83
3. Schaller, A., and Ryan, C. A. (1996) Plant Mol. Biol. 31, 1073–1077
4. D’Hondt, K., Stack, S., Gutteridge, S., Vanderkerekhove, J., Krebbers, E., and Gal, S. (1997) Plant Mol. Biol. 33, 187–192
5. Van den Hazel, H. B., Wolff, A. M., Kielland-Brandt, M. C., and Winther, J. R. (1997) Biochem. J. 326, 339–344
6. Menod, M., Togni, G., Hube, B., and Sanglard, D. (1994) Mol. Microbiol. 13, 357–368
7. Dame, J. B., Reddy, G. R., Yewell, C. A., Dunn, B. M., Kay, J., and Berry, C. (1994) Mol. Biochem. Parasitol. 64, 177–190
8. Wlodawer, A., and Vondrasek, J. (1998) Annu. Rev. Biophys. Biomol. Struct. 27, 249–284
9. Ramalho-Santos, M., Verissimo, P., Cortes, I., Samyn, B., van Beeumen, J., Pires, E., and Faro, C. (1998) Eur. J. Biochem. 255, 133–138
10. Guruprasad, K., Tirmakangas, K., Kervinen, J., and Blundell, T. L. (1994) FEBS Lett. 352, 131–136
11. O’Brien, J. S., Kretz, R. A., Dewji, N., Wenger, D. A., Esch, F., and Fluharty, T. (1998) Trends Biochem. Sci. 23, 131–136
12. Ponting, C. P., and Russell, R. B. (1995) Trends Biochem. Sci. 20, 179–180
13. Sarkkinen, P., Kalkkinen, N., Tilgmann, C., Siuro, J., Kervinen, J., and Mikola, L. (1992) Planta 186, 317–323
14. Hirawa, N., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (1997) Eur. J. Biochem. 246, 1333–1411
15. Mutlu, A., Pfeil, J. E., and Gal, S. (1998) Phytochemistry 47, 1453–1459
16. Asakura, T., Watanabe, H., Abe, K., and Arai, S. (1997) J. Agric. Food Chem. 45, 1670–1675
17. Heimgartner, U., Pietrzak, M., Geertsen, R., Brodelius, P. E., and Saulle, A. (1992) Botan. J. Linnean Soc. 109, 75–123
18. Brodelius, P. E., Cordeiro, M., Mercke, P., Domingos, A., Clemente, A., and Pires, E. (1996) Eur. J. Biochem. 235, 762–768
19. Cordeiro, M., Xue, Z.-T., Pietrzak, M., Pires, E. M., and Brodelius, P. E. (1994) Plant Mol. Biol. 24, 733–741
20. Tigue, N. J., and Kay, J. (1998) J. Biol. Chem. 273, 26441–26446
21. Hill, J., Montgomery, D. S., and Kay, J. (1993) FEBS Lett. 326, 101–104
22. Arnold, D., Keilholz, W., Schild, H., Dumrese, T., Stevanovic, S., and Rammensee, H-G. (1997) Biochem. Biophys. Acta 131–136
23. Hill, J., Montgomery, D. S., and Kay, J. (1993) FEBS Lett. 326, 101–104
24. Arnold, D., Keilholz, W., Schild, H., Dumrese, T., Stevanovic, S., and Rammensee, H-G. (1997) Biochem. Biophys. Acta 131–136
25. Glathe, S., Kervinen, J., Nimtz, M., Li, G. H., Tobin, G. J., Copek, T. D.,
Recombinant Cyprosin

Ashford, D. A., Wlodawer, A., and Costa, J. (1998) *J. Biol. Chem.* 273, 31230–31236
27. Khan, A. R., and James, M. N. G. (1998) *Protein Sci.* 7, 815–836
28. Beyer, B. M., and Dunn, B. M. (1996) *J. Biol. Chem.* 271, 15590–15596
29. Baldwin, E. T., Bhat, T. N., Gulnik, S., Hosur, M. V., Sowder, R. C., Cachau, R. E., Collins, J., Silva, A. M., and Erickson, J. W. (1993) *Proc. Natl. Acad. Sci. U. S. A.* 90, 6796–6800
30. Kay, J., and Dunn, B. M. (1992) *Scand. J. Clin. Invest.* 52, 25–30
31. Deleted in proof
32. Kreft, S., Ravnikar, M., Mesko, P., Pungercer, J., Umek, I., and Strukelj, B. (1997) *Phytochemistry* 44, 1001–1006
33. Valler, M. J., Kay, J., Aoyagi, T., and Dunn, B. M. (1985) *J. Enzyme Inhibition* 1, 77–82
34. Dreyer, T., Valler, M. J., Kay, J., Charlton, P., and Dunn, B. M. (1985) *Biochem. J.* 231, 777–779
35. Marciniszyn, J. P., Jr, Hartsuck, J. A. and Tang, J. (1976) *J. Biol. Chem.* 251, 7088–7094
36. Scarborough, P. E., and Dunn, B. M. (1994) *Protein Eng.* 7, 495–502
37. Jupp, R. A., Dunn, B. M., Jacobs, J. W., Vlasuk, G., Arcur, K. E., Veber, D. F., Perlow, D. S., Payne, I. S., Boger, J., De Laszlo, S., Chakravarty, P. K., Tenhroche, J., Hannot, D. G., Osadnyka, D., Greenlees W. J., and Kay, J. (1990) *Biochem. J.* 265, 871–878
38. Budt, K. H., Peyman, A, Hansen, J., Krolle, J., Meichsner, C., Paessens, A., Ruppert, D., and Stowasser, B. (1995) *Bioorg. Med. Chem. 3*, 559–571
39. Kempf, D. J., Marsh, K. C., Denissen, J. H., McDonald, E., Vasavananoda, S., Flentje, C. A., Green, B. E., Fino, L., Park, C. H., Kong, X.-P., Wideburg, N. E., Saltvar, A., Ruiz, L., Kati, W. M., Sham, H. L., Rohins, T., Stewert, K. D., Hou, A., Plattner, J. J., Leonard, J. M., and Norbeck, D. W. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 2484–2488
40. Roberts, N. A., Martin, J. A., Kingshington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B., Golgin, S. A., Handa, B. K., Kay, J., Krohn, A., Lambert, R. W., Merrett, J. H., Mills, J. S., Parke, K. E. B., Redshaw, S., Ritchie, A. J., Taylor, D. L., Thomas, G. J., and Machin, P. J. (1990) *Science* 248, 358–361
41. Powell, D. J., Bur, D., Wlodawer, A., Gustchina, A., Payne, S. L., Dunn, B. M., and Kay, J. (1996) *Eur. J. Biochem.* 241, 664–674