Biochemical Journal Immediate Publication. Published on 23 Mar 2004 as manuscript BJ20031575

Enzymatic and structural characterization of nepenthesin, a unique member of a novel subfamily of aspartic proteinases

Senarath B. P. ATHAUDA[*,†], Koji MATSUMOTO[*], Sanath RAJAPAKSHE[†], Masayuki KURIBAYASHI[*], Masaki KOJIMA[*], Nobuko KUBOMURA-YOSHIDA[‡], Akihiro IWAMATSU[‡], Chiaki SHIBATA[§], Hideshi INOUE[*] and Kenji TAKAHASHI[*1]

[*]Laboratory of Molecular Biochemistry, School of Life Science, Tokyo University of Pharmacy and Life Science, 1432-1 Horinouchi, Hachioji, Tokyo 192-0392, Japan, [†]Department of Biochemistry, Faculty of Medicine, University of Peradeniya, Peradeniya and Institute of Fundamental Studies, Kandy, Sri Lanka, [‡]Central Laboratories for Key Technology, Kirin Brewery Co., Ltd., Yokohama, Kanagawa 236-0004, Japan and [§]Department of Biology, Nippon Dental University, Chiyoda-ku, Tokyo 102-8159, Japan

Short title: Nepenthesin, a novel plant aspartic proteinase

Abbreviations used: AP, aspartic proteinase; Cya, cysteic acid; DAN, diazoacetyl-D,L-norleucine methyl ester; NAP, nepenthesin-type AP.

[1]To whom correspondence should be addressed (e-mail kenjitak@ls.toyaku.ac.jp)

The nucleotide sequence data of nepenthesins I and II are available from the DDBJ Data Bank under the accession numbers AB114914 (for nepenthesin I) and AB114915 (for nepenthesin II).

Copyright 2004 Biochemical Society

Carnivorous plants are known to secrete acid proteinases to digest preys, mainly insects, for nitrogen uptake. In this study, we have purified, for the first time, to homogeneity two acid proteinases (nepenthesins I and II) from the pitcher fluid of *Nepenthes distillatoria* and investigated their enzymatic and structural characteristics. Both enzymes were optimally active at pH about 2.6 toward acid-denatured hemoglobin; the specificity of nepenthesin I toward oxidized insulin B chain appears to be similar, but slightly wider than those of other aspartic proteinases (APs). Among the enzymatic properties, however, most notable is their unusual stability; at or below 50°C both enzymes were remarkably stable; especially nepenthesin I was extremely stable over a wide range of pH from 3 to 10 for well over 30 days. This suggests an evolutionary adaptation of the enzymes to their specific habitat. We have also cloned the cDNAs and deduced the complete amino acid sequences of the precursors of nepenthesins I and II (437 and 438 residues, respectively) from the pitcher tissue of *Nepenthes gracilis*. Although the corresponding mature enzymes (each 359 residues) are homologous to ordinary pepsin-type APs, both enzymes had a high content of cysteine residues (12 residues per molecule), which are assumed to form six unique disulfide bonds as suggested by computer modeling and are thought to provide with major basis for the remarkable stability of nepenthesins. Moreover, the amino acid sequence identity of nepenthesins with ordinary APs, including plant vacuolar APs, are remarkably low (~20%), and phylogenetic comparison shows that nepenthesins are distantly related to them to form a novel subfamily of APs with a high content of cysteine residues and a characteristic insertion, named 'the nepenthesin-type AP-specific insertion', including a large number of novel, orthologous plant APs emerging in the gene/protein data bases.

Key words: aspartic proteinase; carnivorous plant; characterization; *Nepenthes*; nepenthesin; plant proteinase.

Copyright 2004 Biochemical Society

INTRODUCTION

There are several carnivorous plants of different genera in nature, which catch preys, mainly insects, and digest their proteins primarily by their endogenous proteinase(s) and absorb the digestion products as the nitrogen source [1, 2]. J. D. Hooker [3] was the first to document that *Nepenthes* is carnivorous, inspired directly by Charles Darwin. Since then, it has been the object of studies for nearly 130 years how *Nepenthes* accomplishes this process [4]. Nepenthesin [5, 6] is an acid proteinase secreted in the pitcher of *Nepenthes* species. So far, the acid proteinases from *Nepenthes* and *Drosera* species were only partially purified and poorly characterized [7-12]. Although they were shown to be members of APs. [9, 11], none of the enzymes secreted from carnivorous plants were purified to homogeneity, mainly due to the difficulty to obtain a sufficient amount of their digestive fluids.

APs are widely distributed in living organisms and extensive studies have been performed on mammalian, microbial and viral APs [13, 14]. They are also distributed widely in the plant kingdom, and are present in seeds, leaves and flowers in various plants [15] as well as in the digestive fluids of carnivorous species. Plant APs such as those from barley [16,17] and rice [18] and cyprosins (or cardosins) [19-23] have been purified and well characterized. All these plant proteinases have a plant-specific insertion sequence in the middle of the molecule, and are thought to be intracellular vacuolar enzymes. In contrast, APs from the digestive fluids of carnivorous plants are only known extracellular proteinases of plant origin. These enzymes are, therefore, interesting from various points, such as physiological roles, structure-function relationships, and molecular evolution.

In the present study, we have for the first time purified to homogeneity carnivorous plant APs in the pitcher fluid from *Nepenthes distillatoria* (*i.e.,* nepenthesins I and II) and elucidated their molecular and enzymatic characteristics, including their remarkable stability in a wide range of pH over a long period of incubation time. This stability seems to indicate that they have evolutionally well adapted to their original habitat. We have also cloned cDNAs for the enzymes from the pitcher tissue of *Nepenthes gracilis* to deduce the complete amino acid

Copyright 2004 Biochemical Society

sequences. The results have revealed that they are unique enzymes belonging to a novel subfamily of APs with a high content of cysteine residues, which presumably form disulfide bonds to stabilize the enzymes. This family appears to include a large number of new plant orthologs distantly related with the hitherto-known APs.

EXPERIMENTAL

Materials

Nepenthes pitcher fluid was collected from the plant *Nepenthes distillatoria* in the Singharaja forest, Sri Lanka. *Nepenthes gracillis* was obtained from the Taishoen plantation at Numazu, Shizuoka, Japan. In the present study, *N. distillatoria* was used for studies at the protein level and *N. gracilis* for those at the DNA level. This is because *N. distillatoria* was available only in Sri Lanka, whereas fresh *Nepenthes* tissue was required for cDNA cloning that had to be carried out in Japan. DEAE-cellulose (DE-52) was a product of Whatman Inc. (Kent, UK). Sephacryl S-200 and a Mono Q HR5/5 column were purchased from Amersham Biosciences (Uppsala, Sweden), and pepstatin-Sepharose, diazoacetyl-D,L-norleucine methyl ester (DAN), the B chain of oxidized bovine insulin and porcine pepsin A were obtained from Sigma (St. Louis, MO. USA). Pepstatin A was from Peptide Institute Inc. (Osaka, Japan). Reagents for automated amino acid analysis and sequencing were obtained from Applied Biosystems (Foster City, CA, USA). Other reagents used were of the highest grade available.

Enzyme activity assay

Proteolytic activity of the enzyme was determined essentially as described [24]. In the standard assay, the reaction was performed in a mixture containing 200 μl of the enzyme solution and 400 μl of 2% acid-denatured hemoglobin in 0.1 M formate buffer, pH 3.0, as a substrate at 37 ^{o}C for 2 h, and then stopped by addition of 800 μl of 5% trichloroacetic acid. After centrifugation at 10,000 x *g* for 10 min, the absorbance at 280 nm of the resulting supernatant was measured against a blank sample. One unit of activity was defined as the increase of one

Copyright 2004 Biochemical Society

absorbance unit per hour.

Protein determination

Protein was determined by measuring the absorbance at 280 nm of the sample solution or by the method of Smith *et al.* [25] using the biscinchoninic acid reagent.

Purification procedures

All purification procedures were performed at 4 ^{o}C. The collected fluid (30 liters) from the open pitchers was filtered to remove the insoluble materials and dialyzed against 0.02 M sodium phosphate buffer, pH 7.5, (0.02 M in phosphate) (Buffer A). The proteins in the dialyzed fluid were adsorbed to wet DEAE-cellulose (about 2 liters) equilibrated with the same buffer by batch-wise treatment and the cellulose was put into a glass column (6.0 x 70 cm). The column was washed with Buffer A and the protein was eluted with 0.5 M NaCl. The fractions with proteolytic activity were pooled and applied to a second DEAE-cellulose column (3.0 x 35 cm) equilibrated with Buffer A. The column was washed with the same buffer and the protein was eluted with a linear gradient of 0 to 0.5 M NaCl in 2 l of the same buffer. The active fractions in each activity peak were pooled separately and concentrated by using a small DEAE-cellulose column. Each concentrated sample was applied to a Sephacryl S-200 column (3.1 x 114 cm) equilibrated with Buffer A, 0.2 M NaCl. The active fractions were pooled and dialyzed against 0.02 M sodium acetate buffer, pH 4.0, and applied to a pepstatin-Sepharose column (1.3 x 1.6 cm) equilibrated with the same buffer. The column was washed with the same buffer and the protein was eluted with 0.05 M Tris/HCl buffer, pH 8.0, containing 1 M NaCl, and then with 0.05 M sodium borate buffer, pH 10.0, containing 1 M NaCl. The fraction eluted at pH 10.0 was immediately adjusted to pH 8.0 by the addition of 1 M Tris/HCl buffer, pH 8.0. The active fractions were pooled and dialyzed against 0.02 M Tris/HCl buffer, pH 7.8, and applied to a Mono Q column. The protein was eluted with a linear gradient of 0 to 1 M NaCl in 0.02 M Tris/HCl buffer, pH 7.8, and the active fractions were pooled. The Mono Q chromatography of

Copyright 2004 Biochemical Society

the major enzyme fraction was performed three times with one-third of the sample at a time. The digestive fluid collected from the unopened pitchers was submitted to essentially the same purification procedures.

Purity check and molecular weight determination

SDS-PAGE [26] and gel filtration on a column (3.1 x 114 cm) of Sephacryl S-200 were used for purity check and molecular weight determination of the purified enzymes. In SDS-PAGE, the protein was stained with Coomassie Brilliant Blue and the carbohydrate with Schiff's reagent after periodate oxidation [27].

Determination of the NH_2-terminal and partial internal amino acid sequences

The NH_2-terminal amino acid sequence of each protein was determined by using an Applied Biosystems pulse-liquid protein sequencer model 477A. To determine the partial internal sequences of the enzyme, the protein was reduced and carboxymethylated according to Crestfield *et al*. [28]. The protein (about 150 µg) was then digested at 37^0C with endoproteinase Asp-N (1 µg) for 12 h in 300 µl of 0.1 M ammonium bicarbonate, pH 7.8, with trypsin (3 µg) for 4 h in 300 µl of 0.1 M ammonium bicarbonate, pH 8.1, or with *Staphylococcus aureus* V8 protease (1 µg) for 4 h in 300 µl of 0.1 M ammonium bicarbonate, pH 8.1. The resulting peptides were separated by HPLC using a Hitachi (Tokyo, Japan) 655A-11 system on a column (0.46 x 25 cm) of TSKgel ODS-120T (Tosoh Co., Tokyo, Japan). The peptides were eluted with a linear gradient of acetonitrile (0–50%) in 0.1% trifluoroacetic acid at a flow rate of 0.8ml/min. The effluent was monitored by measuring the absorbance at 215 nm and the peptide peak fractions were collected and lyophilized. An aliquot of each peptide fraction dissolved in water was submitted to automated amino acid sequencing.

Digestion of oxidized insulin B chain and analysis of the cleavage sites

The B chain of oxidized insulin (150 nmol) was digested with the purified nepenthesin I (0.3

Copyright 2004 Biochemical Society

nmol) in 600 μl of 0.1 M formate buffer, pH 3.0, at 37°C for 3h. The resulting peptides were separated by HPLC and analyzed in the same manner as described above.

Reaction of diazoacetyl-D,L-norleucine methyl ester (DAN)

Each purified enzyme (50 μg) was treated with DAN (200 μg) in 3.0 ml of 0.05 M sodium acetate buffer, pH 5.0, at 14 °C in the presence or absence of 2 mM cupric sulfate. Aliquots were withdrawn at appropriate intervals and the remaining activity was determined. Porcine pepsin was treated with DAN under the same conditions for comparison.

Stability studies

To investigate the effect of pH on the stability, each enzyme (50 μg/ml of buffer) as well as the crude pitcher fluid was incubated at different pH values at 37 °C, and after 7 days and 30 days the remaining activity was determined at pH 3.0. The buffers used were 0.05M sodium formate buffer, pH 3.0, 0.05 M sodium acetate buffers, pH 4.0 and 5.0, 0.05 M sodium phosphate buffers, pH 6.0 and 7.0, 0.05 M Tris/HCl buffers, pH 8.0 and 9.0, and 0.05M sodium borate buffer, pH 10.0. Porcine pepsin A (50 μg/ml) in respective buffers was incubated in the same manner for comparison.

In order to investigate the effect of temperature on the stability, each enzyme (50 μg/ml) in 0.02M sodium formate buffer, pH 3.0 as well as the crude pitcher fluid (pH 3.0) was incubated at different temperatures for 2 h, and the remaining activity was determined at 37 °C. Furthermore, each enzyme sample was incubated at different temperatures (4, 25, 37 and 50 °C), and after 7 days and 30 days the remaining activity was determined at 37 °C. Porcine pepsin A solution (50 μg/ml of 0.02M sodium formate buffer, pH 3.0) was incubated in the same manner for comparison.

Preparation of antibody and immunohistochemical staining

The purified nepenthesin I mixed with the complete Freund's adjuvant was injected into rabbits

Copyright 2004 Biochemical Society

(1.0 mg for the primary injection and 0.5 mg for each of three booster injections) and IgG was purified from the antiserum by 40% ammonium sulfate precipitation followed by Protein A-Sepharose affinity chromatography. Longitudinal sections of fresh *N. distillatoria* tissues prepared by using a microtome were fixed by dipping in 3% formaldehyde solution for 1 h. They were successively incubated in phosphate-buffered saline with 1% bovine serum albumin for 2 h, in a 1: 8000 diluted purified primary antibody preparation for 2 h, in a 1:2000 diluted secondary antibody with a VECTASTAIN ABC immunochemical staining kit (Vector Laboratories) for 2 h and in the ABC solution for 1 h to be stained with an alkaline phosphatase substrate as described in the supplier's manual. As a control experiment, the primary antibody preparation preincubated with an excess of nepenthesin I immobilized by polyvinylidene difluoride membrane was used.

cDNA cloning

Total RNA was isolated from the *Nepenthes gracilis* pitchers using the lower ca. 1/3 part of each pitcher possessing the digestive glands according to the modified hot borate method [29] and used for the isolation of poly(A)-RNA using an Oligotex^{TM}-dT30 Super mRNA Purification Kit (Takara, Kyoto, Japan). Single-stranded cDNA was prepared using cDNA Synthesis Kit (Takara, Kyoto, Japan) and an oligo(dT) primer.

Based on the partial amino acid sequences determined at the protein level, degenerate oligonucleotide primers were synthesized as follows (see also Table I): Nep1s, YAGDGEY; Maj3a, PTFVMHF; Maj2s, IWTQCQP; Maj1a, QIPTFVM; Min1a, TCFGEPS; Min2s, IWTQCEP; Min4a, FGCGQTV. PCR was carried out using Ex Taq DNA polymerase (Takara, Kyoto, Japn) and a thermal cycler GeneAmp2400 (Perkin-Elmer Biosystems). The PCR protocol was typically based on 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min and completed by a 7-min extension at 72°C, and then the block temperature was held at 4°C.

The single-stranded cDNA was used as a template for PCR amplification of partial cDNAs using primers Nep1s and Maj3a for nepenthesin I, and Nep1s and Min1a for nepenthesin II.

Copyright 2004 Biochemical Society

The amplified fragments were further subjected to nested PCR using primers Maj2s and Maj1a for nepenthesin I and Min2s and Min4a for nepenthesin II. Sequencing of the fragments obtained was carried out using the terminator cycle sequencing method using BigDye Terminator v3.0 Cycle Sequencing Ready Reaction Kit and a DNA sequencer ABI PRISM 3100 (Applied Biosystems).

The complete 3' end of the cDNA was amplified using a specific forward primer Maj7s (Table I) for nepenthesin I or Min11s for nepenthesin II and oligo(dT) primer as a reverse primer. DNA fragments amplified using primers Nep1s and Maj7a for nepenthesin I and primers Nep1s and Min8a for nepenthesin II were sequenced. 5'-RACE (rapid amplification of cDNA ends) was carried out using 5'-Full RACE Core Set (Takara, Kyoto, Japan) according to the manufacturer's protocol. Single-stranded cDNA was prepared using phosphorylated primer Nep16-RT-P, and was cyclized or concatemerized by ligation. Using the obtained DNA as a template, PCRs were performed with primers Maj12s and Maj13a for nepenthesin I and with Min12s and Min13a for nepenthesin II. Then nested PCRs were carried out with primers Maj2s and Maj14a for nepenthesin I and with Min2s and Min14a for nepenthesin II. The obtained DNAs were sequenced. The sequence between the NH_2- and COOH-termini was completed by PCR using the cDNA and the upstream primer and the downstream primer.

Molecular modeling

The tertiary structures of nepenthesins I and II were predicted by the homology modeling method with the program MODELLER [30], using the crystal structure of porcine pepsin A (PDB ID:5PEP)[31] as a template. The alignment of the amino acid sequences of the target enzyme (nepentheshin I or II) and the template (porcine pepsin) was performed by the program ClustalW [32] so as to conserve the active site architecture [33]. The modeling was carried out by setting all the parameters to their default values and assuming no information about disulfide bonds. Using the resultant structures, the disulfide bonds were assigned, and energy minimization calculation was performed. All the procedures for performing the energy

Copyright 2004 Biochemical Society

minimization and manipulating the atomic coordinates of the cysteine residues were performed by the SYBYL molecular modeling system (Tripos, USA). The energy calculations were performed using a cutoff distance of 10 Å, a dielectric constant of 4r, and the default values for other parameters. The graphic image was produced using the MidasPlus program [34].

Constuction of a phylogenetic tree

The amino acid sequences of nepenthesins and related homologs were aligned and a phylogenetic tree was constructed by the program ClustalW [32]. The amino acid sequences of the homologs of nepenthesins were obtained by FASTA/BLASTP searches of the gene/protein data bases.

RESULTS

Purification of acid proteinases from the pitcher fluid

Upon DEAE-cellulose chromatography of the digestive fluid from the open pitchers of *Nepenthes distillatoria*, two acid proteinase activity peaks were obtained (Fig. 1A); the major peak was eluted at 0.25 M NaCl and the minor peak at 0.42 M NaCl. These enzyme fractions were separately purified by successive steps of chromatography on columns of Sephacryl S-200 (Figs. 1B and 1D), pepstatin-Sepharose and Mono Q (Figs. 1C and 1E). Upon pepstatin-Sepharose chromatography, the major enzyme was completely eluted with 0.05 M Tris/HCl, 1 M NaCl. On the other hand, the minor enzyme was eluted only partially (about 28%) with this buffer, and mostly (about 72%) with 0.05 M sodium borate buffer, pH 10, 1 M NaCl; these fractions were combined for further purification since they were apparently indistinguishable in other properties. Thus, about 1.8 mg of the purified major enzyme was obtained (200-fold purification in a yield of 21%), and about 0.46 mg of the purified minor enzyme (186-fold purification in a yield of 4.8%) (Table 2). Each of the purified enzymes gave a single protein band on SDS-PAGE (Fig. 2). Similar results were obtained with the digestive fluid from the unopen pitchers (data not shown). The major and minor enzymes were

Copyright 2004 Biochemical Society

designated nepenthesins I and II, respectively.

Molecular sizes of nepenthesins

The molecular mass of nepenthesin I was estimated to be approximately 45 kDa and 51 kDa by SDS-PAGE under non-reducing and reducing conditions, respectively, (Fig. 2, lanes 1 and 3) and 58 kDa by gel filtration. The protein band obtained by SDS-PAGE was positively stained with the periodic acid-Schiff's base reagent. On the other hand, the molecular mass of nepenthesin II was estimated to be approximately 35 kDa and 45 kDa by SDS-PAGE under non-reducing and reducing conditions, respectively (Fig. 2, lanes 2 and 4). The protein band on SDS-PAGE was negative to the periodic acid-Schiff's base reagent.

The NH_2-terminal and partial internal amino acid sequences of nepenthesins

The NH_2-terminal 24-residue and 19-residue sequences of nepenthesins I and II, respectively, were determined as follows.

Nepenthesin I : IGPSGVETTVYAGDGEYLMXLSIG
Nepenthesin II: QTVQVEPPYYAGDGEYLMV

In addition, partial internal sequences, including 134 and 158 residues for nepenthesins I and II, respectively, were determined at the protein level, which will be shown later.

Inhibition of nepenthesins by pepstatin and DAN

Nepenthesin I was strongly inhibited by pepstatin under acidic conditions like porcine pepsin A and complete inhibition was obtained at 0.1 mM pepstatin. The acid proteinase activity in the crude pitcher fluid was also inhibited completely by pepstatin under similar conditions. As can be seen from Fig. 3A, pepstatin appeared to bind to nepenthesin I in a 1:1 stoichiometry like porcine pepsin. Nepenthesin II was also shown to be inhibited completely with 0.1 mM pepstatin (data not shown).

Both enzymes were inhibited strongly by DAN in the presence of cupric ions (Fig. 3B) and nearly complete inhibition was obtained after 3 h. They were inactivated at similar rates, but

Copyright 2004 Biochemical Society

much more slowly than porcine pepsin A. In the absence of cupric ions, none of them were inactivated.

Cleavage specificity of nepenthesin toward oxidized insulin B chain

An HPLC pattern of a 3-h digest of oxidized insulin B chain at pH 3.0 by nepenthesin I is shown in Fig. 4A. Several peptide bonds were cleaved and especially the peptide bonds, Phe24-Phe25, Glu13-Ala14, Leu6-Cya7, Leu15-Tyr16 and Tyr16-Leu17, were cleaved to marked extents (Fig. 4B). The extents of cleavage of these bonds were estimated to be 80%, 67%, 50%, 38% and 33%, respectively, under the conditions used.

Effects of pH on the activity and stability of nepenthesins

The pH/activity profiles are shown in Fig 5. The optimal activity of each enzyme toward acid-denatured hemoglobin was observed at pH about 2.6; the crude pitcher fluid showed a similar pH/activity profile with an optimum pH at 2.8. It is notable that the purified enzymes as well as the crude fluid possess some activity at pH around 6.0.

The results obtained when the enzymes were incubated at 37 ^{o}C at various pH values (pH 3.0 - 10.0) are shown in Figs. 6A and 6B. Nepenthesin I was most stable at pH 3.0 and 95% of the original activity was retained after incubation for 30 days. Under the same conditions, the enzyme was considerably stable even at pH 10.0, where it retained 79% of the original activity after incubation for 30 days. The results obtained with the crude pitcher fluid were similar to those obtained with nepenthesin I. On the other hand, nepenthesin II was somewhat less stable. It was most stable at pH 3.0 and retained 85% of the original activity after 30 days, whereas the activity was completely lost at pH 5.0 and above in 30 days. Under the same conditions, porcine pepsin A was extremely unstable; it rapidly lost the activity over a wide range of pH. Porcine pepsin A was only stable at pH 5.0 for 7 days, where the enzyme is known to be most stable.

Copyright 2004 Biochemical Society

Effects of temperature on the activity and stability of nepenthesins

The temperature/activity profiles are shown in Fig.7A. The optimal temperature of nepenthesin I was 55 °C and above this temperature the activity gradually decreased and was lost completely at 80 °C. On the other hand, the optimal temperature of nepenthesin II was 45 °C and the activity was largely lost at 70 °C. The temperature /activity profile of the crude pitcher fluid was rather similar to that of nepenthesin I with an optimum at 50 °C. When the enzymes were incubated at pH 3.0 and at different temperatures for 1 h and then assayed at 37 °C, the results shown in Fig. 7B were obtained. Nepenthesin I as well as the activity in the crude fluid was stable up to 50 °C, and above this temperature it became unstable while under the same conditions nepenthesin II was less stable. These results are in good accord with those shown in Fig. 7A.

The results obtained when the enzymes were incubated at different temperatures at pH 3.0 for 7 days and 30 days, then assayed for the remaining activity, are shown in Figs. 8A and 8B. Nepenthesin I was very stable up to 30 days even at 50 °C. After 30 days at 50 °C, the enzyme retained 60% of the original activity. Similar results were obtained with the crude pitcher fluid. Nepenthesin II was also stable at various temperatures; at 50 °C it retained 44% of the original activity after 30 days. Under the same conditions, porcine pepsin A was markedly unstable; it retained only 10% of the original activity after 7 days at 37 °C, where nepenthesins I and II retained 96% and 90%, respectively, of the original activity.

Cellular localization of nepenthesin

The results of immunohistochemical staining of nepenthesin I are shown in Fig. 9. Parenchymal cells surrounding the secretory glands were stained purple with the antibody (Fig. 9A), which was not observed with the control sample without the primary antibody (Fig. 9B). Positive staining was not observed with the upper part of the pitcher and the bine portion connecting the pitcher with the leaf, both of which lack the secretory glands (data not shown). Under the conditions used, the secretory glands appeared as black dots and it was difficult to see how

Copyright 2004 Biochemical Society

much they were positively stained.

The amino acid sequences of nepenthesins I and II

The complete amino acid sequences of the prepro forms of nepenthesins I and II were deduced by cloning and nucleotide sequencing of the cDNA clones obtained from *Nepenthes gracilis* as shown in Figs. 10A and 10B. In Fig. 10 are also shown the NH_2-terminal and internal sequences determined at the protein level with the enzymes from *Nepenthes distillatoria*. The NH_2-terminus of each mature enzyme was deduced by comparison of the deduced amino acid sequence with the NH_2-terminal sequences of the *N. distillatoria* enzymes. Thus, prepro-nepenthesin I is composed of 437 residues, including a 24-residue putative signal sequence, a 56-residue putative propeptide and a 359-residue mature enzyme (nepenthesin I), and prepro-nepenthesin II is composed of 438 residues, including a 24-residue putative signal sequence, a 55-residue putative propeptide and a 359-residue mature enzyme (nepenthesin II). Nepenthesin I was found to have two variants: one (nepenthesin Ia) having Asp, Asn and Gly and the other (nepenthesin Ib) having Val, Thr and Glu at positions 233, 251 and 392, respectively (the numbering of the prepro-nepenthesin I is used throughout the text unless otherwise specified).

The molecular weights and isoelectric points of the mature enzymes are calculated to be 37,476 and 3.94, respectively, for nepenthesin Ia, and 37,519 and 3.94, respectively, for nepenthesin Ib and 37,511 and 3.09, respectively, for nepenthesin II. Nepenthesin I contains six potential N-glycosylation sites whereas nepenthesin II has none. Each enzyme contains the active site sequence motifs, Asp-Thr-Gly and Asp-Ser-Gly, the so-called flap tyrosine residue (residue 174 corresponding to Tyr75 in the porcine pepsin numbering) and notably twelve cysteine residues per molecule. Each lacks the plant-specific insertion typical of plant vacuolar APs at positions between Asn340 and Leu341, but appears to have an approximately 22-residue insertion (tentatively assigned to residues 148-169) preceding the flap tyrosine residue. This insertion contains four cysteine residues and was named ‘the nepenthesin-type AP

Copyright 2004 Biochemical Society

(NAP)-specific insertion.'

The partial amino acid sequences determined at the protein level for nepenthesin I (total 157 residues) and nephenthesin II (total 177 residues) from *N. distillatoria* were different at 8 and 11 positions from the deduced sequences of nepenthesins I and II from *N. gracilis*, respectively, thus the sequence identity of each enzyme from the two species being 94%. In addition, sequence variations, Ser/Tyr and Thr/Asp, were observed at position 109 in nepenthesin I and at position 148 in nepenthesin II, respectively, from *N. distillatoria*.

The tertiary structures of nepenthesins predicted by molecular modeling

Figure 11 shows the tertiary structure of nepenthesin Ia predicted by molecular modeling. The predicted backbone structures of nepenthesins Ib and II (not shown) were essentially the same as that of nepenthesin Ia. In the predicted structures based on homology modeling, the two cysteine residues in each of the two pairs: Cys45/Cys48 and Cys162/Cys356 were located close enough to form disulfide bonds (the cysteine residue numbering is based on the sequence of the mature enzyme). Therefore, disulfide bonds were created in the above pairs. The pairing of the four cysteine residues, Cys72, Cys77, Cys85 and Cys90, which are closely located in the NAP-specific insertion, could not be predicted since there was little structural information on this region. Therefore, the energy minimization calculation was performed with the two disulfide bonds formed but the four cysteine residues unconnected. In the resultant structure, the locations of the remaining four cysteine residues (Cys51, Cys125, Cys276, and Cys317) were still somewhat distant from each other, but the pair of Cys276/Cys317 was assumed to form a disulfide bond because porcine pepsin has a disulfide bond (Cys249-Cys282) at a similar location. Thus, the disulfide bonds were introduced between Cys276 and Cys317 and between Cys51 and Cys125. After the energy minimization, the location of the two disulfide bonds was shown to be reasonable, and there were no steric or energetic hindrances in the whole molecule. For the cysteine residues in the NAP-specific insertion, the most plausible pairings would be Cys72-Cys90 and Cys77-Cys85 if the insertion sequence be looped out from the rest of the

Copyright 2004 Biochemical Society

enzyme molecule. Figure 11 also shows the predicted disulfide bonds including the tentative pairs within the insertion sequence.

Figure 11B shows the tertiary structure of porcine pepsin A (33) for comparison and Fig. 11C the disulfide bond arrangements in the primary structures of nepenthesin I and porcine pepsin A. As compared with pepsin A, it appears that three additional disulfide bonds (Cys51-Cys125, Cys72-Cys90, and Cyc77-Cys85) are introduced into the NH_2-terminal lobe of nepenthesin, and one additional disulfide bond (Cys162-Cys356) between the NH_2- and COOH-terminal lobes and that one disulfide bond (Cys206-Cys210) in the COOH-terminal lobe of pepsin A is lost from nepenthesin.

Molecular phylogeny

The amino acid sequences of the prepro forms of nepenthesins and some of their typical homologs were aligned with those of some typical homologs to compare them and to construct a phylogenetic tree. For this comparison, we included eight nepenthesin-type aspatic proteinases with twelve conserved cysteine residues at similar positions (nepenthesins and ortholog enzymes from Arabidopsis thaliana, barley (nucellin), rice and tobacco (chloroplast nucleoid DNA binding protein, CND41), which have been identified at the protein and /or cDNA level. As for the Arabidopsis and rice enzymes, two typical ones with much different isoelectric points were selected. We also included six of the hitherto-known typical pepsin-type aspartic proteinases with six or four conserved cysteine residues at similar positions (pepsin A, cathepsin D, rhizopuspepsin and the plant vacuolar enzyme cyprosin, oryzasin and phytepsin). All these enzymes appeared to be roughly similar in size except that the three vacuolar APs had an additional sequence of approximately 100 residues, called a plant-specific insertion. The two active-site aspartic acid residues in the Asp-Thr-Gly and Asp-Ser/Thr/Cys-Gly motifs are conserved among all the enzymes. The tyrosine residue on the flap of porcine pepsin also appears to be conserved among them although we have to consider the presence of the NAP-specific insertion for the nepenthesin-type enzymes, which contains two putative disulfide

Copyright 2004 Biochemical Society

bonds. The sequence identities were calculated for the mature enzymes using the sequences corresponding to those of residue 17 to the COOH-terminal cysteine of nepenthesins after removing the NAP-specific insertion from the nepenthesins and their orthologs and the plant-specific insertion from the three plant vacuolar APs. The identities thus obtained are 67% between nepenthesins I and II, 23-38% (average, 30%) between nepenthesins and the rest of the nepenthesin-type enzymes, and 12-22% (average, 18%) between nepenthesins and the ordinary pepsin-type enzymes. It is also notable that the potential N-glycosylation sites are rich in nepenthesin I (6 sites), one of the Arabidopsis enzyme (accession no., AY088536) (5 sites) and one of the rice enzyme (accession no., AK106097) (5 sites). Based on the sequence information, a phylogenetic tree was constructed as shown in Fig. 12.

DISCUSSION

Two acid proteinases, nepenthesins I and II, were purified to homogeneity as examined by SDS-PAGE and NH_2-terminal amino acid sequencing from the pitcher fluid of *Nepenthes distillatoria*. Moreover, the primarry structures of the two enzymes were deduced by cloning and sequencing the corresponding cDNAs from *Nepenthes gracilis*. This is the first case of complete purification and primary structure determination of extracellular proteinases in the digestive fluid of a carnivorous plant. Very recently, Ann et al. have reported the cloning of APs from the pitcher tissue of *Nepenthes alata* [35]. However, these enzymes clearly belong to vacuolar APs since they share a so-called plant-specific insersion and lack the primary structure features characteristic of nepenthesins. In the present work, we used *N. distillatoria* for studies at the protein level and *N. gracilis* for those at the DNA level. Nepenthesins I and II from the former are thought to be essentially the same enzymes as the corresponding enzymes from the latter; the partial amino acid sequences of nepenthesins I and II from *N. distillatoria* determined at the protein level were 93-94% identical with the corresponding sequences of nepenthesins I and II from *N. gracilis*, respectively.

Nepenthesins I and II from *N. distillatoria* are significantly different from each other in

Copyright 2004 Biochemical Society

properties as judged from the differences in chromatographic behavior, especially on DEAE-cellulose, molecular mass, NH_2-terminal and partial internal amino acid sequence and effects of temperature and pH on the activity and stability. These results are consistent with the amino acid sequences deduced from the cDNAs for the enzymes from *N. gracilis*. The amino acid sequence identity between nepenthesins I and II from *N. gracilis* is 66.6%. Since nepenthesin I was eluted from the DEAE-cellulose column before nepenthesin II, the former should be less acidic than the latter. Indeed, the isoelectric points calculated from the amino acid composition of each enzyme are 3.94 and 3.09 for *N. gracilis* nepenthesins I and II, respectively. The approximate molecular masses of *N. distillatoria* nepenthesins I and II were estimated from SDS-PAGE to be 45 kDa and 35 kDa, respectively, under non-reducing conditions, and 51 kDa and 45 kDa, respectively, under reducing conditions. On the other hand, the calculated molecular masses of *N. gracilis* nepenthesins I and II were both 37.5 kDa. This discrepancy is thought to be due to the presence of carbohydrate in nepenthesin I, but not in nepenthesin II. This is consistent with the fact that six potential N-glycosylation sites are present in *N. gracilis* nepenthesin I, but none in *N. gracilis* nepenthesin II (see Fig. 10). The difference in molecular mass of 6-10 kDa between *N. distillatoria* nepenthesins I and II indicates that the carbohydrate content in nepenthesin I should be over 10% by weight. In the partial amino acid sequencing of *N. distillatoria* nepenthesin I, the Asn residues in the first and second potential N-glycosylation sites (residues 98 and 131) could not be positively identified at the protein level, while the Asn residues at the third and fourth sites (residues 98 and 167) were identified, although no information has yet been obtained for the remaining three sites at the protein level. Therefore, the results seem to suggest that at least the first and second sites are glycosylated and that the third and fourth sites are not.

As shown by immunohistochemical staining (Fig. 9), the enzymes appear to be synthesized in the parenchymal cells surrounding the secretory glands. However, the pitcher tissue was not so strongly stained. Furthermore, the extract of the pitcher tissues failed to show a significant level of acid proteinase activity. It is unlikely that these results are due to the occurrence of the

Copyright 2004 Biochemical Society

enzymes in the zymogen form in the tissue since they are thought to react with the polyclonal antibody like the mature enzyme and to be rapidly activated under the acidic assay conditions. Therefore, these results seem to indicate that the zymogens which are thought to be synthesized in the pitcher tissue are rapidly secreted into the pitcher fluid and activated without accumulation in the tissue although other possibilities cannot be completely excluded.

The AP-specific inhibitor pepstatin and DAN were shown to strongly inhibit the purified enzymes as porcine pepsin A. Previously, we showed that the acid proteinase activity in the crude pitcher fluid of *Nepenthes ampullaria* was strongly inhibited by pepstatin and DAN [11]. Thus, the previous finding was confirmed with the purified enzymes, indicating that the enzymes belong to the typical pepstatin-sensitive aspartic proteinase family. On pepstatin-Sepharose chromatography, nepenthesin II was more firmly bound to pepstatin-Sepharose than nepenthesin I. This suggests that nepenthesin II should be more sensitive to pepstatin than nepenthesin I.

The specificity of nepenthesin I toward oxidized insulin B chain appears to be similar, but somewhat wider than those of other APs such as pepsin A and cathepsin D; human pepsin A cleaved mainly the Leu 11-Val12, Leu15-Tyr16 and Phe25-Tyr26 bonds and rat cathepsin D the Leu15-Tyr16, Phe 24-Phe25 and Phe25-Tyr26 bonds [36]. Most peptide bonds susceptible to these enzymes were more or less cleaved by nepenthesin I. Interestingly, the Leu6-Cya7 bond was found to be one of the major sites of cleavage by nepenthesin I; the cleavage of this bond by other APs has not been reported. Leucine residue seems to be one of the preferred P1 site of the enzyme since three of the four Leu-X bonds in oxidized insulin B chain were more or less cleaved.

At or below 50 ^{o}C nepenthesin I was found to be extraordinarily stable in a wide range of pH for a long period. Nepenthesin II was also fairly stable, but less stable than nepenthesin I. Such an unusual stability, especially of nepenthesin I, has never, to our knowledge, been reported for other proteinases. The cDNA sequencing revealed that both nepenthesins have a high content of cysteine residues, twelve residues per molecule of protein, which would form six disulfide

Copyright 2004 Biochemical Society

bonds. APs with such a high disulfide bond content have not been known before except for plant vacuolar APs that have three additional disulfide bonds within the plant-specific insersion [38] which is absent in nepenthesins. The high content and specific pairing of the disulfide bonds should contribute greatly to the stability of both nepenthesins. Indeed, both nepenthesins are predicted to have three disulfide bonds, each linking two cysteine residues distantly located in the primary structure, i.e., one in the NH_2-terminal lobe, the second between the NH_2- and COOH-terminal lobes the third in the COOH-terminal lobe (Fig. 11). On the other hand, most of the ordinary pepsin-type APs have only one such disulfide bond in the COOH-terminal lobe. In nepenthesins, one disulfide bond present in the COOH-terminal lobe of porcine pepsin A is absent; this would not much affect the stability since this disulfide bond is formed between two nearby cysteine residues and is not always conserved in ordinary pepsin-type APs, such as rhizopuspepsin and penicillopepsin.

The presence of carbohydrate and a higher isoelectric point of nepenthesin I are thought to contribute to render nepenthesin I much more stable than nepenthesin II. The carbohydrate moieties may help increase the stability by reducing the possibility of autolysis and/or denaturation. On the other hand, nepenthesin I (isoelectric point, 3.94) has twenty acidic residues (12 Asp and 8 Glu) and five basic residues (1 Lys, 1 His and 3 Arg) while nepenthesin II (isoelectric point, 3.09) has thirty three acidic residues (18 Asp and 15 Glu) and one basic residue (1 Lys). Therefore, the charge repulsion among the dissociating carboxylate groups that will lead to denaturation should be less pronounced in nepenthesin I than in nepenthesin II as the pH is raised. Thus, these enzymes, especially nepenthesin I, appear to be so designed that they are able to remain stable in a wide range of pH at relatively high temperatures to work for a long period in the pitcher fluid of the *Nepenthes* plants in tropical habitat, indicating an evolutionary adaptation of the enzymes to their specific environments. To extend further the structure/function studies, it is absolutely necessary to elucidate the three-dimensional structure of the enzymes.

As can be seen from the phylogenetic tree (Fig. 12), nepenthesins clearly belong to a novel

Copyright 2004 Biochemical Society

subfamily of APs. Previously, Chen and Foolad [38] reported that the barley gene encoded an AP-like protein with no plant-specific insertion, named it nucellin and suggested that it may be involved in nucellar cell death, although no studies were performed at the protein level. On the other hand, Nakano et al. [39] isolated a similar 41-kD AP-like protein with DNA binding activity from the chloroplast nucleoids of tobacco cells, named it CND41 and suggested its possible function as a negative regulator of chloroplast gene expression. Later, this protein was purified and shown to have proteolytic activity optimally at acidic pH, which, however, was reported to be only weakly inhibited by pepstatin. [40]. The elucidation of the primary structures of nepenthesins in this study has shown that nepenthesins, nucellin and CND41 belong to the same type of APs, and facilitated us to make further search for the same types of enzymes. FASTA and BLASTP homology searches for nepenthesins in various data bases revealed that at the genomic DNA level, there are nearly ninety and thirty orthologs in *Arabidopsis thaliana* and *Oryza sativa*, respectively, two in *Nicotinia tabacum*, one in *Zea mays*, and one in *Hordeum vulgar*. So far the cDNAs have been obtained for sixteen orthologs in *Arabidopsis thaliana*, six in *Oryza sativa*, one in *Nicotinia tabacum* (CND41) and one in *Hordeum vulgar* (barley nucellin).

These nepenthesin-type APs have the NAP-specific insertion, but no plant-specific insertion, and many of them have 12 cysteine residues at the positions corresponding to those in nepenthesins, while some others lack one or few of these cysteine residues. The NAP-specific insertion mostly contains four (and occasionally a fewer) cysteine residues, and those of the six nepenthesin-type enzymes (Fig. 12) are shown below. Among them, there are significant variations in sequence and it is tempting to assume that they may play some specific role(s) as involved in intracellular targeting or functional regulation of the enzymes.

(1) Nepenthesin Ia: LP**C**SSQL**C**QALSSPT**C**SNNF**C**Q
(2) Nepenthesin II: LP**C**ESQY**C**QDLPSET**C**NNNE**C**Q
(3) Arabidopsis 1: LT**C**SAPQ**C**SLLETSA**C**RSNK**C**L
(4) Arabidopsis 2: VF**C**NSST**C**QDLVAATSNSGP**C**GGNNGVVKTP**C**E
(5) Nucellin: VV**C**GSPL**C**VAVRRDVPGIPE**C**SRNDPHR**C**H
(6) Rice 1: VP**C**ANSI**C**TALHSGSSPNKK**C**TTQQQ**C**D

Copyright 2004 Biochemical Society

(7) Rice 2: VP**C**NASA**C**EASLKAATGVPGS**C**ATVGGGGGGGKSER**C**Y
(8) CND41: IS**C**TSAA**C**SSLKSATGNSPG**C**SSSN**C**V

It is also interesting to note that the putative mature forms of these enzymes can be classified into two groups with different isoelectric points as in the case of nepenthesins although their isoelectric points are generally higher than those of nepenthesins: one group is weakly acidic with an isoeletric point of ca.4-6 and the other group is basic with an isoelectric point of ca.8-10. These differences might be correlated with their physiological roles.

Construction of a phylogenetic tree including all of these over hundred orthologs (K. Takahashi, unpublished) have shown that they all belong to the nepenthesin-type APs distinct from hitherto-known ordinary pepsin-type APs. Thus, nepenthesins and those orthologs constitute a novel subfamily of APs, suggesting their ubiquitous distribution and multiple roles in the plant kingdom. More APs of this type will be found in other plants as the gene sequencing progresses. In so far as examined, an orthologous enzyme appears to be present even in *Chlamydomanas*, though not in Cyanobacteria. Among these enzymes, nepenthesins are the only known extracellular enzymes, and the others seem to be intracellular enzymes. Therefore, nepenthesins are thought to have adapted specifically to become extracellular digestive enzymes during the course of molecular evolution. The physiological roles of the intracellular nepenthesin-type APs remain to be elucidated.

The authors are grateful to Dr. Masao Tanji (Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo) for his help in the preliminary studies on carnivorous plant acid proteinases, and to Prof. Natalia S. Andreeva (Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow) and Prof. Carlos Faro (Department of Biochemistry, University of Coimbra, Portugal) for helpful discussions on the sequence alignment regarding the flap tyrosine. This study was supported in part by grants-in-aid for scientific research from the Ministry of Education, Science, Culture and Sports of Japan and the Japan Society for Promotion of Science and by a grant (NSF/BT/97/03) from National Science Foundation, Sri Lanka.

Copyright 2004 Biochemical Society

REFERENCES

1 Darwin, C. (1875) Insectivorous Plants, John Murray, London

2 Lloyd, F. E. (1942) The Carnivorous Plants, Ronald Press, New York

3 Hooker, J. D. (1874) The carnivorous habits of plants. Nature **10**, 366-372

4 Frazer, C. K. (2000) The enduring controversies concerning the process of protein digestion in *Nepenthes* (Nepenthaceae). Carniv. Pl. Newslett. **29**, 56-62

5 Woessner, J. F. (1998) Nepenthesin. In Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D. and Woessner, J. F., eds.) pp. 846-848, Academic Press, San Diego

6 Nakayama, S. and Amagase, S. (1968) Acid protease in *Nepenthes*: Partial purification and properties of the enzyme. Proc. Japan Acad., **44**, 358-362

7 Amagase, S., Nakayama, S. and Tsugita, A. (1969) Acid protease in *Nepenthes* . II. Study on the specificity of nepenthesin. J. Biochem. (Tokyo) **66**, 431-439

8 Amagase, S. (1972) Digestive enzymes in the insectivorous plants. III. Acid proteases in the genus *Nepenthes* and *Drosera peltata*. J. Biochem. (Tokyo) **72**, 73-81

9 Lobareva, L. S., Rudenskaya, G. N. and Stepanov, V. M. (1973) Pepsin-like protease from the insectivorous plant *Nepenthes*. Biokhimiya **38**, 640-642

10 Jenttsch, J. (1972) Isolation of the peotease nepthacin. FEBS Lett. **21**, 273-276

11 Takahashi, K., Chang, W.-J. and Ko, J.-S. (1974) Specific inhibition of acid proteases from brain, kidney, skeletal muscle and insectivorous plants by diazoacetyl-D,L-norleucine methyl ester and by pepstatin. J. Biochem (Tokyo) **76**, 897-899

12 Tökés, Z. A., Woon, W. C. and Chambers, S. M. (1974) Digestive enzymes secreted by the carnivorous plant *Nepenthes macfarianei*. Planta **119**, 39-46

13 Rawlings, N. D. and Barrett, A. J. (1995) Families of aspartic peptidases, and those of unknown catalytic mechanism. Methods Enzymol. **248**, 105-120

14 Barrett, A. J., Rawlings, N. D. and Woessner, J. F. (eds.) (1998) Aspartic peptidases. In Handbook of Proteolytic Enzymes, pp. 799-986, Academic Press, San Diego

Copyright 2004 Biochemical Society

15 Kervinen, J., Törmäkangas, K., Runeberg-Roos, P., Guruprasad, K, Blundell, T. and Teeri, T.H. (1995) Structure and possible function of aspartic proteinases in barley and other plants. In Aspartic Proteinases: Structure, Function, Biology, and Biomedical Implications (Takahashi, K., ed.) pp. 241-254, Prenum Press, New York

16 Kervinen, J. (1998) Phytepsin. In Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D. and Woessner, J. F., eds.) pp. 836-839, Academic Press, San Diego

17 Runeberg-Roos, P., Törmäkangas, 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

18 Asakura, T., Watanabe, H., Abe, K. and Arai, S. (1995) Rice aspartic proteinase, oryzasin, expressed during seed ripening and germination, has a gene organization distinct from those of animal and microbial aspartic proteinases. Eur. J. Biochem. **232**, 77-83

19 Cordeiro, C. M., Brodelius, P. E. and Pais, S. M. (1998) Cyprosin. In Handbook of Proteolytic Enzym*es* (Barrett, A. J., Rawlings, N. D., and Woessner, J. F., eds.) pp. 839-842, Academic Press, San Diego

20 Cordeiro, M., Xue, Z-T., Pietrzak, M., Pais, M.S. and Brodelius, P. (1995) Plant aspartic proteinases from *Cynara cardunculus* spp. *flavescens* cv. Cardoon: nucleotide sequence of a cDNA encoding cyprosin and its organ-specific expression. In Aspartic Proteinases: Structure, Function, Biology and Biomedical Implications (Takahashi, K., ed.) pp.367-372, Plenum Pess, New York

21 Pires, E. M. V. (1998) Cardosin A. In Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D.and Woessner, J. F., eds.) pp. 843-844, Academic Press, San Diego

22 Pires, E. M. V. (1998) Cardosin B. In Handbook of Proteolytic Enzymes (Barrett, A. J., Rawlings, N. D. and Woessner, J. F., eds.) pp. 844-846, Academic Press, San Diego

23 Faro, C., Verissimo, P., Lin, Y., Tang, J. and Pires, E. (1995) Cardosin A and B, aspartic proteinases from the flowers of cardoon. In Aspartic Proteinases. Structure, Function, Biology and Biomedical Implications (Takahashi, K., ed.) pp.374-377, Plenum Pess, New

Copyright 2004 Biochemical Society

York

24 Athauda, S. B. P., Tanji, M., Kageyama, T. and Takahashi, K. (1989) A comparative study on the NH_2-terminal amino acid sequences and some other properties of six isozymic forms of human pepsinogens and pepsins. J. Biochem.(Tokyo) **106**, 920-927

25 Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. Anal. Biochem. **150**, 76-85

26 Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**, 680-685

27 Zacharius, R. M., Zell, T. E., Morrison, J. H. and Woodlock, J. J. (1969) Glycoprotein staining following electrophoresis on acrylamide gels. Anal. Biochem. **30**, 148-152

28 Crestfield, A. M., Moore, S. and Stein, W. H. (1963) The preparation and enzymatic hydrolysis of reduced and S-carboxymethylated proteins. J. Biol. Chem. **238**, 622-627

29 Wan, C.-Y. and Wilkins, T. A. (1994) A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). Anal. Biochem. **223**, 7-12

30 Sali, A. and Blundell, T.L. (1993) Comparative protein modeling by satisfaction of spatial restraints. J. Mol. Biol. **234**, 779-815

31 Cooper, J.B., Khan, G., Taylor, G., Tickle, I.J. and Blundell, T.L. (1990) X-ray analyses of aspartic proteinases. II. Three-dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 A resolution. J. Mol. Biol. **214**, 199-222.

32 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**, 4673-4680

33 Andreeva, N. S. and Rumsh, L. D. (2001) Analysis of crystal structures of aspartic proteinases: On the role of amino acid residues adjacent to the catalytic site of pepsin-like

Copyright 2004 Biochemical Society

enzymes. Prot. Sci. 10, 2439-2450

34 Ferrin, T. E., Huang, C. C., Jarvis, L. E. and Langridge, R. (1988) The MIDAS display system. J. Mol. Graph. **6**, 13-27.

35 An, C.-I., Fukusaki, E. and Kobayashi, A. (2000) Aspartic proteinases are expressed in pitchers of the carnivorous plant *Nepenthes alata* Blanco. Planta **214**, 661-667

36 Matsuzaki, O. and Takahashi, K. (1988) Improved purification of slow moving protease from human gastric mucosa and its action on the B chain of oxidized bovine insulin. Biomed. Res. **9**, 515-523

37 Kervinen, J., Tobin, G. J., Costa, J., Waugh, D. S., Wlodawer, A. and Zdanov, A. (1999) Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. EMBO J. **18**, 3947-3955

38 Chen, F. and Foolad, M. R. (1997) Molecular organization of a gene in barley Which encodes a protein similar to aspartic protease and its specific expression in nucellar cells during degeneration. Plant Mol. Biol. **35**, 821-831

39 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

40 Murakami, S., Kondo, Y., Nakano, T. and Sato, F. (2000) Protease activity of CND41, a chloroplast nucleoid DNA-binding protein, isolated from cultured tobacco cells. FEBS Lett., **468**, 15-18

Copyright 2004 Biochemical Society

Table 1 Primers used for cDNA cloning of nepenthesins I and II

Primer	Orientation	Sequence (5' -> 3')	Target	Location
Nep1s	forward	TAYGCNGGNGAYGGNGARTA	I	265-284
			II	268-287
Maj3a	reverse	AARTGCATNACRAANGTNGG	I	1099-1118
Maj2s	forward	ATHTGGACNCARTGYCARCC	I	355-374
Maj1a	reverse	CATNACRAANGTNGGDATYTG	I	1093-1113
Maj7a	reverse	TGATCCCAATAGAAGAGTGC	I	749-768
Maj7s	forward	ATGACCCCCATTGGTAGC	I	721-738
ASP16-RT-P	reverse	AAGGRAGGGTRGAGAAGGA	I	430-448
			II	433-451
Maj12s	forward	TGGATACCGGTAGCGATCTT	I	335-354
Maj13a	reverse	TTGCGGAGAAAGGTTGTGCT	I	312-331
Maj14a	reverse	GTTCCAATCGATAAGTTCATC	I	288-308
Min1a	reverse	SWNGGYTCNCCRAARCANGT	II	1057-1076*
Min2s	forward	ATHTGGACNCARTGYGARCC	II	358-377

Copyright 2004 Biochemical Society

Min4a	reverse	ACNGTYTGNCCRCANCCRAA	II	604-623*
Min8a	reverse	GTGAAGGTCTCGGTTGCC	II	555-572
Min11s	forward	AAACCTGCAATAATAATGAA	II	485-504
Min12s	forward	TGGATACCGGCAGTGATCTC	II	338-357
Min13a	reverse	TGGCCGAGAAAGAACTATCC	II	315-334
Min14a	reverse	GGAGTACCAATTGCTACGTTC	II	294-314

D, H, N, R, S, W, and Y indicate degenerate sites (D, A/G/T; H, A/C/T; N, A/C/G/T; R, A/G S, C/G; W, A/T; and Y, C/T). *There are some differences in sequence between the designed primer and the obtained cDNA.

Copyright 2004 Biochemical Society

Table 2 Purification of acid proteinases from Nepenthes pitcher fluid

Step	Enzyme	Total protein	Total activity	Specific activity	Yield	Purification
		mg	units	units/mg	%	-fold
Crude fluid		1770	7710	4.36	100	1
1st DE-52		500	5580	11.2	72.0	2.6
2nd DE-52	I	99.0	2600	26.3	34.0	6.0
	II	18.0	1140	63.3	14.8	14.5
Sephacryl S-200	I	13.6	2080	153	27.0	35.1
	II	3.70	790	214	9.1	49.1
Pepstatin-Sepharose	I	2.52	1810	718	23.5	165
	II	0.98	500	510	6.5	117
FPLC (Mono Q)	I	1.83	1600	874	21.0	200
	II	0.46	372	809	4.8	186

Copyright 2004 Biochemical Society

Legends to Figures

Figure 1 Purification of *Nepenthes* acid proteinases

(A) DEAE-cellulose chromatography. The enzyme fraction eluted from the first DEAE-cellulose column was rechromatographed Fraction size, 10 ml. (B) Sephacryl S-200 chromatography of the major enzyme from the second DEAE-cellulose column. Fraction size, 11 ml. (C) Mono Q chromatography of the major enzyme. The enzyme fraction obtained by successive chromatography on Sephacryl S-200 and pepstain-Sepharose was chromatographed. The figure shows the result obtained with one-third of the enzyme sample. (D) Sephacryl S-200 chromatography of the minor enzyme. (E) Mono Q chromatography of the minor enzyme. The conditions were the same as in (C) except that the whole sample was used. In each chromatography, the fractions under the bar were pooled. The purified major and minor enzymes were designated nepenthesins I and II, respectively.

Figure 2 SDS-PAGE of the purified nepenthesins I and II

Electophoresis was performed under non-reducing (lanes 1 and 2) and reducing (lanes 3 and 4) conditions. The protein was stained with Coomassie Brilliant Blue. Lanes 1 and 3, nepenthesin I; lanes 2 and 4, nepenthesin II.

Figure 3 Inhibition of nepenthesins by pepstatin and DAN

(A) Pepstatin inhibition of nepenthesin I (●) at pH 3.0. Inhibition of porcine pepsin A (□) by pepstatin was also examined under the same conditions. (B) DAN inhibition of nepenthesin I in the presence (−●−) and absence (--●--) of cupric ions, and of nepenthesin II in the presence (−○−) and absence (--○--) of cupric ions. Porcine pepsin A was treated with DAN in the same manner in the presence (−□−) and absence (--□--) of cupric ions.

Figure 4 Action of nepenthesin I on oxidized insulin B chain

(A) HPLC of a 3-h digest of oxidized insulin B chain with the purified enzyme at pH 3.0 and 37

Copyright 2004 Biochemical Society

^{o}C. The numbers on each peak indicate the residue positions. (B) Cleavage sites and yields of peptides. The number on each horizontal bar indicates the yield (%) of the corresponding peptide. Large and small arrowheads show the major and minor cleavage sites, respectively. The number under each arrowhead indicates an estimated extent (%) of cleavage. Cya, cysteic acid.

Figure 5 pH dependence of the activity nepenthesins toward hemoglobin

The activity was measured at various pH values at 37°C toward acid-denatured hemoglobin.

Figure 6 Effects of temperature on the activity and stability of nepenthesins

(A) The activity was measured at various temperatures at pH 3.0. (B) The activity was measured at pH 3.0 and 37°C after incubation of each enzyme at various temperatures at pH 3.0 for 2 h. −●−, nepenthesin I; −○−, nepenthesin II; --▲--, crude pitcher fluid.

Figure 7 Stability of nepenthesins at different pH values

The activity was measured after incubation at 37°C for 7 days (A) and 30 days (B) at different pH values. The first (closed) bar, crude pitcher fluid; the second (light) bar, nepenthesin I; the third (dark) bar, nepenthesin II; and the fourth (open) bar, porcine pepsin A.

Figure 8 Stability of nepenthesins at different temperatures

The activity was measured after incubation at pH 3.0 for 7 days (A) and 30 days (B) at different temperatures. The same notations as in Fig. 7 are used.

Figure 9 Immunohistochemical staining of *Nepenthes* pitcher tissue with an anti-nepenthesin I antibody

Longitudinal sections of the lower 1/3 part of the *Nepenthes* pitchers tissue were stained with the specific antibody against nepenthesin I (A) and without the specific antibody (control) (B).

Copyright 2004 Biochemical Society

Magnification, x 400.

Figure 10 The nucleotide and the deduced amino acid sequences of nepenthesin Ia (A) and nepenthesin II (B) from *N. gracilis*

The putative active site aspartic acids and the flap tyrosine are shown with an asterisk. The potential *N*-glycosylation sitemotifs are underlined. Under the sequences are also shown the partial amino acid sequences of peptides obtained from the *N. distillatoria* enzymes determined at the protein level. Peptide numbers are shown in parenthesis before or after each sequence. N-Term, *N*-terminal sequence determined with the whole protein; DN, endopeptidase Asp-N; T, trypsin; and V, *Staphylococcus aureus* V8 protease.

Figure 11 The backbone structure and disulfide bond pairing of nepenthesin Ia predicted by molecular modeling and comparison with porcine pepsin A

The side chains of the cysteine residues and the putative catalytic aspartic acid residues are also presented in (A) and (B). (A) The backbone structure of *N. gracilis* nepenthesin Ia. All the cysteine residues are assumed to form disulfide bonds. Cys45-Cys48 is colored in orange, Cys51-Cys125 in green, Cys72-Cys90 in light green, Cys77-Cys85 in blue, Cys162-Cys356 in cyan, and Cys276-Cys317 in magenta. The catalytic residues, Asp35 (left) and Asp237 (right), are colored in red. (The numbering used is for the mature nepenthesin Ia.) (B) The backbone structure of porcine pepsin A. Cys45-Cys50, Cys206-Cys210, and Cys249-Cys282 (pepsin numbering) are shown in orange, light green, and magenta, respectively, and the catalytic residues, Asp32 (left) and Asp215 (right), are shown in red. (C) Predicted arrangements of the disulfide bonds in nepenthesin Ia (upper row) and porcine pepsin A (lower row). The cysteine residues and the disulfide bonds are shown in the same colors as in (A) and (B). The catalytic aspartic acid residues are shown with an asterisk.

Figure 12 A phylogenetic tree for nepenthesins and related typical aspartic proteinases

Copyright 2004 Biochemical Society

The branches of the nepenthesin-type APs are shown in bold line and the ordinary pepsin-type APs in thin line. The phylogenetic tree was produced by ClustalW using the sequence of residues 95-434 of prepro-nepenthesin Ia and the corresponding sequences of the other APs, after removing the NAP-specific insertion and the plant-specific insertion. The sequences of the following twelve APs are compared: (1) Nepenthesin Ia, *N. gracilis* AP Ia (AB114914); (2) Nepenthesin II, N. gracilis AP II (AB114915); (3) Arabidopsis 1, *A. thaliana* AP (AF37029) (Ip, 4.3); (4) Arabidopsis 2, *A. thaliana* AP (AY088536) (Ip, 9.5): (5) Nucellin, *Hordeum vulgare* (barley) AP (U87148); (6) Rice 1, *Oryza sativa* AP (AK068348) (Ip, 8.7); (7) Rice 2, *O. sativa* AP (AK165097) (Ip, 5.3); (8) CND41, *Nicotiana tabacum* (tobacco) AP (D26015); (9) Cyprosin, *Cynara caldunclus* AP (X81984); (10) Oryzasin, *O. sative* AP (D32114); (11) Phytepsin, *H. vulgare* (barley) AP (X56136); (12) Cathepsin D, porcine cathepsin D (M11233); (13) Pepsin A, porcine pepsin A (M20920); and (14) Rhisopuspepsin, *Rhizopus chinensis* AP (P06026). The number in parenthesis is the accession number. Ip, isoelectric point of the mature enzyme.

Copyright 2004 Biochemical Society

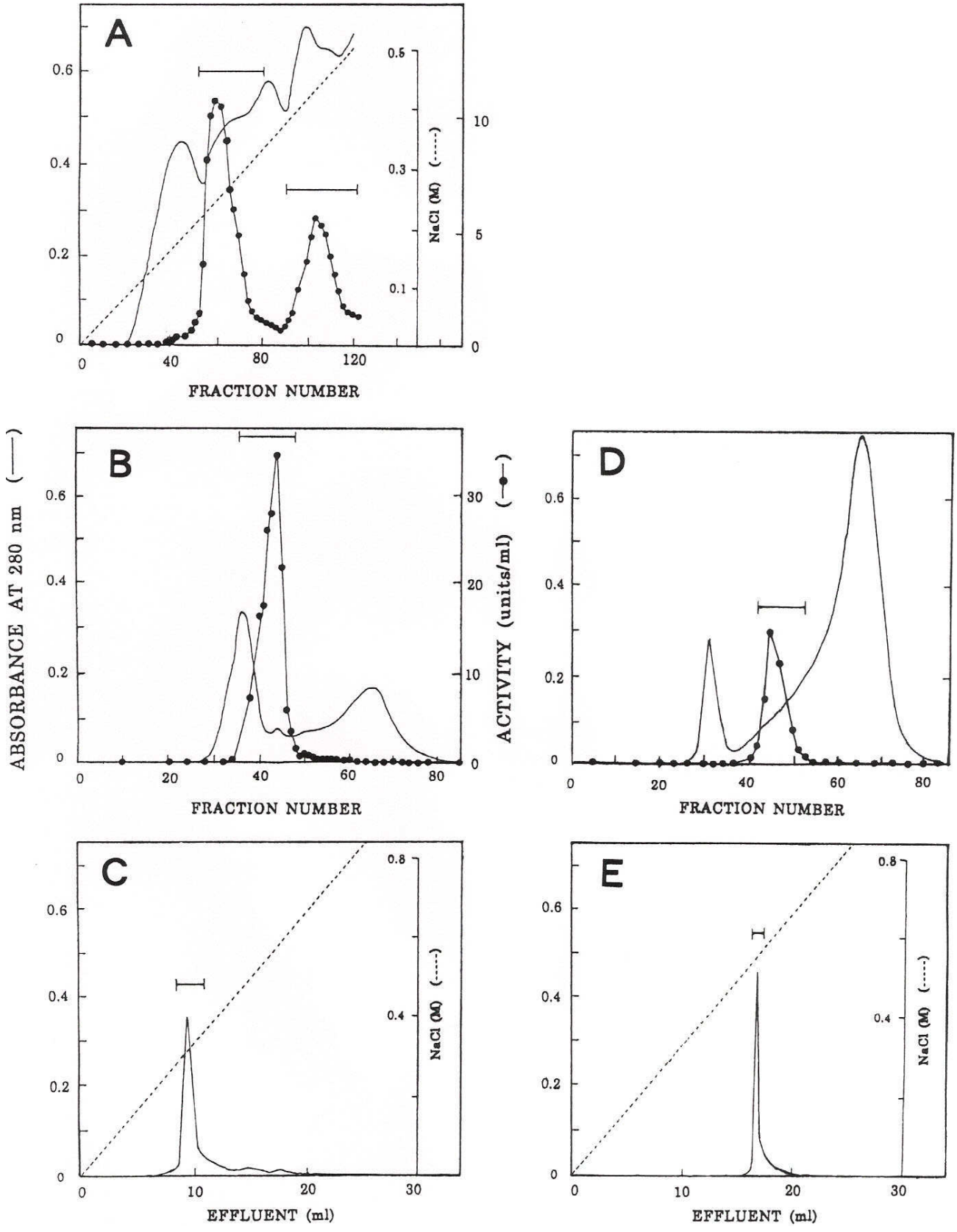


Fig. 1

Copyright 2004 Biochemical Society

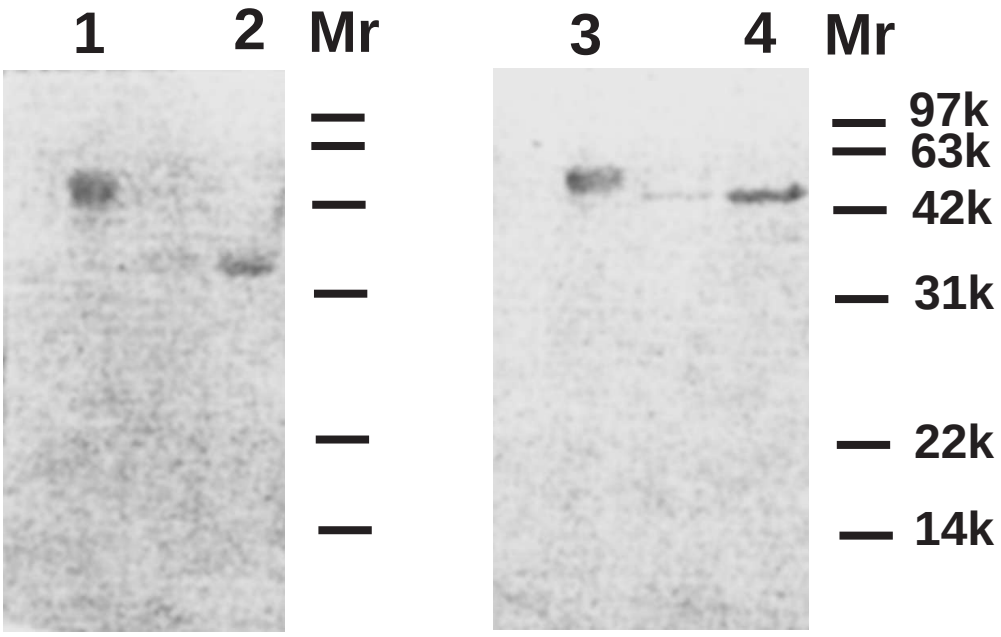
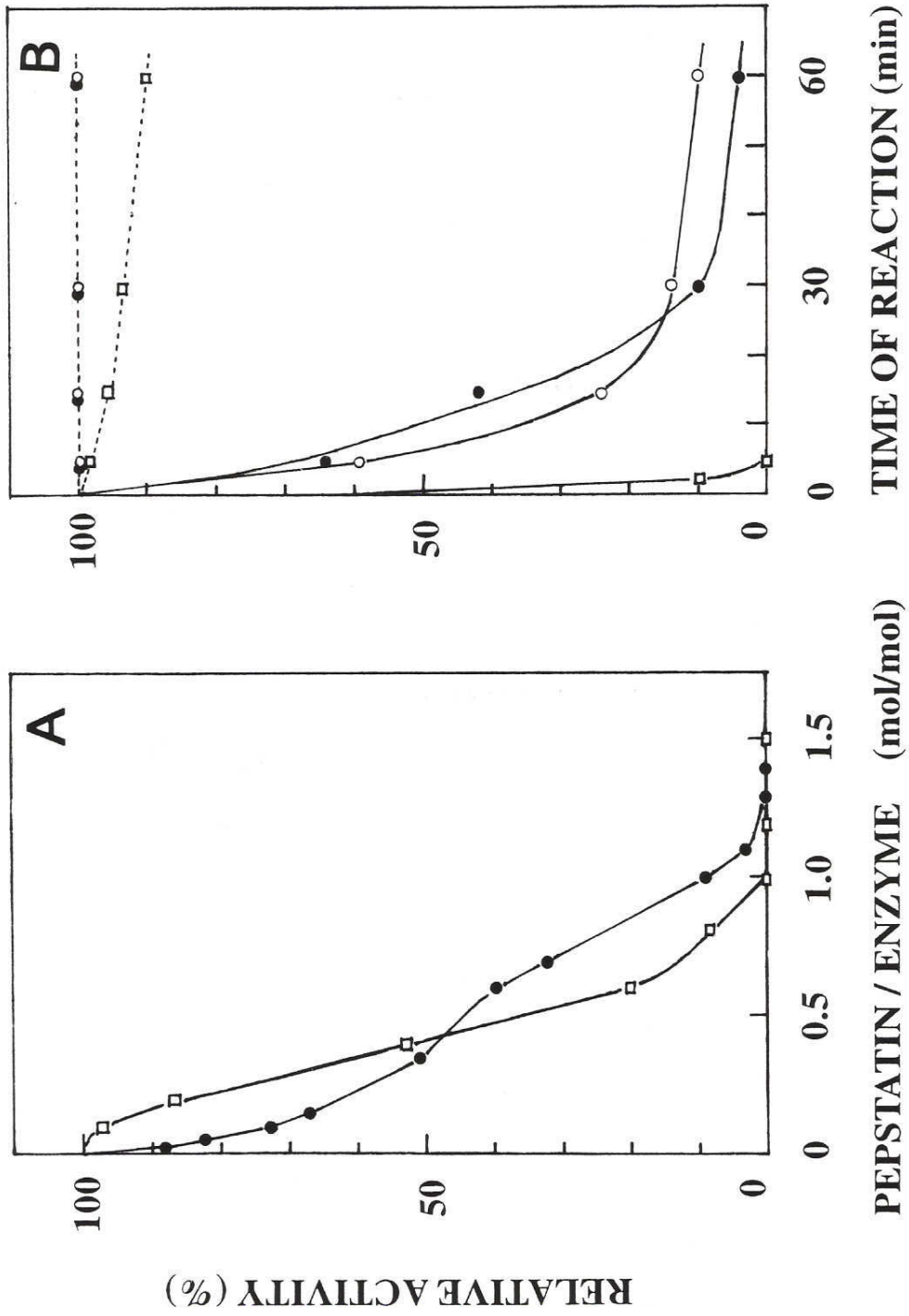


Fig. 2

Copyright 2004 Biochemical Society

Fig. 3



Copyright 2004 Biochemical Society

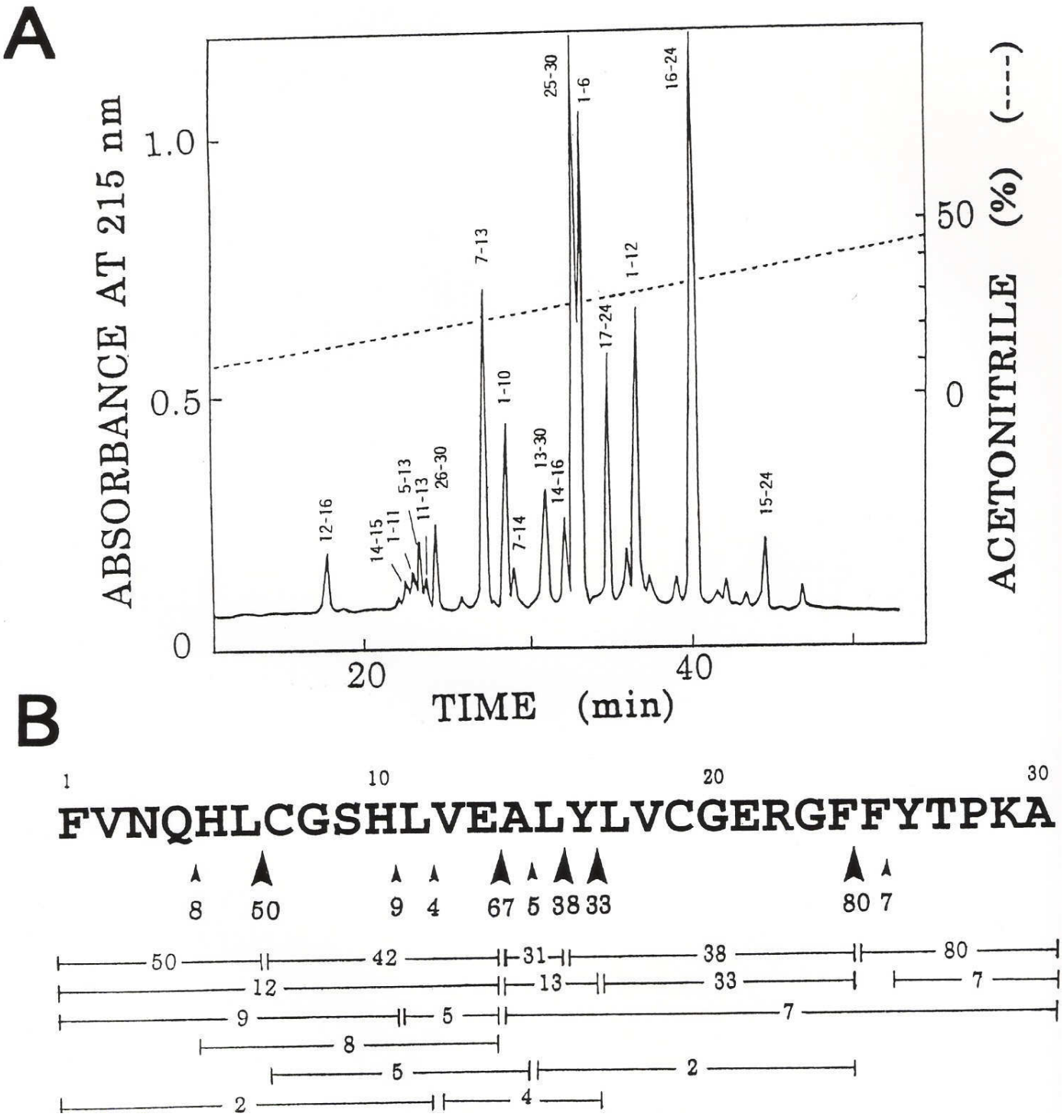


Fig. 4

Copyright 2004 Biochemical Society

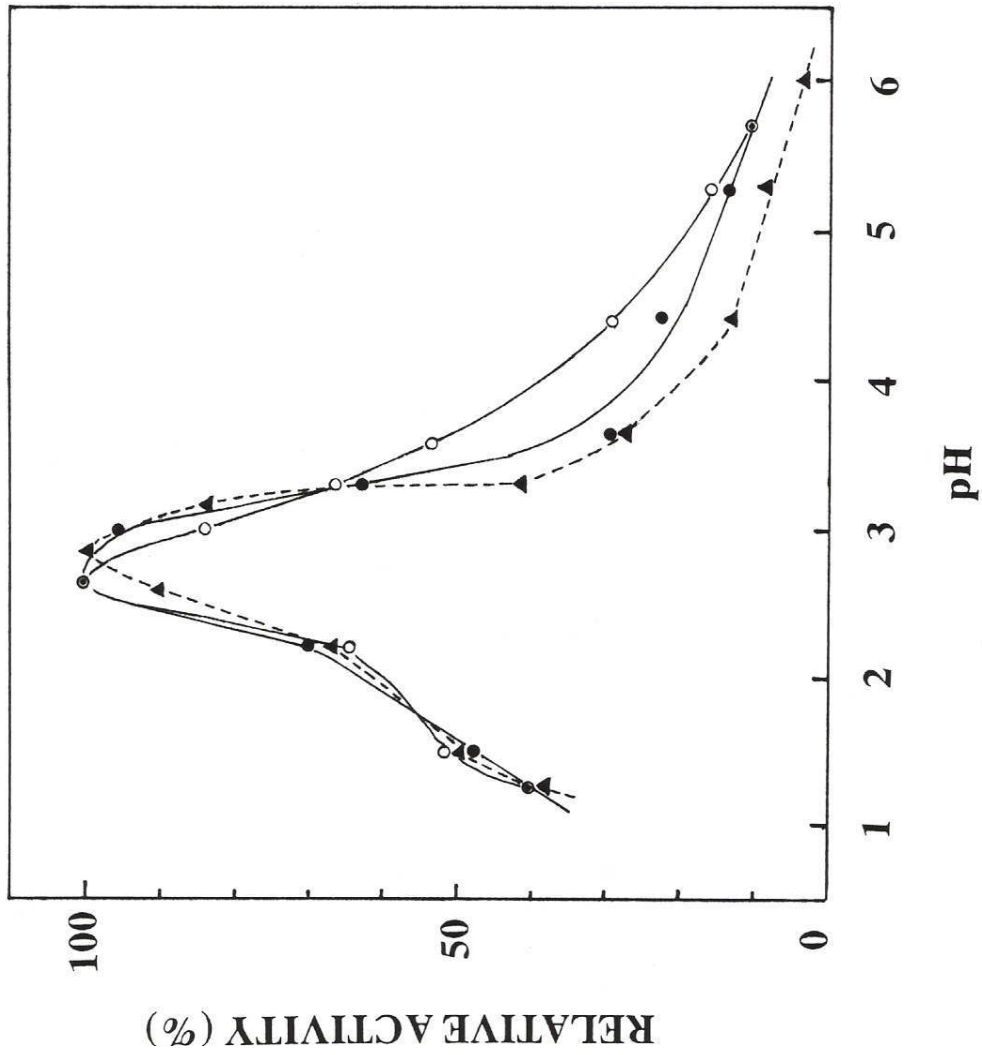


Fig. 5

Copyright 2004 Biochemical Society

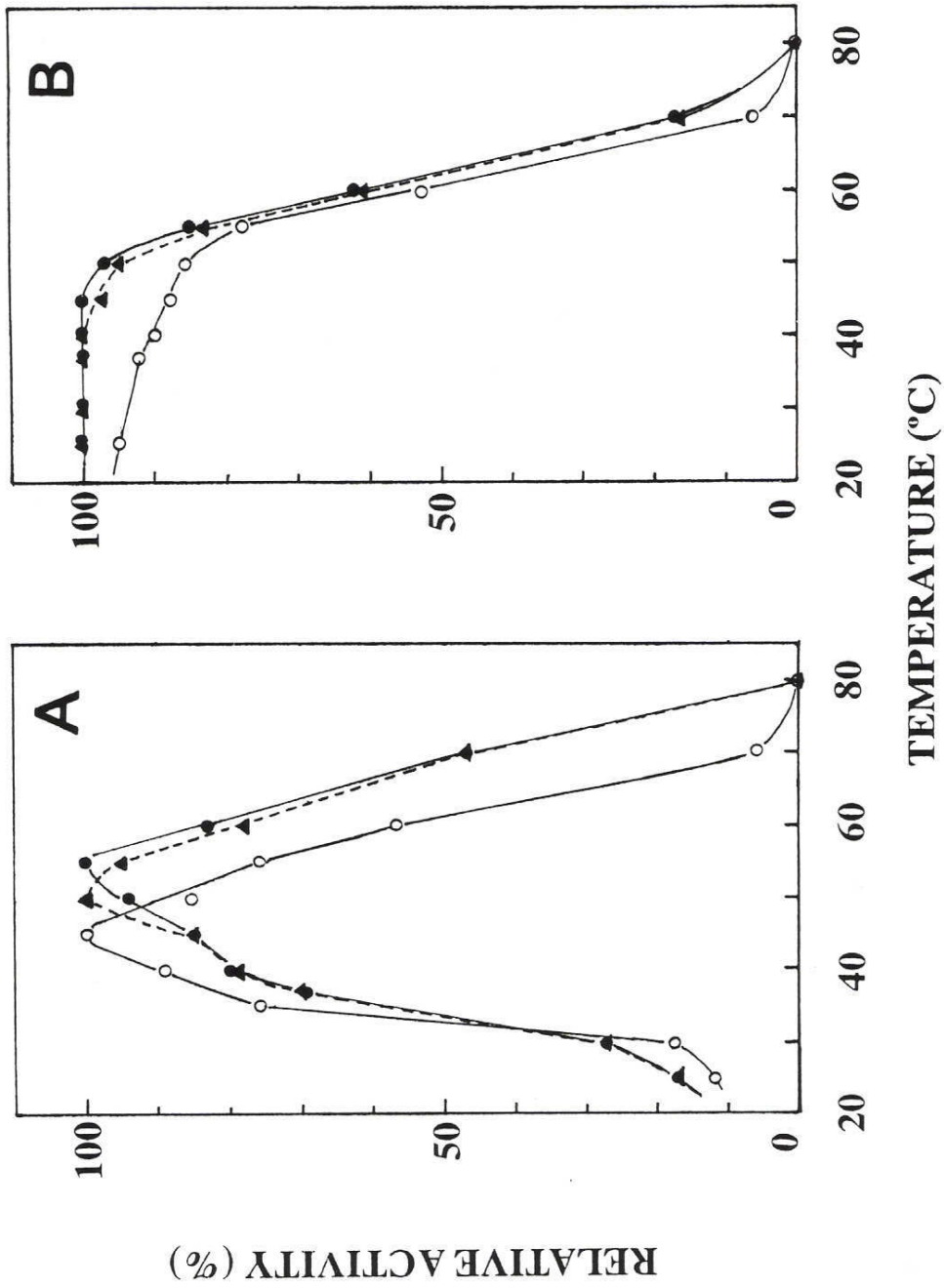


Fig. 6

Copyright 2004 Biochemical Society

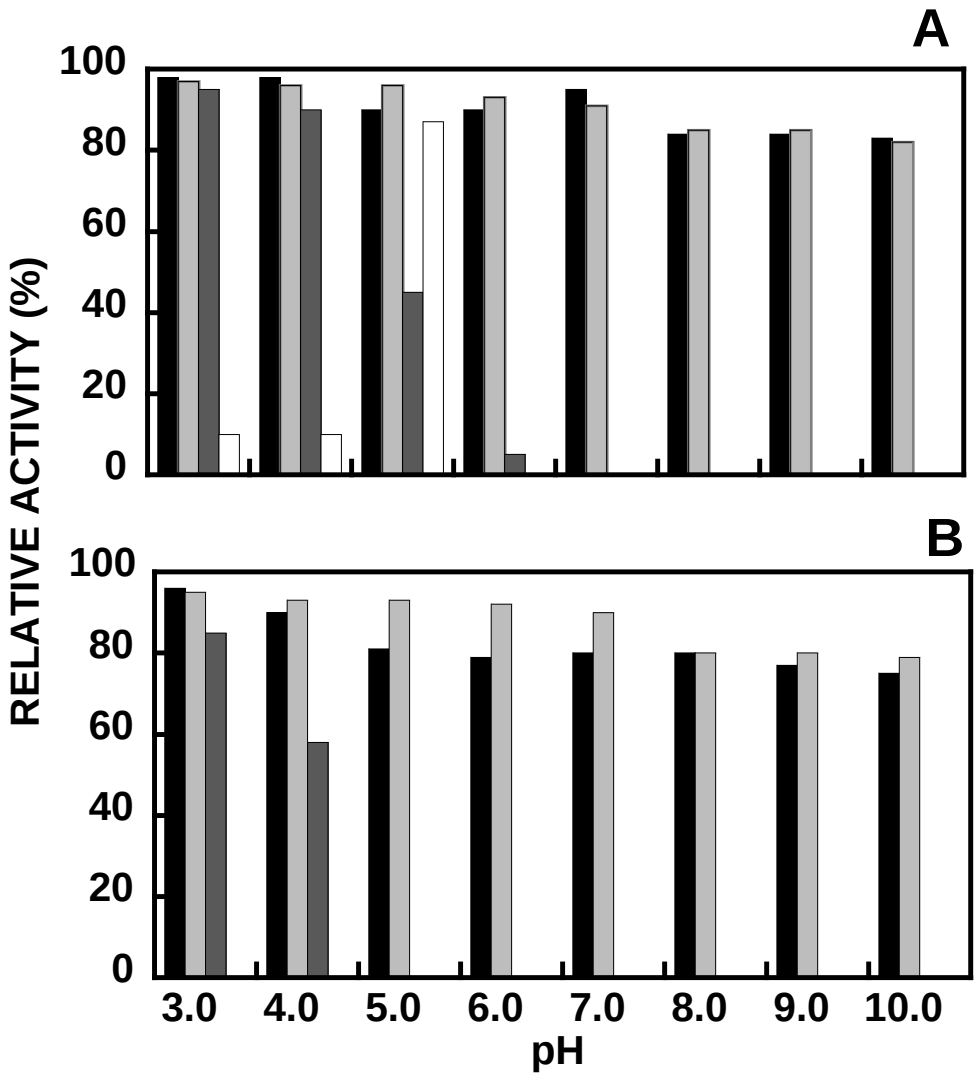


Fig. 7

Copyright 2004 Biochemical Society

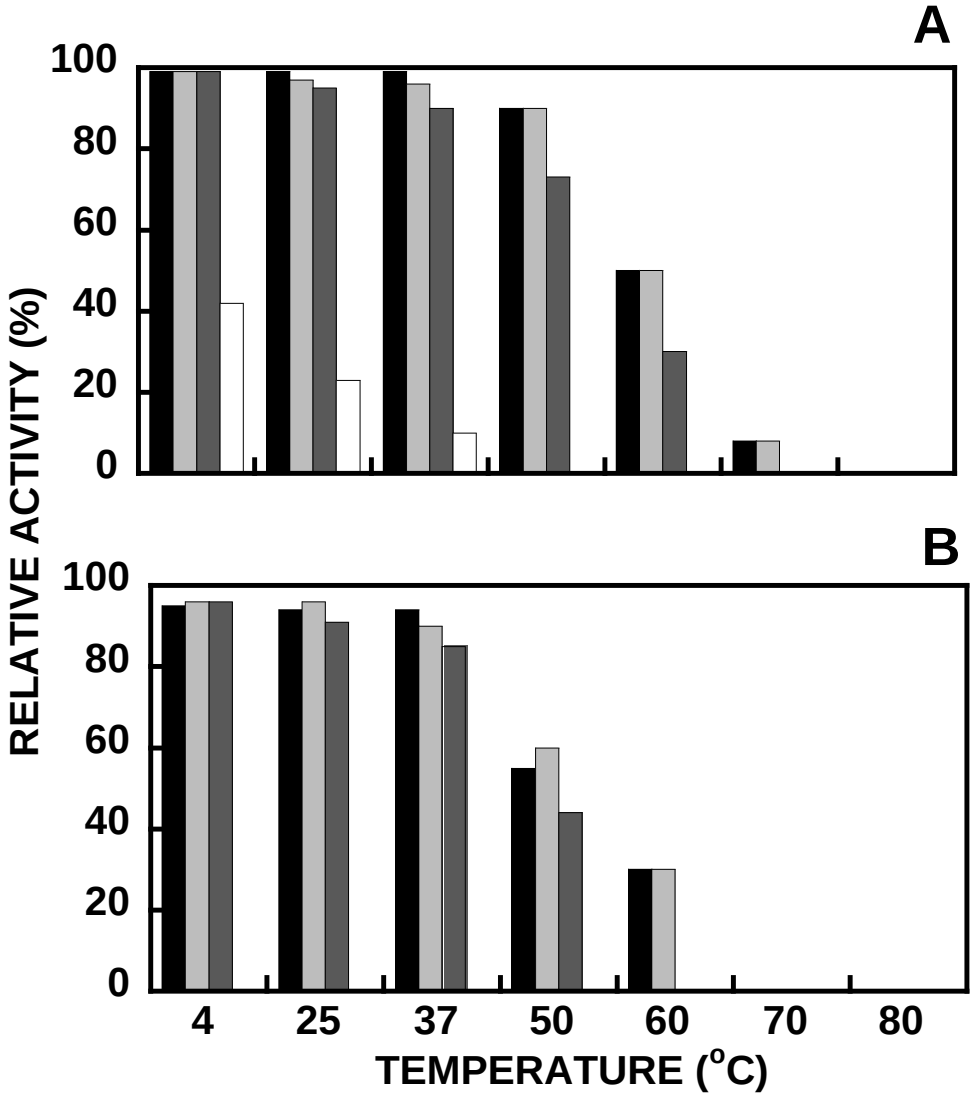


Fig. 8

Copyright 2004 Biochemical Society

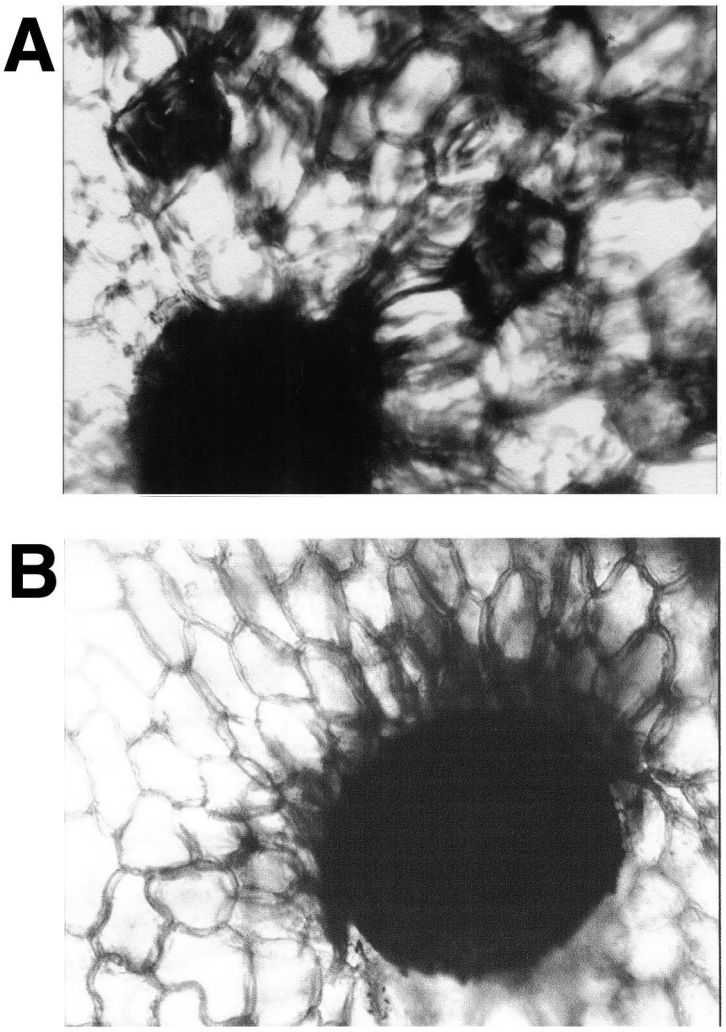


Fig. 9

Copyright 2004 Biochemical Society

A

```
ATGGCCTCGTCGCTGTATTCTTTCTTACTCGCATTATCGATAGTTTACATCTTTGTTGCC   60
 M  A  S  S  L  Y  S  F  L  L  A  L  S  I  V  Y  I  F  V  A    20
CCAACACATTCTACGTCAAGAACAGCTCTCAATCACCGTCACGAAGCCAAAGTAACGGGC  120
 P  T  H  S  T  S  R  T  A  L  N  H  R  H  E  A  K  V  T  G    40
TTTCAGATAATGCTTGAACATGTTGATTCGGGCAAAAACTTAACCAAATTCCAGCTCTTA  180
 F  Q  I  M  L  E  H  V  D  S  G  K  N  L  T  K  F  Q  L  L    60
GAACGTGCTATCGAAAGGGGTAGTCGTAGATTGCAGAGGCTCGAAGCCATGTTAAATGGC  240
 E  R  A  I  E  R  G  S  R  R  L  Q  R  L  E  A  M  L  N  G    80
                                              (N-Term) I  G
                                                (DN-1) I  G
CCCTCCGGTGTGGAAACTTCCGTCTACGCCGGAGATGGCGAATATCTGATGAACTTATCG  300
 P  S  G  V  E  T  S  V  Y  A  G  D  G  E  Y  L  M  N  L  S   100
 P  S  G  V  E  T  T  V  Y  A  G  D  G  E  Y  L  M  X  L  S
 P  S  G  V  E  T  T  V  Y  A  G/ D  G  E  Y  L  M  X  L  S
ATTGGAACTCCGGCACAACCTTTCTCCGCAATCATGGATACCGGTAGCGATCTTATCTGG  360
 I  G  T  P  A  Q  P  F  S  A  I  M  D* T  G  S  D  L  I  W   120
 I  G           (T-66) Y  A  I  M  D  T  G  S  D  L  I  W
 I  G  T  P  A  Q  P  F  S  A  I (DN-6)
ACGCAGTGCCAGCCTTGCACTCAGTGTTTTAATCAATCAACGCCCATATTTAATCCTCAA  420
 T  Q  C  Q  P  C  T  Q  C  F  N  Q  S  T  P  I  F  N  P  Q   140
 T  Q  X  Q  P  X  T  Q  X  F  X  Q  S      (DN-4) D  P  Q
GGATCATCCTCCTTCTCCACCCTCCCTTGCTCAAGCCAACTCTGTCAAGCCCTTTCAAGC  480
 G  S  S  S  F  S  T  L  P  C  S  S  Q  L  C  Q  A  L  S  S   160
 G  S  S  S  F  S  T  L  P  C
CCGACATGCTCTAATAATTTCTGCCAATACACCTACGGGTATGGGGACGGGTCCGAAACC  540
 P  T  C  S  N  N  F  C  Q  Y  T  Y  G  Y* G  D  G  S  E  T   180
                                                (V-59) T
                                 (T-48) G  Y  G  D  -  S  E  T
CAAGGATCCATGGGCACTGAGACTCTCACTTTCGGGTCGGTTTCCATCCCTAATATCACA  600
 Q  G  S  M  G  T  E  T  L  T  F  G  S  V  S  I  P  N  I  T   200
 Q  G  S  M  G  T  E/ T  F  T  F  G  S  V  S  I  P  N  I  T
 Q  G  S  M  G  T  E  T  F
TTCGGCTGCGGGGAAAACAACCAAGGGTTTGGGCAAGGAAACGGGGCAGGCTTGGTTGGG  660
 F  G  C  G  E  N  N  Q  G  F  G  Q  G  N  G  A  G  L  V  G   220
 F  G  X  G  E (V-100)
ATGGGTCGGGGCCCTCTGTCGCTTCCTTCTCAACTCGACGTGACCAAATTCTCTTACTGC  720
 M  G  R  G  P  L  S  L  P  S  Q  L  D  V  T  K  F  S  Y  C   240
   (T-60) G  P  L  P  L  P  X  Q  L  D  V  A  K
ATGACCCCCATTGGTAGCTCAACCCCTAGCAATCTTCTATTGGGATCACTGGCTAATTCT  780
 M  T  P  I  G  S  S  T  P  S  N  L  L  L  G  S  L  A  N  S   260
GTCACCGCCGGTAGTCCTAATACAACCCTAATCCAAAGCTCTCAAATACCAACTTTCTAT  840
 V  T  A  G  S  P  N  T  T  L  I  Q  S  S  Q  I  P  T  F  Y   280
TATATTACTCTCAACGGGTTGAGTGTTGGTTCAACTCGCTTGCCCATTGATCCGAGTGCT  900
 Y  I  T  L  N  G  L  S  V  G  S  T  R  L  P  I  D  P  S  A   300
 Y  I  T  L  D (V-81)           (V-96) L  P  I  D  P  S  A
TTTGCACTTAATAGCAATAATGGAACAGGAGGGATAATAATAGACTCTGGAACGACACTT  960
 F  A  L  N  S  N  N  G  T  G  G  I  I  I  D* S  G  T  T  L   320
 F
ACTTACTTCGTTAACAATGCTTATCAATCTGTAAGGCAAGAGTTCATCTCCCAGATTAAT 1020
 T  Y  F  V  N  N  A  Y  Q  S  V  R  Q  E  F  I  S  Q  I  N   340
CTACCCGTCGTAAATGGTTCCTCCTCCGGCTTTGATCTGTGCTTCCAGACGCCTTCTGAT 1080
 L  P  V  V  N  G  S  S  S  G  F  D  L  C  F  Q  T  P  S  D   360
                        (DN-2) D  L  C  F  Q  T  P  S
                                                (DN-7) D
CCGTCAAACCTGCAGATACCCACCTTTGTGATGCATTTTGACGGTGGAGATTTGGAGTTG 1140
 P  S  N  L  Q  I  P  T  F  V  M  H  F  D  G  G  D  L  E  L   380
 P  S  N  L  Q  I  P  T  F  V  M  H  F
CCCAGTGAGAATTATTTCATCTCCCCAAGCAACGGGCTGATTTGCTTGGCGATGGGGAGT 1200
 P  S  E  N  Y  F  I  S  P  S  N  G  L  I  C  L  A  M  G  S   400
TCGTCGCAGGGGATGTCCATTTTTGGGAATATTCAGCAGCAAAACATGCTAGTCGTTTAC 1260
 S  S  Q  G  M  S  I  F  G  N  I  Q  Q  Q  N  M  L  V  V  Y   420
GACACCGGAAATTCGGTGGTTTCATTCGCTTCTGCTCAATGTGGTGCGTCGTAA       1314
 D  T  G  N  S  V  V  S  F  A  S  A  Q  C  G  A  S stop       437
 D  T  G  N  S  V  V  S  F  V  S  A  Q  C  G  A (DN-3)
```

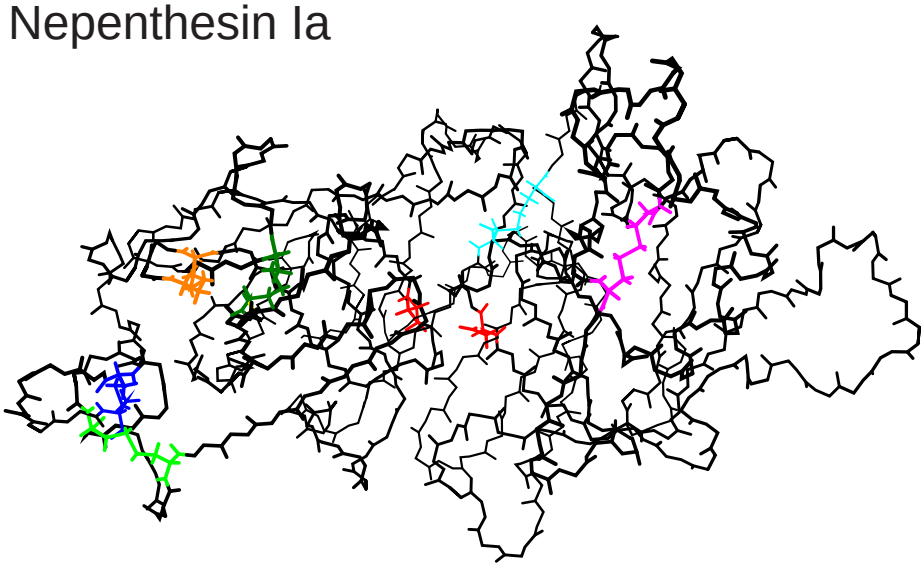
B

```
ATGGCCTCACCACTATACTCTGTGGTACTTGGCTTAGCAATAGTTTCTGCCATTGTTGCA   60
 M  A  S  P  L  Y  S  V  V  L  G  L  A  I  V  S  A  I  V  A    20
CCAACAAGCTCCACCTCAAGAGGAACCCTTCTTCATCATGGTCAGAAAAGGCCACAACCC  120
 P  T  S  S  T  S  R  G  T  L  L  H  H  G  Q  K  R  P  Q  P    40
GGCCTTCGTGTTGATCTCGAGCAGGTCGATTCGGGCAAGAATTTGACCAAATACGAGCTC  180
 G  L  R  V  D  L  E  Q  V  D  S  G  K  N  L  T  K  Y  E  L    60
ATCAAACGTGCTATCAAGCGTGGGGAGAGGAGGATGCGAAGCATTAATGCTATGTTGCAG  240
 I  K  R  A  I  K  R  G  E  R  R  M  R  S  I  N  A  M  L  Q    80
                                                (N-Term) Q
AGCTCCTCCGGTATTGAAACTCCTGTTTATGCGGGAGACGGTGAATATCTAATGAACGTA  300
 S  S  S  G  I  E  T  P  V  Y  A  G  D  G  E  Y  L  M  N  V   100
 T  V  Q  -  V  E  P  P  Y  Y  A  G  D  G  E  Y  L  M  V
GCAATTGGTACTCCGGATAGTTCTTTCTCGGCCATTATGGATACCGGCAGTGATCTCATT  360
 A  I  G  T  P  D  S  S  F  S  A  I  M  D* T  G  S  D  L  I   120
                                          (DN-14) D  L  I
TGGACGCAATGCGAGCCATGTACGCAGTGCTTCAGTCAACCTACGCCCATTTTCAACCCA  420
 W  T  Q  C  E  P  C  T  Q  C  F  S  Q  P  T  P  I  F  N  P   140
 W  T  Q  C  E  P  C  T  Q  C  F  S  Q
CAGGACTCGTCTTCCTTCTCTACCCTTCCTTGCGAGAGCCAGTATTGCCAAGATCTTCCG  480
 Q  D  S  S  S  F  S  T  L  P  C  E  S  Q  Y  C  Q  D  L  P   160
    D  S  S  S  F  S  T  L  P  C  E  S  Q (DN-7)    D  L  P
              (DN-3) D  L  P  C  E  S  Q  Y  C  Q
AGCGAAACCTGCAATAATAATGAATGCCAATATACATACGGATACGGAGACGGTTCCACA  540
 S  E  T  C  N  N  N  E  C  Q  Y  T  Y  G  Y* G  D  G  S  T   180
 S  E  T  C                            (DN-12a) D  G  S  S
              (DN-2) D  C  Q  Y  T  Y  G  Y  G
ACCCAAGGTTATATGGCAACCGAGACCTTCACTTTCGAGACGAGCTCCGTGCCGAATATC  600
 T  Q  G  Y  M  A  T  E  T  F  T  F  E  T  S  S  V  P  N  I   200
 T  Q  G  Y  M  A  X  E     (DN-10) D  G  S  S  V  P  N  I
GCGTTCGGTTGCGGGGAAGACAACCAGGGATTCGGGCAAGGCAACGGGGCTGGCCTGATC  660
 A  F  G  C  G  E  D  N  Q  G  F  G  Q  G  N  G  A  G  L  I   220
 A  F  G  C  G
GGGATGGGTTGGGGCCCGTTATCGCTTCCTTCTCAACTCGGCGTGGGTCAGTTCTCTTAC  720
 G  M  G  W  G  P  L  S  L  P  S  Q  L  G  V  G  Q  F  S  Y   240
TGCATGACCTCCTATGGAAGCTCCTCACCCAGCACTCTCGCACTTGGATCCGCAGCCAGT  780
 C  M  T  S  Y  G  S  S  S  P  S  T  L  A  L  G  S  A  A  S   260
GGAGTGCCTGAAGGCTCCCCGAGTACGACCCTCATCCATAGTTCTTTGAATCCAACGTAC  840
 G  V  P  E  G  S  P  S  T  T  L  I  H  S  S  L  N  P  T  Y   280
TATTATATTACGCTCCAAGGTATAACGGTTGGTGGCGATAATTTGGGTATTCCATCGAGT  900
 Y  Y  I  T  L  Q  G  I  T  V  G  G  D  N  L  G  I  P  S  S   300
                            (DN-12b) D  N  L  Q  I
ACTTTTCAACTTCAAGACGATGGAACTGGCGGGATGATAATTGACTCCGGGACAACGCTC  960
 T  F  Q  L  Q  D  D  G  T  G  G  M  I  I  D* S  G  T  T  L   320
                                 (DN-8) D  S  G  T  T  L
ACTTATCTTCCACAAGACGCTTACAATGCGGTAGCACAAGCGTTCACTGACCAGATAAAT 1020
 T  Y  L  P  Q  D  A  Y  N  A  V  A  Q  A  F  T  D  Q  I  N   340
 T  Y  L  P  Q                         (DN-11) D  Q  I  N
       (DN-9a) D  A  Y  N  A  V  A  Q  A  F  T
CTCCCCACCGTCGATGAATCCTCGAGCGGCCTCAGTACGTGCTTCCAGCAACCGTCCGAC 1080
 L  P  T  V  D  E  S  S  S  G  L  S  T  C  F  Q  Q  P  S  D   360
 L  P  T  V                                      (DN-13) D
     (DN-6) D  E  S  S  S  G  L  S  T  C  F  Q  E  P  S
GGATCAACCGTGCAAGTTCCGGAGATTTCAATGCAGTTTGATGGTGGGGTGCTGAACTTA 1140
 G  S  T  V  Q  V  P  E  I  S  M  Q  F  D  G  G  V  L  N  L   380
 G  S  T  V  Q  V  P  E  I  S  M  Q     D  G  G  V  L  N (DN-9b)
GGGGAACAGAATATATTGATCTCTCCAGCTGAAGGGGTGATATGCTTGGCGATGGGAAGT 1200
 G  E  Q  N  I  L  I  S  P  A  E  G  V  I  C  L  A  M  G  S   400
TCATCGCAGCTGGGAATTTCCATTTTTGGGAATATCCAGCAGCAAGAAACGCAGGTGCTC 1260
 S  S  Q  L  G  I  S  I  F  G  N  I  Q  Q  Q  E  T  Q  V  L   420
TATGACCTTCAGAATTTGGCCGTGTCGTTCGTTCCTACTCAGTGTGGTGCGTCGTAG    1317
 Y  D  L  Q  N  L  A  V  S  F  V  P  T  Q  C  G  A  S stop    438
    D  L  Q  N  L  A  V  S  F  F  P  T  Q  C  G  A  S (DN-16)
```

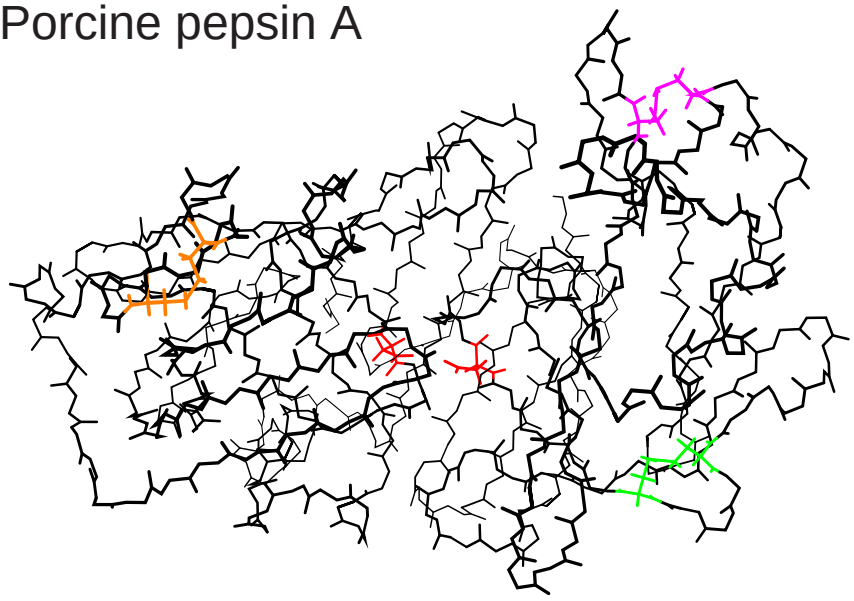
Fig. 10

Copyright 2004 Biochemical Society

A Nepenthesin Ia



B Porcine pepsin A



C

```
                                                                  *
Nepenthesin Ia:   NGPSGVETSVYAGDGEYLMNLSIGTPAQPFSAIMDTGSDLIW
Porcine pepsin A: --IGDEPLENYL-DTEYFGTIGIGTPAQDFTVIFDTGSSNLW

TQCQPCTQCFNQSTPIFNPQGSSSF-STLPCSSQLCQALSSPTCSNNFCQYTYGYGDGSE
VPSVYCSSLACSDHNQFNPDDSSTFEAT----SQE---------------LSITYGTG-S

TQGSMGTETLTFGSVSIPNITFGCGENNQGFGQGNG--AGLVGMGRGPL--S--LP--SQ
MTGILGYDTVQVGGISDTNQIFGLSETEPGSFLYYAPFDGILGLAYPSISASGATPVFDN

L-DVTK-----FSYCMT-PI--GSSTPSNLLLGSLANSVTAGSPNTTLIQSSQIPTFYYI
LWDQGLVSQDLFSVYLSSNDDSG----SVVLLGGIDSSYYTGSLNWVPVS---VEGYWQI

                                          *
TLNGLSVGSTRLPIDPSAFALNSNNGTGGIIIDSGTTLTYFVNNAYQSVRQEFISQINLP
TLDSITMDG----------ETIACSGGCQAIVDTGTSLLTGPTSAIANIQSDIGASE---

VVNGSSSGFDLCFQTPSDPSNLQIPTFVMHFDGGDLELPSENYFISPSNGLICLAMGSSS
--NSDGEMVISCSSIASLPDIVFTINGVQYPLSPSAYILQDDDSCTSGFEGMDVPTSSGE

QGMSIFGNIQQQNMLVVYDTGNSVVSFASAQCGAS
--LWILGDVFIRQYYTVFDRANNKVGLAPVA----
```

Fig. 11

Copyright 2004 Biochemical Society

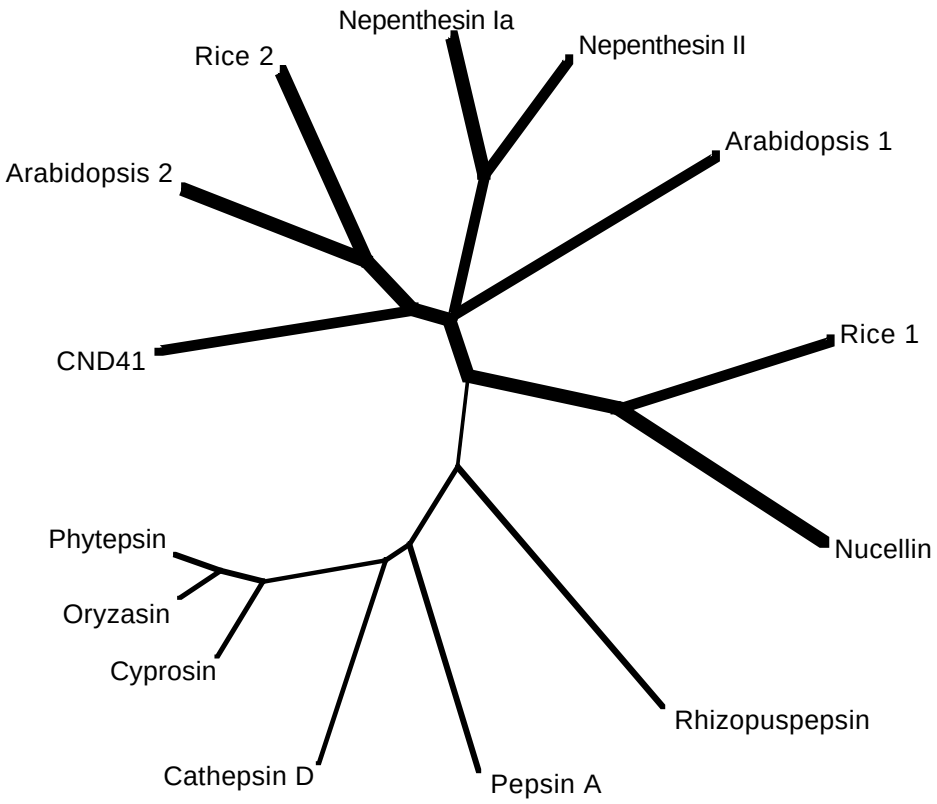


Fig. 12

Copyright 2004 Biochemical Society