The EMBO Journal Vol.17 No.1 pp.1–9, 1998

Structure of proline iminopeptidase from *Xanthomonas campestris* pv. *citri*: a prototype for the prolyl oligopeptidase family

F.J.Medrano[1,2,3], J.Alonso[4], J.L.García[4], A.Romero[4], W.Bode[1] and F.X.Gomis-Rüth[1,2,5]

[1]Max-Planck-Institut für Biochemie, Abteilung Strukturforschung, Am Klopferspitz 18, D-82152 Martinsried, Germany and [4]Centro de Investigaciones Biologicas, C.S.I.C., Velazquez 144, 28006 Madrid, Spain

[3]Present address: Centro de Investigaciones Biologicas, C.S.I.C., Velazquez 144, 28006 Madrid, Spain

[5]Present address: Centre d'Investigacions i Desenvolupament, C.S.I.C., Jordi Girona 18–26, 08034 Barcelona, Spain

[2]Corresponding authors
e-mail: cibj154@fresno.csic.es; xgrcri@cid.csic.es

The proline iminopeptidase from *Xanthomonas campestris* pv. *citri* is a serine peptidase that catalyses the removal of N-terminal proline residues from peptides with high specificity. We have solved its three-dimensional structure by multiple isomorphous replacement and refined it to a crystallographic *R*-factor of 19.2% using X-ray data to 2.7 Å resolution. The protein is folded into two contiguous domains. The larger domain shows the general topology of the α/β hydrolase fold, with a central eight-stranded β-sheet flanked by two helices and the 11 N-terminal residues on one side, and by four helices on the other side. The smaller domain is placed on top of the larger domain and essentially consists of six helices. The active site, located at the end of a deep pocket at the interface between both domains, includes a catalytic triad of Ser110, Asp266 and His294. Cys269, located at the bottom of the active site very close to the catalytic triad, presumably accounts for the inhibition by thiol-specific reagents. The overall topology of this iminopeptidase is very similar to that of yeast serine carboxypeptidase. The striking secondary structure similarity to human lymphocytic prolyl oligopeptidase and dipeptidyl peptidase IV makes this proline iminopeptidase structure a suitable model for the three-dimensional structure of other peptidases of this family.

Keywords: crystal structure/proline iminopeptidase/serine proteinase/*Xanthomonas campestris*

Introduction

Xanthomonas campestris is a Gram-negative bacterium belonging to the family Pseudomonaceae which is phytopathogenic for cruciferous plants. *Xanthomonas campestris* pv. *citri* is associated with various citrus bacterial diseases (Verniere *et al.*, 1993).

Proline iminopeptidase [PIP, EC 3.4.11.5] activity was reported for the first time by Sarid *et al.* (1959). Different sources of the enzyme have been described, implying its wide distribution in nature. However, this enzyme has been found mainly in bacteria (Yoshimoto *et al.*, 1983; Ehrenfreund *et al.*, 1992; Kitazono *et al.*, 1992, 1994a; Albertson and Koomey, 1993; Allaker *et al.*, 1994; Atlan *et al.*, 1994; Gilbert *et al.*, 1994; Klein *et al.*, 1994) and in some plants (Ninomiya *et al.*, 1982).

Several bacterial PIP genes have been cloned and sequenced to date: *Bacillus coagulans* (Kitazono *et al.*, 1992), *Neisseria gonorrhoeae* (Albertson and Koomey, 1993), *Lactobacillus delbrueckii* subsp. *bulgaricus* (Atlan *et al.*, 1994), *L.delbrueckii* subsp. *lactis* (Klein *et al.*, 1994), *Aeromonas sobria* (Kitazono *et al.*, 1994a) and *Mycoplasma genitalium* (Fraser *et al.*, 1995). These enzymes show a high degree of sequence similarity, share similar characteristics with regard to substrate specificity and molecular weight, and can therefore be grouped together in a PIP family (Figure 1).

PIP from *X.campestris* (XCPIP) consists of 313 residues (34 kDa) and catalyses the removal of N-terminal proline from peptides with fairly high specificity. Other amino acids are also cleaved off with much lower efficiency, however, with alanine representing the next best substrate. D-Alanine is also removed, with much lower efficiency than the L-isomer (Alonso and García, 1996). Another interesting activity of this enzyme is the removal of the alanine moiety from the aminoacyl nucleoside antibiotic ascamycin (Sudo *et al.*, 1996), converting it into a dealanylated derivative, which displays a broad antibacterial activity (Takahashi and Beppu, 1982; Osada and Isono, 1985).

Due to its inhibition by *p*-chloromercuribenzoate and other metal compounds, and because of the fact that other inhibitors, specific for serine peptidases, have little or no influence on the enzymatic activity, XCPIP has been suggested to be a cysteine peptidase. However, sequence similarity with 2-hydroxy-6-oxohepta-2,4-dienoate hydrolase, 2-hydroxymuconic semialdehyde hydrolase, 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase and atropinesterase from *Pseudomonas putida* (Atlan *et al.*, 1994; Alonso and García, 1996), and the localization of the catalytic serine residue by site-directed mutagenesis in the related *B.coagulans* and *A.sobria* forms (Kitazono *et al.*, 1994b) have helped to identify PIP as a serine peptidase.

The inhibition of this enzyme by inhibitors specific for cysteine peptidases suggested an important role for one of the cysteine residues present in the XCPIP sequence (Figure 1). This residue was believed to be situated in or close to the active site; alternatively, its modification could affect activity through long-range allosteric effects leading to strong inhibition.

Here we report the general structural features of XCPIP and the detailed structural information pertaining to the enzyme active site. In addition, we compare XCPIP with

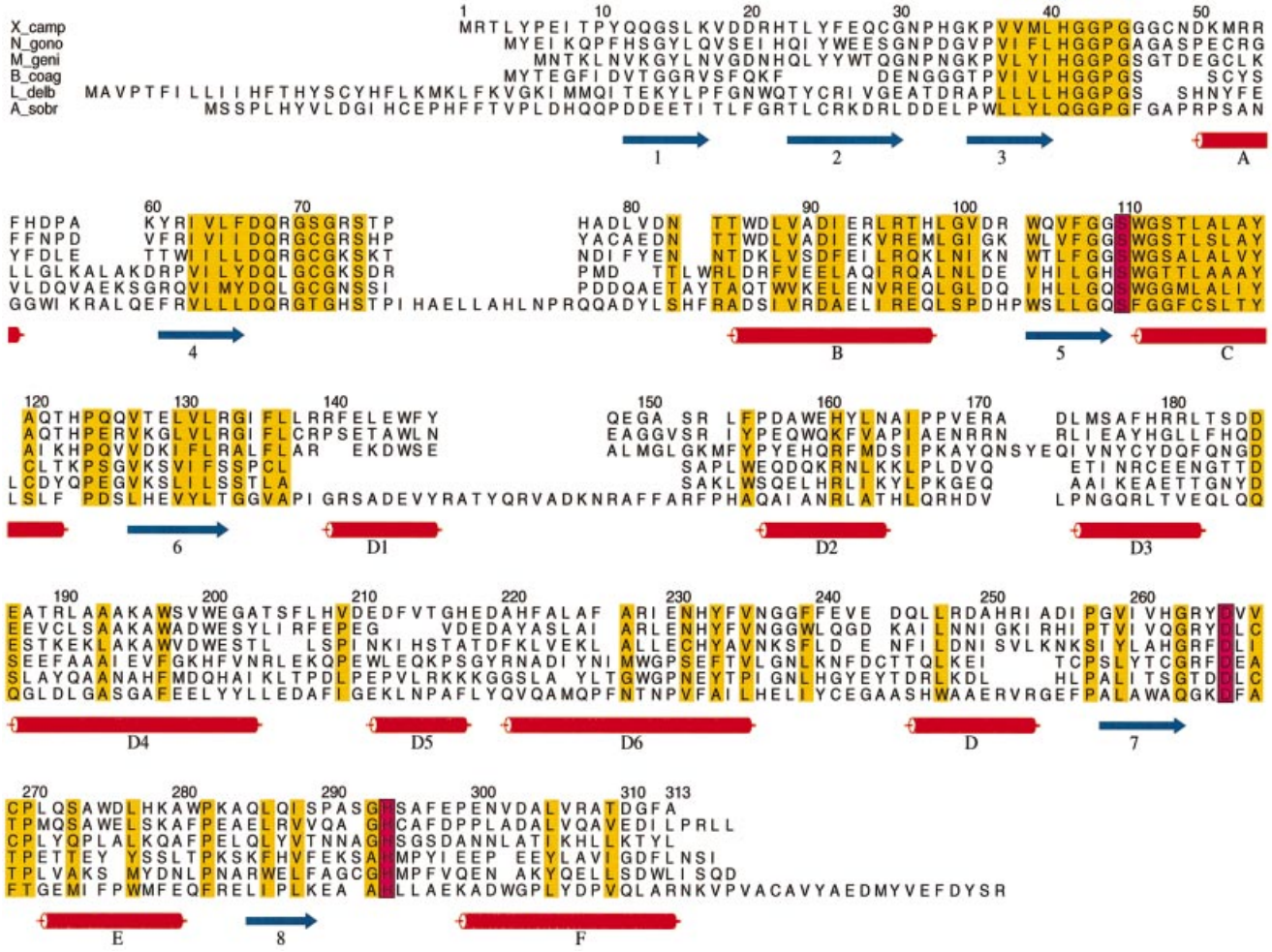

Fig. 1. Alignment of the amino acid sequences of some proline iminopeptidases: *Xanthomonas campestris* (X_camp), *Neisseria gonorrhoeae* (N_gono), *Mycoplasma genitalium* (M_geni), *Bacillus coagulans* (B_coag), *Lactobacillus delbrueckii* subsp. *bulgaricus* (L_delb), and *Aeromonas sobria* (A_sobr). Conserved residues are highlighted in yellow. Catalytic residues are highlighted in purple. Cylinders represent α-helices and arrows represent β-strands, as present in X_camp. This figure was prepared with ALSCRIPT (Barton, 1993).

the known X-ray structures of the yeast and wheat serine carboxypeptidases, *Aeromonas proteolytica* aminopeptidase, bovine lens leucine aminopeptidase and *Escherichia coli* methionine aminopeptidase. This comparison shows a particularly striking similarity with yeast serine carboxypeptidase. The comparison of the secondary structure elements of PIP with those of the prolyl oligopeptidase family shows that PIP can be used as a model for the catalytic domain of this family of peptidases.

Results and discussion

Polypeptide chain fold

XCPIP folds into two contiguous domains and has an ellipsoidal shape of approximate dimensions 65 Å× 50 Å×40 Å. A ribbon drawing of the Cα backbone is shown in Figure 2A. The 'lower' domain exhibits a secondary structure pattern characteristic for α/β hydrolases and is made up of the N-terminal segment Met1–Leu138 and by the C-terminal segment from Gln245 to Ala313. This domain is formed by a central eight-stranded β-sheet, with all strands except the second strand aligned in a parallel manner (Figure 2B). This mixed β-sheet is twisted, with the first and the last strand arranged almost perpendicular to each other, and is flanked on one side by the 11 N-terminal residues, a short 3_{10}-helix of six residues (helix A) and the C-terminal α-helix F, and on the other side by a five residue α-helix continuing into a four-residue 3_{10}-helix (helix D), and three other long α-helices (B, C and E). The topology of the β-strands is +1, +2, -1x, +2x, $(+1x)_3$ (Richardson, 1981). Between strand 6 and helix D, the chain deviates, forming the 'upper' domain. This domain, comprising the central segment Arg139–Asp244 and consisting of six α-helices, labelled D1–D6, is placed on the upper edge of the central sheet acting like a cap on top of the lower domain. A topology diagram is shown in Figure 2B. None of the three cysteine residues present in the sequence of PIP participates in a disulfide bridge.

The active site

The known catalytic Ser110 residue (Kitazono *et al.*, 1994b) and the putative Asp266 and His294 catalytic residues assigned due to sequence similarity with known serine peptidases (Atlan *et al.*, 1994; Alonso and Garcıa, 1996) are indeed clustered in the centre of the XCPIP molecule (Figure 3) and form a typical catalytic triad as

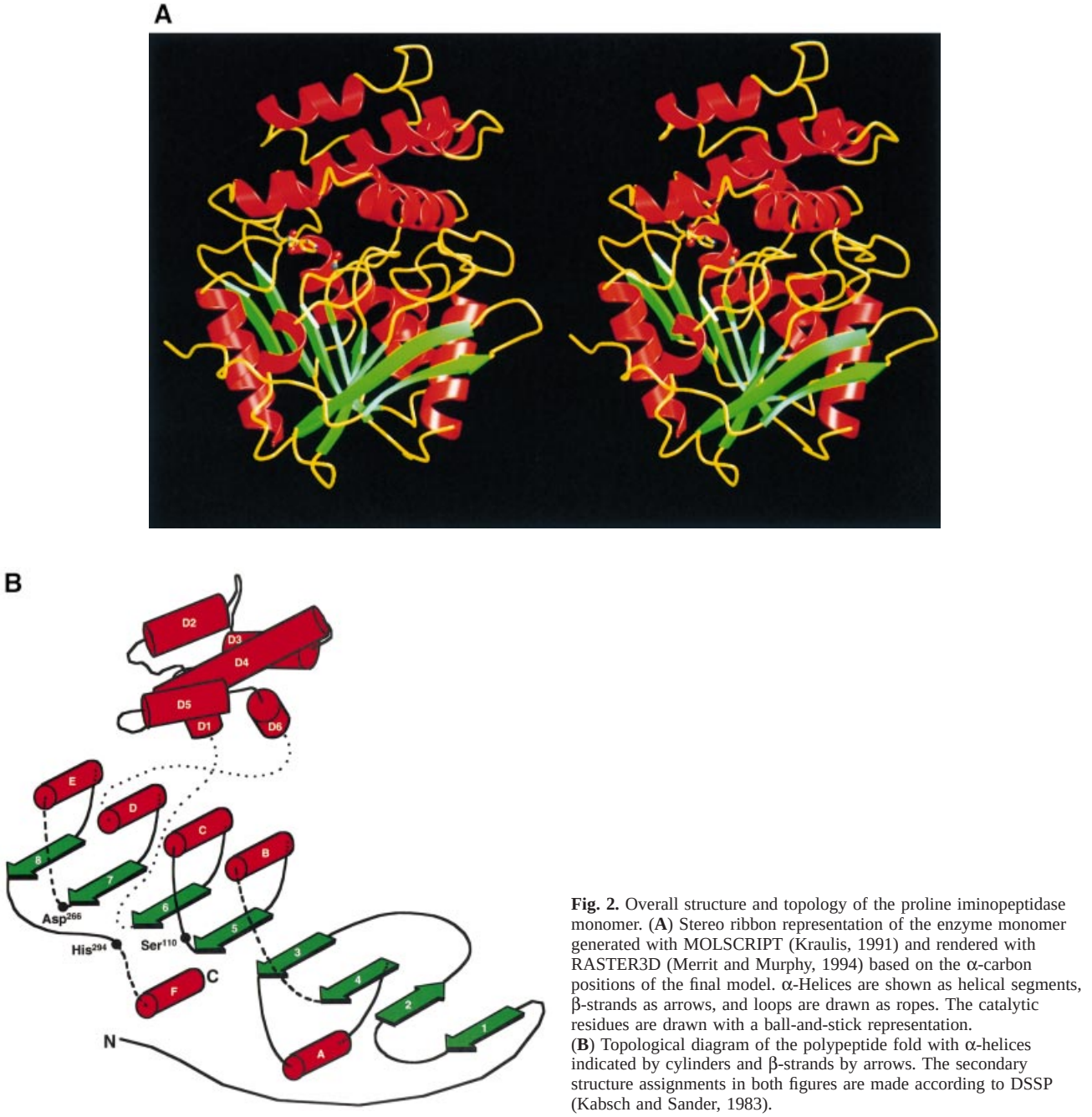

Fig. 2. Overall structure and topology of the proline iminopeptidase monomer. (**A**) Stereo ribbon representation of the enzyme monomer generated with MOLSCRIPT (Kraulis, 1991) and rendered with RASTER3D (Merrit and Murphy, 1994) based on the α-carbon positions of the final model. α-Helices are shown as helical segments, β-strands as arrows, and loops are drawn as ropes. The catalytic residues are drawn with a ball-and-stick representation. (**B**) Topological diagram of the polypeptide fold with α-helices indicated by cylinders and β-strands by arrows. The secondary structure assignments in both figures are made according to DSSP (Kabsch and Sander, 1983).

observed in several serine α/β hydrolases (Sussman *et al.*, 1991; Liao *et al.*, 1992).

The active site is located at the interface between both domains, which leave a funnel-shaped deep cavity (see Figures 2A and 4). The bottom of the pit is made up of hydrophobic residues (Phe136, Phe146, Phe234, Val268 and Cys269). The active site residues (Ser110, Asp266 and His294) lie in the centre of the pit, with the side chains of the catalytic Ser and His directed towards the bulk solvent. The catalytic Ser110 is spatially flanked by residues Glu201 and Arg133 and by two amide groups provided by Gly42 and Gly43. The entrance to the active site pit is wide open and is lined by a more hydrophobic part made up of Phe55, Phe206, Phe213, Phe222, Phe226 and Phe297, and by a more hydrophilic part made up by Asn49, His217 and of the carbonyl groups of Gly46, Gly47 and Cys48.

Ser110 is embedded in the sequence G108-X109-S110-B111-G112-Z113 (B, bulky, Z, small side chain and X, any amino acid) and spatially flanked by G42–G43 and G134 conserved in all known sequences of PIPs and in most of the α/β hydrolases (Ollis *et al.*, 1992). In the immediate vicinity of Ser110, a number of residues with small side chains occupy key positions (Gly42, Gly43, Gly108, Gly112, Gly134 and Ser113) to avoid steric hindrance. Ser110 is arranged in a sharp γ-like turn (the nucleophile elbow), which connects strand 5 and helix C. This is the strand–nucleophile–helix feature typical of the α/β hydrolase fold (Ollis *et al.*, 1992). At Ser110, the main chain exhibits an energetically unfavourable con-

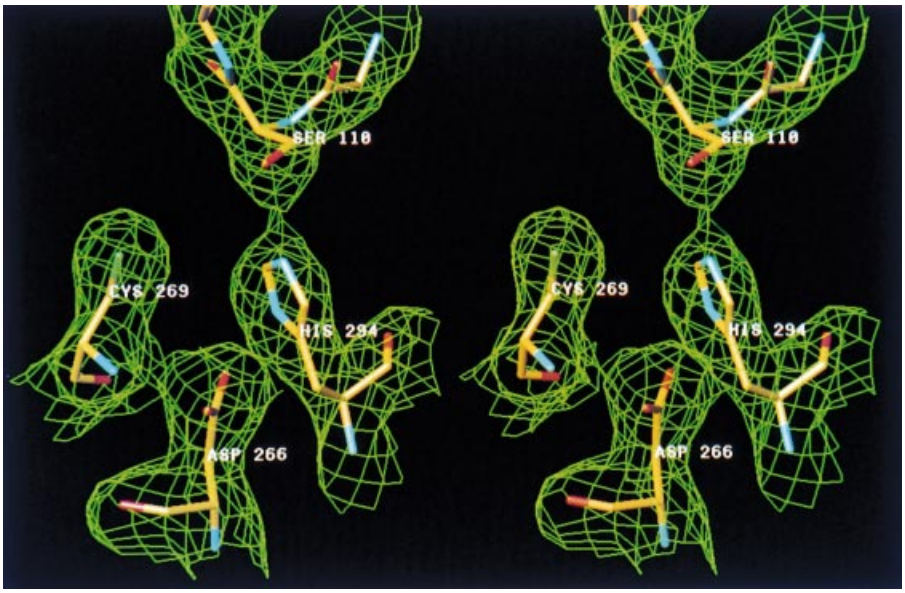

Fig. 3. Stereo view of the enzyme active site including the final $2F_o–F_c$ electron density contoured at 1 σ. The cysteine located at the back of the active site pocket is also shown. This figure was prepared with TURBO-FRODO (Roussel and Cambilleau, 1992).

formation; in the Ramachandran plot this corresponds to only a generously allowed region ($\phi = 54°$ and $\psi = -131°$), but is defined by clear density (Figure 3).

His294 is located in the middle of a nine-residue loop situated between strand 8 and helix F. The plane of this loop lies perpendicular to the plane formed by the strand and the helix, placing the imidazole group of His294 in the proper position to make hydrogen bonds with the flanking Ser110 and Asp266 residues (Figure 3).

Asp266 is the joining residue between two reverse turns situated between strand 7 and helix E. Its carboxylate groups form hydrogen bonds with the amide group of residues Val268 and Cys269, as well as with the side chains of His262 and His294. In contrast to the trypsin or subtilisin families of serine endopeptidases, however, the carboxylate group is arranged out of the plane of the histidine imidazole ring. Thus, the His–Asp hydrogen bond makes an angle of ~60° to the plane of the carboxylate (Figure 3).

Cys269 is located at the bottom of the pit, 5 Å from the catalytic triad (Figure 3). The observed inhibition of the hydrolytic activity of XCPIP by thiol-specific reagents such as iodoacetic acid or mercury compounds could be due to the presence of this cysteine in the active site. This inhibitor bound to the cysteine will block the space for the P1 residue.

Comparison of the active site features with those of other members of the α/β hydrolase fold family leads us to suggest that an 'oxyanion-binding site' is made by the main chain NH groups of residues Trp111 and Gly43 (Ollis *et al.*, 1992). These NH groups point into a small cavity between the loop where the P1 carbonyl of a bound substrate should be placed. This cavity appears to be well designed to stabilize the tetrahedral intermediate by binding the scissile carbonyl oxygen atom. The residue following Gly43, Pro44, is arranged in a *cis* conformation, allowing Gly43 to be placed correctly to shape the 'oxyanion hole' (Figure 5).

Specificity pocket

The S1 specificity pocket has approximate dimensions 7×7 Å at its base, providing good access to the active site (Figure 4). It is lined by the side chains of Phe136, Phe146, Phe243, Val268 and Cys269. Model building attempts show that this small cavity could accommodate the P1 residue of a bound peptide substrate. The small size and the hydrophobic character of this pocket are suitable to accommodate a proline or an alanine residue, in agreement with the substrate preference of XCPIP for these residues. A longer substrate with N-terminal proline or alanine will probably be bound in an extended conformation, with its primed side residues aligned along Gly42–Gly43. These residues are part of a loop that connects strand 3 with helix A (Figure 5).

Residues Glu201 and Glu230 are located close to the catalytic Ser110. Both residues are hydrogen bonded to a solvent molecule (Wat584 in molecule A and Wat609 in molecule B, Figure 5). This solvent molecule is, furthermore, in hydrogen bond contact with Gly43 O (2.5 Å), and more loosely with the Oγ of the catalytic Ser110 (3.7 Å). Therefore, the role of this glutamate pair Glu230/Glu201 (with an electrostatically unfavourable interaction between their charged side chains) would be binding and neutralization of the charged amino-terminal group of the P1 residue of a bound peptide substrate (Figure 5).

Another prominent residue located very close to the active site is Arg133. The side chain of this residue is located in the putative S1′ position and could bind the carbonyl oxygen of the P1′ residue (Figure 5).

Comparison with known X-ray structures of aminopeptidases

There are, up to now, three other aminopeptidases whose three-dimensional structure is known, namely bovine lens leucine aminopeptidase (LAP, Burley *et al.*, 1990), *A.proteolytica* aminopeptidase (AAP, Chevrier *et al.*, 1994) and methionine aminopeptidase from *E.coli* (MAP,

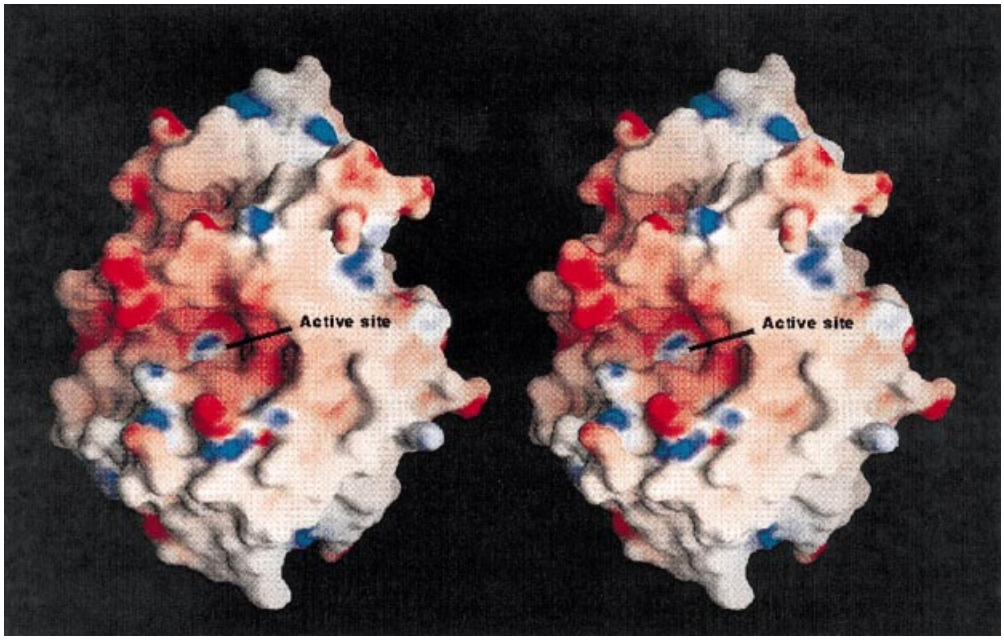

Fig. 4. Solid surface stereo representation of the electrostatic potential of the proline iminopeptidase monomer. The colour coding is according to the surface potential from negative (red) to positive (blue). The view is down the active site funnel, at the interface between the lower and the upper domain. The positive charge of the histidine of the catalytic triad is clearly visible in the middle of the active site. This figure was prepared with GRASP (Nicholls *et al.*, 1993).

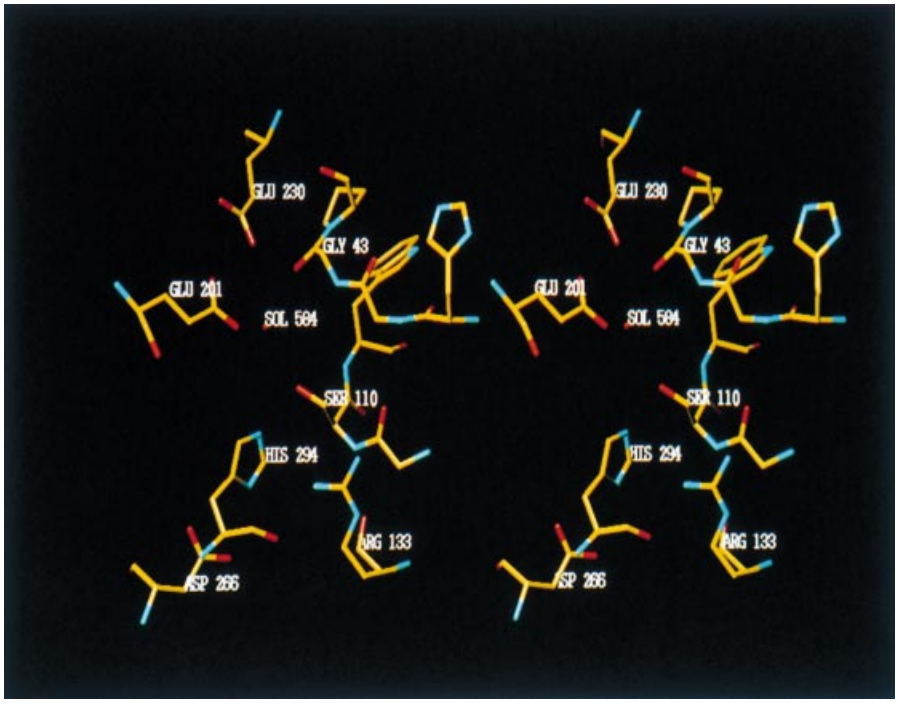

Fig. 5. Stereo view of the active site showing the catalytic triad residues (Ser110, Asp266 and His294), Glu201 and Glu230, probably involved in the binding of the amino-terminal group of the substrate, Arg133, located in the putative S1′ position, Trp111, Gly42 and Gly43, shaping the oxyanion-binding site, and the solvent molecule, Sol584, in close contact with the two glutamates and the catalytic serine. This figure was prepared with TURBO-FRODO (Roussel and Cambilleau, 1992).

Roderick and Matthews, 1993). These three aminopeptidases are metallo-proteases, with LAP and AAP having two zinc ions and MAP two cobalt ions at their active centres. Both LAP and AAP exhibit a similar proteolytic specificity. The carboxy-terminal domains of LAP and AAP exhibit homologous folds. In both enzymes, the catalytic domain includes an eight-stranded β-sheet and seven structurally equivalent helices. MAP has an original fold displaying internal pseudo 2-fold symmetry.

XCPIP is a low molecular weight serine protease and exhibits proteolytic specificity different from the other aminopeptidases with known three-dimensional structure (Alonso and Garcıa, 1996). The lower domain of PIP includes an eight-stranded β-sheet flanked by six helices

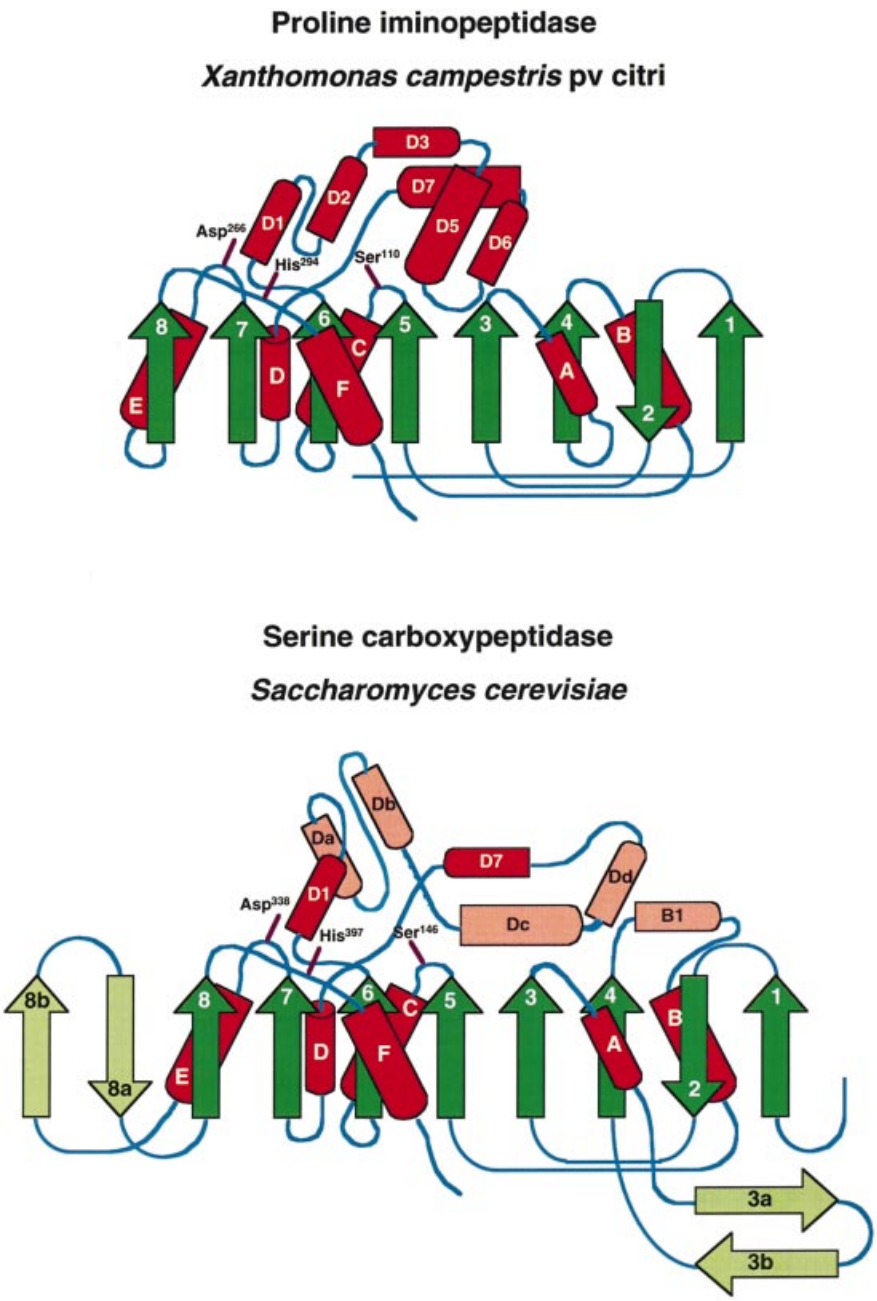

Fig. 6. Topology diagram of the polypeptide fold of XCPIP (top) and YSC (bottom), with α-helices indicated by cylinders and β-strands by arrows. The green arrows and red cylinders in the YSC diagram represent structurally equivalent secondary structure elements between both proteins.

and displays a low degree of similarity with LAP and AAP, as predicted by the program DALI. Nevertheless, these enzymes differ in the detailed topology of the central β-sheet and the surrounding helices. Thus, these aminopeptidases not only differ in their catalytic mechanisms, but also exhibit only distantly related folds.

Comparison with serine carboxyepeptidases

Despite a very low sequence identity, varying length and specificity, XCPIP displays significant topological similarity with yeast serine carboxypeptidase (YSC, Endrizzi *et al.*, 1994). Both enzymes have 243 structurally equivalent residues out of 313 for PIP and 421 for YSC. Superposition of both structures shows a difference of <4 Å, with an r.m.s. deviation of 2.50 Å for the common Cα-residues (Figure 6). A comparison with wheat serine carboxypeptidase II (WSC, Liao *et al.*, 1992) gives very similar results.

Both serine carboxypeptidases (SCs) as well as XCPIP fold into two contiguous domains. Superposition of the structures of these enzymes after least-squares fitting of the residues comprising the catalytic triad shows that the eight strands of the central β-sheet and the flanking helices of the lower domain have equivalent structural elements with identical directions and connectivities. SCs have two additional strands inserted in the central β-sheet between helix E and strand 8, and a short β-hairpin located between helix A and strand 4.

The SCs, like XCPIP, have an upper domain inserted between strand 6 and helix D. This domain likewise

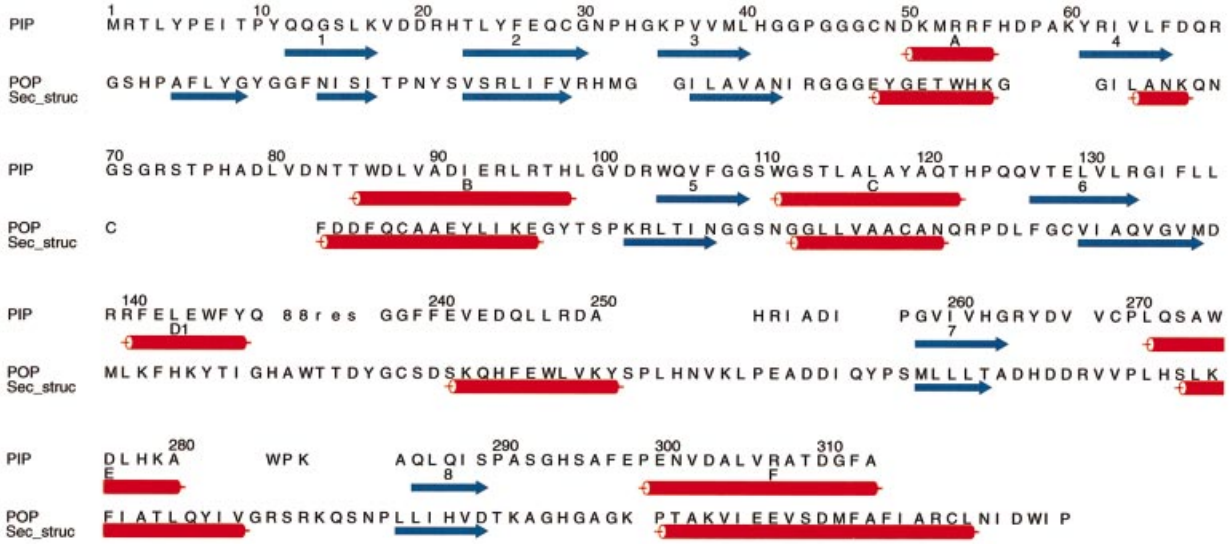

Fig. 7. Sequence alignment of the peptide chain segments forming the lower domain of XCPIP (PIP) with the catalytic domain of human lymphocytic prolyl oligopeptidase (POP). The alignment was made according to the known and proposed secondary structure elements. Blue arrows and red cylinders represent β-sheets and α-helices, respectively. Predicted secondary structure of POP (Sec_struc) as shown in Goossens *et al.* (1992). This figure was prepared with ALSCRIPT (Barton, 1993).

contains six helices, but includes an additional helix inserted between strand 4 and helix B. The detailed topologies of the upper domains are different, however, located in structurally non-equivalent positions (Figure 6).

The topology and localization of the catalytic triad and of the loops which provide these residues are very similar. SCs and XCPIP display a similar consensus sequence around the catalytic serine (see above), with a likewise unfavourable main chain conformation. The segment His41-Gly42-Gly43-Pro44-Gly45-Gly46 of XCPIP, playing a putative role in substrate binding at the primed subsites, is virtually identical in conformation to the segment Asp51-Gly52-Gly53-Pro54-Gly55-Cys56 of WSC. In both cases, the proline residue exhibits a *cis* conformation. In WSC, in contrast to XCPIP, this loop is buried in the protein structure and does not seem to play the role proposed for XCPIP, in agreement with the different proteinase classes to which these enzymes belong.

The localization of the entrance to the active site is very different. In the case of XCPIP, the entrance to the active site pit is located at the interface between both domains. The upper part of the entrance is made up by the upper domain, and the lower part by the lower domain. In YSC, there is, on top of the upper domain, a large central opening giving access to the active site.

Both enzymes seem to have a common ancestor, with a hydrophobic core made up of an eight-stranded β-sheet, and flanked by helices. The upper domain has been adapted to the different enzymatic activities and specificities.

Implications for the prolyl oligopeptidase family of peptidases

Prolyl oligopeptidase (POP) and dipeptidyl peptidase IV (DPPIV), members of the prolyl oligopeptidase family (S9) (Rawlings and Barret, 1994), specifically cleave internal prolyl-X bonds, like XCPIP. POP, *in vitro*, catalyses the cleavage of several biologically active peptides such as angiotensin II, oxytocin, vasopresin and bradykinin. Its activity in plasma correlates with different stages of depression (Maes *et al.*, 1994). It may also be involved in the regulation of blood pressure by participating in the renin–angiotensin system (Kohara *et al.*, 1991). This enzyme can decrease the amount of angiotensin II and its precursor. Moreover, the peptide generated by cleavage through POP, angiotensin (1–7), has vasodilator activity itself (Kohara *et al.*, 1991). POP has gained pharmaceutical interest in the past years, since injection of specific inhibitors in rats reverses scopolamine-induced amnesia (Yoshimoto *et al.*, 1987). DPPIV cleaves X-proline dipeptides from the amino-terminus of several bioactive peptides, such as growth hormone-releasing hormone (Frohmam *et al.*, 1989) and substance P (Ahmad *et al.*, 1992). In addition, it has been identified as CD26, a surface differentiation marker involved in the transduction of mitogenic signals in thymocytes and T lymphocytes in mammals (Vivier *et al.*, 1991), and in cell matrix adhesion through specific interactions with fibronectin and collagen (Piazza *et al.*, 1989).

No structure of any member of this family of peptidases is available to date. Out of the >700 residues that comprise these enzymes, the last 250 residues have been identified as the catalytic peptidase domain. A secondary structure prediction performed by means of a neural network approach (Goossens *et al.*, 1995) reveals this catalytic peptidase domain to be made up of eight sheets and five helices. This pattern is strongly in accordance with that of the lower domain of XCPIP. A comparison of the secondary structure reveals that strands 1–3, strands 5–8, helices A–C and helices D–F are, very probably, arranged in a similar fashion (Figure 7). Furthermore, this proposed XCPIP-based model would equally fulfil the predicted positions of the catalytic triad residues: the serine nucleophile would be located in a turn between strand 5 and helix C, the acid between strand 7 and helix E, and the base between strand 8 and the C-terminal helix F. Therefore, the lower subdomain of XCPIP comprising the eight-stranded central twisted β-sheet would represent a valid model for the structure of the catalytic peptidase domain, in

Table I. Concentration, soaking time, conditions and data collection statistics for heavy-atom derivatives

Compound		Concentration	Time	Limiting resolution (Å)	No. of collected/ unique reflections	Completeness of data (%)	R_{merge}[d]	R_{iso}[e]	Phasing power[f]	Sites
Native				2.7	154 900/29 854	93.9	9.6	–	–	–
HG1[a]	HgS	Saturated	3 days	3.6	27 028/9781	78.4	7.4	0.311	1.54	3
HG2[b]	1-(4-chloromercuriphenylazo)-2-naphthol ($C_{16}H_{11}ClHgNO_2$)	Saturated	2 days	3.0	64 607/18 837	84.6	12.8	0.256	0.16	2
PT1[b]	Pt(II)-(2,2′-6′,2′′-terpyridine)-chloride	Saturated	2 days	3.0	65 116/20 036	93.8	14.7	0.177	0.88	2
PT2[a]	K_2PtCl_4	10 mM	16 h	3.0	52 678/19 170	88.7	8.9	0.242	0.29	3
URA[c]	$Na_2U_2O_7$	Saturated	16 h	3.0	71 979/14 608	68.4	12.9	0.101	0.12	1

[a]0.1 M sodium citrate, pH 6.0 and 4.0 M NaCl.
[b]0.2 M Tris–HCl, pH 8.1 and 15% polyethylene glycol monomethyl ether 5000.
[c]0.1 M sodium citrate, pH 5.0 and 4.0 M NaCl.
[d]$R_{merge} = \Sigma|I-<I>|/\Sigma I$, where I is the measured intensity and <I> is the average intensity obtained from multiple measurements of symmetry-related reflections.
[e]$R_{iso} = \Sigma||F_P|-|F_{PH}||/\Sigma|F_P|$, where $|F_P|$ is the protein structure-factor amplitude and $|F_{PH}|$ is the heavy-atom derivative structure-factor amplitude.
[f]Phasing power = r.m.s. ($|F_H|/E$), where $|F_H|$ is the heavy-atom structure-factor amplitude and E is the residual lack of closure calculated for acentric data (20–3.0 Å).

accordance with previous studies (Goossens *et al.*, 1995) grouping these enzymes to those presenting the α/β-hydolase fold (Ollis *et al.*, 1992).

Materials and methods

Crystal data

XCPIP was purified and crystallized using NaCl as precipitating agent as previously described (Medrano *et al.*, 1997). These crystals belong to the orthorhombic space group C222, contain one homodimer per asymmetric unit and have cell constants $a = 147.2$ Å, $b = 167.8$ Å and $c = 85.6$ Å. These crystals diffract beyond 2.7 Å resolution.

Heavy-atom derivatives of the enzyme were prepared at room temperature by soaking crystals in heavy-atom solutions (Table I). One native and several derivative X-ray diffraction data sets were collected on a 300 mm MAR-Research image plate detector attached to a Rigaku RU200 rotating anode generator providing graphite monochromatized CuKα radiation at –15 to –20°C. Data were processed with the MOSFLM package (Leslie, 1991), and loaded, scaled and merged with the CCP4 package (CCP4, 1994). Statistics for native and heavy-atom derivative data are given in Table I.

Structure determination and refinement

The structure was solved by multiple isomorphous replacement (MIR). Heavy-atom positions were localized by difference Patterson maps, vector verification calculations and difference Fourier maps using the program PROTEIN (Steigemann, 1974). The correct handedness was established using the procedure described by Hoeffken (1988). Heavy-atom positions, occupancies and isotropic *B*-factors were refined with the program MLPHARE as implemented in the CCP4 package. The mean figure of merit at 3.0 Å resolution was 0.33 (see Table I for heavy-atom refinement and phasing statistics).

The initial 3.0 Å Fourier map was improved by density modification, including solvent flattening, solvent flip and skeletonization in two cycles with DM as implemented in the CCP4 package. After this step, the average figure of merit increased to 0.80, in particular due to the high solvent content (65%). The boundaries of both crystallographically independent molecules were clearly visible. An initial discontinuous poly-alanine polypeptide model was built into this modified electron density using the program TURBO-FRODO (Roussel and Cambilleau, 1992). Atomic models were refined applying NCS restraints using the program X-PLOR (Brunger, 1992). The electron density was additionally improved performing 2-fold averaging with the RAVE package (Kleywegt and Jones, 1994).

Calculated phases, obtained from intermediate models in successive cycles of manual rebuilding, were combined with the phases obtained after density modification using the program SIGMAA as implemented in the CCP4 package; these combined phases were used for calculation of $2F_o–F_c$ and $F_o–F_c$ maps. The amino acid sequence (Alonso and Garcıa, 1996) was successively built in to completion of the final model containing residues 1–313 for each crystallographically independent molecule. Additionally, 199 solvent molecules placed at stereochemically reasonable positions were added with the help of the program WATPEAK of CCP4 and checked visually. Positional and individual constrained isotropic *B*-factor refinement was performed with X-PLOR. A summary of the refinement statistics is shown in Table II. A Ramachandran plot of the main chain torsion angle pairs, calculated with the program PROCHECK (Laskowski *et al.*, 1993), show all residues but Ser110 situated in allowed regions, with only the latter in a generously allowed position. Pro44 and Pro76 exhibit a *cis* conformation. No significant differences are observed between both crystallographically independent molecules, reflected by the low r.m.s. deviation of 0.11 Å for all 313 common Cα atoms. The polypeptide chains of both molecules could be entirely traced; only six side chains of molecule A and five of molecule B had to be inactivated due to partially discontinuous density.

Comparison of the XCPIP structure with other known three-dimensional structures was carried out with the program DALI (Holm and Sander, 1993).

The X-ray coordinates of XCPIP will be deposited at the Brookhaven Protein Data Bank.

Table II. Refinement statistics

Space group	C222
Cell constants	
a	147.2 Å
b	167.8 Å
c	85.6 Å
Limiting resolution	2.7 Å
Non-hydrogen protein atoms	6140
Solvent molecules	199
Reflections used for refinement	26 672
Resolution range	7.0–2.7 Å
R-factor	19.2%
R_{free}	25.3%
R.m.s deviations from target values	
bond length	0.010 Å
bond angles	1.530°
bonded *B*-factors	2.613 Å^2

Acknowledgements

We thank Professor R.Huber for continuous encouragement, and the SFB 469 and the EU Biotechnology programme ERBB104CT960464 for financial support.

References

Ahmad,S., Wang,L. and Ward,P.E. (1992) Dipeptidyl (amino) peptidase IV and aminopeptidase M metabolize circulating substance P *in vivo*. *J. Pharmacol. Exp. Ther.*, **260**, 1257–1261.

Albertson,N.H. and Koomey,M. (1993) Molecular cloning and characterization of a proline iminopeptidase gene from *Neisseria gonorrhoeae*. *Mol. Microbiol.*, **9**, 1203–1211.

Allaker,R.P., Young,K.A. and Hardie,J.M. (1994) Production of hydrolytic enzymes by oral isolates of *Eikenella corrodens*. *FEMS Microbiol. Lett.*, **123**, 69–74.

Alonso,J. and Garcıa,J.L. (1996) Proline iminopeptidase gene from *Xanthomonas campestris* pv. *citri*. *Microbiology*, **142**, 2951–2957.

Atlan,D., Gilbert,C., Blanc,B. and Portalier,R. (1994) Cloning, sequencing and characterization of the pepIP gene encoding a proline iminopeptidase from *Lactobacillus delbrueckii* subsp. *bulgaricus* CNRZ 397. *Microbiology*, **140**, 527–535.

Barton,G.J. (1993) ALSCRIPT: a tool to format multiple sequence ąlignments. *Protein Eng.* **6**, 37–40.

Brunger,A. (1992) *X-PLOR Version 3.1, A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.

Burley,S.K., David,P.R., Taylor,A. and Lipscomb,L. (1990) Molecular structure of leucine aminopeptidase at 2.7 Å resolution. *Proc. Natl Acad. Sci. USA*, **87**, 6878–6882.

Chevrier,B., Schalk,C., D'Orchymont,H., Rondeau,J.M., Moras,D. and Tarnus,C. (1994) Crystal structure of *Aeromonas proteolytica* aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. *Structure*, **2**, 283–291.

Collaborative Computational Project, number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Ehrenfreund,P., Mollay,C. and Kreil,G. (1992) Purification and properties of an iminopeptidase from culture media of *Streptomyces plicatus*. *Biochem. Biophys. Res. Commun.*, **184**, 1250–1255.

Endrizzi,J.A., Breddam,K. and Remington,S.J. (1994) 2.8 Å Structure of yeast serine carboxypeptidase. *Biochemistry*, **33**, 11106–11120.

Fraser,C.M. *et al*. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science*, **270**, 397–403.

Frohmam,L.A., Downs,T.R., Heimer,E.P. and Felix,A.M. (1989) Dipeptidyl peptidase IV and trypsin-like enzymic degradation of human growth hormone-releasing hormone in plasma. *J. Clin. Invest.*, **83**, 1533–1540.

Gilbert,C., Atlan,D., Blanc,C. and Portalier,R. (1994) Proline iminopeptidase from *Lactobacillus delbrueckii* subsp. *bulgaricus* CNRZ 397: purification and characterization. *Microbiology*, **140**, 537–542.

Goosens,F., De Meester,I., Vanhoof,G., Hendriks,D., Vriend,G. and Scharpe,S. (1995) The purification, characterization and analysis of primary and secondary-structure of prolyl oligopeptidase from human lymphocytes. *Eur. J. Biochem.*, **233**, 432–441.

Hoeffken,H.W., Knof,S.H., Bartlett,P.A., Huber,R., Moellering,H. and Schumacher,G. (1988) Crystal structure determination, refinement and molecular model of creatine amidinohydrolase from *Pseudomonas putida*. *J. Mol. Biol.*, **204**, 417–433.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–128.

Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

Kitazono,A., Yoshimoto,T. and Tsuru,D. (1992) Cloning, sequencing, and high expression of the proline iminopeptidase gene from *Bacillus coagulans*. *J. Bacteriol.*, **174**, 7919–7925.

Kitazono,A., Kitano,A., Tsuru,D. and Yoshimoto,T. (1994a) Isolation and characterization of the prolyl aminopeptidase gene (pap) from *Aeromonas sobria*: comparison with the *Bacillus coagulans* enzyme. *J. Biochem.*, **116**, 818–825.

Kitazono,A., Ito,K. and Yoshimoto,T. (1994b) Prolyl aminopeptidase is not a sulphydryl enzyme: identification of the active serine residue by site directed mutagenesis. *J. Biochem.*, **116**, 943–945.

Klein,J.R., Schmidt,U. and Plapp,R. (1994) Cloning, heterologous expression, and sequencing of a novel proline iminopeptidase gene, pepI, from *Lactobacillus delbrueckii* subsp. *lactis* DSM 7290. *Microbiology*, **140**, 1133–1139.

Kleywegt,G. and Jones,T.A. (1994) Halloween... mask and bones. In Bailey,S., Hubbard,R. and Waller,R. (eds), *Proceedings of the CCP4 Study Weekend 1994*. SERC Daresbury Laboratory, Warrington, UK, pp. 59–66.

Kohara,K., Brosnihan,K., Chapell,M., Khosla,M. and Ferrario,C. (1991) Angiotensin (1–7). *Hypertension (Dallas)*, **17**, 131–138.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Laskowski,R.A., McArthur,M.W., Moss,D.S. and Thornton,J. (1993) PROCHECK: a program to check the quality of protein structures. *J. Appl. Crystallogr.*, **26**, 282–291.

Leslie,A.G.W. (1991) Macromolecular data processing. In Moras,D., Podjarny,A.D. and Tierry,J.P. (eds), *Crystallographic Computing V*. Oxford University Press, Oxford, pp. 27–38.

Liao,D.-I., Breddam,K., Sweet,R.M., Bullock,T. and Remington,S.J. (1992) Refined atomic model of wheat serine carboxypeptidase II at 2.2 Å resolution. *Biochemistry*, **31**, 9796–9812.

Maes,M., Goossens,F., Scharpe,S., Meltzer,H.Y., D'Hondt,P. and Cosyns,P. (1994) Lower serum prolyl endopeptidase enzyme activity in major depression: further evidence that peptidases play a role in the pathophysiology of depression. *Biol. Psychiatr.*, **35**, 545–552.

Medrano,F.J., Alonso,J., Garcıa,J.L., Bode,W. and Gomis-Ruth,F.X. (1997) Crystallization and preliminary X-ray diffraction analysis of proline iminopeptidase from *Xanthomonas campestris* pv. *citri*. *FEBS Lett.*, **400**, 91–93.

Merritt,E.A. and Murphy,M.E.P. (1993) Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr.*, **D50**, 869–873.

Nicholls,A., Bharadwaj,R. and Honig,B. (1993) Grasp—graphical representation and analysis of surface properties. *Biophys. J.*, **64**, A166.

Ninomiya,K., Kawatani,K., Tanaka,S., Kawata,S. and Makisumi,S. (1982) Purification and properties of a proline iminopeptidase from apricot seeds. *J. Biochem.*, **92**, 413–421.

Ollis,D.L. *et al*. (1992) The α/β hydrolase fold. *Protein Eng.*, **5**, 197–211.

Osada,H. and Isono,K. (1985) Mechanism of action and selective toxicity of ascamycin, a nucleoside antibiotic. *Antimicrob. Agents Chemother.*, **27**, 230–233.

Piazza,G.A., Callanan,H.M., Mowery,J. and Hixson,D.C. (1989) Evidence for a role of dipeptidyl peptidase IV in fibronectin-mediated interactions of hepatocytes with extracellular matrix. *Biochem. J.*, **262**, 327–334.

Rawlings,N.D. and Barret,A.J. (1994) Families of serine peptidases. *Methods Enzymol.*, **244**, 19–61.

Richardson,J.S. (1981) The anatomy and taxonomy of protein structure. *Adv. Protein Chem.*, **34**, 167–339.

Roderick,S.L. and Matthews,B.W. (1993) Structure of the cobalt-dependent methionine aminopeptidase from *Escherichia coli*: a new type of proteolytic enzyme. *Biochemistry*, **32**, 3907–3912.

Roussel,A. and Cambilleau,C. (1992) TURBO-FRODO, Biographics, LCCMB, Marseille, France.

Sarid,S., Berger,A. and Katchalski,E. (1959) Proline iminopeptidase. *J. Biol. Chem.*, **234**, 1740–1744.

Steigemann,W. (1974) Die Entwicklung und Anwendung von Rechtsverfahren und Rechenprogrammen zur Strukturanalyse von Proteinen. Ph.D. thesis, Technische Universitat, Munchen, Germany.

Sudo,T., Shinohara,K., Dohmae,N., Takio,K., Usami,R., Horikoshi,K. and Osada,H. (1996) Isolation and characterization of the gene encoding an aminopeptidase involved in the selective toxicity of ascamycin toward *Xanthomonas campestris* pv. *citri*. *Biochem. J.*, **319**, 99–102.

Sussman,J.L., Harel,M., Frolow,F., Oefner,C., Goldman,A., Toker,L. and Silman,I. (1991) Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science*, **253**, 872–879.

Takahashi,T. and Beppu,T. (1982) A new nucleosidic antibiotic AT-265. *J. Antibiot.*, **35**, 939–947.

Verniere,C., Provost,O., Civerolo,E.L., Gambin,O., Jacquemoud-Collet,J.P. and Luisette,J. (1993) Evaluation of the biological substrate utilization system to identify and assess metabolic variation among strains of *Xanthomonas campestris* pv. *citri*. *Appl. Environ. Microbiol.*, **59**, 243–249.

Vivier,I., Marguet,D., Naquet,P., Bonicel,J., Black,D., Li,C.X., Bernard,A.M., Gorvel,J.P. and Pierres,M. (1991) Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase IV). *J. Immunol.*, **147**, 447–454.

Yoshimoto,T., Saeki,T. and Tsuru,D. (1983) Proline iminopeptidase from *Bacillus megaterium*: purification and characterization. *J. Biochem.*, **93**, 469–477.

Yoshimoto,T., Kado.,K., Matsubara,F., Koriyama,N., Kaneto,H. and Tsuru,D. (1987) Specific inhibitors for prolyl endopeptidase and their anti-amnesic effect. *J. Pharmacobio-Dyn.*, **10**, 730–735.

Received September 19, 1997; revised October 21, 1997; accepted October 23, 1997