Structures of the tricorn-interacting aminopeptidase F1 with different ligands explain its catalytic mechanism

Peter Goettig¹, Michael Groll,
Jeong-Sun Kim, Robert Huber and
Hans Brandstetter

Max-Planck-Institut für Biochemie, Abteilung Strukturforschung,
Am Klopferspitz 18a, D-82152 Martinsried, Germany

¹Corresponding author
e-mail: goettig@biochem.mpg.de

F1 is a 33.5 kDa serine peptidase of the α/β -hydrolase family from the archaeon *Thermoplasma acidophilum*. Subsequent to proteasomal protein degradation, tricorn generates small peptides, which are cleaved by F1 to yield single amino acids. We have solved the crystal structure of F1 with multiwavelength anomalous dispersion (MAD) phasing at 1.8 Å resolution. In addition to the conserved catalytic domain, the structure reveals a chiefly α -helical domain capping the catalytic triad. Thus, the active site is accessible only through a narrow opening from the protein surface. Two structures with molecules bound to the active serine, including the inhibitor phenylalanyl chloromethylketone, elucidate the N-terminal recognition of substrates and the catalytic activation switch mechanism of F1. The cap domain mainly confers the specificity for hydrophobic side chains by a novel cavity system, which, analogously to the tricorn protease, guides substrates to the buried active site and products away from it. Finally, the structure of F1 suggests a possible functional complex with tricorn that allows efficient processive degradation to free amino acids for cellular recycling.

Keywords: biological rectifier/caged active site/gating mechanism/prolyl peptidases/substrate channelling

Introduction

Regulated protein breakdown is a major task of every cell and, therefore, all organisms have developed machineries to enable and control this process. Proteins that are fated to be disassembled are substrates of the ubiquitous proteasomes, which represent the major protein degradation pathway in all domains of life. Proteasomes produce in a largely unspecific manner peptides showing a broad size distribution with a peak length of 6–12 amino residues that have to be processed further, in order to yield molecules reusable for the metabolic processes of the cell (Bochtler *et al.*, 1999; Voges *et al.*, 1999).

The archaeon *Thermoplasma acidophilum* is employed as a model organism to study such fundamental processes of life for several reasons, in particular its evolutionary relationship to eukaryotes (Ruepp *et al.*, 2000). In *Thermoplasma*, components of a complete proteolytic machinery have been identified and structurally character-

ized, including the proteasome (Löwe *et al.*, 1995) and the tricorn protease (Brandstetter *et al.*, 2001). These structures are examples of cage-forming proteases inside cells (Lupas *et al.*, 1997).

Tricorn turns over proteasomal products to peptides of 2–4 amino residues, combining tryptic and chymotryptic activity patterns. Three so-called tricorn-interacting factors F1, F2 and F3 possess complementary substrate specificities and are sufficient to cleave any tricorn product *in vitro* to release single amino acids (Tamura *et al.*, 1996a, 1998).

Higher organisms, from yeast to mammals, utilize proteasomes as well as large enzyme complexes such as the cytosolic and lysosomal dipeptidyl and tripeptidyl peptidases (DPP and TPP) with functional analogies to tricorn (Tomkinson, 1999). On the sequential level, tricorn homologues have been identified in several archaea and bacteria. Moreover, most of the higher eukaryotic organisms contain genes encoding protein homologues of tricorn domains (Pallen *et al.*, 2001). Therefore, one can assume that these gene products function as modules in non-covalent complexes, corresponding to tricorn. In general, those proteins have differentiated in eukaryotes to a very great extent, but still adhere to principles of protein degradation pathways that evolved more than three billion years ago. For example, a mammalian homologue of F2 and F3 is responsible for the N-terminal trimming of oligopeptides that are generated by immuno-proteasomes (Stoltze *et al.*, 2000).

The tricorn-associated F2 and F3 are 89 kDa metallo-peptidases with an identity of 56%. While they exhibit preferences for positively and negatively charged residues in the P1 position of peptide substrates, respectively, they also accept polar and some hydrophobic residues (Tamura *et al.*, 1998).

F1 completes the specificity spectrum by its ability to cleave most hydrophobic N-terminal residues from peptides, such as alanine, proline, phenylalanine and leucine, but also glycine and tyrosine. In contrast, charged and polar substrates with arginine, glutamate or glutamine in the P1 position are excluded from turnover. The enzyme consists of 293 residues (33.5 kDa) and has been described as prolyl iminopeptidase (Tamura *et al.*, 1996a,b). It belongs to the α/β -hydrolase superfamily (PIP_THEAC, EC 3.4.11.5, peptidase family S33) and is not related to F2 or F3.

The closest relative is F1 from *Thermoplasma volcanium*, which shares 78% identical amino acid residues with the *T. acidophilum* protein. Also, the crenarchaeote *Sulfolobus solfataricus*, closely related to eukaryotes, has an F1-encoding gene, as well as a tricorn protease. Human homologues of F1 are a liver epoxidase with 37% identity and a putative protein of 42% identity, encoded by the gene CGI-58. The X-ray structures of three close F1

Fig. 1. Sequence alignment of F1 and homologues from the euryarchaeotes *Thermoplasma acidophilum* and *Thermoplasma volcanium*, and the crenarchaeote *Sulfolobus solfataricus*. Bacterial PIPs of *Bacillus coagulans* and *Lactobacillus delbrueckii*, PIP from *Xanthomonas campestris* and PAP from *Serratia marcescens* are also closely related. From porcine POP, only the catalytic and the cap sequence parts are shown; lower case letters indicate POP sequence segments with weak homology to F1. Secondary structure elements of F1 are indicated using the same colour coding as in Figure 2. Homologies are boxed and shaded grey. The catalytic triad residues are highlighted in yellow (S105, D244 and H271), the oxyanion pocket-forming residues in blue (G37 and Y106), and the two N-terminus binding glutamates together with the hydrogen-bonded tyrosine in red (E213, E245 and Y205). In PIP and PAP, a sequentially distinct glutamate pair recognizes the N-terminus (shown in red), while the endopeptidase POP lacks similar glutamates.

relatives have been solved, namely prolyl iminopeptidase (PIP) from the phytopathogenic bacterium *Xanthomonas campestris* (Medrano *et al.*, 1998), prolyl aminopeptidase (PAP) from the bacterium *Serratia marcescens* (Yoshimoto *et al.*, 1999) and the mammalian prolyl oligopeptidase (POP), sometimes termed prolyl endopeptidase (PEP) (Fülöp *et al.*, 1998). Despite sharing 25% identical residues in the catalytic domain with F1, POP has evolved beyond protein degradation pathways to become a regulator of proline-containing neuropeptides, which affect blood pressure, memory and psychic states (Fülöp *et al.*, 2001) (Figure 1).

Nevertheless, the structure of mammalian POP provides a detailed model of protein domain recognition in the tricor-F1 system. Attached to its F1-like catalytic domain is a large seven-bladed β -propeller with an open velcro topology, similar to those of the tricor protease, which has a six- and a seven-bladed β -propeller per monomer (Brandstetter *et al.*, 2001).

Here we present crystal structures of F1 in its free and inhibitor-bound forms revealing the key elements in the

catalytic mechanism. Moreover, the structure suggests how F1 might assemble further with tricor to form a peptide-degrading supercomplex.

Results

Biochemical properties of recombinant F1

F1 from *T. acidophilum* was overexpressed in *Escherichia coli* and purified as a single chain and fully active protease. The concentrated enzyme could be stored at 4°C for several weeks without significant degradation, consistent with mass spectroscopy (ESI-MS) showing one major single peak of 33 488 Da. F1 exists as a monomer in solution, as has been confirmed by gel filtration and dynamic light scattering experiments. The recombinant protein cleaved fluorogenic substrates such as H-Pro-AMC (H-proline aminomethyl coumarin) or H-Ala-Ala-Phe-AMC, but failed to cleave after charged amino acids in H-Arg-AMC and H-Asp-AMC, in accordance with reported data (Tamura *et al.*, 1996a,b). The cleavage of H-Pro-AMC was reduced to <4% when F1 was incubated

Fig. 2. (A) Overall structure of F1. Helices and loops in the catalytic domain are coloured yellow, the β -strands are coloured green and the cap domain is coloured red. The molecular surface and the termini are indicated. (B) Topology of F1. The eight-stranded β -sheet comprises seven parallel strands with the second one running antiparallel. The colour coding corresponds to that in (A), with the first and the last residue of every secondary structure element specified. The green to yellow gradient colouring of Ser105 reflects the involvement of its backbone atoms in the β -sheet (green) and the subsequent α -helix (yellow), explaining its strained conformation. (C) Stereo diagram of the smoothed C- α trace of F1. The colour coding is as in the preceding figures, but F1 rotated by 180° around the vertical y-axis. The position of the active Ser105, together with a covalently bound MES molecule, is represented by ball-and-sticks, as is the disulfide bridge between Cys5 and Cys22 near the N-terminus. The cavity system is integrated as semi-transparent surfaces, with the major access tunnel to the active site indicated by an arrow.

prior to the peptidase assay with phenylalanyl chloromethylketone (PCK), whereas other serine protease inhibitors, such as phenylmethylsulfonyl fluoride (PMSF), exhibited only ~50% inhibition of efficiency.

The overall structure of F1

Crystals of F1 were analysed in the free form and in complex with the substrate analogues morpholinoethanesulfonic acid (MES) and PCK. We collected data up to 1.8 Å resolution, and traced the protein sequence in a well-defined electron density, starting with residue Cys5 that forms a disulfide bond to Cys22 and ending with the C-terminal Leu293.

F1 is a pear-shaped two-domain protein with dimensions of ~60 and 45 Å, respectively (Figure 2A). The central eight-stranded β -sheet (shown in green in Figure 2A and B) is twisted by ~90° and forms the scaffold of the catalytic domain, seen for the related prolyl peptidases (Medrano *et al.*, 1998; Yoshimoto *et al.*, 1999). The F1-specific cap domain is inserted between residues 132 and 225 (shown in red in Figure 2A and B).

A remarkable structural feature of the F1 peptidase is the architecture of its active site environment, which is located in the interior of the protein; in fact, the centre of mass coincides with the catalytic Ser105. Substrates appear to access the active site via a cavity system

Fig. 3. (A) The catalytic residues and the hydrogen-bonding network in the active site of uninhibited F1. Ser105, Asp244 and His271 form the latent catalytic triad. A water molecule binds to the amide of Gly37, which together with the amide nitrogen of Tyr106 form the oxyanion pocket. However, free access to the oxyanion pocket is blocked by the O- γ of Ser105 binding to Tyr106. The residues Glu213, Glu245 and Tyr205 belong to a hydrogen-bonding network that plays a role in substrate binding. The $2F_o - F_c$ electron density is contoured at 1.0σ . (B) The active site of MES-bound F1 in the same orientation as in (A). The backbone amide groups of Gly37 and Tyr106 fix one of the sulfonyl oxygens of the MES molecule. Its tetrahedral sulfur is covalently bound to O- γ of Ser105 and mimics a trapped transition state of the catalytic reaction. The morpholino ring of MES is located in the larger hydrophobic cavity (S1 site). (C) Stereo representation of the covalently bound inhibitor PCK in the active site with the final $2F_o - F_c$ density contoured at 0.7σ . PCK was omitted from phasing in calculation of the electron density. Note the purple electron density contoured at 3.0σ , which identified the unhydrolysed chlorine. Glu213 and Gly37 function as the N-terminus recognition elements, whereas the amides of Gly37 and Tyr106 form the oxyanion pocket. The orientation corresponds to that in (A) and (B).

(Figure 2C). The entrance tunnel (indicated by an arrow in Figure 2C) branches at a cross-section into a hydrophobic pocket, a polar extension and a constricted region containing the catalytic Ser105. This region connects to another large hydrophobic cavity, which is completely encapsulated in the free F1 structure. In contrast, in the two complexed F1 structures, this cavity opens with a tunnel to the surface of the cap domain.

Latent catalytic triad conformation in the unliganded F1

The active site in F1 contains the elements of a classic serine protease catalytic triad: the nucleophile Ser105, the

proton donor His271 and the acceptor Asp244 (Figure 3A). However, their sequential and, in particular, spatial positions are completely different as compared with proteinases or peptidases such as trypsin and subtilisin. Significantly, Ser105 adopts an unfavourable stereochemical conformation, as was observed in related structures, including those of PIP, PAP and POP (Fülöp *et al.*, 1998; Medrano *et al.*, 1998; Yoshimoto *et al.*, 1999). Hence the high-energy conformation of Ser105 is very likely to be critical for its catalytic function.

Similarly to the related structures, the NH groups of Gly37 and Tyr106 are positioned to form the oxyanion pocket. Intriguingly, the O- γ of Ser105 binds into the

oxyanion pocket with a distance of 2.6 Å to the backbone amide nitrogen of Tyr106. It is stabilized further by a water molecule, which forms another hydrogen bond to the NH group of Gly37 (Figure 3A). This side chain rotamer of Ser105 precludes the formation of a hydrogen bond between its O-γ and the N-ε of His271, given the distance of 3.9 Å and their unfavourable relative orientation. Thus, the O-γ of Ser105 is not activated by the catalytic histidine and represents a catalytic triad in a latent state.

Two glutamates (Glu213 and Glu245) contribute further to the active site composition and have a similar orientation to that observed earlier in PIP and PAP, where they have been proposed to bind the imino group of a substrate's N-terminal proline (Yoshimoto *et al.*, 1999). The arrangement of the carboxylate of Glu245 between the OH group of Tyr205 and the carboxylate of Glu213 can be explained easily by a hydrogen network, involving a protonated Glu245 (Figure 3A).

The structures with MES and PCK bound to Ser105

By analysing an F1 crystal, grown in the presence of 100 mM MES, we unexpectedly found continuous electron density extending from the O-γ of Ser105, which indicated the covalent binding of a MES molecule (Figure 3B). This serendipitous finding implies an extremely strong nucleophilicity for Ser105. Accompanying the MES binding, we observed simultaneous conformational changes of Ser105 and His271, while the C-α positions of the catalytic triad residues remain unchanged.

The O-γ of Ser105 of the unliganded F1, together with the water molecule, is displaced from the oxyanion hole upon its binding by one of the sulfonic oxygen atoms of MES. O-γ moves towards N-ε of His271 (reducing the distance from 3.9 to 2.8 Å), which in turn orients itself for optimal activation of the catalytic serine (Figure 3B). The second sulfonic oxygen is hydrated by a water pair. Notably, MES had no inhibitory effect on F1 when tested in a fluorogenic activity assay.

In an attempt to obtain additional information about catalysis in F1, we investigated the structure of its complex with PCK, an efficient inhibitor of F1 *in vitro*. The electron density map showed PCK to be covalently linked to the O-γ of the catalytic Ser105, thus resembling a transition state of a peptidic substrate (Figure 3C). The carbonyl oxygen of PCK occupies the oxyanion pocket of Gly37 and Tyr106, and the amino group of PCK forms close hydrogen bonds with the carbonyl oxygen of Gly37 at a distance of 2.2 Å, and with both carboxylate oxygens of Glu213 (Figure 3C). The position of the phenyl ring of PCK superimposes with the morpholino ring of MES, indicating that MES functions as a P1 residue analogue. PCK, like MES, induced the intact state of the catalytic triad by displacing the Ser105 O-γ from the oxyanion hole and pushing it towards N-ε of His271 where it is oriented ideally and activated for catalysis (Figure 4).

Functional mapping of the F1 cavities

The structures of both complexes identified the larger cavity in F1 as the S1 specificity site and thus provided a rationale for the further assignment of the remaining cavity structures.

Fig. 4. Activation switch of the latent catalytic triad. In the unliganded F1 (shown in grey), the catalytic Ser105 O-γ forms a hydrogen bond to the backbone amide of Tyr106, but not to N-ε of His271. Upon activation of the triad by binding of the P1 carbonyl oxygen of PCK (and similarly by MES), the O-γ is pushed from the oxyanion amide towards the N-ε of His271. This movement is accompanied by a re-orientation of His271 and results in an ideally oriented N-ε–O-γ hydrogen bond at 2.8 Å distance.

Fig. 5. The cavity system of the F1 peptidase. The cap domain that encompasses the major part of the cavities is represented with darker shading. E1 (red) forms the entrance to the active site chambers, namely the S1' site in green, the S1 site in orange and the connecting junction C harbouring the catalytic residues. E2 in violet designates the putative exit; P in yellow indicates a polar water-filled cavity, which is contained mainly within the catalytic domain.

With the exception of the polar pocket (P), all cavities are contained mostly within the cap domain (Figure 5). The large entrance channel (E1 in Figure 5) from the surface of F1 to the active site region is ~6 Å long and wide. It is lined with polar and hydrophobic residues of the cap domain (Tyr178, Trp188, Leu199, Tyr205 and Asn212). The channel broadens to a hydrophobic cavity (S1' in Figure 5) of ~8 Å length and 8 Å diameter, and is comprised mainly by the cap domain (Leu182, Leu183, Trp188, Val192 and Ser195) and some residues of the catalytic domain (Met40, Tyr44 and Leu272). This pocket is positioned C-terminal to the PCK ligand, which suggests a role as an S1' site. In accordance with this interpretation, the size, shape and hydrophobic character of this substructure explain the turnover of fluorogenic substrates

with AMC as P1' analogue. The non-polar nature of this region is emphasized by the resistance of the PCK chlorine against hydrolysis, as unambiguously resolved by the electron density (Figure 3C). The arrangement of the catalytic triad and the oxyanion hole precludes the alternative substitution of the chlorine by the catalytic His271, analogously to trypsin-like proteases (Bode *et al.*, 1989).

A narrow hydrophilic pocket (P in Figure 5) extends from the S1' cavity opposite the entrance channel (E1), which is composed of side chains of the catalytic domain (Ser104, Ser128, Thr239, Asp244, His271 and Thr273) and harbours three water molecules.

The cavity system is completed by the voluminous S1 site with dimensions of 8, 10 and 12 Å. It consists predominantly of hydrophobic side chains, which are provided by both the cap domain (Val134, Thr137, Met141, Ile216, Asn209, Ile220 and Trp223) and the catalytic domain (Pro38, Ile79, Tyr106, Ala109, Leu131, Val246 and Val250) (Figure 5). The absence of water molecules in this pocket reflects its hydrophobicity.

In the liganded F1 structures, a channel connects the S1 cavity to the exterior of the protein (E2 in Figure 5). In the MES complex, the channel is ~5 Å in diameter and 8 Å long, and lined by residues of the cap domain (Asp142, Val138, Ile145, Phe214, Thr215 and Ile216), but does not contain crystallographically defined waters. Interestingly, this channel is closed in the unliganded protein and, therefore, the access of peptides to the free active site seems possible only through the major entrance (E1).

Discussion

Substrate access and recognition in F1

The location of the active residues in the interior of a protein is a critical characteristic of F1. As with functionally related caged active sites (Fülöp *et al.*, 1998; Brandstetter *et al.*, 2001; Jozic *et al.*, 2002), the shielding represents a regulation element of these proteases, because these enzymes typically are expressed in their fully active form.

The access of substrates to F1 is governed by three determinants. First, the primary driving force appears to be electrostatic attraction of the positively charged N-terminus of the peptide. Electrostatic calculations of the surface potential with GRASP (Nicholls, *et al.*, 1993) showed a balanced distribution of positively and negatively charged residues on the solvent-exposed parts of F1. However, the active site residues Glu213 and Glu245 generate a strong negative potential, which extends into the entrance channel and attracts the positively charged N-terminus of a small peptide. Secondly, only short and unstructured peptides can pass through the narrow and bent entrance channel. Thirdly, the unprimed substrate recognition site S1 of F1 tolerates only uncharged residues.

The substrate induces an activation switch of the latent catalytic triad

The unliganded F1 represents an apo-enzymatic state, because neither the oxyanion hole nor the catalytic triad is functional. The Ser105 O-γ blocks direct access to the oxyanion pocket by binding the amide of Tyr106. The

inaccessibility of the oxyanion pocket is witnessed by the water molecule of the free enzyme which, unlike in other serine proteases (Bode, 1979), binds to one of the amides (Gly37) only. The distance and orientation of the Ser105 O-γ to the N-ε of His271 result in a broken catalytic triad (Figures 3A and 4). Upon substrate binding, however, the latent catalytic triad switches to the fully active state (Figures 3B and C, and 4). This conformational change is triggered by the P1 carbonyl oxygen binding to the oxyanion hole, which displaces the Ser105 O-γ towards the N-ε of His271 with a resulting distance of 2.8 Å.

Intriguingly, Ser105 adopts an energetically unfavourable geometry in all three structures, as has been observed in the related peptidases PIP, PAP and POP (Medrano *et al.*, 1998). The conformationally stored energy is likely to be used during catalysis.

In fact, the nucleophilicity of the activated Ser105 must be exceptionally high, since it is able to attack the sulfonic group of a MES molecule. The oxygens shield the sulfur in a way comparable with the carbon in the carboxylate moiety of the P1 product. Such a reaction has been described for classic serine proteinases as exemplified by resynthesis of cleaved bovine trypsin inhibitor (Tschesche and Kupfer, 1976).

Product exit, 'rectifier' mechanism and comparison with tricorn

The processivity of the tricorn-F1 complex (Tamura *et al.*, 1998) implies that tri- or tetrapeptides do not leave the active site of F1 but are completely degraded to single amino acids. This functional property suggests that the cleaved P1 product cannot exit the active site via the entrance E1 (Figure 5). Instead, the free amino acid has to exit the S1 cavity through the smaller E2 channel (Figure 5). If the P1' product is a single amino acid, it is likely to leave the active site directly through the larger opening (E1); otherwise it will be pulled towards Glu213 and Glu245 for further processing, if it fulfils the S1 specificity. The directionality of the substrate flow is maintained by the E2 exit opening mechanism, which is triggered by the loading of the S1 cavity. On a mechanistic level, this E2 gating presumably is caused by the P1 carbonyl binding to the oxyanion hole, which switches the state of the active site. The one-way flow directionality represents a biological rectifier.

Taken together, we find remarkable analogies in the enzymatic mechanisms of F1 and tricorn: both enzymes degrade their substrates processively by employing separate entrance and exit channels to and from the active site, respectively. Their active sites are buried in cavities and utilize a mirrored substrate recognition mechanism. Two glutamates are required at the N-terminus of the substrate in F1, while two arginines of tricorn bind the C-terminus of the peptide (Brandstetter *et al.*, 2001).

A comparison of F1 with PAP, PIP and POP

The distribution of the surface-exposed amino acids of F1 (Karshikoff and Ladenstein, 2001), together with the intramolecular disulfide bridge formed by Cys5 and Cys22, may contribute to its thermostability (Vielle and Zeikus, 2001).

Despite these thermophilic adaptations, the overall α/β-hydrolase structure in the catalytic domain of F1,

Fig. 6. Comparison of the related prolyl peptidases PIP, PAP and POP with F1. The catalytic domains (green and yellow) are structurally conserved within all four enzymes, whereas the cap domain (red) is reduced in POP and largely replaced by a β -propeller (cyan) structure.

PIP (Medrano *et al.*, 1998), PAP (Yoshimoto *et al.*, 1999) and even POP is quite similar (Figure 6), in accordance with their sequence homology (Figure 1).

An overlay of the active sites shows few deviations for the serine, histidine and aspartate of the catalytic triad in the four enzymes, with the remarkable exception of the inactive, latent catalytic triad of F1 in the unliganded form. The oxyanion hole-forming residues that bind the carbonyl oxygen of the P1 residue are conserved in PIP and PAP, and are still present in the mammalian POP, where the OH group of a tyrosine substitutes the backbone amide of the oxyanion pocket-forming glycine.

Both structure and function of the cap domain deviate significantly between the four prolyl peptidases. The hydrogen-bonded network of Glu213, Glu245 and Tyr205 in F1, which is involved in the fixation of the substrate's N-terminus, is only partially conserved in sequence and is structurally rearranged in PIP and PAP. Nonetheless, two glutamates and one tyrosine have the capability to bind the imide nitrogen of an N-terminal proline symmetrically. The carbonyl oxygen of Gly37 in F1, which binds the amino group of PCK, has an exact structural analogue in PIP, whereas in PAP the carbonyl of the corresponding glycine adopts a β -turn. Moreover, F1 lacks the arginine, which is conserved in the active sites of PIP (Arg133) and PAP (Arg136) and was found to be essential for catalysis in PAP (Ito *et al.*, 2000). This arginine is located close to the serine nucleophile, suited to bind the negatively charged C-terminus of dipeptides. In PIP and PAP, the preference for dipeptidic substrates is achieved by the simultaneous presence of a glutamate and an arginine at the N- and C-terminus of the dipeptide, respectively (Yoshimoto *et al.*, 1999). Therefore, PIP and PAP may be considered as both amino- and carboxypeptidases, while the absent arginine explains why F1 tolerates substrates with more than two residues. Finally, the mammalian POP is able to bind and turn over octapeptides, since residues for the binding of the N- and C-termini of the substrate in the active site are missing, explaining its endopeptidase activity (Fülöp *et al.*, 1998).

The narrow entrance (E1 in Figure 5), the S1' cavity and the polar pocket (P) have no counterparts in PIP and PAP, where less homologous parts of the cap and catalytic domains occupy the corresponding regions. On the other

hand, the broad mouth of the access openings of PIP and PAP is filled by a helix of the cap domain in F1. The funnels of PIP and PAP allow direct binding of peptides with the N-terminus ahead into the S1 site. In contrast, the substrate has to be threaded through the bent cavity system of F1 before it reaches the S1 site in a proper orientation (Figure 5). Finally, the open velcro propeller domain serves as substrate access and filter for POP.

The functional F1 complex with the tricorn protease

One of the still unresolved problems of the protein degradation pathway in *Thermoplasma* is the processivity and the possibility of a direct substrate transfer from tricorn to its interacting factors (Tamura *et al.*, 1998). In both F1 and tricorn, a separate entrance to and exit from the active site implement enzymatic processivity. As in the tricorn protease, the size of the connecting channels of F1 reflects that of the substrate and product: the entrance channel is wider than the exit channel (Figures 2C and 5) as the substrates are larger than the single amino acid products.

Two major models for tricorn–F1 interactions have been proposed: the first model is based on an icosahedral supercomplex of tricorn hexamers, discovered in electron micrographs, which contain the peptidases F1, F2 and F3 in a cage-like structure (Walz *et al.*, 1997, 1999). The second model focuses on the tricorn hexamer as the core of presumably transient heterocomplexes, to be formed with F1, F2 and F3 (Brandstetter *et al.*, 2001).

Up to now, the only evidence for the physical formation of those complexes was by dynamic light scattering measurements, where we observed an increase in particle size, when F1 was added to tricorn in solution. The results indicate that up to three F1 molecules may bind per tricorn hexamer. However, BIACORE experiments failed to show binding of tricorn at 40 μ M concentration to immobilized F1 (see <http://www.biacore.com> for a reference). The interpretation of these data is difficult, because of the multi-dispersity of the tricorn protease, which exists in different oligomerization states, as observed in gel filtration and ultracentrifugation experiments (data not shown).

The structure of mammalian POP provides a model of F1 recognition by the tricorn protease: it contains both a

Fig. 7. Model complex of F1 and tricorn derived from the homologous domains in POP. With the catalytic domain of F1 being aligned with that of POP, we superimposed the β 6-propeller of tricorn on the propeller domain of POP. The resulting F1–tricorn complex is oriented identically to POP in Figure 6.

domain homologous to the catalytic domain of F1 and an open topology propeller domain with unique similarity to the tricorn propeller domains. The channel of the β 6-propeller is proposed as the product exit in the tricorn protease and should, therefore, serve as the docking site for F1. By superimposing the β 6-propeller of tricorn and the catalytic domain of F1 on the propeller and catalytic domain of POP, respectively, we obtained a molecular model of the tricorn–F1 complex (Figure 7).

In an attempt to cross-validate the described complex computationally, we employed an automatic docking routine, where all possible F1–tricorn complexes were ranked by a scoring function using shape and charge complementarity. The top ranking solution qualitatively coincides with the complex suggested by POP. Therefore, we assume that there is a low affinity interaction between F1 and tricorn, which is consistent with functional requirements and our observations. Moreover, the often underestimated effects of total macromolecular concentration inside cells, which amounts to an occupancy of up to 30% of the cellular volume, are thought to promote the formation of weakly interacting complexes (Ellis, 2001).

This model extends the concept of substrate channeling, postulated for the tricorn protease, to a multi-enzyme complex. Further experiments will address the structural and functional organization of the components in the proteolytic machinery of *Thermoplasma* and other archaea, and may even provide a first glimpse of the more sophisticated processes in eukaryotes.

Materials and methods

Cloning and protein purification

Molecular biological and biochemical procedures were performed according to standard techniques (Sambrook *et al.*, 1989). The F1 gene

was amplified in a PCR from genomic *T. acidophilum* DNA, and the products were purified with the QIAprep kit (Qiagen). *Nco*I and *Xho*I (New England Biolabs) restriction sites were used to insert the gene into the plasmid pRset6c (Schoepfer, 1993). The expression vector was cloned in *E. coli* DH5 α cells, and subsequently directly transformed in BL21(DE3)RIL cells (Stratagene). The native protein was expressed using LB medium, while for selenomethionine incorporation, bacteria were grown in minimal medium (Budisa *et al.*, 1995).

Cells were harvested by centrifugation and resuspended in 50 mM Tris–HCl pH 8.0, 1 mM EDTA, 0.02% NaN₃ (TEN), 1 mM dithiothreitol (DTT). Cell lysis was achieved by sonification, and heat-labile proteins were removed by heating at 58°C. The soluble protein was purified by anion exchange chromatography using Q-Sepharose and hydroxyapatite, followed by hydrophobic interaction chromatography using octyl-Sepharose, and finally with gel filtration using Superose-12 (Pharmacia). F1 was obtained in quantities of 10 mg from 4 l of culture medium and at >95% purity for crystallization, as judged by Coomassie-stained SDS–gels. The identity of the tricorn-interacting factor F1 was confirmed by N-terminal sequencing of the first 10 residues and by mass spectrometry that gave a molecular mass of 33 488.0 Da (theoretical value 33 487.1 Da).

Crystallization and data collection

The starting concentration for native and selenomethionine protein was 20 mg/ml in drops mixed with equal volumes of reservoir buffer. Suitable crystallization buffers were 100 mM MES, Bis-Tris–HCl, both at pH 6.0 with 7–12% polyethyleneglycol (PEG) 6000 and, alternatively, 100 mM HEPES pH 7.5 with 20% PEG 8000. Within 2 days, orthorhombic crystals of space group $P2_12_12_1$ were grown by the hanging drop vapour diffusion method. The inhibitor PCK (BACHEM) was dissolved to 100 mM in dimethylsulfoxide and added to F1 crystals containing drops with a 30-fold molar excess of the inhibitor. After 30 min soaking time, the crystals were harvested and mounted for X-ray measurements.

MAD experiments with synchrotron radiation were performed at the BW6 beamline of DESY in Hamburg. After a wavelength scan for the selenium absorption in F1 single crystals, data sets were collected on a MAR CCD detector for peak (0.9788 Å), inflection point (0.9795 Å) and remote (0.9500 Å) wavelength. Images in frames of 1° were recorded over a range of 80°, and, in the case of occurring anomalous contributions, additionally with the crystal rotated by 180°, to measure reflections of the Friedel partners at 2.0 Å resolution. The images were processed with DENZO and SCALEPACK (Otwinowsky and Minor, 1997), and scaled further with TRUNCATE, CAD and SCALEIT of the CCP4 software (CCP4, 1994). X-ray diffraction data for the unliganded protein were collected at the Swiss beamline at Zürich (SLS) with synchrotron radiation, and inhibitor-complexed protein was measured with Cu–K α radiation generated by a rotation anode and collected on a MAR 345 image plate, processed with MOSFLM (Leslie, 1991), SCALA and CAD of the CCP4 suite (Table I).

Heavy atom search and phasing

Heavy atom positions were searched with difference Patterson methods, and resulting maps of the peak wavelength showed several distinct positions of strong anomalous scatterers that were employed to identify further sites by heavy atom search procedures using the CNS program Version 1.1 (Brünger *et al.*, 1998). After cross-vector verification in MLPHARE (CCP4, 1994), two mutually inverted sets of eight selenium positions were refined and used for MAD phasing in CNS. Fourier transforms and application of solvent flattening resulted in electron density maps at low resolution that allowed the interpretation of secondary structure elements. The chirality ambiguity was solved finally by matching the interpreted protein with a model of the F1 core, built with homology modelling using the structures of PIP, PAP and POP (SWISSPROT, <http://www.expasy.ch/sprot/>). Improved experimental phases and electron densities up to 2.0 Å resolution were obtained with the program SHARP (de La Fortelle and Bricogne, 1997).

Model building and refinement

Model building was performed with the interactive three-dimensional graphics program MAIN (Turk, 1992). Most of the side chains could be identified unambiguously and, with the exception of the first three N-terminal amino acids, all residues sequentially fit the electron density maps. The resulting model was refined and water molecules were added by the water_pick routine in CNS, iteratively corrected in model-phased $2F_o - F_c$ maps and in maps with combined experimental and model phases. The quality of the model was cross-validated by using 5% of independent reflections in a test data set (Table II). Ligand coordinates

Table I. X-ray diffraction data and MAD phasing statistics

	SeMetF1–MES			F1–PCK	F1
Wavelength	0.9788 Å	0.9795 Å	0.9500 Å	1.5418 Å	0.9794 Å
Resolution	20–2.0 Å	20–2.0 Å	20–2.0 Å	20–2.4 Å	20–1.8 Å
Unique reflections	33 131	33 102	17 462	11 344	22 463
Completeness	99.0%	98.9%	98.0%	85.7%	91.1%
<i>I</i> / σ	24.6	27.7	14.3	3.2	9.4
No. of selenium sites	8	8	8		
Phasing power	1.625	1.487			

Table II. Refinement statistics

	SeMetF1–MES	F1–PCK	F1
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell constants	<i>a</i> = 52.14 Å, <i>b</i> = 60.25 Å, <i>c</i> = 80.39 Å	<i>a</i> = 57.58 Å, <i>b</i> = 61.72 Å, <i>c</i> = 80.74 Å	<i>a</i> = 54.96 Å, <i>b</i> = 57.10 Å, <i>c</i> = 82.27 Å
Non-hydrogen atoms	2360	2345	2328
Water molecules	77	36	179
Resolution range	20.0–2.0 Å	20.0–2.4 Å	20.0–1.8 Å
Reflections in refinement	31 976	9958	22 449
Crystallographic <i>R</i> -factor ^a	25.5%	36.6%	19.9%
Free <i>R</i> -factor ^b	28.0%	31.4%	23.6%
R.m.s. bond lengths	0.012 Å	0.004 Å	0.007 Å
R.m.s. bond angles	1.594°	0.948°	1.210°

^aCrystallographic *R*-factor $R = \sum_{\text{hkl}} \|F_{\text{obs}}\| - kF_{\text{calc}} / \sum_{\text{hkl}} \|F_{\text{obs}}\|$.

^bFree *R*-factor $R_{\text{free}} = \sum_{\text{hkl}} \|F_{\text{obs}}\| - kF_{\text{calc}} / \sum_{\text{hkl}} \|F_{\text{obs}}\|$.

were obtained on the HICUP-server (<http://xray.bmc.uu.se/hicup/>) and refined together with standard topology and parameter files (Engh and Huber, 1991) in CNS. The model of selenomethionine (SeMet) F1 was used as a starting point for the refinement against the native F1 data at 1.8 Å and the PCK–F1 data at 2.4 Å resolution. Probably due to high mosaicity of the PCK-soaked crystal, the refinement of the F1–PCK complex stalled at relatively high *R*-factors ($R_{\text{cryst}} = 31.4\%$, $R_{\text{free}} = 36.6\%$). Nevertheless, the electron density clearly revealed the orientation of the bound inhibitor at the active site. The stereochemistry of the final models was analysed with PROCHECK (Laskowski *et al.*, 1993). Coordinates have been deposited in the Protein Data Bank under accession codes 1MT3, 1MTZ and 1MU0.

Sequence comparisons, model docking and graphic representations

The search for sequential and structural analogues of F1 was performed with the BLAST search tools, available on the EXPASY site (<http://www.expasy.ch>), and with the DALI-server (<http://croma.ebi.ac.uk/dali>). Sequence alignment was done by the SEQLAB[®] (GCG, Wisconsin) and LALIGN software (<http://www.ch.embnet.org>), and depicted with ALSCRIPT (Barton, 1993). Surface potentials were calculated with the GRASP program (Nicholls, *et al.*, 1993). Model complexes of F1 with tricorn were found by FTDOCK (Gabb *et al.*, 1997). Figures were created with MOLSCRIPT (Kraulis, 1991), BOBSCRIPT (Esnouf, 1999) and Raster3D (Merrit and Bacon, 1997).

Acknowledgements

We wish to thank Ravishankar Ramachandran for critical reading of the manuscript, Gerd Bader for support in model docking programs, and Walter Göhring for BIACORE experiments. We also thank Hans Bartunik and Gleb Bourenkov of the Deutsches Elektronensynchrotron (DESY) at Hamburg, as well as Clemens Schulz-Briese and Takashi Tomizaki of the Swiss light source (SLS) at Zürich, for help with measurements using synchrotron X-ray radiation.

References

- Barton, G.J. (1993) ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.*, **6**, 37–40.
- Bochtler, M., Ditzel, L., Groll, M., Hartmann, C. and Huber, R. (1999) The proteasome. *Annu. Rev. Biophys. Biomol. Struct.*, **28**, 295–317.
- Bode, W. (1979) The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. II. The binding of the pancreatic trypsin inhibitor and of isoleucine–valine and of sequentially related peptides to trypsinogen and to *p*-guanidinobenzoate–trypsinogen. *J. Mol. Biol.*, **127**, 357–374.
- Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S.R. and Hofsteenge, J. (1989) The refined 1.9 Å crystal structure of human α -thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.*, **8**, 3467–3475.
- Brandstetter, H., Kim, J.-S., Groll, M. and Huber, R. (2001) Crystal structure of the tricorn protease reveals a protein disassembly line. *Nature*, **414**, 466–470.
- Brünger, A.T. *et al.* (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D*, **54**, 905–921.
- Budisa, N., Steipe, B., Demange, P., Eckes, C., Kellermann, J. and Huber, R. (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminoheptanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur. J. Biochem.*, **230**, 788–796.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
- de LaFortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.
- Ellis, R.J. (2001) Macromolecular crowding: obvious but underestimated. *Trends Biochem. Sci.*, **26**, 597–604.
- Engh, R.A. and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein-structure refinement. *Acta Crystallogr. A*, **47**, 392–400.
- Esnouf, R. (1999) Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D*, **55**, 938–940.
- Fülöp, V., Böcskei, Z. and Polgar, L. (1998) Prolyl oligopeptidase: an unusual β -propeller domain regulates proteolysis. *Cell*, **94**, 161–170.
- Fülöp, V., Böcskei, Z. and Polgar, L. (2001) Structures of prolyl oligopeptidase: substrate/inhibitor complexes. *J. Biol. Chem.*, **276**, 1262–1266.
- Gabb, H.A., Jackson, R.M. and Sternberg, M.J.E. (1997) Modelling docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.*, **272**, 106–120.
- Ito, K., Inoue, T., Kabashima, T., Kanada, N., Huang, H.-S., Ma, X., Azmi, N., Azab, E. and Yoshimoto, T. (2000) Substrate recognition mechanism of prolyl aminopeptidase from *Serratia marcescens*. *J. Biochem.*, **128**, 673–678.
- Jozic, D., Bourenkov, G., Bartunik, H., Scholze, H., Dive, V., Henrich, B., Huber, R., Bode, W. and Maskos, K. (2002) Crystal structure of the dinuclear zinc aminopeptidase PepV from *Lactobacillus delbrueckii* unravels its preference for dipeptides. *Structure*, **10**, 1097–1106.
- Karshikoff, A. and Ladenstein, R. (2001) Ion pairs and thermotolerance of proteins from hyperthermophiles: a 'traffic rule' for hot roads. *Trends Biochem. Sci.*, **26**, 550–556.
- Kraulis, P.J. (1991) MOLSCRIPT—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–940.

- Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Leslie,A.G.W. (1991) Macromolecular data processing. In Moras,D., Pordjany,A.D. and Thierry,J.D. (eds), *Crystallographic Computing V*. Oxford University Press, Oxford, pp. 27–38.
- Löwe,J., Stock,D., Jap,B., Zwickl,P., Baumeister,W. and Huber,R. (1995) Crystal structure of the 20S proteasome from the archaeon *Thermoplasma acidophilum* at 3.4 Å resolution. *Science*, **268**, 533–539.
- Lupas,A., Flanagan,J.M., Tamura,T. and Baumeister,W. (1997) Self-compartmentalizing proteases. *Trends Biochem. Sci.*, **22**, 399–404.
- Medrano,F.J., Alonso,J., Garcia,J.L., Romero,A., Bode,W. and Gomis-Rüth,F.X. (1998) Structure of proline iminopeptidase from *Xanthomonas campestris* pv. *citri*: a prototype for the prolyl oligopeptidase family. *EMBO J.*, **17**, 1–9.
- Merrit,E.A. and Bacon,D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.*, **277**, 505–524.
- Nicholls,A., Bharadwaj,R. and Honig,B. (1993) Grasp—graphical representation and analysis of surface properties. *Biophys. J.*, **64**, A166.
- Otwinowsky,A. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Pallen,M.J., Larn,A.C. and Loman,N. (2001) Tricorn-like proteases in bacteria. *Trends Microbiol.*, **9**, 518–521.
- Ruepp,A. *et al.* (2000) The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature*, **407**, 508–513.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Schoepfer,R. (1993) The pRSET family of T7 promotor expression vectors for *Escherichia coli*. *Gene*, **124**, 83–85.
- Stoltze,L. *et al.* (2000) Two new proteases in the MHC class I processing pathway. *Nat. Immunol.*, **1**, 413–418.
- Tamura,T., Tamura,N., Cejka,Z., Hegerl,R., Lottspeich,F. and Baumeister,W. (1996a) Tricorn protease—the core of a modular proteolytic system. *Science*, **274**, 1385–1389.
- Tamura,T., Tamura,N., Lottspeich,F. and Baumeister,W. (1996b) Tricorn protease (TRI) interacting factor 1 from *Thermoplasma acidophilum* is a proline iminopeptidase. *FEBS Lett.*, **398**, 101–105.
- Tamura,N., Lottspeich,L., Baumeister,W. and Tamura,T. (1998) The role of tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation. *Cell*, **95**, 637–648.
- Tomkinson,B. (1999) Tripeptidyl peptidases: enzymes that count. *Trends Biochem. Sci.*, **24**, 355–359.
- Tschesche,H. and Kupfer,S. (1976) Hydrolysis–resynthesis equilibrium of the lysine–15–alanine–16 peptide bond in bovine trypsin inhibitor. *Hoppe Seylers Z. Physiol. Chem.*, **357**, 769–776.
- Turk,D. (1992) Weiterentwicklung eines Programms für Molekulargraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen. PhD Thesis, Technische Universität, München.
- Vielle,C. and Zeikus,G.J. (2001) Hyperthermophilic enzymes: sources, uses and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.*, **65**, 1–43.
- Voges,D., Zwickl,P. and Baumeister,W. (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.*, **68**, 1015–1068.
- Walz,J., Tamura,T., Tamura,N., Grimm,R., Baumeister,W. and Koster,A. (1997) Tricorn protease exists as an icosahedral supermolecule *in vivo*. *Mol. Cell*, **1**, 59–65.
- Walz,J., Tamura,T., Tamura,N., Grimm,R., Baumeister,W. and Koster,A. (1999) Capsids of tricorn protease studied by electron cryomicroscopy. *J. Struct. Biol.*, **128**, 65–68.
- Yoshimoto,T., Kabashima,T., Uchikawa,K., Inoue,T., Tanaka,N., Nakamura,K.T., Tsuru,M. and Ito,K. (1999) Crystal structure of prolyl aminopeptidase from *Serratia marcescens*. *J. Biochem.*, **126**, 559–565.

Received April 22, 2002; revised August 8, 2002;
accepted August 29, 2002