doi:10.1016/S0022-2836(02)01153-1 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 324, 1041–1050

JMB

Navigation Inside a Protease: Substrate Selection and Product Exit in the Tricorn Protease from *Thermoplasma acidophilum*

Jeong-Sun Kim*, Michael Groll, Hans-Jürgen Musiol Raymond Behrendt, Markus Kaiser, Luis Moroder, Robert Huber and Hans Brandstetter

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

*Corresponding author

The proposed pathway and mechanism of substrate entry and product egress in the hexameric D3 symmetric tricorn protease from *Thermoplasma acidophilum* were explored by crystallographic studies of ligand complexes and by structure-based mutagenesis. Obstruction of the pore within the 7-bladed β-propeller (β7) domain by alkylation or oxidation of an engineered double cysteine mutant strongly decreased enzymatic activities. In line herewith, the crystal structure of the tricorn protease in complex with a trideca-peptide inhibitor modifying the catalytic Ser965 revealed part of the peptide trapped inside the channel of the β7 domain. The cysteine mutation widening the lumen of the 6-bladed β-propeller (β6) domain enhanced catalytic activity, which was restored to normal values after its alkylation. A charge reversal mutant at the putative anchor site of the substrate C terminus, R131E-R132E, drastically reduced the proteolytic activity. The complex crystal structure of a peptide inhibitor with a diketo group at the cleavage site mapped the substrate recognition site and confirmed the role of Arg131-Arg132 as an anchor site. Our results strongly suggest the wider β7 domain to serve as a selective filter and guide of the substrate to the sequestered active site, while the narrower β6 domain routes the product to the surface. Moreover, we identified the role of Arg131-Arg132 in anchoring the substrate C terminus.

Keywords: crystal structure; enzymatic mechanism; hydrolase; β-propeller; tricorn protease

Introduction

Protein synthesis and degradation in cells are balanced to maintain homeostasis of the required cellular components. Initial intracellular protein degradation is carried out mostly by unspecific proteases with sieving mechanisms for substrate selection. Commonly, they form molecular cages by oligomerization to exclude tertiary folded substrates from their central cavities harboring the active sites. Proteasomes are representatives of this class of proteases, which form large multi-subunit complexes reaching molecular weights of 2.0 MDa. They take part in the degradation of most cytosolic proteins[1] and at least some proteins of the endoplasmic reticulum.[2–4] Their central entry ports are not wide enough to allow folded substrates to reach the active sites.[5–7] Accessory proteins form the regulatory particle of the proteasome and are attached to the core particle to regulate the substrate access to the active sites in an ATP-dependent way.[8] The proteasome removes

Abbreviations used: β6, 6-bladed β-propeller; β7, 7-bladed β-propeller; FMI, fluorescein-5-maleimde; NEM, *N*-ethylmaleimde; FITC, fluorescein isothiocyanate; RP, reverse-phase; Bz, benzoyl; AMC, 7-amino-4-methylcoumarin; Z, benzyloxycarbonyl; TLCK, *N*-α-tosyl-L-lysine chloromethyl ketone; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; inh1, decanoyl-Arg-Val-Arg-Lys-chloromethyl ketone; inh2, H-Asp-Gln-Thr-Gln-Lys-Gln-Tyr-Gln-Glu-Leu-Thr-Phe-Phe-chloromethyl ketone; inh3, Z-PheΨ[CO–CONH]Arg-Glu-Phe-OH; inh4, D-Phe-Pro-Arg-chloromethyl ketone; inh5, H-Ala-Ala-Pro-Val-chloromethyl ketone.

E-mail address of the corresponding author: jkim@biochem.mpg.de

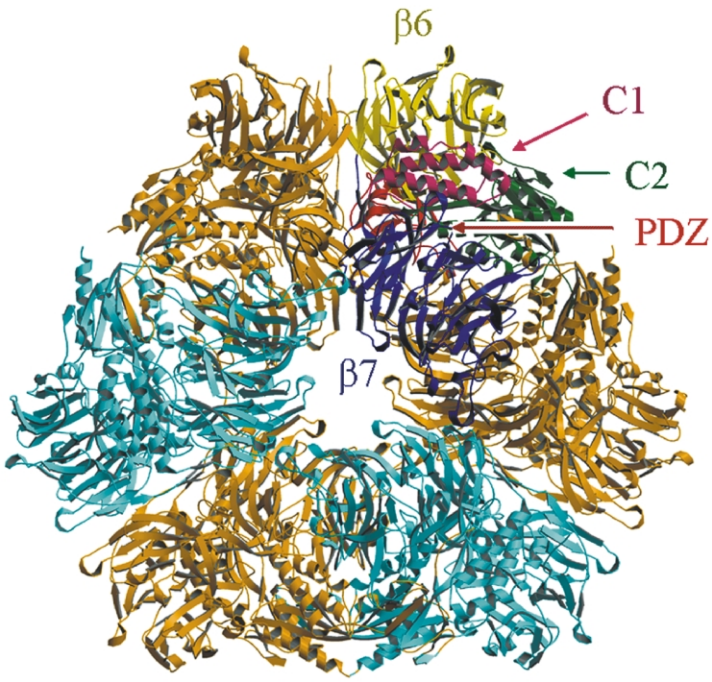

Figure 1. Ribbon representation of the overall hexameric structure of the tricorn protease viewed along the molecular 3-fold axis. Individual subunits are drawn in alternating colors. One subunit is colored in yellow, blue, purple, red, and dark-green for the β6, β7, C1, PDZ, and C2 subdomain from the N terminus. Figures 1, 2, 4 and 5(a),(c) were prepared using Molscript,[26] GRASP,[27] Povscript[28] and rendered with Raster3D.[29]

misfolded or unneeded proteins in eukaryotes and archaea. Peptides, with a preferred length between 7 and 9 amino acid residues, are the major products of the proteasome. For reuse in protein synthesis or energy production, these peptides must be further degraded into free amino acid residues. Recently, a different and ATP-independent substrate filtering mechanism has been suggested for prolyl oligopeptidase from human and porcine, which makes use of an open Velcro β-propeller structure for substrate selection and guidance to the active site.[9,10]

In the archaeon *Thermoplasma acidophilum*, the degradation of proteasomal products is performed by the tricorn protease to preferentially yield di- and tripeptides. The further and final degradations to free amino acid residues are accomplished by the tricorn interacting factors, a proline iminopeptidase termed F1 and two metalloproteases termed F2 and F3.[11–13]

Tricorn is a hexameric protease of 720 kDa and can assemble into a giant icosahedral capsid, which might serve as the organizing center of a multi-proteolytic complex.[14,15] The 120 kDa subunits have a mosaic structure, two open Velcro β-propeller structures of 6 and 7 blades (Met39-Asp310 and Ala326-Lys675), a helical bundle (Val679-Ser745, C1), a PDZ domain (Arg761-Asp855) and an αβ sandwich structure (Arg856-Asn1061, C2) (Figure 1). The active site residues, Ser965 and His746, are located in the C-terminal half. The β7 structure and the PDZ domain, together with two arms (Leu520-Val535 and Asn930-Asn949), provide the contact interfaces in the tight dimer subcomplex of the tricorn protease.[16]

On the basis of the topology of the domains and the location and accessibility of the active site, we had suggested that the β7 domain serves as a substrate filter and access channel to the active site, while the β6 domain allows the product egress from the active site. A prominent segment with two adjacent arginines (Arg131-Arg132) offered itself as a docking site for the substrate C terminus.[16] In this article, we present biochemical and structural results for detailing

Table 1. Summary of enzymatic activities of the wild-type and mutants

	Wild-type	β7 R414C, A643C	Anchor R131E-R132E	β6 L184C
AMC	[a]	[b]	[a]	[c]
Insulin B	[a]	[d]	[e]	[c]
Casein	[a]	NM	[f]	NM
After oxidation				
AMC	[a]	[e]	NM	NM
After alkylation				
AMC	[a]	[e]	NM	[b]
Insulin B	[a]	[e]	NM	[b]

NM: not measured, AMC: substrate, Bz-Val-Gly-Arg-AMC, alkylating agent: NEM or FMI, oxidizing agent: oxidized glutathione. Fluorogenic AMC substrate was purchased from Bachem, FMI from Molecular Probes and FITC-labeled casein, NEM, oxidized insulin B-chain and oxidized glutathione from Sigma-Aldrich.

[a] 80–100%.
[b] 55–80%.
[c] > 150%.
[d] 35–55%.
[e] 15–35%.
[f] < 15%.

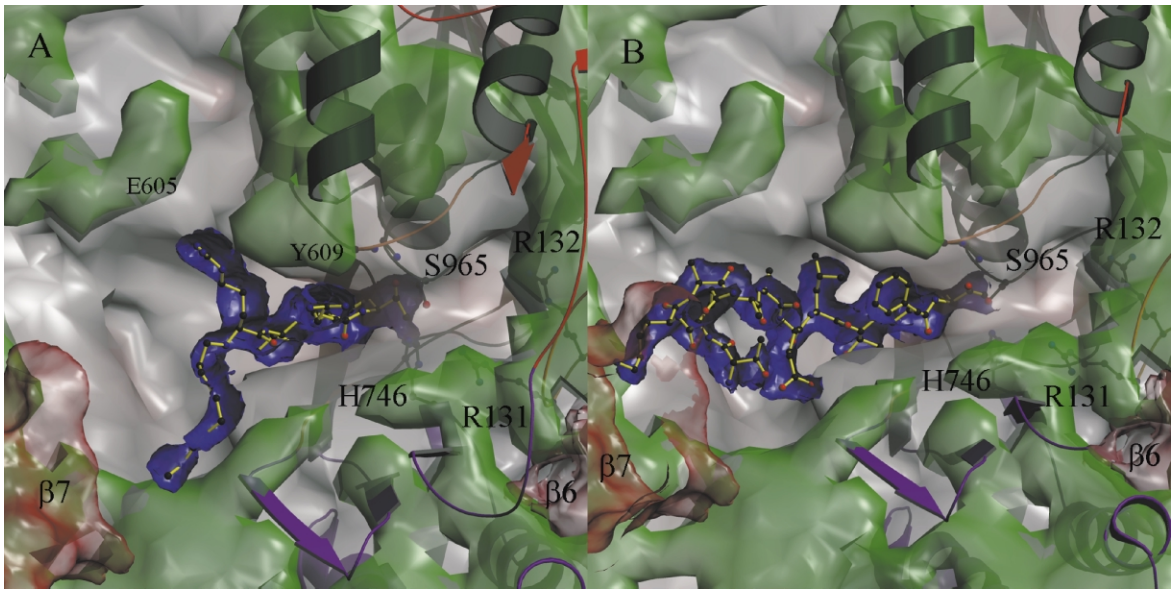

Figure 2. Cut-open surface representation with the averaged electron density map around the bound inh1 at 2.8 Å and the inh2 at 2.6 Å, respectively. The lumen of the β7 structure is colored in red in the left lower part, the β6 in purple in the right lower part, and the experimental electron density in blue. Tricorn segments are drawn as a ribbon diagram and the inhibitor, Ser965, His746, Arg131, and Arg132 as colored stick models. The inhibitor is covalently bound to Ser965. (a) The deca-hydrocarbon chain is not long enough to reach the β7 domain although it was clearly traced. (b) Towards the N terminus of the inhibitor, the density does not allow positioning the side-chains, but the main-chain can be unequivocally traced. Undefined side-chains have been truncated to alanines. The inhibitor conformation is extended from Glu1109 to Phe1113 and forms a hairpin-like structure from Gln1104 to Tyr1107.

these suggestions and elucidate the enzymatic mechanism of the tricorn protease.

Results

Substrate enters the active site through the β7 domain

To examine the role of the β7 domain, we designed a mutant obstructing its central propeller channel by replacing two residues with cysteine, which are located at the entrance to the β7 domain, R414C and A643C. Compared with the wild-type enzyme, this double mutant enzyme itself exhibited 70% and 50% residual activities towards fluorogenic substrate and insulin B-chain, respectively, prior to any modification or oxidation. After modification with fluorescein-5-maleimide (FMI), the enzymatic activity with fluorogenic substrates was further decreased to 50%, with insulin B-chain to less than 40%, resulting in relative residual activities of 35% and 20% of the wild-type enzyme towards fluorogenic substrate and insulin B-chain, respectively. Oxidation of both cysteine residues by oxidized glutathione exhibited similar results, with 30% activities of the wild-type. As expected, no effect on the wild-type protein was observed by alkylation or oxidation (Table 1).

In order to visualize directly the role of the β7 domain during the proteolysis, we analyzed the crystal structure in complex with chloromethyl ketone-based peptide inhibitors of various lengths covalently bound to the active site Ser965. In the complex crystal structure of decanoyl-Arg-Val-Arg-Lys-chloromethyl ketone (inh1), we could observe electron density for the N-terminal deca-hydrocarbon tail. However, the inhibitor was not long enough to reach the β7 tunnel (Figure 2(a)). The H-Asp-Gln-Thr-Gln-Lys-Gln-Tyr-Gln-Glu-Leu-Thr-Phe-Phe-chloromethyl ketone (inh2) inhibitor complex structure clearly showed electron density for the peptide backbone trapped in the tunnel of the β7 domain, even though the density of the inhibitor segment Gln1106-Gln1108 was defined only for its main-chain. The N-terminal part of this inhibitor is located in the funnel-shaped end of the β7 tunnel and forms a hairpin structure. The chloromethyl ketone group alkylates the catalytic Ser965 *via* an O^{γ}-methyl covalent bond. The carbonyl oxygen of P1 residue is trapped in the oxyanion hole formed by the backbone amides of Gly918 and Asp966 (Figure 2(b)).

The cleaved products exit through the β6 domain

The role of the β6 domain was examined by the mutation L184C. This residue is located midway in the pore of the β6 domain. The mutant showed a doubling of both peptidolytic and proteolytic activities. After modification of L184C with *N*-ethylmaleimide (NEM), activities compared to

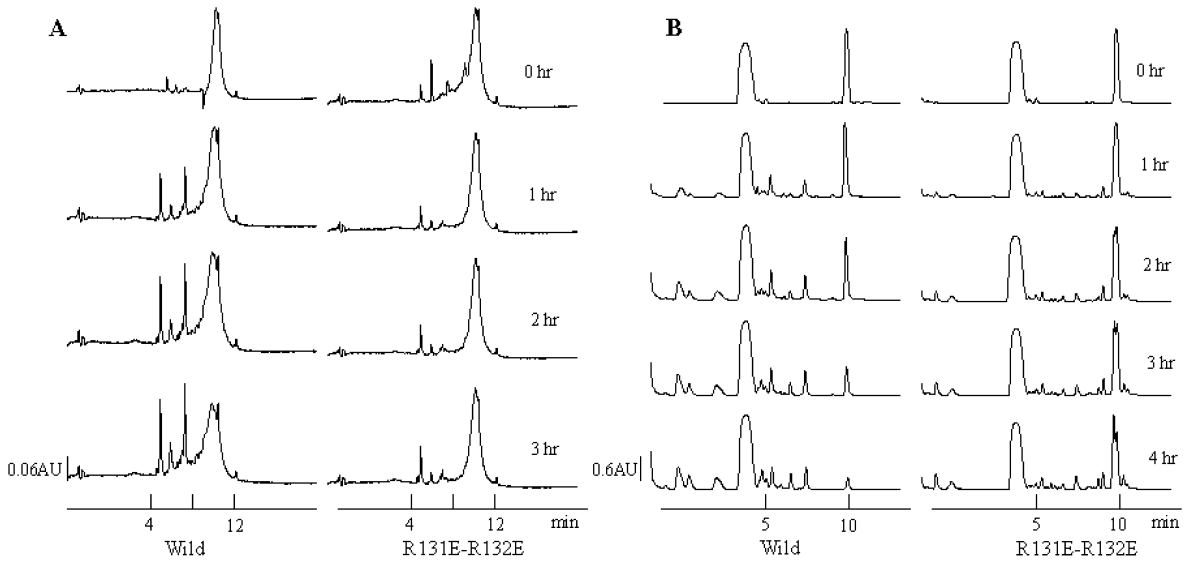

Figure 3. Protein substrate digestions with the wild-type and a charge reversal mutant at the substrate C-terminal docking site (R131E-R132E). Time-dependent degradation results show that a charge reversal mutant at the substrate C terminus anchoring site has about 10% activities of the wild-type towards FITC-labeled casein (a), and about 30% towards oxidized insulin B-chain (b). Moreover, the cleavage patterns of a charge reversal mutant differ from those of the wild-type, but remain time-invariant, indicating processivity. The high peaks in (a) around 11 minutes retention time correspond the undegraded FITC-labeled caseins. The uneven base lines of (a) at 0 hour appear to be due to insufficient column equilibration. The high peaks in (b) around five and ten minutes retention time originate from the inh5, which was added to stop the reaction after specific times, and undegraded oxidized insulin B-chain, respectively. The inh5 was purchased from Bachem.

the wild-type were reduced for both fluorogenic substrates and insulin B-chain (Table 1).

The substrate C-terminal is anchored at Arg131-Arg132

On the basis of the structural features of the tricorn protease, we had suggested that a peptide loop across the inner funnel into the tunnel of the β6 domain would be the anchoring site for the substrate C terminus. This peptide loop carries Arg131-Arg132. When mutated to glutamates, R131E-R132E, the enzyme exhibited the same activity towards fluorogenic substrates as the wild-type. However, the mutated enzyme showed significantly lowered enzymatic activity for protein substrates, about 10% for fluorescein isothiocyanate (FITC)-labeled casein compared with the wild-type, and about 30%, for oxidized insulin B-chain (Figure 3, Table 1). This mutant enzyme degraded insulin B-chain in a processive manner, in the same way that has been described for the wild-type enzyme.[13] The cleavage patterns, however, are different (Figure 3(b)).

The crystal structure of the benzyloxycarbonyl (*Z*)-PheΨ[CO–CONH]Arg-Glu-Phe-OH (inh3) with a diketo group at the cleavage site showed that the side-chains of Arg131-Arg132 are hydrogen-bonded to the carbonyl oxygen of the Glu-P2′ and Phe-P3′ peptide. While they displayed no hydrogen bonds to the C terminus of the Phe-P3′ residue, the electrostatic charges are partially balanced (Figure 4). However, a free carboxy terminus is not required and the tricorn protease processes also a substrate with an amidated C terminus (data not shown).

Two β-strands sandwich the substrate at the active site

Earlier structures in complex with short inhibitors, *N*-α-tosyl-L-lysine chloromethyl ketone (TLCK) and *N*-tosyl-L-phenylalanine chloromethyl ketone (TPCK), identified His746 and Ser965 as the catalytic residues of the tricorn protease. The methylene group of the inhibitor ketone is covalently bound to Ser965 with its carbonyl oxygen trapped in the oxyanion hole. This mode of interaction differs from thrombin which is alkylated at His57 and a tetrahedral adduct formed with Ser195.[17] Asp936 of the tricorn protease at the bottom of the S1 pocket, originating from a neighboring monomer, occurs in two conformations adapting to the characteristics of the substrate side-chains bound at P1. It makes an ionic interaction with the positively charged lysine residue, but rotates away when a phenylalanine side-chain is located in this pocket, which is shaped by Ile969, Val991, and Phe1013.[16]

In order to elucidate the substrate-binding mode of the tricorn protease in more detail, several crystal structures in complex with inhibitors of different lengths and residue characteristics were examined. Inh1–inh3, and a chloromethyl ketone

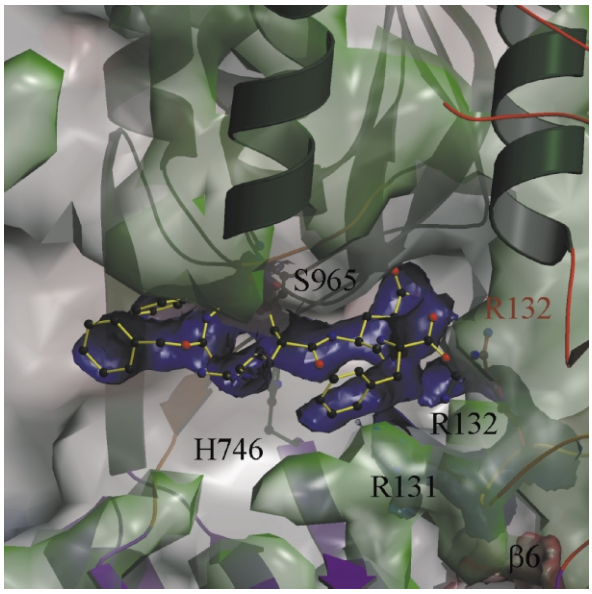

Figure 4. Cut-open surface representation with the averaged electron density map of the bound inh3 between the active site and the Arg131-Arg132 anchoring segment at 2.7 Å. Tricorn segments are drawn as a ribbon diagram, side-chains of key residues at the active site (Ser965 and His746), the anchor segment (Arg131-Arg132), and the inhibitor as colored stick models. The conformation of the Arg132 side-chain as seen in the chloromethyl ketone-based inhibitor and inhibitor-free structure was drawn in red sticks and labeled with same color. The inhibitor is bound *via* the first carbonyl carbon of the diketo group to the O^{γ} of Ser965 as a tetrahedral adduct.

derivative of tripeptide, D-Phe-Pro-Arg-chloromethyl ketone (inh4), which is a specific inhibitor of thrombin, clearly showed that two strands, Gly993-Thr995 and Gly916-Gly918, define both the unprimed and primed substrate recognition site at the active cleft (Figure 5). Upon binding, the substrate inserts between these strands to form a local anti-parallel β-sheet by providing hydrogen donors and acceptors to the protein matrix. The interactions are mainly between the backbone atoms of protease and substrate (Figure 5). Like in the complex crystal structures of TLCK and TPCK, the inhibitors are covalently bound to Ser965 by alkylation of the hydroxyl group in the case of the chloromethyl ketones and tetrahedral adduct formation with the −PheΨ[CO−CONH]− moiety in the case of inh3 (Figures 2, 4). The carbonyl oxygen of the conserved Gly918 serves as a hydrogen bond acceptor of the backbone amide of the P1 residue. Ile994 is on the opposite side and acts as a hydrogen bond donor with its backbone amide, to the carbonyl oxygen of the substrate P2 residue and as an acceptor with its carbonyl oxygen, to the peptide amide of the P2 residue (Figure 5). The P1 residue of the inh1, Lys-P1, binds to Asp936 in the S1 pocket analogously as in the TLCK−tricorn complex. Intriguingly, Arg-P2 is interacting with the peptide oxygen of Tyr609 while the P3 residue, Val-P3, is trapped in the neutral S2 pocket of Tyr609, Phe919 and Phe1011. Arg-P4 interacts with the side-chain of Glu605 (Figure 2(a)). The inh2 is bound to the active site keeping the same interactions between protein and ligand residues as the other chloromethyl ketone-based inhibitors. In particular, its Phe-P1, Phe-P2 and Thr-P3 residues fit well into the corresponding S1, S2 and S3 pocket, as shown by the clear electron density in Figure 2(b). The P1 residue of the inh4, Arg-P1, maintains the same conformation as the TLCK inhibitor in the S1 pocket. Pro-P2 and Phe-P3 are positioned inside neutral pockets (data not shown).

At the primed sites observed in the inh3 complex, only two hydrogen bonds contribute besides the interactions with Arg131-Arg132, as described below. The peptide nitrogen of the Glu-P2′ is forming a hydrogen bond with the carbonyl oxygen of Gly916. Also, the side-chain of the Arg-P1′ is fixed by the peptide oxygen of Tyr609 (Figure 4).

Discussion

β-propeller structures are generally rigidified by clamping their terminal blades either through the exchange of β-strands between them,[18,19] or in the case of 4-bladed β-propeller, by a disulfide bond to covalently connect them.[20] In contrast, prolyl oligopeptidase has an open Velcro β7 domain with regular sequential strand arrangement but lacking a clamp for its terminal blades. The lack of strand exchange between the terminal blades has been suggested to allow structural flexibility.[9]

Similarly, both the β6 and the β7 structures of the tricorn protease are unconstrained. Their central pores provide the shortest path of substrate to, and product from, the active site. Arg414 and Ala643 are located at the pore entry to the β7 domain (Figure 5(a)) and showed reductions of activity with fluorogenic and protein substrates, when mutated to cysteine residues. After alkylation or oxidation of both cysteine residues, the activity was strongly reduced. The introduction of bulky side groups into the lumen of the propeller pore is likely to hinder substrate passage, as does disulfide bond formation. The latter modification might also reduce a breathing motion of the propeller blades by formation of a disulfide bond between the two introduced cysteine residues or modification by alkylation. Disulfide bond formation was not examined, however. As the channel diameter measured between the Cα carbon atoms of Arg414 and Ala643 is only 6.4 Å in the free enzyme, transient widening would facilitate substrate passage. It is interesting to note that the entry part of the blade containing Ala643 appears more flexible, as indicated by relatively high crystallographic *B*-values in the free enzyme (above 42 $Å^2$, compared to 30 $Å^2$ of the other blades). Complete inhibition of enzyme activity had been observed in prolyl oligopeptidase when a disulfide bond was introduced between the two

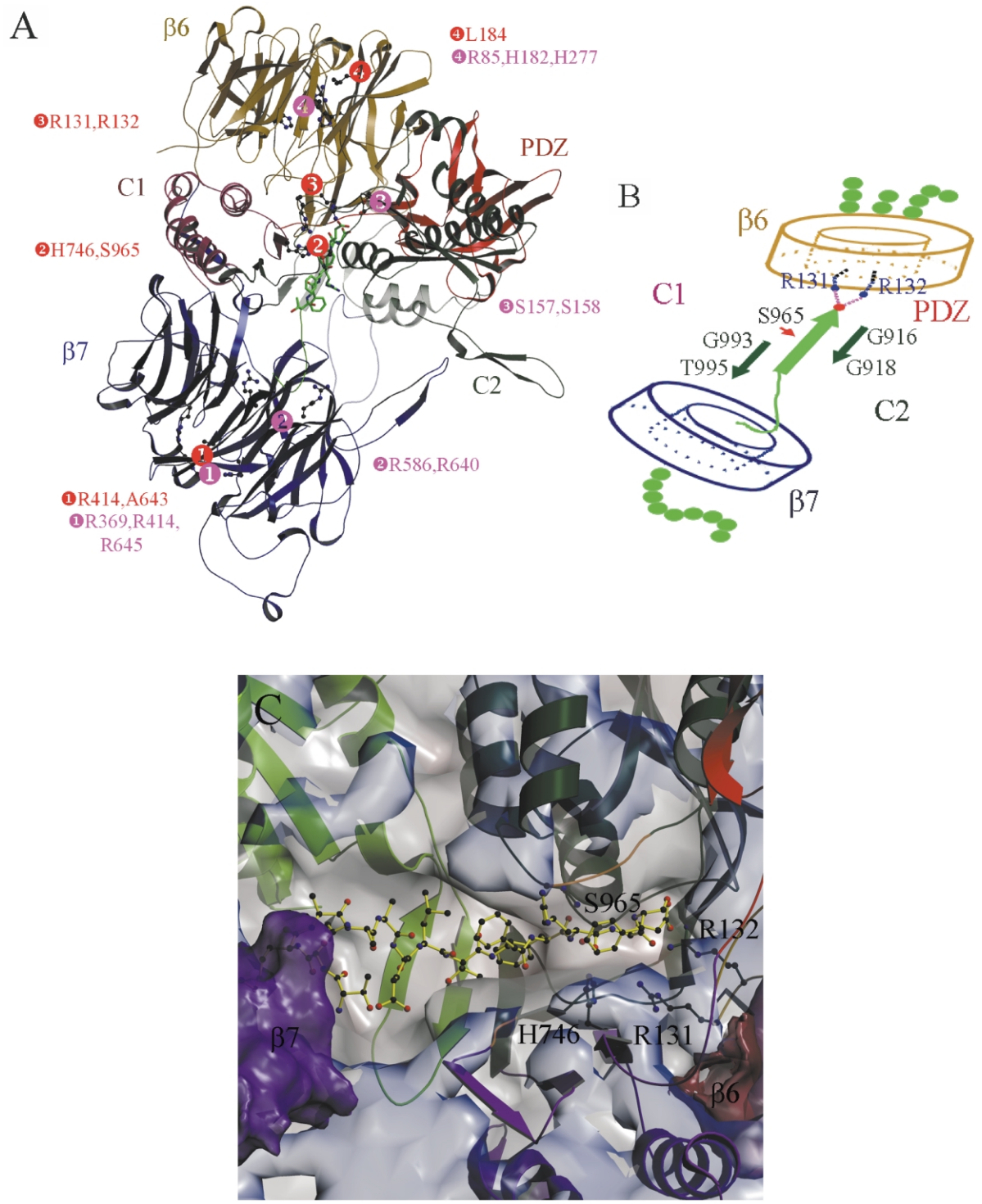

Figure 5. Substrate binding mode derived from the experimental data. (a) The tricorn monomer with a model of a physiological substrate at the active site. A tricorn monomer is drawn in a ribbon representation. Each domain is displayed by different colors. The modeled substrate is highlighted by thick sticks, whereas its N-terminal tail that lacks defined interactions with the protease is drawn as a thin green coil in the β7. A helix (Phe606-Tyr610) was simplified as a blue coil. The substrate strand and strands for forming a local anti-parallel β-sheet with substrate (Gly993-Thr995 and Gly916-Gly918) were displayed transparently, to show their residues. Residues, which are participating in substrate recognition (Arg369, Arg414 and Arg645) at the entrance part of, and transfer (Arg368 and Arg640) within, the β7 domain, are drawn by thin ball-and-stick models as well as residues involved in product release (Arg131, Arg132, Ser157, Ser158, His277, His182 and Arg85) through the β6 domain. The substrate modeled on the basis of two inhibitor complexes is bound to the O^{γ} of Ser965 and was represented as a tetrahedral adduct. The unprimed residues were derived from the inh2 complex structure and the primed residues from the inh3, respectively. (b) A schematic diagram of the tricorn protease during the proteolysis step. Two cylinders, color-coded as in (a), represent the 7- and 6-bladed β-propeller domains, respectively. Substrate to the β7 domain and products from the β6 domain

terminal blades 1 and 7 of the open Velcro propeller structures.[10] The residual activity of the modified tricorn protease may be in part due to an alternative path through the central pore along the molecular 3-fold axis, which is much longer and tortuous as compared to the direct pathway through the β7 domain. The notion that the pore of the β7 domain is the main substrate route to the active site is strongly supported by the structure in complex with the inh2 as it directly showed the peptide in the lower part of the propeller pore (Figure 2(b)).

The chloromethyl ketone-based inhibitors and the inh3 are fixed by formation of a local anti-parallel β-sheet with the two strands (Gly993-Thr995 and Gly916-Gly918) at the unprimed and the primed site, respectively (Figure 5). The main interactions are accomplished between backbone atoms. Three hydrogen bonds are observed at the unprimed site, one at the primed site and two at the oxyanion hole. Two more hydrogen bonds are provided by the side-chains of Arg131-Arg132 to the peptide carbonyl oxygen of substrate to fix the last peptide group. Arg132 reorients away from Ser157-Ser158 with which it forms hydrogen bonds in the inhibitor-free and the chloromethyl ketone-based inhibited enzyme (Figure 4). The substrate-enzyme interaction, which is dominated by main-chain rather than side-chain hydrogen bonds, is in accord with the broad substrate specificity of the tricorn protease.[11]

The backbone interaction of the P1 residue provided by Gly918 resembles the anti-parallel S1-P1 strand recognition in the tricorn associated F1 protease by Gly37, which is strictly conserved in F1 homologous proteins.[21] In addition, two negatively charged residues (Glu213 and Glu245) of F1 serve to recognize the N terminus of substrate, analogous to the recognition of the C terminus of substrate by Arg131-Arg132 in the tricorn protease.

While the complex crystal structures of chloromethyl ketone-based inhibitors confirmed the overall backbone interaction pattern at the unprimed site, the side-chain of Arg-P2 in the inh1 formed an additional hydrogen bond with the peptide oxygen of Tyr609, which was defined as the S1′ site in the complex structure of the inh3. The crystal structure of TPCK had shown that the geometry and charge of the S1 site is able to adapt *via* Asp936 to the P1 residue.[16] The inh3 structure indicates similar adaptation of the Arg131-Arg132 anchor to the primed substrate residues (Figure 4). The observed geometry and interaction explains how tri- and even tetrapeptide segments can be accommodated at the primed side. In fact, it appears even better suited for dipeptides which are expected to directly bind with their C termini to the Arg131-Arg132 anchor.

The catalytic nucleophile, Ser965 activated by His746, attacks the peptide carbonyl carbon of substrate. The exact cleavage positions are determined by the oxyanion hole of Gly918 and Asp966 and the S1 pocket (Figure 5). A tetrahedral adduct is formed and the peptide bond cleaved by subsequent acyl enzyme formation and hydrolysis.

Once substrates are cleaved at the active site, products must dissociate from the active site for the next round of proteolysis or for the recovery of free enzyme state. A direct and short pathway of the di- or tripeptide product to the surface is offered by the central pore of the β6 domain. The Arg131-Arg132 anchoring segment spans across its inner entry (Figure 5(a)). By slight rearrangements, a path is opened for the diffusion of the peptide product, which may be facilitated by His277, His182 and Arg85 lining the pore wall. The result of the mutation L184C is in accord with this model. Leu184 is located in the middle of the β6 domain (Figure 5(a)). The smaller cysteine side group may facilitate product diffusion leading to increased enzyme activities with protein substrates as well as fluorogenic substrates. Alkylation of the L184C mutant enzyme restricts the channel through the β6 domain, thus leading to reduced enzyme activities, slightly less than wild-type (Table 1). Dissociation of the di- or tripeptide product generates the room necessary for the unprimed product to move forward by two or three residues towards the Arg131-Arg132 anchor.

Substrate preferences are weak as already indicated by the S1 site adaptable for hydrophobic and basic P1 residues.[16] S1′ seems to prefer basic residues which are long enough to interact with the peptide oxygen of Tyr609 as shown with the inh3 complex structure. Alternatively, this basic residue may be provided by the P2 position of the substrate, as seen with the inh2 complex structure. The tricorn protease also favors a basic residue at the P4 site as shown by the crystal structure in complex with inh1. The observed structural features explain the processive cleavage of substrates from the carboxy- to the amino terminus and classify the tricorn protease as a di-(tri-) peptidyl carboxypeptidase.

The processive cleavage was indicated by an invariant product profile of insulin B-chain as

are displayed by continuous green dots in which a single green dot means single amino acid. The substrate bound to the active site is indicated by a green strand leaving its N-tail coil in the β7 channel. The red dot on the end of substrate strand stands for the substrate C terminus. The catalytic Ser965 is indicated as a red arrow at the cleavage site. Hydrogen bonds between the substrate C terminus and the Arg131-Arg132 are shown as dotted pink lines. Two β-strands, which sandwich the substrate, are displayed with the same color as C2 domain. The rest parts of a tricorn monomer are symbolized as C1, PDZ and C2. (c) Cut-open surface representation of a close-up view at the active site.

shown by the time-course HPLC elution pattern. Analysis of these degradation products further revealed the exopeptidase activity of the tricorn protease, with a clear preference for di- and tripeptidyl cleavage.[13] Similarly, the R131E-R132E anchor mutant shows a time-independent product profile, which, however, differs from the wild-type products (Figure 3(b)). This observation assigns a marked role of the primed side in the alignment of the substrate with the active site.

Several independent lines of arguments provide evidence for the unidirectionality of the substrate-to-product flow. A first simple, yet fundamental one is that the wider β7 domain is better suited to guide the larger substrate than the β6 domain, which reflects the size distribution of substrates (~10 amino acid residues) and products (2–3 amino acid residues). An opposite substrate flow where substrate enters the active site through the β6 domain is very unlikely based on these size considerations. Second, in the case of high substrate excess, the catalytic turnover is expected to be limited by the product exit rather than the substrate entry. In accord with expectation, the widening of the β6 channel increased the catalytic turnover rate of the tricorn protease over the wild-type (Table 1). Third, the filtering mechanism of the β7 and the β6 channel is likely to distinguish substrates upon the size and the conformation, whereas it is impossible to distinguish di- or tripeptidic products upon these criteria. We found indeed that the mutation in the β7 domain has different effects towards protein and fluorogenic substrates, contrasting the mutation in the β6 domain (Table 1). A fourth, direct argument to support the role of β7 and β6 domain in substrate access and product egress is provided by the crystal structure of the tricorn protease in complex with a trideca-peptide substrate analogue, which clearly defines the substrate binding to the active site, and suggests the β6 domain as an exit route for the cleaved di- or tripeptidic product (Figure 2(b)). Use of the alternative exit routes, the β7 domain or the central pore, would require displacement of the unprimed product, conflicting the processivity of the tricorn protease which implies that the unprimed substrate should be contained at the active site until it is completely degraded.

Materials and Methods

Site-directed mutagenesis, protein expression, and purification

The tricorn wild-type construct (pRSet6c-Tricorn with the *Nde* I and the *Xho* I restriction enzyme sites for insertion) was used as a template for all performed mutations by the Quickchange site-directed mutagenesis kit from Stratagene. Initial denaturation and denaturation for elongation was set for 5 and one minute at 95 °C, annealing for one minute at 55 and elongation for 14 minutes at 68 °C, respectively. The introduced mutations were confirmed by DNA sequencing. Individual constructs were transformed into the B834 (DE3) or BL21 (DE3) *Escherichia coli* strain and protein expression was induced with 1 mM isopropyl β-D-thiogalactoside at a cell density $OD_{600\ nm}$ 0.8 at 30 °C. Cells were harvested after five hours and stored at −20 °C. Tricorn protein was purified by heat treatment and sequential chromatographic steps of Q-Sepharose, Hydroxyapatite, and Superose 6 gel filtration as previously described.[16] Purified proteins were stored at 4 °C in a buffer containing 20 mM Tris·HCl, pH 7.5, 100 mM NaCl.

Synthesis of the inh2

The C-terminal H-Phe-CH_2Cl was obtained by reaction of Boc-Phe-OH *via* the mixed anhydride with trimethylsilyldiazomethane[22] followed by treatment of the resulting Boc-Phe-CHN_2 with 6 M HCl in dioxane. The side-chain protected N^{α}-Boc-dodecapeptide, which was synthesized on chlorotrityl-resin by standard Fmoc/*t* Bu procedures and cleaved from the resin by exposure to CH_2Cl_2/trifluoroethanol/AcOH (8:1:1), was coupled with H-Phe-CH_2Cl with EDC/HOAt/DIEA. Upon exposure of the fully protected intermediate to TFA/H_2O/TIPS (95:4:1) for two hours at room temperature, the crude compound of the inh2 was purified by preparative reverse-phase (RP) HPLC.

Synthesis of the inh3

2-(*R*,*S*)-Hydroxy-3(*S*)-amino-4-phenylbutanoic acid[23] was converted to the N^{α}-benzyloxycarbonyl (Z) derivative Z-PheΨ[CH(OH)–CO]OH and then coupled with H-Arg(Pbf)-Glu(O*t* Bu)-Phe-O*t* Bu with EDC/HOBt. The resulting protected tetrapeptide derivative was oxidized at the hydroxy group with Dess–Martin periodinane reagent to produce the corresponding α-ketoamide derivative. Upon exposure to 95% TFA for two hours at room temperature, the crude product was purified by preparative RP-HPLC.

Crystallization, inhibitor soaking, data collection, and refinement

For crystallizations, the purified wild-type protein was concentrated to 5 mg/ml in a buffer containing 20 mM Tris·HCl, pH 7.5, 100 mM NaCl and 2 mM β-mercaptoethanol. Crystals were obtained by hanging-drop vapor-diffusion at room temperature by mixing 1.5 μl of 5 mg/ml protein solution and 1.0 μl of reservoir solution, consisting of 12% isopropanol and 100 mM 2-morpholinoethanesulphonic acid, pH 6.4. Crystals suitable for X-ray diffraction grew within 2–14 days. The inhibitory efficacy of each inhibitor was checked by a fluorogenic activity assay before soaking. After that, inhibitors were added to the drop as a solution or a powder, and then equilibrated for 2 days at room temperature. 25% of 2-methyl-2,4-pentanediol (final concentration) was added as a cryoprotectant for cryocrystallography. Diffraction data were indexed, integrated and scaled using the DENZO and SCALEPACK package.[24] Refinements were done with the CNS package[25] using the wild-type tricorn protease structure as a model (Table 2).

Peptidolytic and inhibitory assays

Peptidolytic activities of the tricorn protease were checked by mixing enzyme with fluorogenic benzoyl

Table 2. Data collection and refinement statistics

	Inh1	Inh2	Inh3
Radiation source	DESY	SLS	DESY
Wavelength (Å)	1.05	0.9794	1.05
Space group	$P2_1$	$P2_1$	$P2_1$
Resolution (Å)	20–2.8	30–2.6	20–2.7
Cell parameters			
a (Å)	95.44	95.47	95.67
b (Å)	245.43	245.10	244.55
c (Å)	159.40	157.89	158.32
β (deg.)	104.79	105.19	104.59
Unique/total reflections	172151/372512	209855/645775	178808/295654
Redundancy	2.16	3.07	1.65
Completeness (%)	88.9 (87.0[a])	97.7 (85.1[a])	93.2 (83.6[a])
R_{sym}[b] (%)	9.0 (39.3[a])	8.4 (30.5[a])	6.6 (26.3[a])
R_{factor}[c] (%)	28.4	25.4	24.7
R_{free}[d] (%)	31.5	28.8	28.0
rmsd bond lengths (Å)	0.009	0.014	0.010
rmsd bond angles (deg.)	1.25	1.49	1.27
rmsd NCS protein (Å)	0.034	0.026	0.012

SLS, swiss light source; DESY, Deutsche elektronen synchrotron; rmsd, root-mean-square deviation.

[a] Values for the reflections in the highest-resolution shell (2.84–2.80 Å for the inh1 data set; 2.64–2.60 Å for the inh2 data set; 2.75–2.70 Å for the inh3 data set).

[b] $R_{sym} = \sum_{hkl} \sum_j |I_j - \langle I \rangle| / \sum_{hkl} \sum_j I_j$; where $\langle I \rangle$ is the mean intensity of reflection *hkl*.

[c] $R_{factor} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$; where F_{obs} and F_{calc} are, respectively, the observed and calculated structure factor amplitude for reflections *hkl* included in the refinement.

[d] R_{free} is the same as R_{factor} but calculated over a randomly selected fraction of reflection data not included in the refinement (8.8% for the inh1 data set; 8.6% for the inh2 data set; 8.5% for the inh3 data set).

(Bz)-Val-Gly-Arg-7-amino-4-methylcoumarin (AMC) substrates in a 100 µl reaction volume. For inhibitory assays, 1 mM (final concentration) of each inhibitor was premixed with the enzyme solution and preincubated for one hour at 4 °C. After that, 10 mM (final concentration) of the above mentioned fluorogenic substrate was added, and the reaction mix was incubated for 30 minutes or 60 minutes at 65 °C. After adding of 900 µl reaction buffer, fluoroscences were measured at a wavelength of 460 nm by exciting UV at 360 nm.

Proteolytic activity assays and RP HPLC analysis of degradation products of casein, insulin B-chain, secretin and secretinamide

Oxidized insulin B-chain, FITC-labeled casein, secretin, or secretinamide were incubated with each purified wild-type and mutant protein sample at 65 °C in a reaction volume of 400 µl in a buffer composed of 20 mM Tris·HCl, pH 7.5 and 100 mM NaCl. For control reactions, a 100 µl aliquot was taken immediately after substrates and the protease were mixed. At different times, further 100 µl aliquots were taken and mixed with the H-Ala-Ala-Pro-Val-chloromethyl ketone inhibitor (inh5) to stop the further cleavage reaction. 80 µl of reaction mixture was analyzed by RP HPLC on a X-Terra-MS C8 RP column. The column was equilibrated with 5% of acetonitrile (AcCN) and 95% of 2% H_3PO_4 and eluted with a linear gradient of 5–90% of AcCN and 95–10% of 2% H_3PO_4 in 20 minutes at a flow rate of 1.5 ml/min. Degradation products of oxidized insulin B-chain, secretin and secretinamide were detected by UV at 210 nm and of FITC-labeled casein at 440 nm, respectively.

Alkylation and oxidation

Excess amounts (about ten times of the protein molar concentration) of NEM or FMI were added to the wild-type and the cysteine mutants, and modification reactions were carried out overnight at 4 °C. Unreacted maleimide derivatives were removed by overnight dialysis. For oxidation, 1 mM (final concentration) oxidized glutathione was added to the wild-type and double cysteine mutants, R414C, A643C, and incubated overnight at 4 °C. Residual glutathione was removed by extensive dialysis.

Protein Data Bank accession numbers

The final coordinates of the structure have been deposited in a Data Bank with the PDB codes (1N6D, 1N6E and 1N6F).

Acknowledgements

We thank Gleb Bourenkov, Hans Bartunik of BW6 at DESY (Hamburg, Germany), and Takashi Tomizaki, Clemens Schulze-Briese of the SLS beamline X06SA at PSI (Villigen, Switzerland) for helping with synchrotron data collection. H.B. is supported by Roche Diagnostics, Penzberg; M.G. by SFB594; L.M. by SFB533; R.H. by EU grants QLK3-CT-2001-01976, HPRN-CT-1999-00084 and SFB496.

References

1. Rock, K. L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D. & Goldberg, A. L. (1994). Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell*, **78**, 761–771.
2. Sommer, T. & Jentsch, S. (1994). A protein translocation defect linked to ubiquitin conjugation at the endoplasmic reticulum. *Nature*, **365**, 176–179.
3. Sommer, T. & Wolf, D. H. (1997). Endoplasmic reticulum degradation: reverse protein flow of no return. *FASEB J.* **11**, 1227–1233.
4. Hiller, M. M., Finger, A., Schweiger, M. & Wolf, D. H. (1996). ER degradation of a misfolded luminal protein by the cytosolic ubiquitin-proteasome pathway. *Science*, **273**, 1725–1728.
5. Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. & Huber, R. (1995). Crystal structure of the 20 S proteasome from archeon *T. acidophilum* at 3.4 Å resolution. *Science*, **268**, 533–539.
6. Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D. & Huber, R. (1997). Structure of 20 S proteasome from yeast at 2.4 Å resolution. *Nature*, **386**, 463–471.
7. Bochtler, M., Hartmann, C., Song, H. K., Bourenkov, G., Bartunik, H. D. & Huber, R. (2000). The structure of HslU and the ATP-dependent protease HslU-HslV. *Nature*, **403**, 800–805.
8. Groll, M., Bajorek, M., Kohler, A., Moroder, L., Rubin, D. M., Huber, R. *et al.* (2000). A gated channel into the proteasome core. *Nature Struct. Biol.* **7**, 1062–1067.
9. Fülöp, V., Böcskei, Z. & Polgár, L. (1998). Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. *Cell*, **94**, 161–170.
10. Fülöp, V., Szeltner, Z. & Polgár, L. (2000). Catalysis of serine oligopeptidase is controlled by a gating filter mechanism. *EMBO Rep.* **1**, 277–281.
11. Tamura, T., Tamura, N., Cejka, Z., Hegerl, R., Lottspeich, F. & Baumeister, W. (1996). Tricorn protease—the core of a modular proteolytic system. *Science*, **274**, 1385–1389.
12. Tamura, T., Tamura, N., Lottspeich, F. & Baumeister, W. (1996). Tricorn protease (TRI) interacting factor 1 from *Thermoplasma acidophilum* is a proline iminopeptidase. *FEBS Letters*, **398**, 101–105.
13. Tamura, N., Lottspeich, F., Baumeister, W. & Tamura, T. (1998). The role of tricorn protease and its aminopeptidase-interacting factors in cellular protein degradation. *Cell*, **95**, 637–648.
14. Walz, J., Tamura, T., Tamura, N., Grimm, R., Baumeister, W. & Koster, A. J. (1997). Tricorn protease exists as an icosahedral supermolecule *in vivo*. *Mol. Cell.* **1**, 59–65.
15. Walz, J., Koster, A. J., Tamura, T. & Baumeister, W. (1999). Capsids of tricorn protease studied by electron cryomicroscopy. *J. Struct. Biol.* **128**, 65–68.
16. Brandstetter, H., Kim, J.-S., Groll, M. & Huber, R. (2001). Crystal structure of the tricorn protease reveals a protein disassembly line. *Nature*, **414**, 466–470.
17. Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R. & Hofsteenge, J. (1989). The refined 1.9 Å crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethyl ketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **8**, 3467–3475.
18. Neer, J. N. & Smith, T. F. (1996). G-protein heterodimers: new structures propel new questions. *Cell*, **84**, 175–178.
19. Baker, S. C., Saunders, N. F. W., Willis, A. C., Ferguson, S. J., Hajdu, J. & Fülöp, V. (1997). Cytochrome cd_1 structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to β-propeller folds. *J. Mol. Biol.* **269**, 440–455.
20. Faber, H. R., Groom, C. R., Baker, H. M., Morgan, W. T., Smith, A. & Baker, E. N. (1995). 1.8 Å structure of the C-terminal domain of rabbit serum haemopexin. *Structure*, **3**, 551–559.
21. Goettig, P., Groll, M., Kim, J.-S., Huber, R. & Brandstetter, H. (2002). Crystal structure of tricorn-interacting aminopeptidase F1 with different ligands explain its catalytic mechanism. *EMBO J.* **21**, 5343–5352.
22. Cesar, J. & Sollner-Dolenc, M. (2001). Trimethylsilyldiazomethane in the preparation of diazoketones *via* mixed anhydride and coupling reagent methods: a new approach to the Arndt–Eistert synthesis. *Tetrahedron Letters*, **42**, 7099–7102.
23. Harbeson, S. L., Abelleira, S. M., Akiyama, A., Barrett, R., Carroll, R. M., Straub, J. A. *et al.* (1994). Stereospecific synthesis of peptidyl alpha-keto amides as inhibitors of calphain. *J. Med. Chem.* **37**, 2918–2929.
24. Otwinowski, Z. & Minor, W. (1997). Procession of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
25. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
26. Kraulis, P. J. (1991). A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
27. Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP—graphical representation and analysis of surface properties. *Biophys. J.* **64**, A166.
28. Esnouf, R. M. (1997). An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132–134.
29. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **227**, 505–524.

Edited by A. R. Fersht

(Received 15 July 2002; received in revised form 11 October 2002; accepted 16 October 2002)