Calcium–mediated thermostability in the subtilisin superfamily: The crystal structure of Bacillus Ak.1 protease at 1.8 Å resolution

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · JANUARY 2000

Impact Factor: 4.33 · DOI: 10.1006/jmbi.1999.3291 · Source: PubMed

5 AUTHORS, INCLUDING:

Clyde A Smith
Stanford University
92 PUBLICATIONS **3,214** CITATIONS
SEE PROFILE

Helen S Toogood
The University of Manchester
55 PUBLICATIONS **1,107** CITATIONS
SEE PROFILE

Edward N Baker
University of Auckland
154 PUBLICATIONS **8,226** CITATIONS
SEE PROFILE

Article No. jmbi.1999.3291 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 294, 1027–1040

JMB

Calcium-mediated Thermostability in the Subtilisin Superfamily: The Crystal Structure of *Bacillus* Ak.1 Protease at 1.8 Å Resolution

Clyde A. Smith[1]*, Helen S. Toogood[2], Heather M. Baker[1] Roy M. Daniel[2] and Edward N. Baker[1]

[1]*School of Biological Sciences University of Auckland Auckland, New Zealand*

[2]*Thermophile Research Unit University of Waikato Hamilton, New Zealand*

**Corresponding author*

Proteins of the subtilisin superfamily (subtilases) are widely distributed through many living species, where they perform a variety of processing functions. They are also used extensively in industry. In many of these enzymes, bound calcium ions play a key role in protecting against autolysis and thermal denaturation. We have determined the crystal structure of a highly thermostable protease from *Bacillus* sp. Ak.1 that is strongly stabilized by calcium. The crystal structure, determined at 1.8 Å resolution ($R = 0.182$, $R_{free} = 0.247$), reveals the presence of four bound cations, three Ca^{2+} and one Na^{+}. Two of the Ca^{2+} binding sites, Ca-1 and Ca-2, correspond to sites also found in thermitase and the mesophilic subtilisins. The third calcium ion, however, is at a novel site that is created by two key amino acid substitutions near Ca-1, and has not been observed in any other subtilase. This site, acting cooperatively with Ca-1, appears to give substantially enhanced thermostability, compared with thermitase. Comparisons with the mesophilic subtilisins also point to the importance of aromatic clusters, reduced hydrophobic surface and constrained N and C termini in enhancing the thermostability of thermitase and Ak.1 protease. The Ak.1 protease also contains an unusual Cys-X-Cys disulfide bridge that modifies the active site cleft geometry.

Keywords: calcium binding; thermostability; crystal structure; serine protease; subtilase family

Introduction

How enzymes maintain their structural integrity and retain functionality at elevated temperatures remains an intriguing question of structural biology. Enzymatic function requires some flexibility, but some rigidity is also essential for ligand binding sites to remain intact, or for the residues responsible for catalysis to retain the same relative orientations. A delicate balance is thus required; a high level of rigidity implies increased stability, albeit at the expense of enzyme activity, while flexibility is essential for function, with excessively flexible proteins being susceptible to proteolysis or denaturation.

Enzymes from thermophilic organisms employ a range of strategies to maintain their structural integrity at the elevated temperatures at which they are required to operate. Some of these are chemical, by minimizing their content of amino acids that are prone to such phenomena as oxidation (Cys, Met) or deamidation (Asn, Gln). In other cases enhanced stability is achieved by a variety of structural means, including increased numbers of salt bridges, hydrophobic or hydrogen bonding interactions, enhanced subunit associations, shortened N and C termini or surface loops, extended secondary structure, metal binding, and various residue substitutions that enhance rigidity. These factors have different levels of prominence in different proteins (Knapp *et al.*, 1997; Vogt *et al.*, 1997; Wallon *et al.*, 1997).

Enzymes of the subtilisin superfamily, also referred to as the subtilases (Siezen *et al.*, 1991; Siezen & Leunissen, 1997), are an extremely widely distributed family of serine proteases, found in

Abbreviations used: PMSF, phenylmethyl sulfonyl fluoride; SAAPFpNA, succinyl-Ala-Ala-Pro-Phe-pNA; rms, root-mean-square; AkP, *Bacillus* sp. Ak.1 protease.

E-mail address of the corresponding author: ca.smith@auckland.ac.nz

0022-2836/99/491027–14 $30.00/0

great numbers among micro-organisms, as well as in plants, insects, nematodes, fish, mammals and many other species. Some mediate processes of exquisite specificity, for example, the kex family of pro-protein convertases that activate peptide hormones, growth factors, and so on (Barr, 1991). Others, including the archetypal subtilisin BPN′ from *Bacillus amyloliquefaciens* and subtilisin Carlsberg from *Bacillus licheniformis*, have much broader specificities, and are important industrial enzymes. These enzymes have been the subject of many protein engineering studies aimed at increasing their thermostability for industrial use (Wells & Estell, 1988).

The crystal structures of a number of subtilases have been determined, including subtilisin BPN′ (Bott *et al.*, 1988; Drenth *et al.*, 1972), subtilisin Carlsberg (Bode *et al.*, 1987; McPhalen *et al.*, 1985; Neidhart & Petsko, 1988), proteinase K (Betzel *et al.*, 1988), savinase (Betzel *et al.*, 1992), M-protease (Shirai *et al.*, 1997), mesentericopeptidase (Dauter *et al.*, 1991) and thermitase (Gros *et al.*, 1989, 1991; Teplyakov *et al.*, 1990). A common feature of these enzymes is that bound calcium ions play an important structural role; in the subtilisins, for example, there are two calcium sites (Pantoliano *et al.*, 1989), and in thermitase there are three (Frömmel & Höhne, 1981). These calcium ions serve to tie together surface loops and have been shown to give enhanced thermostability and resistance to autolysis.

Bacillus sp. Ak.1, isolated as a contaminant on solid medium, produces an extracellular, highly thermostable serine protease. The gene for this protease has been cloned, sequenced and expressed in *Escherichia coli* (MacIver *et al.*, 1994; Peek *et al.*, 1993). Analysis of the sequence in relation to other subtilases shows that it falls into a sub-group of enzymes found only in micro-organisms, and including a number of extremophiles (thermophiles and halophiles) (Siezen & Leunissen, 1997). This group is typified by thermitase, with which the Ak.1 protease (AkP) has 61% sequence identity (Figure 1). Despite this relatively high sequence identity, AkP does have some quite striking differences from thermitase. In the absence of Ca^{2+}, AkP is slightly less stable than thermitase (half-life at 80 °C of less than one minute, compared with nine minutes for thermitase), but in the presence of Ca^{2+}, its stability is dramatically enhanced (half-life at 80 °C of 15 hours, compared with 19 minutes for thermitase) (unpublished results). The Ak.1 protease also lacks the free Cys residue at position 75 in thermitase (replaced by a Val residue in AkP), but instead has two Cys residues at positions 137 and 139. It has quite a limited substrate specificity, preferentially cleaving substrates with amino acids such as phenylalanine, valine or alanine at the P1 site, and proline at the P2 site (Toogood *et al.*, 1999). Maximal activity was observed for the substrate succinyl-Ala-Ala-Pro-Phe-pNA (SAAPFp-NA), with somewhat lower activity for analogues with Phe replaced by Leu (~33%) or Ala (~ 8%).

Here, we present the three-dimensional structure of the Ak.1 protease, determined at high resolution by X-ray crystallography. The structure reveals a novel, additional calcium binding site, not shared by any of the other enzymes of this family. This site arises as a result of amino acid substitutions in an external loop and with the other metal binding sites confers a markedly increased thermostability. We also note an unusual Cys-X-Cys disulfide bond that impacts on the active site of the enzyme.

Results

The crystal structure of *Bacillus* sp. Ak.1 protease (AkP) was determined by molecular replacement, using search models derived from the 1.4 Å model for thermitase (Teplyakov *et al.*, 1990). The structure was refined at 1.8 Å resolution, to a crystallographic *R*-factor of 0.182 ($R_{free} = 0.247$), with good geometry (Table 1). The final model consists of 271 out of 280 residues (2029 protein atoms), with one loop missing (residues 166-174), 225 water molecules, three calcium ions and one sodium ion. A Ramachandran plot (Ramakrishnan & Ramachandran, 1965) indicates that virtually all non-glycine and non-proline residues lie within the allowed regions of conformational space; 83% within the most favoured regions, as defined by PROCHECK (Laskowski *et al.*, 1993), and 99.7% within the allowed regions. Representative sections of the final $2F_o - F_c$ electron density map are shown in Figure 2.

Overall structure

The overall molecular structure of the *Bacillus* sp. Ak.1 protease is shown in Figure 3. The polypeptide fold is essentially the same as in the subtilisins (McPhalen & James, 1988; Neidhart & Petsko, 1988), thermitase (Teplyakov *et al.*, 1990), savinase (Betzel *et al.*, 1992) and proteinase K (Betzel *et al.*, 1988). The molecule comprises 11 β-strands and eight α-helices arranged in a single domain, with a topology shown schematically in Figure 4. The

Table 1. Refinement statistics

Resolution limits (Å)	20.0-1.8
R-factor [a]	0.182
R_{free}	0.247
Number of reflections used	19,473
Reflections used for R_{free}	1028
Number of protein atoms	2029
Number of solvent molecules	225
Other species in final model	3 Ca^{2+}, 1 Na^{+}
Average *B* value (Å^2)	24.4
rms deviations	
Bond lengths (Å)	0.012
Bonds angles (deg.)	1.49
Ramachandran plot	
Most favored regions (%)	83
Number of outliers	2

[a] $R = \Sigma|F_o - F_c|/\Sigma F_o$ where F_o and F_c are the observed and calculated structure factors, respectively.

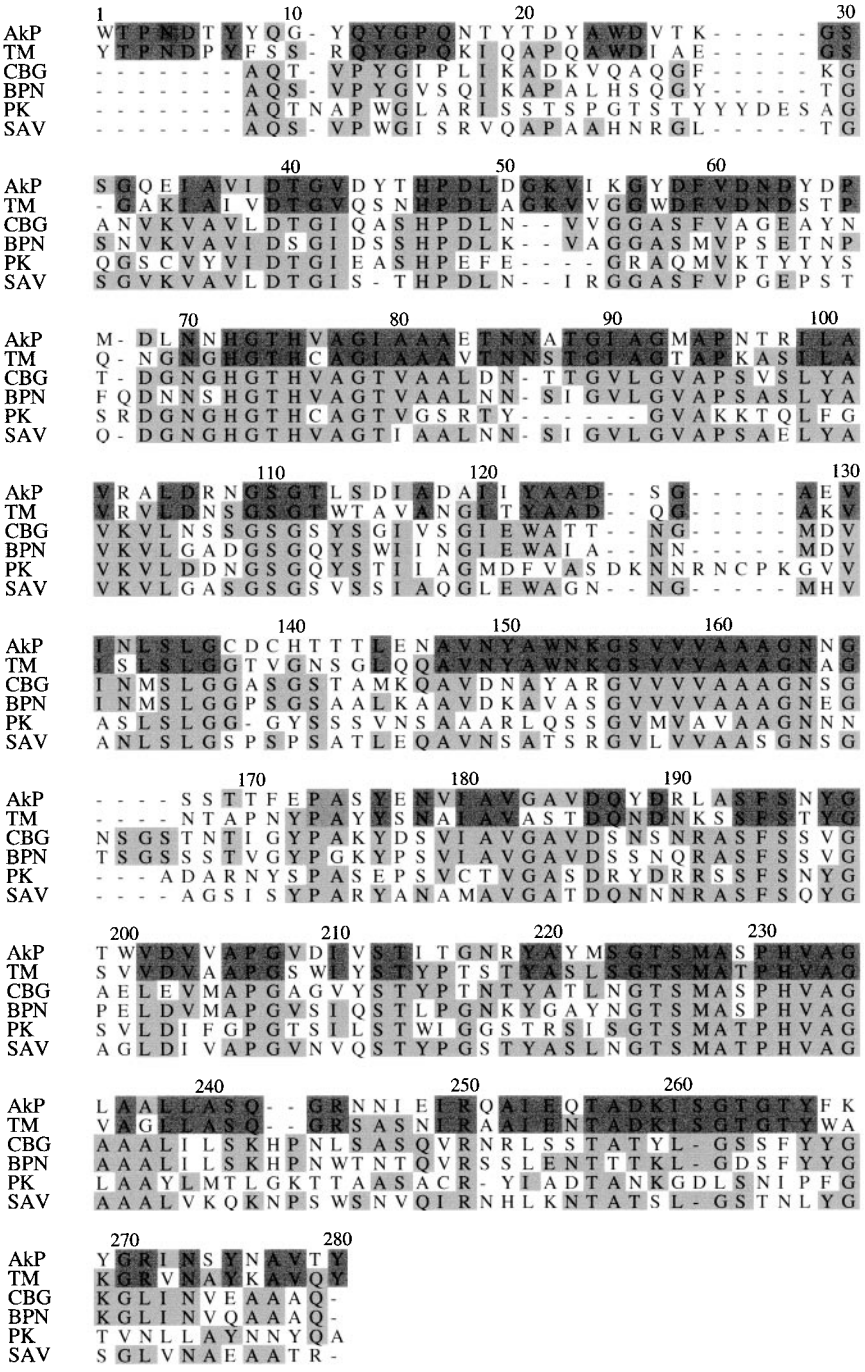

Figure 1. Structure-based alignment of the amino acid sequences of Ak.1 protease (AkP), thermitase (TM), subtilisin Carlsberg (CBG), subtilisin BPN′ (BPN), proteinase K (PK) and savinase (SAV). Grey shading indicates residues that are the same in at least three of the six proteins. Residues that are identical between AkP and TM are highlighted with darker shading.

core of the molecule is composed of an eight-stranded parallel β-sheet (β1-β7 and β11) flanked by five α-helices (α3-α7). There are also three additional strands: β10 which runs antiparallel to the core β-sheet, and a two-stranded antiparallel β-hairpin (strands β8 and β9). As expected from sequence comparisons (Siezen & Leunissen, 1997), the polypeptide conformation of AkP is most similar to that of thermitase. When the two structures are superimposed, 257 C^{α} atoms can be matched with a root-mean-square (rms) difference of 0.55 Å. The only places at which the two structures deviate are around the single insertion (Ser31 in AkP); at residues 54-55 (AkP numbering) close to the new Ca^{2+} binding site; and at the active site rim where the new disulfide bridge 137-139 is associated with bodily displacement of residues 134-142.

A feature of the thermitase subfamily, to which AkP belongs, is that the N and C termini are very tightly associated with the main body of the structure; this may be a significant contributor to their thermostability (see below). The N-terminal five residues fold back into a shallow crevice between the high-affinity calcium binding loop (residues 83-89, see below) and the side of helix α2, such that the N-terminal residue, Trp1, butts up against the

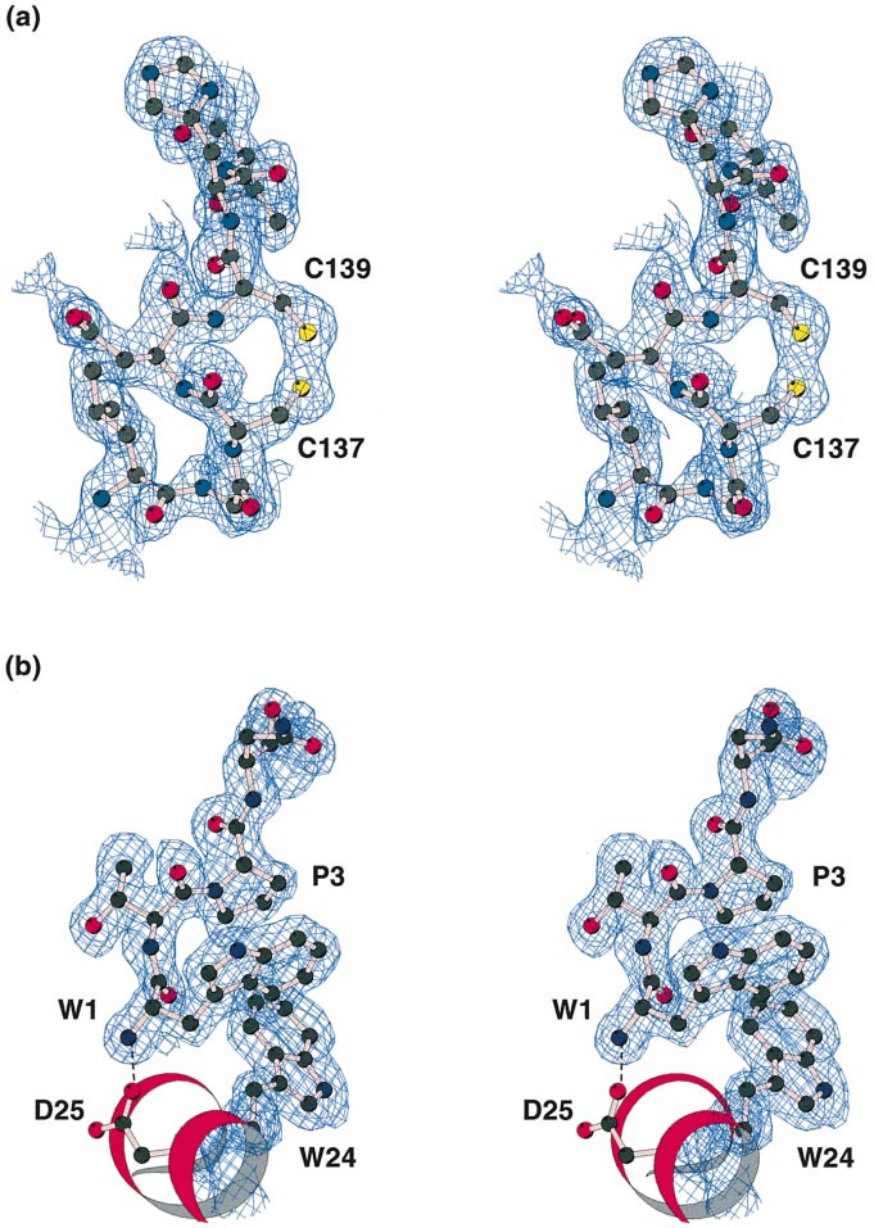

Figure 2. Portions of the AkP structure shown in the final $2F_o - F_c$ electron density map. (a) Residues 135-141, showing the disulfide bond formed between Cys137 and Cys139. (b) The environment of the N-terminal residue, Trp1, also showing its packing against helix α2.

side of α2. The α-amino group forms a salt bridge with Asp25 and hydrogen bonds to the carbonyl oxygen atom of Asp21, and the side-chain of Trp1 is oriented such that its indole ring is packed against the plane of the pyrrolidine ring of Pro3 and makes an approximate edge-to-face interaction with the side-chain of Trp24 (see Figure 2). Such interactions are very favourable for aromatic residues (Brocchieri & Karlin, 1994; Burley & Petsko, 1985; Hunter *et al*., 1991), and the overall result is that Trp1 is extremely well defined (Figure 2). The C terminus is constrained in a similar way. The α-carboxyl group forms a double salt bridge with Arg244 and the side-chain of the C-terminal residue, Tyr280, is hydrogen bonded to Gln251.

A surprising feature of the AkP structure was the absence of the loop comprising residues 166-174. In thermitase this loop is well ordered, running alongside the active site rim residues 135-140. It was for this reason that interpretation of the density for the 137-139 disulfide was approached with caution initially. Inspection of the AkP sequence, however, suggested that the missing residues represented the site of an autolytic cleavage. The enzyme has a pronounced specificity for cleavage following a Pro-X sequence, where X is Phe, Leu or Ala (unpublished results). Given that the sequence of residues 173-175 is Pro-Ala-Ser there is a potential cleavage site between residues Ala174 and Ser175. In confirmation, the N-terminal sequence

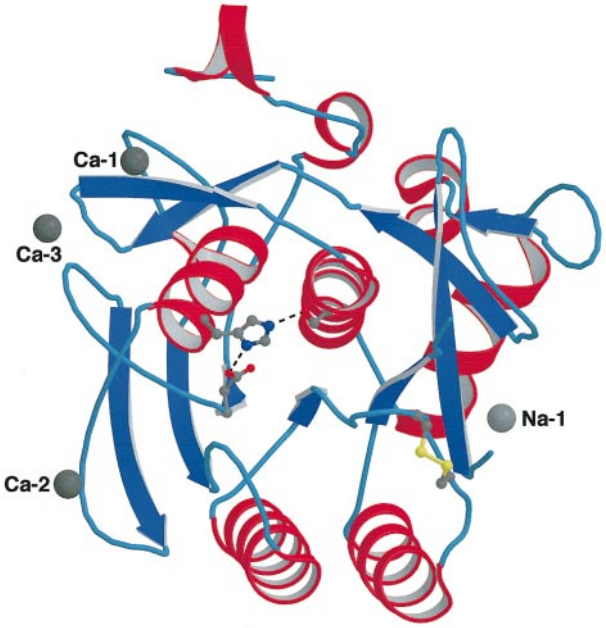

Figure 3. Ribbon representation of the overall molecular structure of Ak.1 protease, with the four bound cations (three Ca^{2+}, one Na^{+}) shown as spheres and the side-chains of the catalytic triad (Ser226. . .His72. . .Asp39) and disulfide bond 137-139 shown in ball-and-stick representation. This Figure and Figures 4 to 7 were produced with the programs MOLSCRIPT (Kraulis, 1991) and RASTER3D (Merritt & Murphy, 1994).

analysis of the dissolved crystals gave two clear N-terminal sequences, one corresponding to the true N terminus (WTPNDT), and the second with the sequence SYENVI, corresponding to residues 175-180 (Figure 1). We conclude that residues 166-174 are disordered as a result of cleavage between residues 174 and 175.

Disulfide bond 137-139

The disulfide bridge joining Cys137 to Cys139 completes a rather unusual 11-membered ring (Figure 2). A search of the Protein Data Bank (Bernstein *et al.*, 1977) shows that disulfide bonds involving Cys-X-Cys sequences are rare, though not unprecedented. Similar disulfide bonds are found in the L40C mutant of NADH peroxidase (Miller *et al.*, 1995) and the Mengo virus coat protein (Krishnaswamy & Rossmann, 1990), but in other cases such Cys-X-Cys pairs do not form disulfide bonds. The disulfide geometry in AkP is generally regular, with just one unusual feature in that the C-S-S-C dihedral angle, X_3, is 124°, significantly higher than the expected value of 90° (Thornton, 1981). The X_3 angles for the Cys-X-Cys disulfide bonds in NADH peroxidase L40C and Mengo virus coat protein are 122° and 120°, respectively, indicating that such disulfide bonds cannot form without some expansion of this angle. This deviation from ''normal'' geometry may perhaps indicate some strain.

The 137-139 disulfide bond has a significant impact on the active site cleft, since residues 133-136 form a promontory that projects between the S1 and S4 sites (Figure 5), such that the displacement of the chain that results from disulfide bond formation affects both sites.

The active site

As in all the other proteases of the subtilisin superfamily, Ak.1 protease contains a catalytic triad located at the C-terminal end of the core β-sheet, comprising Asp39, His72 and Ser226. The overall geometry of the catalytic triad is similar to that observed in substrate-free forms of subtilisin BPN′ (PDB code 2ST1) (Bott *et al.*, 1988) and thermitase (1THM) (Teplyakov *et al.*, 1990), with His72 hydrogen bonded to the side-chains of Asp39

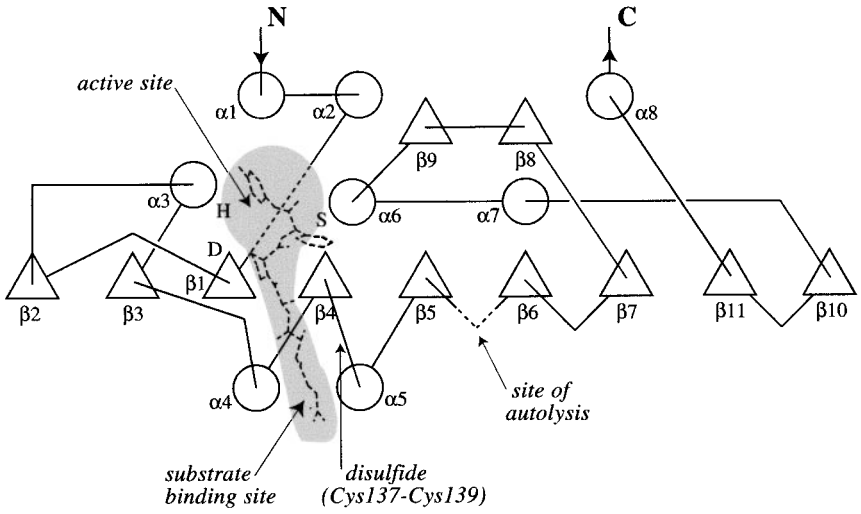

Figure 4. Schematic representation of the topology of Ak.1 protease. Helices are identified by circles and β-strands by triangles. The location of the active site is indicated with grey shading, with the location of bound substrate shown by a molecule with broken lines.

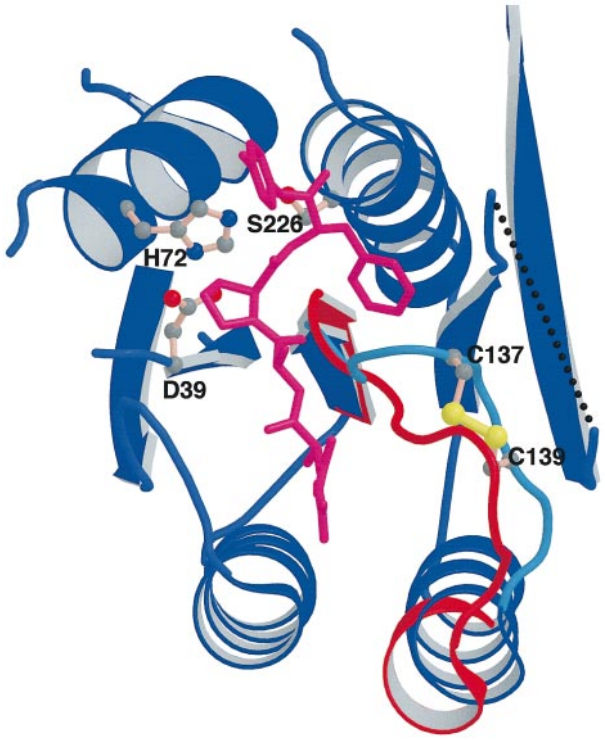

Figure 5. Substrate binding in the active site cleft of Ak.1 protease. The AkP structure is shown in blue, with the substrate analogue succinyl-Ala-Ala-Pro-Phe-pNA (magenta) modeled into the active site, based on the structure of the thermitase-eglin c complex (Gros *et al.*, 1989; PDB code 2TEC). A dotted line indicates where residues 166-174 are missing from the AkP structure. The altered course of residues 133-141 in thermitase is shown in red; the equivalent residues 134-142 in AkP are displaced, although the disulfide bridge 137-139 (yellow) partially compensates.

(72N$^{\delta1}$-39O$^{\delta1}$ 2.59 Å) and Ser226 (72N$^{\varepsilon2}$-226O$^{\gamma}$ 3.09 Å). Although the Asp and His side-chains remain in the same positions, the side-chain of Ser226 is rotated such that it points towards the S1 specificity pocket rather than away from it as in the other structures.

The active site cleft, similar in overall shape to that seen in other subtilases (Siezen & Leunissen, 1997), runs along the surface of the molecule between two extended loops which join two strand/helix pairs; β3 and α4, and β4 and α5 (Figures 4 and 5). The major specificity pocket (S1) in AkP is formed by the side-chain of Thr225 and the main-chain atoms of residues 133-136 and 161-164, and is predominantly uncharged and hydrophobic. This pocket is changed somewhat, relative to the other subtilases, as a result of the 137-139 disulfide bond which kinks the polypeptide chain between 136 and 140, forcing a change in the main chain conformation of residues Ser134 and Leu135; the carbonyl oxygen atoms of both residues now point in towards the S1 pocket. Surface calculations for the S1 specificity pockets in AkP and thermitase indicate that the site is slightly more constricted in AkP, although a phenyl ring can still be accommodated.

Of the other specificity pockets, S2 is flanked by helix α3 and the catalytic His72 residue on one side and the C terminus of β3 (Gly109 and Ser110) on the other, and has the C terminus of strand β1 at the bottom. Although there is a preference for a proline residue at the S2 position (unpublished results), inspection of the substrate binding cleft in AkP does not suggest why this should be the case, unless it is for conformational reasons. The S4 pocket in the subtilases is delineated on one side of the cleft by the main-chain segment equivalent to 135-136 (AkP numbering), and on the other side by residues 111-113 and the side-chains of three hydrophobic residues, Ile107, Leu113 and Leu144 in AkP. The displacement of residues 134-142, as a result of the 137-139 disulfide bond, has the effect of slightly enlarging the S4 site, although the disulfide bond itself partially compensates (Figure 5).

Calcium binding sites

Bound calcium ions have been found in the crystal structures of all subtilases characterized so far, and predictions from sequence comparisons suggest that similar sites will be found in many of the enzymes of the entire subtilase superfamily (Siezen & Leunissen, 1997). There are distinct variations among the different subfamilies, however. The subtilisins, from family A (Siezen & Leunissen, 1997), generally have two Ca^{2+} sites, a strong site and a weak site (Pantoliano *et al.*, 1989). These are shared by thermitase (family B), which also has a third site of intermediate binding strength (Gros *et al.*, 1991). On the other hand, proteinase K, from family C, has two sites, but one of them is different from those in either the subtilisins or thermitase. The current model for Ak.1 protease contains four cation binding sites (Figure 3), three occupied by Ca^{2+} (Ca-1, Ca-2, Ca-3) and the fourth modeled as sodium (Na-1). Details of these sites are in Table 2.

The first site, Ca-1, corresponds to the high-affinity site in the subtilisins and the highest-affinity site in thermitase. The Ca^{2+} is coordinated to six protein residues, three carbonyl oxygen atoms from Glu83, Thr88 and Ile90, and the side-chains of Asn86 (O$^{\delta1}$), Asp48 (O$^{\delta1}$ and O$^{\delta2}$) and Asp5 (O$^{\delta2}$), giving a pentagonal bipyramidal coordination geometry (Figure 6(a)). Four of the ligands are in a surface loop (residues 83-90), with the other two ligands coming from the N terminus of the enzyme (Asp5) and a neighbouring surface loop (Asp48). Between them, these residues totally enclose the bound Ca^{2+}. The liganding side-chains are also held in place by a network of hydrogen bonds; Asp5 O$^{\delta1}$ is hydrogen bonded to the peptide nitrogen atom of residue 85, Asp48 O$^{\delta2}$ is hydrogen bonded to the peptide nitrogen atom of residue 90 and Asp48 O$^{\delta1}$ is hydrogen bonded to Asn86 N$^{\delta2}$. The tight enclosure of this site explains why Ca^{2+} cannot be removed from thermitase

Table 2. Cation binding sites

	Residue	Distance (Å)	B (Å^2)	Thermitase dist. (Å) (ligand)	Subtilisin BPN′ dist. (Å) (ligand)
Ca-1	Asp5 $O^{\delta2}$	2.25	15.29	2.37 (D5 $O^{\delta2}$)	2.38 (Q2 $O^{\varepsilon1}$)
	Asp48 $O^{\delta1}$	2.44	9.64	2.42 (D47 $O^{\delta1}$)	2.40 (D41 $O^{\delta1}$)
	Asp48 $O^{\delta2}$	2.48	9.98	2.56 (D47 $O^{\delta2}$)	2.61 (D41 $O^{\delta2}$)
	Glu83 O	2.12	12.37	2.35 (V82 O)	2.27 (L75 O)
	Asn86 $O^{\delta1}$	2.31	10.59	2.46 (N85 $O^{\delta1}$)	2.37 (N77 $O^{\delta1}$)
	Thr88 O	2.09	12.79	2.36 (G88 O)	2.35 (I79 O)
	Ile90 O	2.54	13.31	2.36 (I89 O)	2.30 (V81 O)
Ca-2	Asp58 $O^{\delta2}$	2.17	13.99	2.42 (D57 $O^{\delta2}$)	
	Asp63 $O^{\delta1}$	2.35	19.90	2.46 (D62 $O^{\delta1}$)	
	Asp63 $O^{\delta2}$	2.78	23.90	2.64 (D62 $O^{\delta2}$)	
	Asp65 O	2.34	12.04	2.35 (P64 O)	
	Wat301 O	2.33	14.14	2.37 (Q66 $O^{\varepsilon1}$)	
	Wat302 O	2.42	20.64	2.25 (W370 O)	
	Wat303 O	2.36	14.60	2.30 (W353 O)	
Ca-3	Pro47 O	2.16	16.20		
	Asp50 $O^{\delta1}$	2.23	14.52		
	Glu83 $O^{\varepsilon1}$	2.40	19.42		
	Glu83 $O^{\varepsilon2}$	2.58	9.63		
	Wat307 O	2.38	17.13		
	Wat308 O	2.30	30.01		
	Wat309 O	2.28	19.74		
Na-1	Tyr176 O	2.97	11.60	2.96 (Y175 O)	2.98 (Y171 O)
	Val179 O	2.61	11.70	2.80 (A178 O)	2.76 (V174 O)
	Asp202 $O^{\delta2}$	2.88	18.31	2.97 (D201 $O^{\delta2}$)	2.85 (D197 $O^{\delta2}$)
	Wat304 O	2.50	21.32	2.83 (A173 O)	2.82 (G169 O)
	Wat305 O	2.72	25.01	2.95 (Wat335 O)	3.05 (Glu195 O)
	Wat306 O	2.45	38.37	2.90 (Wat420 O)	3.00 (Wat313 O)
					3.07 (Wat368 O)

without destroying the structure. Structurally, this binding site is virtually identical in AkP, thermitase and the subtilisins; the Ca^{2+} position in AkP is displaced by only 0.4 Å from that in thermitase and 0.5 Å from that in subtilisin BPN′ (Figure 6(a)). There is one significant difference in the mesophilic subtilisins, however, in that the residue from the N terminus is glutamine, rather than aspartate as in AkP and thermitase, leading to weaker binding than in thermitase (Briedigkeit & Frömmel, 1989) and presumably also AkP. Proteinase K, on the other hand, lacks this binding site completely, as a result of a three-residue deletion in the central calcium-binding loop.

The second site, Ca-2, corresponds with the second site observed in thermitase, a site that has not been observed in the subtilisin structures. As in thermitase, this site is created by residues from an external loop, 58-68, that is rich in Asp residues. The Ca^{2+} is bound by Asp58 ($O^{\delta2}$), Asp63 ($O^{\delta1}$ and $O^{\delta2}$), the main-chain carbonyl oxygen atom of Asp65, and three water molecules (Figure 6(b)). The resultant geometry is a distorted pentagonal bipyramid. This site is also relatively enclosed and constrained by hydrogen bonding interactions; the carboxyl group of Asp58 is held by a salt bridge with Arg103, that of Asp63 is held by a hydrogen bond between $O^{\delta1}$ and the peptide nitrogen atom of Asp65, and the three water molecules are hydrogen bonded to the carboxyl groups of Asp61 and Asp65 and the carbonyl oxygen atom of residue 67. The site is qualitatively different in thermitase where the Ca^{2+} is six-coordinate with Gln66 $O^{\varepsilon1}$ occupying the position of one of the water molecules; in AkP Gln66 becomes Met67 whose side-chain is oriented away from the calcium binding site. The sites still superimpose closely, however, with the Ca^{2+} positions displaced by only 0.7 Å (Figure 6(b)). In subtilisins this site is absent because the loop has a different conformation and lacks the Asp residues, whereas in proteinase K it is absent as a result of deletion of three residues from the loop.

The third site, Ca-3 (Figure 7), is unique to Ak.1 protease, among the subtilases that have been structurally characterised to date. It is closely linked to the first site, Ca-1, taking ligands from the surface loops 45-50 and 83-90. The Ca^{2+} is coordinated to the carboxyl groups of Asp50 ($O^{\delta1}$) and Glu83 ($O^{\varepsilon1}$ and $O^{\varepsilon2}$), the carbonyl oxygen atom of residue 47 and three water molecules, in a pentagonal bipyramidal arrangement. Although sites Ca-1 and Ca-3 strictly do not have any ligands in common, the carbonyl oxygen atom of Glu83 is a ligand for Ca-1 and the side-chain of Glu83 is a ligand for Ca-3 (Figure 8). This third site is more exposed than either Ca-1 or Ca-2, as the three water ligands and the Asp50 side-chain are fully in contact with bulk solvent.

The fourth cation binding site (Na-1) corresponds with a weak binding site found in both subtilisin BPN′ and thermitase. It is highly solvent-exposed, with the cation bound by three protein ligands, Asp202 $O^{\delta2}$ and the carbonyl oxygen atoms of residues 176 and 179, together with three water molecules that are in contact only with bulk solvent. The geometry is distorted octahedral. In thermitase this site appears to be occupied by a

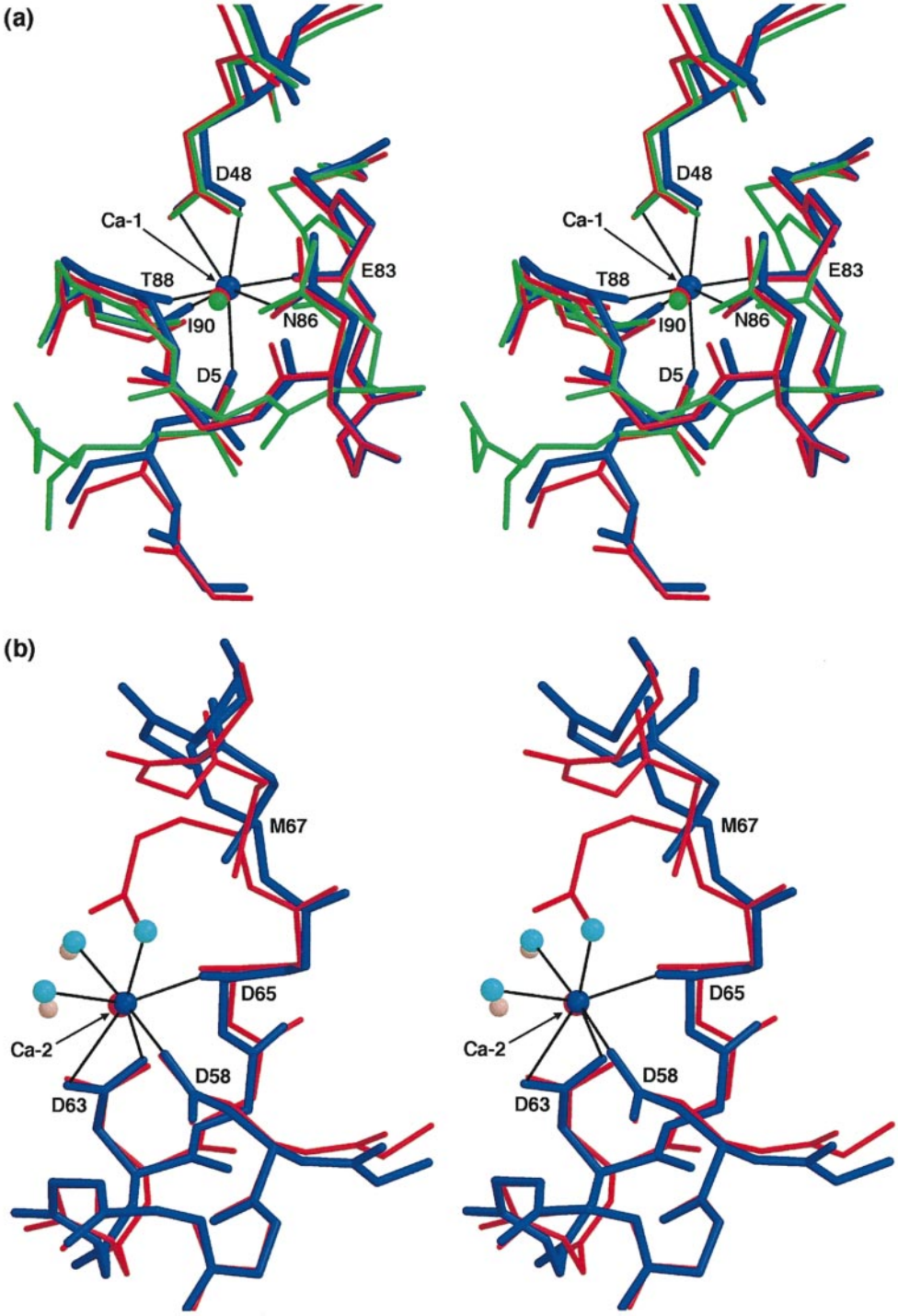

Figure 6. Stereo diagrams showing the Ca^{2+} binding sites Ca-1 and Ca-2. (a) Site Ca-1, shown with the equivalent sites for AkP, thermitase and subtilisin BPN′ superimposed; AkP is shown in blue, thermitase in red and subtilisin BPN′ in green. (b) Site Ca-2, shown with the equivalent sites for AkP (blue, with cyan spheres for water molecules) and thermitase (red, with pink spheres for water molecules) superimposed. In both Figures residues are labelled as for AkP.

sodium ion when crystals are grown without calcium (Teplyakov *et al.*, 1990), but by calcium for crystals grown in 100 mM Ca^{2+} (Gros *et al.*, 1989, 1991); it is clearly a highly exchangeable site. In AkP we have tentatively identified the cation as Na^{+}. Crystallographic refinement is equivocal; when refined as Ca^{2+} the *B*-factor is 23.6 $Å^2$, somewhat higher than the surrounding protein atoms (10-20 $Å^2$), but when refined as Na^{+} it is 9.9 $Å^2$. The bond distances, in the range 2.5-3.0 Å, favour Na^{+}, however. We cannot exclude that it could be a water molecule, although its coordination geome-

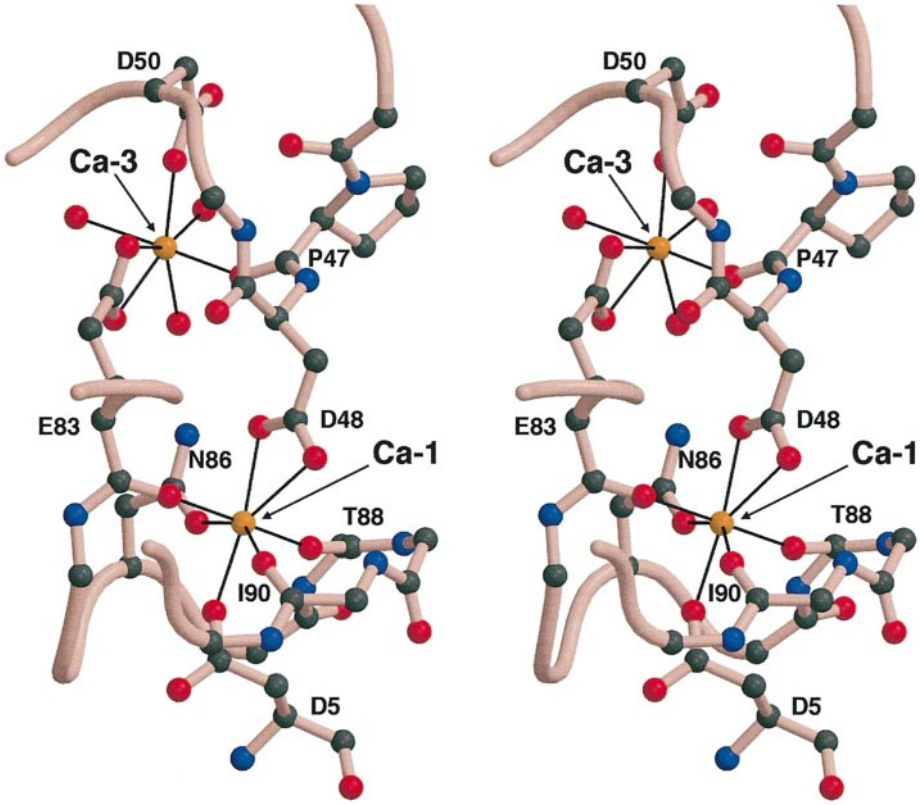

Figure 7. Stereo diagram showing the relationship between the generic, high-affinity, Ca^{2+} binding site Ca-1 (lower) and the novel Ca^{2+} binding site Ca-3 in AkP (upper). The Ca^{2+} are represented by orange balls, oxygen atoms by red balls, nitrogen blue and carbon black.

try and environment favour a cation. Consistent with the relative weakness of the binding site, the cation position is more variable in different enzymes; the Na^{+} position in AkP is displaced by 1.5 Å from that in thermitase (PDB code 1THM) and 2.0 Å from that in subtilisin BPN′ (PDB code 2ST1).

Determinants of thermostability

Other factors which may influence the thermostability of Ak.1 protease, relative to thermitase (Teplyakov *et al.*, 1990) (PDB code 1THM) and to the mesophilic subtilisins, represented by subtilisin BPN′ (Bott *et al.*, 1988) (PDB code 2ST1) and savinase (Betzel *et al.*, 1992) (PDB code 1SVN), have been analyzed, and are summarized in Table 3. Two significant differences are apparent between the thermophilic and mesophilic proteases. First, thermitase and AkP have significantly more aromatic residues (24 and 28, respectively) compared to the mesophilic enzymes. A large proportion of these (greater than 60 %) are near the surface of the molecule and are involved in the formation of aromatic clusters. This has previously been noted for thermitase (Teplyakov *et al.*, 1990), where there are 13 aromatic interactions involving 16 residues on the enzyme surface. Similar aromatic clusters are also a feature of AkP, with 14 of the 16 aromatic residues conserved, plus three additional aromatic residues, resulting in 13 interactions. The mesophilic enzymes not only have fewer aromatic residues (16 in subtilisin BPN′ and 12 in savinase), but are also almost entirely lacking in aromatic clusters. Second, the total exposed hydrophobic surface in the two thermophilic enzymes is considerably less (12-14 %) than in the mesophilic enzymes (22 %). This should also contribute to the increased stability of the thermophilic enzymes, since decreased hydrophobic surface is strongly correlated with increased thermostability (Fersht & Serrano, 1993).

Other factors that have been proposed to influence protein thermostability show no obvious trends in the enzymes compared here. Thermitase has more salt bridges than AkP (defined as oppositely charged residue pairs separated by less than 3.5 Å) and this may be a factor in its higher intrinsic thermostability in the absence of calcium. On the other hand, the number of salt bridges in AkP is no greater than in subtilisin BPN′ or savinase, suggesting that salt bridges are not an important determinant of thermostability in these enzymes. The thermophilic enzymes show slightly increased helix N-capping, but there is no overall difference in side-chain hydrogen bonding, either in long-range interactions (defined as between groups more than four residues apart along the polypeptide chain) or in local interactions. Apart from the

Table 3. Analysis of determinants of thermostability

	AkP	Thermitase	BPN′	Savinase
Bound cations	4	3	2	2
Arginine/lysine ratio	1.6	0.5	0.2	1.6
Asparagine + glutamine	29	33	28	31
Proline residues	8	12	14	12
Cysteine residues	(2)	1	0	0
Aromatic residues	28	24	16	12
Aromatic interactions	13	13	3	3
Salt bridges	8	12	9	9
Side-chain hydrogen bonds				
Long range	59	59	58	65
Local	38	37	40	34
Helix end caps	7	6	4	5
Accessible surface (Å^2)				
Total	9990	9885	10,060	9370
Hydrophobic	1480	1275	2230	2050
Exposed hydrophobic (%)	14.8	12.9	22.2	21.9

aromatic residues, compositional differences do not seem to show any obvious trend.

The novel disulfide bond in AkP is something of an enigma. It may be expected to stabilize AkP somewhat, relative to thermitase, which has a single free Cys residue exposed in its active site region. However, disulfide bonds are expected to make the greatest contribution to thermostability when they join residues far apart in the primary structure, whereas Cys137 and Cys139 could hardly be closer. Reduction of the disulfide bond in AkP does decrease its half-life tenfold (unpublished results), but this may reflect destabilization resulting from the presence of two free Cys residues in the reduced form rather than any intrinsic stability in the native protein resulting from the disulfide.

Discussion

Determination of the three-dimensional structure of Ak.1 protease provides a second structure for proteins of the ''extremophile'' family of proteins from the subtilase superfamily (Siezen *et al*., 1991; Siezen & Leunissen, 1997). This has allowed us to draw comparisons both between these two thermophilic enzymes and the mesophilic subtilisins, and between the two thermophilic enzymes, thermitase and Ak.1 protease, themselves.

Comparing first the thermophilic and mesophilic enzymes, we note four main factors that differentiate the two groups: the thermophilic enzymes have more bound calcium ions, have extensive aromatic clusters near their protein surface, have a markedly reduced proportion of hydrophobic surface (and increased polar surface) and have their N and C termini tied down by multiple interactions with the rest of the protein. All of these factors should enhance the thermostability of thermitase and AkP relative to the mesophilic subtilisins. Comparison of the two thermophilic enzymes, thermitase and AkP, with each other, emphasises the key role played by calcium binding. In the absence of added calcium, AkP is slightly less stable than thermitase (half-life at 80 °C of less than one minute, compared with nine minutes for thermitase), possibly due to its smaller number of stabilizing salt bridges (Table 3). In the presence of Ca^{2+}, however, its stability is greatly enhanced (half-life at 80 °C of 15 hours, compared with 19 minutes for thermitase) (unpublished results).

Metal ions have the ability to play a significant role in stabilizing protein structures against denaturation or proteolysis, by reducing the flexibility of the polypeptide chain and thereby inhibiting local unfolding. This is particularly true of calcium, because of its preference for binding to carboxylate and other oxygen ligands, which are the metal binding groups most likely to be presented by external loops. In enzymes of the subtilase superfamily, calcium binding contributes to thermal stability (Pantoliano *et al*., 1989) and helps to stabilize them against autolysis (Briedigkeit & Frömmel, 1989). In proteinase K, at least, it is also necessary for full enzymatic activity (Bajorath *et al*., 1988), probably by stabilizing the active conformation.

The calcium binding sites in the subtilase superfamily vary in their likely contributions to stability. Ca-1 in Ak.1 protease is a high-affinity site that is shared with thermitase and the subtilisins. Protein engineering, thermodynamic and kinetic studies (Bryan *et al*., 1992; Strausberg *et al*., 1993) show that this site plays a unique role in the stabilization of the subtilisins, since the site is apparently formed only after removal of the 77 amino acid residue pro-peptide, on which correct folding depends. This means that once calcium is bound, there is a large enthalpic barrier to unfolding (Strausberg *et al*., 1993). The contribution of this site to stabilization of the folded structure is likely, therefore, to be very high in all these enzymes. The Ca^{2+} is completely buried, and its ligands include two carboxylate groups and no water molecules, giving a formal zero charge to the site. Structurally it is formed primarily by the 83-90 loop that protrudes from helix α3 and provides four neutral oxygen ligands, but it also involves the carboxylate ligands Asp5 from the N-terminal peptide, on one

side, and Asp48 from the β1-β2 loop (43-51), on the other side, and thus ties together three separate parts of the structure.

Ca-2 is formed primarily by a single, Asp-rich, loop (58-68) that is part of the β2-α3 connection. The Ca^{2+} ligands all come from this loop, with the addition of three water molecules, and it therefore cannot have the same cross-linking role as Ca-1; indeed the corresponding site in thermitase is regarded as being of medium affinity. As for Ca-1, however, the Ca^{2+} has two carboxylate groups, from Asp58 and Asp63, directly bound (again giving a zero net charge) and also has two more carboxylate groups, Asp61 and Asp65, in its outer shell. These must enhance the affinity for Ca^{2+}. Between them the four Asp side-chains and the three water ligands generate an extensive hydrogen bonding network that extends to the surrounding structure and must be strongly enhanced by the rigidification resulting from Ca^{2+} binding. The sole positively charged group in this network, the guanidium group of Arg103, makes a big contribution, too, by bridging between Asp58 and Asp61. Site Ca-2 must contribute to the enhanced thermostability of thermitase and AkP, compared with the mesophilic subtilisins, in which it does not occur.

The major new feature in Ak.1 protease is the third Ca^{2+} binding site, Ca-3, which has not been seen in any other member of the subtilase superfamily. This site is intimately associated with Ca-1 (Figure 8) and may act cooperatively with it. It is generated by the appearance of two carboxylate side-chains that are not present in thermitase, Asp50 (Ala in thermitase) and Glu83 (Val in thermitase) and results in an additional constraint between the 83-90 loop and the β1-β2 loop (43-51). Glu83 contributes to both Ca-3 (through its carboxylate group) and Ca-1 (through its carbonyl oxygen atom) and Asp50 and Pro47 which contribute to Ca-3 are on either side of Asp48 which contributes to Ca-1, emphasising the likely cooperativity between these two binding sites. We conclude that Ca-3 must account for the dramatically enhanced effect of Ca^{2+} in stabilizing AkP, compared with thermitase. This particular combination of Asp50 and Glu83 does not appear to occur in any other subtilase from among the more than 120 sequences compared by Siezen & Leunissen (1997), and might provide a useful means of engineering enhanced thermostability; such an approach has been used previously to stabilise subtilisin BPN′ (Braxton & Wells, 1992), but with much more extensive mutagenesis required.

Conclusions

Both Ak.1 protease and thermitase, the two structurally characterized members of the ''extremophile'' family B of the subtilases (Siezen & Leunissen, 1997), have features in common that seem likely to contribute to their enhanced thermostability. These include the clustering of aromatic residues, reduced hydrophobic character in their surfaces, and the manner in which both their N and C termini are tied down by salt bridges and other interactions, so helping to prevent these chain termini from becoming nuclei for local unfolding. A significant factor, however, is the enhanced calcium binding in the two proteins. It has been suggested that in these extracellular proteases the dependence on calcium binding for conformational stability is an evolutionary development that protects against undesirable intracellular proteolysis; only when the enzyme reaches the extracellular environment, where the calcium level is higher, does calcium binding stabilize the structure sufficiently for activation of the enzymes (Betzel *et al.*, 1990). Thus it seems reasonable that in these thermostable enzymes the existing calcium dependence becomes enhanced to cope with their more extreme environment. Ak.1 protease has taken this further, with the evolution of the novel additional calcium site revealed in the structural analysis.

Materials and Methods

Preparation and purification of Ak.1 protease

E. coli clone PB5517 (MacIver *et al.*, 1994), containing the AkP gene, was grown aerobically in a fed-batch manner at 35 °C for 17 hours. The culture supernatant was heated to 70 °C for 15 minutes to precipitate the majority of the *E. coli* proteins, and the supernatant was then removed and heat-treated for a further two hours at 70 °C to activate the protease. The activated enzyme was purified using a modification of the method of Peek *et al.* (1993). This involved passing the enzyme through a phenyl-Sepharose column, followed by MonoQ and a second pass down the phenyl-Sepharose column.

Crystallization and data collection

Crystallization conditions were found using an incomplete factorial system based on orthogonal arrays, using polyethylene glycols (PEGs) as precipitant (Kingston *et al.*, 1994). Crystals were grown in hanging drops, with a protein solution comprising 10 mg/ml of the PMSF-inhibited enzyme in 10 mM TES buffer at pH 8.0 containing 5 mM $CaCl_2$ and 0.01 % Triton X-100 and a reservoir solution of 14 % PEG 4000 in 100 mM acetate buffer (pH 4.9). The crystals grew as clusters of small, very thin plates which could not be improved by alterations of pH, protein concentration, drop size or PEG concentration.

Single crystals of dimensions 0.3 mm × 0.3 mm × < 0.01 mm were separated from the clusters using a thin fibre. These crystals are monoclinic, space group $P2_1$, with cell dimensions $a = 44.1$ Å, $b = 51.7$ Å, $c = 52.8$ Å, $\beta = 96.1°$. Based on a molecular mass of 27,800 Da, calculated from the sequence (MacIver *et al.*, 1994), V_M was determined to be 2.32 Da/Å^3, indicating a solvent content of 45 % (Matthews, 1968). For data collection, the crystals were flash frozen at 110 K with a cryoprotectant comprising 35 % MPD, added to 15 % PEG 4000 in 100 mM acetate buffer at pH 4.9. These crystals diffracted to better than 2.0 Å. Because of the extreme

thinness of the crystals, the data quality was critically dependent upon the amount of cryoprotectant in the loop surrounding the crystal. Only when most of the cryoprotectant was removed, by dragging the loop along the surface of a microscope cover-slip such that the crystal was supported by an essentially parallel film of cryoprotectant, could the background scattering be minimized, and good quality high-resolution data obtained. A very fine (0.1 mm) collimator was also used to maximize the signal-to-noise ratio of the diffracted spots.

X-ray diffraction data to 1.75 Å resolution were collected from a frozen crystal using an R-Axis IIc imaging plate detector on a Rigaku RU-200 rotating anode generator. Diffraction data were processed with DENZO (Otwinowski, 1993) and scaled using programs in the CCP4 suite (Collaborative Computing Project 4, 1994). Data collection statistics are summarized in Table 3.

Structure determination and refinement

The structure was solved by molecular replacement using the automated version of AMoRe (Navaza, 1994), with data in the resolution range 10.0 to 3.5 Å, and starting models taken from the refined 1.4 Å structure of thermitase (Teplyakov *et al.*, 1990; PDB code 1THM). Two search models based on this structure were employed: a truncated model (TRUNC) consisting of the conserved residues, as determined from a sequence alignment (Figure 1), with the variable residues set to alanine; and a polyalanine model (POLYALA). The rotation solution in $P2_1$ was identical for both models at a level of 14σ for TRUNC (~ 9σ higher than the next peak) and 10σ for POLYALA (~ 5σ above the second peak). The translation solution was also identical for both models, and after rigid body fitting, the correlation coefficient and *R*-factor were 48.7 and 0.42 for TRUNC, and 36.1 and 0.48 for POLYALA.

The initial model (TRUNC) was first subjected to rigid-body refinement and simulated annealing as implemented in the program X-PLOR (Brünger, 1990; Brunger *et al.*, 1989), using data between 10.0 and 2.5 Å. For the R_{free} calculation (Brünger, 1990), in this and subsequent refinement, 5% of reflections were randomly selected. The model was then rebuilt into σ_A-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps (Read, 1986), using the molecular graphics program TURBO (Roussel, A., Inisan, A.G. & Cambillau, C., AFMB and Biographics, Marseilles, France). A total of 50 out of the 86 side-chains that had been truncated to alanine were built into the electron density during this phase. In addition, several disordered or poorly defined areas of the model were omitted. Following this first rebuilding phase, the model consisted of six fragments, 1-27, 32-53, 58-133, 145-165, 173-214 and 218-280, with the omitted regions all being in external loops connecting pieces of secondary structure.

Further refinement was with TNT (Tronrud *et al.*, 1987) using all data in the resolution range 20.0 to 1.8 Å. The high resolution limit was chosen because of a marked drop-off in data completeness beyond 1.85 Å, along with an increase in the R_{merge} for the 1.8 Å-1.75 Å shell (Table 4). Most of the remaining truncated residues and omitted loops were readily rebuilt from $2F_o - F_c$ and $F_o - F_c$ electron density maps. Solvent molecules (regarded as water) were added using the CCP4 programs PEAKMAX (with a sigma cutoff of 4.0σ) and WATPEAK (with a potential hydrogen bonding distance range of 2.5 Å to 3.4 Å); these were only retained in the model, however, if their locations made chemical sense and were not close to the omitted or poorly ordered parts of the structure. Two neighbouring regions, 135-139 and 166-175, remained problematical, however. Following further refinement, the polypeptide backbone for residues 135-139 could be modeled, but detailed interpretation was confused by a large piece of density running almost parallel with the main-chain and forming a closed ring between residues 137 and 139. Inspection of the sequence showed that these two positions were both cysteine residues, and it immediately became clear that Cys137 and Cys139 formed a disulfide bond.

Table 4. X-ray data processing

Temperature (K)	110
Oscillation range (deg.)	1.5
Number of images	105
Maximum resolution (d_{min}) (Å)	1.75
R_{merge} [a] (in highest resolution shell[b]) (%)	7.6 (37.2)
Average I/σ (in highest shell)	8.5 (1.9)
Total reflections to d_{min}	97,032
Unique reflections to d_{min}	20,847
Completeness (in highest shell) (%)	87.4 (47.0)
Redundancy (in highest shell)	2.6 (1.4)

[a] $R = \Sigma|I - \bar{I}|/\Sigma\bar{I}$.
[b] 1.80-1.75 Å.

Inspection of the temperature factors of the water molecules following 30 cycles of positional and *B*-factor refinement showed that two had very low temperature factors. When these were checked they were found to be adjacent to carboxylate and carbonyl groups, and to have several $2F_o - F_c$ and $F_o - F_c$ peaks nearby. From their environments they were recognized as possible Ca^{2+} and were included as such in the subsequent refinement. During further refinement, two more metal ions were located, one designated as Ca^{2+} and the other as Na^+.

Quality of the final model

The final model, comprising 2029 protein atoms (from 271 out of 280 residues), 225 water molecules, three Ca^{2+} and one Na^+, had an *R*-factor of 0.182 for all data (19,473 reflections) in the range 20-1.8 Å, with an R_{free} of 0.247 (1028 reflections). Analysis of the geometry showed overall rms deviations from standard bond lengths and angles (Engh & Huber, 1991) of 0.012 Å and 1.49° respectively. The average coordinate error estimated from a Luzzati plot (Luzzati, 1952) is between 0.20 and 0.25 Å, and a σ_A plot (Read, 1986) suggests an rms error in the atomic coordinates of 0.24 Å.

Biochemical analysis of the crystals

Gel electrophoresis showed that the enzyme had undergone some degree of autolysis despite the addition of PMSF. To characterize the enzyme species present in the crystals, a cluster of crystals from a single drop was washed thoroughly with several changes of synthetic mother liquor (100 mM acetate (pH 4.9), 15% (w/v) PEG 4000) to remove all protein remaining in solution. The wash solution was then removed, the crystals were dissolved in 50% formic acid, and N-terminal sequence analysis was carried out on an Applied Biosystems model 476A protein sequencer.

Miscellaneous

Silicon Graphics O2 workstations were used for all refinement, graphical manipulations and calculations. The structural superpositions were done with the program LSQKAB from the CCP4 program suite (CCP4, 1994) and the quality of the final model was analyzed with PROCHECK (Laskowski *et al*., 1993). Molecular surfaces were calculated with GRASP (Nichols *et al*., 1991) using a probe sphere radius of 1.4 Å to define the solvent-accessible surface. Aromatic interactions were inferred where the center-to-center distance between the two rings was in the range 3.5-8 Å.

Protein Data Bank accession number

Coordinates and structure factor amplitudes have been deposited in the Protein Data Bank and assigned the codes 1DBI and R1DBISF, respectively.

Acknowledgments

We thank Malcolm MacArthur for help with the search for disulfide bonds, and we gratefully acknowledge research support from the Marsden Fund of New Zealand, and the Foundation for Research, Science and Technology in New Zealand (award of a Postdoctoral Fellowship to C.A.S.).

References

Bajorath, J., Hinrichs, W. & Saenger, W. (1988). The enzymatic activity of proteinase K is controlled by calcium. *Eur. J. Biochem.* **176**, 441-447.

Barr, P. J. (1991). Mammalian subtilisins: the long-sought dibasic processing endoproteases. *Cell,* **66**, 1-3.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. J., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). Protein Data Bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.

Betzel, C., Pal, G. P. & Saenger, W. (1988). Three-dimensional structure of proteinase K at 0.15-nm resolution. *Eur. J. Biochem.* **178**, 155-171.

Betzel, C., Teplyakov, A. V., Harutyunyan, E. H., Saenger, W. & Wilson, K. S. (1990). Thermitase and proteinase K: a comparison of the refined three-dimensional structures of the native enzymes. *Protein Eng.* **3**, 161-172.

Betzel, C., Klupsch, S., Papendorf, G., Hastrup, S., Branner, S. & Wilson, K. S. (1992). Crystal structure of the alkaline proteinase Savinase from *Bacillus lentus* at 1.4 Å resolution. *J. Mol. Biol.* **223**, 427-445.

Bode, W., Papamokos, E. & Musil, D. (1987). The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech *Hirudo medicinalis*. Structural analysis, subtilisin structure and interface geometry. *Eur. J. Biochem.* **166**, 673-692.

Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B. & Power, S. (1988). The three-dimensional structure of *Bacillus amyloliquefaciens* subtilisin at 1.8 Å and an analysis of the structural consequences of peroxide inactivation. *J. Biol. Chem.* **263**, 7895-7906.

Braxton, S. & Wells, J. A. (1992). Incorporation of a Ca^{2+}-binding loop into subtilisin BPN′. *Biochemistry,* **31**, 7796-7801.

Briedigkeit, L. & Frömmel, C. (1989). Calcium ion binding by thermitase. *FEBS Letters,* **253**, 83-87.

Brocchieri, L. & Karlin, S. (1994). Geometry of interplanar residue contacts in protein structures. *Proc. Natl Acad. Sci. USA,* **91**, 9297-9301.

Bryan, P., Alexander, P., Strausberg, S., Schwarz, F., Wang, L., Gilliland, G. & Gallagher, D. T. (1992). Energetics of folding subtilisin BPN′. *Biochemistry,* **31**, 4937-4945.

Brünger, A. T. (1990). *X-PLOR Version 3.1: A System for Crystallography and NMR*, Yale University, New Haven, CT.

Brünger, A. T., Karplus, M. & Petsko, G. A. (1989). Crystallographic refinement by simulated annealing: application to crambin. *Acta Crystallog. sect. A,* **45**, 50-61.

Burley, S. K. & Petsko, G. A. (1985). Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science,* **229**, 23-28.

Collaborative Computing Project 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.

Dauter, Z., Betzel, C., Genov, N., Pipon, N. & Wilson, K. S. (1991). Complex between the subtilisin from a mesophilic bacterium and the leech inhibitor eglin-C. *Acta Crystallog. sect. B,* **47**, 707-730.

Drenth, J., Hol, W. G., Jansonius, J. N. & Koekoek, R. (1972). Subtilisin novo. The three-dimensional structure and its comparison with subtilisin BPN′. *Eur. J. Biochem.* **26**, 177-181.

Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A,* **47**, 392-400.

Fersht, A. R. & Serrano, L. (1993). Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**, 75-83.

Frömmel, C. & Höhne, W. E. (1981). Influence of calcium binding on the thermal stability of 'thermitase', a serine protease from *Thermoactinomyces vulgaris*. *Biochim. Biophys. Acta,* **670**, 25-31.

Gros, P., Betzel, C., Dauter, Z., Wilson, K. S. & Hol, W. G. (1989). Molecular dynamics refinement of a thermitase-eglin-c complex at 1.98 Å resolution and comparison of two crystal forms that differ in calcium content. *J. Mol. Biol.* **210**, 347-367.

Gros, P., Kalk, K. H. & Hol, W. G. (1991). Calcium binding to thermitase. Crystallographic studies of thermitase at 0, 5, and 100 mM calcium. *J. Biol. Chem.* **266**, 2953-2961.

Hunter, C. A., Singh, J. & Thornton, J. M. (1991). π-π interactions: the geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J. Mol. Biol.* **218**, 837-846.

Kingston, R. L., Baker, H. M. & Baker, E. N. (1994). Search designs for protein crystallization based on orthogonal arrays. *Acta Crystallog. sect. D,* **50**, 429-440.

Knapp, S., de Vos, W. M., Rice, D. & Ladenstein, R. (1997). Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. *J. Mol. Biol.* **267**, 916-932.

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

Krishnaswamy, S. & Rossmann, M. G. (1990). Structural refinement and analysis of Mengo virus. *J. Mol. Biol.* **211**, 803-844.

Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Luzzati, V. (1952). Traitement statistiques des erreurs dans la determination des structures cristallines. *Acta Crystallog.* **5**, 802-810.

MacIver, B., McHale, R. H., Saul, D. J. & Bergquist, P. L. (1994). Cloning and sequencing of a serine proteinase gene from a thermophilic *Bacillus* species and its expression in *Escherichia coli*. *Appl. Env. Microbiol.* **60**, 3981-3988.

Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.

McPhalen, C. A. & James, M. N. (1988). Structural comparison of two serine proteinase-protein inhibitor complexes: eglin-c-subtilisin Carlsberg and CI-2-subtilisin novo. *Biochemistry,* **27**, 6582-6598.

McPhalen, C. A., Schnebli, H. P. & James, M. N. (1985). Crystal and molecular structure of the inhibitor eglin from leeches in complex with subtilisin Carlsberg. *FEBS Letters,* **188**, 55-58.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D Version 2.0. A Program for photorealistic molecular graphics. *Acta Crystallog. sect. D,* **50**, 869-873.

Miller, H., Mande, S. S., Parsonage, D., Sarfaty, S. H., Hol, W. G. & Claiborne, A. (1995). An L40C mutation converts the cysteine-sulfenic acid redox center in enterococcal NADH peroxidase to a disulfide. *Biochemistry,* **34**, 5180-5190.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.

Neidhart, D. J. & Petsko, G. A. (1988). The refined crystal structure of subtilisin Carlsberg at 2.5 Å resolution. *Protein Eng.* **2**, 271-276.

Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281-296.

Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N. & Bailey, S., eds), pp. 56-62, SERC Daresbury Laboratory, Daresbury, England.

Pantoliano, M. W., Whitlow, M., Wood, J. F., Dodd, S. W., Hardman, K. D., Rollence, M. L. & Bryan, P. N. (1989). Large increases in general stability for subtilisin BPN′ through incremental changes in the free energy of unfolding. *Biochemistry,* **28**, 7205-7213.

Peek, K., Veitch, D. P., Prescott, M., Daniel, R. M., MacIver, B. & Bergquist, P. L. (1993). Some characteristics of a proteinase from a thermophilic *Bacillus* sp. expressed in *Escherichia coli*: comparison with the native enzyme and its processing in *E. coli* and in vitro. *Appl. Environ. Microbiol.* **59**, 1168-1175.

Ramakrishnan, C. & Ramachandran, G. N. (1965). Stereochemical criteria for polypeptide and protein chain conformation. *Biophys. J.* **5**, 909-933.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog. sect. A,* **42**, 140-149.

Shirai, T., Suzuki, A., Yamane, T., Ashida, T., Kobayashi, T., Hitomi, J. & Ito, S. (1997). High-resolution crystal structure of M-protease: phylogeny aided analysis of the high-alkaline adaptation mechanism. *Protein Eng.* **10**, 627-634.

Siezen, R. J. & Leunissen, J. A. M. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* **6**, 501-523.

Siezen, R. J., de Vos, W. M., Leunissen, J. A. M. & Dijkstra, B. W. (1991). Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteinases. *Protein Eng.* **4**, 719-737.

Strausberg, S., Alexander, P., Wang, L., Schwarz, F. & Bryan, P. (1993). Catalysis of a protein folding reaction: thermodynamic and kinetic analysis of subtilisin BPN′ interactions with its propeptide fragment. *Biochemistry,* **32**, 8112-8119.

Teplyakov, A. V., Kuranova, I. P., Harutyunyan, E. H., Vainshtein, B. K., Frömmel, C., Höhne, W. E. & Wilson, K. S. (1990). Crystal structure of thermitase at 1.4 Å resolution. *J. Mol. Biol.* **214**, 261-279.

Thornton, J. M. (1981). Disulphide bridges in globular proteins. *J. Mol. Biol.* **151**, 261-287.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallog. sect. A,* **43**, 489-501.

Vogt, G., Woell, S. & Argos, P. (1997). Protein thermal stability, hydrogen bonds and ion pairs. *J. Mol. Biol.* **269**, 631-643.

Wallon, G., Kryger, G., Lovett, S. T., Oshima, T., Ringe, D. & Petsko, G. A. (1997). Crystal structure of *Escherichia coli* and *Salmonella typhimurium* 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from *Thermus thermophilus*. *J. Mol. Biol.* **226**, 1016-1031.

Wells, J. A. & Estell, D. A. (1988). Subtilisin - an enzyme designed to be engineered. *Trends Biochem. Sci.* **13**, 291-297.

Edited by D. C. Rees

(Received 26 July 1999; received in revised form 27 September 1999; accepted 11 October 1999)