of the reactive group explains well the reaction process of the present crystal at both temperatures.

The crystal of *R*-ce-3mepy also belongs to the space group $P2_12_12_1$ and has two crystallographically independent molecules *A* and *B*. However, the two molecules are related by a pseudo inversion center. Moreover, the volumes of the *A* and *B* cavities are 10·24 and 14·29 Å^3, respectively. Since the two ce groups face each other around a pseudo inversion center, not only the *A* but also the *B* ce groups became disordered racemates at early stages. After 400 h exposure, the configuration of the *A* ce group gradually inverted to the original one whereas that of the *B* ce group was inverted to the opposite one. Then the pseudo inversion center became a local inversion center. The cooperative motion of the two ce groups may bring about such a complicated reaction process.

For the present *R*-ce-pip, the cooperative motion of the two reactive groups would be impossible and the two groups should be changed independently by X-ray exposure, since they are isolated from each other. Therefore, the racemization mode of *R*-ce-pip is essentially the same as mode I except that the crystal has two reactive groups.*

* Recently we have prepared the racemic crystals, *rac*-ce-pip, which have the space group $P2_12_12_1$ and have an isomorphous structure to the present one, *R*-ce-pip. At high temperatures, the crystal revealed a racemic-to-chiral conversion on exposure to X-rays. The reaction process will be reported in detail elsewhere (Osano, Uchida & Ohashi, 1991).

This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture, Japan.

References

MAIN, P., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J. P. & WOOLFSON, M. M. (1978). *MULTAN78. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.

OHASHI, Y. (1988). *Acc. Chem. Res.* **21**, 268–274.

OHASHI, Y. & SASADA, Y. (1977). *Nature (London)*, **262**, 142–144.

OHASHI, Y., SASADA, Y. & OHGO, Y. (1978*a*). *Chem. Lett.* pp. 457–460.

OHASHI, Y., SASADA, Y. & OHGO, Y. (1978*b*). *Chem. Lett.* pp. 743–746.

OHASHI, Y., TOMOTAKE, Y., UCHIDA, A. & SASADA, Y. (1986). *J. Am. Chem. Soc.* **108**, 1196–1202.

OHASHI, Y., UCHIDA, A., SASADA, Y. & OHGO, Y. (1983). *Acta Cryst.* B**39**, 54–61.

OHASHI, Y., YANAGI, K., KURIHARA, T., SASADA, Y. & OHGO, Y. (1981). *J. Am. Chem. Soc.* **103**, 5805–5812.

OHASHI, Y., YANAGI, K., KURIHARA, T., SASADA, Y. & OHGO, Y. (1982). *J. Am. Chem. Soc.* **104**, 6353–6359.

OHGO, Y., TAKEUCHI, S., NATORI, Y., YOSHIMURA, J., OHASHI, Y. & SASADA, Y. (1981). *Bull. Chem. Soc. Jpn*, **54**, 3095–3099.

OSANO, Y. T., UCHIDA, A. & OHASHI, Y. (1991). *Nature (London)*. In the press.

SHELDRICK, G. M. (1976). *SHELX76*. Program for crystal structure determination. Univ. of Cambridge, England.

TAMURA, T. (1987). MSc Thesis, Tokyo Institute of Technology, Japan.

TOMOTAKE, Y., UCHIDA, A., OHASHI, Y., SASADA, Y., OHGO, Y. & BABA, S. (1985). *Isr. J. Chem.* **25**, 327–333.

UCHIDA, A., OHASHI, Y., SASADA, Y., OHGO, Y. & BABA, S. (1984). *Acta Cryst.* B**40**, 473–478.

Acta Cryst. (1991). B**47**, 707–730

Complex Between the Subtilisin from a Mesophilic Bacterium and the Leech Inhibitor Eglin-C

BY ZBIGNIEW DAUTER AND CHRISTIAN BETZEL

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, *D*-2000 *Hamburg* 52, *Germany*

NICOLAY GENOV

Institute of Organic Chemistry, Bulgarian Academy of Sciences, Sofia 1040, *Bulgaria*

AND NATHALIE PIPON AND KEITH S. WILSON

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, *D*-2000 *Hamburg* 52, *Germany*

(*Received* 20 *November* 1990; *accepted* 5 *April* 1991)

Abstract

The alkaline proteinase from the mesophilic bacterium *Bacillus mesentericus* has been crystallized in a 1:1 complex with the inhibitor eglin-C from the medical leech. The crystals have cell dimensions of $a = 43{\cdot}0$, $b = 71{\cdot}9$, $c = 48{\cdot}3$ Å and $\beta = 110{\cdot}0°$ and are in the space group $P2_1$. Three-dimensional data to

0108-7681/91/050707-26$03.00 © 1991 International Union of Crystallography

2·0 Å have been recorded on film from a single crystal. The orientation and position of the complex in the unit cell have been established using the refined coordinates of subtilisin Carlsberg and of eglin-C as independent models. The structure of the complex has been refined by restrained least-squares minimization. The crystallographic R factor $(=\sum||F_o| - |F_c||/\sum|F_o|)$ is 15·1% including two Ca^{2+} ions and 312 water molecules. The structure is discussed in terms of its physicochemical properties in solution and its relation to other *Bacillus* subtilisins.

Introduction

The subtilisins are a group of extracellular alkaline proteinases produced by various species of *Bacillus* (Markland & Smith, 1971). The increasing interest during the last decades in these enzymes has been due to the biological significance of their proteolytic activity and therefore many studies of their structure and function have been carried out. For example subtilisin BPN′/NOVO was the first enzyme to be studied by protein engineering as reviewed by Wells & Estell (1988). Genetically engineered mutants of subtilisin BPN′ have been extensively investigated (*e.g.* Russell & Fersht, 1987; Russell, Thomas & Fersht, 1987; Pantoliano, Ladner, Bryan, Rollence, Wood & Poulos, 1987; Pantoliano, Whitlow, Wood, Rollence, Finzel, Gilliland, Poulos & Bryan, 1988). In addition the subtilisins have extensive practical uses, mainly as additives to laundry detergents to improve their cleaning power by hydrolysis of protein stains. They have also found applications in medicine as components of proteolytic creams for removal of necrotic material in wounds leading to their faster closure, for the preparation of collagen implants *etc.* It is therefore not suprising that there was a period when these proteinases were produced in far greater quantities for commercial use than all other enzymes together.

Mesentericopeptidase (henceforth SBMEP) (peptidyl peptide hydrolase, EC 3.4.21) is an alkaline extracellular serine proteinase produced by a strain of *Bacillus mesentericus*, a mesophilic bacterium (Karadjova, Bakurdjieva & Velcheva, 1970) with an optimum growth temperature of 310 K and is a member of the subtilisin family of proteinases (Genov & Shopova, 1974*b*). SBMEP has been isolated in a homogeneous state and its complete amino-acid sequence determined (Svendsen, Genov & Idakieva, 1986). The protein contains 275 amino-acid residues in a single polypeptide chain.

Subtilisins have been extracted from many species of *Bacillus*. Those which have been most studied are shown in Table 1, which gives the name of the bacterial source, the trivial name in general use for the subtilisin secreted by the organism, and the abbreviation used throughout this paper for the enzyme. The amino-acid sequences of several *Bacilli* subtilisins have been determined to date including subtilisin NOVO or BPN′ (the same protein: SBPN) (Markland & Smith, 1967), SBCARL (Smith, De Lange, Evans, Landon & Markland, 1968), SBAS (Kurihara, Markland & Smith, 1972; Yoshimoto, Oyama, Honda, Tone, Takeshita, Kamiyama & Tsuru, 1988), SBDY (Nedkov, Oberthür & Braunitzer, 1983) and SBI168 (Stahl & Ferrari, 1984). In addition several three-dimensional structures have been determined by X-ray diffraction. The first crystal structure to be studied was that of SBPN, first as subtilisin BPN′ (Wright, Alden & Kraut, 1969) and then as the identical subtilisin NOVO (Drenth, Hol, Jansonius & Koekoek, 1972). This structure was subsequently refined at 1·8 Å resolution with phenyl methyl sulfonyl fluoride (PMSF) inhibitor covalently bound (Bott, Ultsch, Kossiakoff, Graycar, Katz & Power, 1988), in complex with *Streptomyces* inhibitor at 2·6 Å (Hirono, Akagawa, Mitsui & Iitaka, 1984) and at 2·1 Å as a complex with the chymotrypsin inhibitor 2 from barley seed (McPhalen, Svendsen, Jonassen & James, 1985; McPhalen & James, 1988). High-resolution structures of site-directed mutants of SBPN (Pantoliano *et al.*, 1987, 1988) have also been reported. The three-dimensional structures of other *Bacilli* subtilisins have been published: the SBCARL–eglin-C complex at 1·8 Å (McPhalen, Schnebli & James, 1985; McPhalen & James, 1988) and independently at a nominal resolution of 1·2 Å (Bode, Papamokos, Musil, Seemüller & Fritz, 1986; Bode, Papamokos & Musil, 1987), and native SBCARL at 2·5 Å (Neidhart & Petsko, 1988).

The sequences of homologous subtilisin-like proteinases from at least three organisms other than *Bacilli* have also been determined: thermitase from *Thermoactinomyces vulgaris* (Melhoun, Baudys, Kostka, Hausdorf, Frömmel & Höhne, 1985), proteinase K from *Tritirachium album* Limber (Jany & Meyer, 1985) and aqualysin from *Thermus aquaticus* (Terada, Kwon, Miyata, Matsuzawa & Ohta, 1990). The first two enzymes belong to the subgroup containing a single free cysteinyl residue. The three-

Table 1. *Bacilli subtilisin proteinases*

The names of the bacterial species producing the various subtilisins, the 'trivial' names (if any) used for the subtilisins from these *Bacilli*, and the abbreviations used throughout the text for these subtilisins.

Bacterium	Trivial name	Abbreviation
Bacillus mesentericus	Mesentericopeptidase	SBMEP
Bacillus subtilis I168	Subtilisin I168	SBI168
Bacillus amylosaccharitícus		SBAS
Bacillus amyloliquefaciens	Subtilisin BPN′/NOVO	SBPN
Bacillus subtilis DY	Subtilisin DY	SBDY
Bacillus licheniformis	Subtilisin Carlsberg	SBCARL

dimensional crystal structures of proteinase K (Betzel, Pal & Saenger, 1988), of native thermitase (Teplyakov, Kuranova, Harutyunyan, Vainshtein, Frömmel, Höhne & Wilson, 1990) and of thermitase complexed with eglin-C (Dauter, Betzel, Höhne, Ingelman & Wilson, 1988; Gros, Betzel, Dauter, Wilson & Hol, 1989; Gros, Fujinaga, Dijkstra, Kalk & Hol, 1989) have been refined at high resolution. Although the overall tertiary fold is similar, and the active-site geometry very similar, these enzymes are substantially less homologous to the *Bacilli* subtilisins than are the latter among themselves. Comparisons in this paper will be generally restricted to the *Bacilli* subtilisins.

Ca^{2+} ions play an important role in the biological action of subtilisins (Voordouw, Milo & Roche, 1976). They stabilize the protein molecule and prevent (but not totally) autolysis. *Bacilli* subtilisins have two calcium binding sites (Bode *et al.*, 1987; McPhalen & James, 1988; Bott *et al.*, 1988). The first one binds Ca^{2+} very tightly so that even in the presence of EDTA the metal ion cannot be completely removed. In contrast the second Ca^{2+} can be easily removed with EDTA.

Three Ca^{2+} binding sites are present in thermitase and two in proteinase K (Betzel *et al.*, 1988; Gros, Betzel *et al.*, 1989). In proteinase K the first (stronger) calcium site is located in a position close, but not structurally equivalent, to the second (weaker) Ca^{2+} site in the *Bacilli* subtilisins. The second site is quite different, and as yet has not been observed in other subtilisin or subtilisin-like proteinases. Thermitase has two major calcium sites, plus a third one occupied by calcium in one of the eglin-C complexes but empty in the native structure and the other eglin-C complex. The first and third sites are equivalent to those in the *Bacilli* subtilisins, but the second site is different (Gros, Betzel *et al.*, 1989). The major Ca^{2+} sites in proteinase K and thermitase are different from one another. This emphasizes the difference between the detailed structures of these enzymes and the *Bacilli* subtilisins and may be one of the contributing factors in the enhanced thermostability of thermitase and proteinase K (Betzel, Teplyakov, Harutyunyan, Saenger & Wilson, 1990). In all these proteinases the Ca^{2+} sites are located close to the surface in loop regions.

The sequence data show a high homology of SBMEP to SBPN (38 substitutions out of 275 residues) and to SBCARL (80 substitutions, one deletion), and a near identity to SBI168 and SBAS with only three and five substitutions, respectively. However there are certain interesting differences between the physicochemical properties of SBMEP and other subtilisins. In particular three 'global' properties have been shown to be different for SBMEP, namely low solubility, low thermal stability and susceptibility to denaturation by urea. Some specific differences in kinetic properties have been reported, and a good deal of information on the spectroscopic behaviour and chemical reactivity of SBMEP in solution. These will be discussed in terms of the refined structure in later sections, where full references are provided.

In this paper we describe the 2·0 Å resolution analysis of the crystal structure of SBMEP in a 1:1 complex with the proteinase inhibitor eglin-C from the leech, *Hirudo medicinalis*. This is the first elucidation of the three-dimensional structure of a subtilisin from a mesophilic bacterium. The structure is compared to those of the other *Bacilli* subtilisins.

Purification and crystallization

Attempts to crystallize native SBMEP remained unsuccessful after experimenting with a wide range of conditions. SBMEP–eglin-C complex crystals were grown from polyethylene glycol (PEG) 4000 using the hanging-drop vapour-diffusion method. Crystals grew in 2–3 days at a PEG concentration of 18% in the well. The droplets contained initially 5 mg ml^{-1} of enzyme, 9% PEG 4000, 25 m*M* PIPES buffer at pH 7·0 and an enzyme/eglin-C ratio of 1:1·3. To avoid precipitation, enzyme and PEG solutions were mixed prior to the addition of eglin-C. Only two large crystals, with maximum dimension about 0·4 mm, were obtained with the limited amount of enzyme and inhibitor available. Both had smaller crystals growing from the various faces and these were cut off with a fragment of razor blade before data collection.

Both crystals were mounted in 1·0 mm diameter thin-walled glass capillaries. The first was used for determination of the cell dimensions and symmetry using an Enraf–Nonius precession camera with an Elliott GX18 rotating-anode generator. The space group was identified as $P2_1$ with cell dimensions a = 43·0, b = 71·9, c = 48·3 Å and β = 110°. Assuming a molecular weight of 28 kD for the protein and 7·5 kD for the eglin-C the packing density, V_m, was 2·1 D Å^{-3}, with one complex per asymmetric unit. This corresponds to a water content of 41% (Matthews, 1968).

Data collection and processing

Three-dimensional diffraction data were recorded from the second large crystal using synchrotron radiation from the EMBL beamline X31 at the DORIS storage ring, DESY, Hamburg. A modified Enraf–Nonius Arndt–Wonacott oscillation camera was used, with flat-plate cassettes each containing three sheets of CEA Reflex-25 X-ray film. The wavelength of the radiation was 1·488 Å and the crystal-

to-film distance was 56 mm. The crystal was rotated around the axis approximately parallel to the crystal c^* axis through a total of 90°, with a rotation 2° and an average exposure time of about 10 min per pack.

The films were digitized on an Optronics P-1000 photoscan microdensitometer controlled by a VAX 11/750 computer, in the absorbance range 0–3 OD and with a raster size of 50 μm. The images were processed using the *MOSFLM* package of programs (Leslie, Brick & Wonacott, 1986) to give the integrated intensities. A total 32 235 reflections were measured with 25 625 being fully recorded and 6610 partially recorded. After scaling the partially recorded intensities from adjacent films were summed and used with the exception of those at the two ends of the data-collection range or those spread over more than two films. The data reduced to 17 300 independent reflections, which is 92·3% of the available data in the range from 10·0–2·0 Å. The overall completeness and the percentage of the theoretically measurable intensities recorded above the 3σ level are shown as a function of resolution in Fig. 1. A significant percentage of reflections is missing

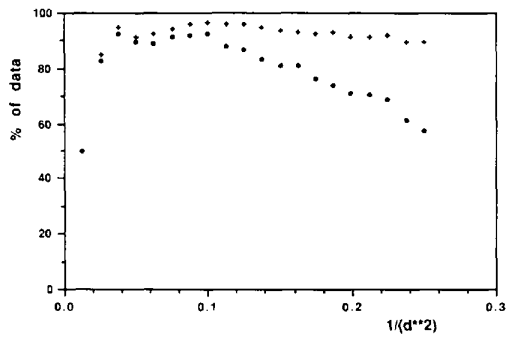


Fig. 1. The percentage completeness (crosses) of the X-ray diffraction data for SBMEP as a function of resolution. The percentage of the measured X-ray intensities greater than 3σ is also shown (dots).

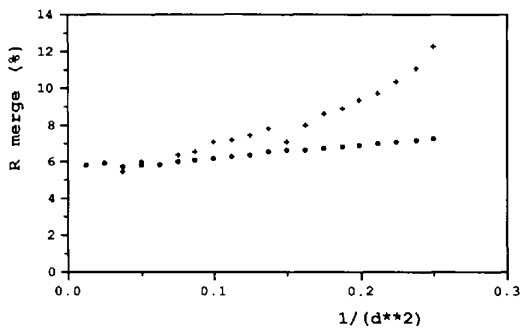


Fig. 2. The merging R factor, as defined in the text, as a function of resolution. The crosses correspond to values in the individual resolution ranges, and the circles to the cumulative values to that resolution.

from the lowest resolution range: these were overloaded on all three films of the peaks. Including all data, with no σ cutoff, the merging R factor ($\sum|I - \langle I\rangle|/\sum I$) for the symmetry-equivalent reflections was 7·3%, rising to 12·3% for the highest resolution shell. The percentage of reflections with intensity greater than 3σ in the highest resolution shell is 64%. The variation of the merging R factor with resolution is shown in Fig. 2. The Wilson plot (Wilson, 1949) for the data gives an overall temperature factor of 11·9 Å^2.

All further calculations were carried out using the units of a MicroVAX cluster and the *CCP*4 suite of programs provided by the SERC Daresbury Laboratory.

Structure solution

The structure was solved by molecular replacement. The program *ALMN* was used for the rotation-function calculations and the program *SEARCH* for the R-factor translation searches. Initially the unrefined coordinates of the homologous SBPN (Drenth *et al.*, 1972) from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977; coordinate set reference 2SBT) were used. However with these coordinates it was not possible to produce a satisfactory solution to the rotation function, after many attempts using various reflection ranges and program parameters.

Subsequently the coordinates of the SBCARL–eglin-C complex refined at 1·8 Å (McPhalen & James, 1988), which were kindly provided by Michael James and colleagues prior to their release from the Brookhaven Protein Data Bank (coordinate set 2SEC), were used. These gave a straightforward solution to the rotation function immediately, followed by a simple solution to the translation problem. The results are shown in Figs. 3(*a*,*b*). There was a single large peak in the rotation function, nine times the σ level of the map and 2·5 times higher than the next peak. Data in the resolution range 10·0 to 3·0 Å were used in the calculation with the 1795 strongest reflections from the X-ray data and 1664 ones computed from the model. In the R-factor search, restricted to the (x,z) plane, the lowest R value of 35·0%, based on 1975 reflections in the same resolution range, corresponded to the correct solution. The next lowest R value was 38·7% and average value in the map 40·3%. These results confirm what was found in studies of thermitase–eglin-C (Dauter *et al.*, 1988): it is most important to have accurate refined coordinates as a model for molecular replacement.

Restrained least-squares refinement was started using this model of the SBMEP–eglin-C. After an encouraging start when the R factor fell quite rapidly

from 46·0 to 28·0% in the first six cycles using data to 2·5 Å, the refinement including all data between 10·0 and 2·0 Å converged after more than 50 cycles of least squares to an R factor of about 28%. The temperature factors were relatively high for most of the eglin-C atoms. Inspection of the $(3F_o - 2F_c)$ synthesis showed very poor density for most of the inhibitor molecule, excepting those parts in direct contact with SBMEP. The most likely explanation for this lay in the flexible binding of the eglin-C molecule to subtilisins: it had previously been noted that the relative orientation of the eglin-C is different in its two complexes with thermitase (Gros, Betzel *et al.*, 1989) and in that with SBCARL (McPhalen & James, 1988; Bode *et al.*, 1987). Omission of the eglin-C from the structure-factor calculation and computation of 'omit' difference Fourier syntheses did not allow interpretation of the true eglin-C position.

The solution to this problem involved an independent solution of the rotation and translation functions for the SBMEP and eglin-C separately. The functions were first calculated for the SBCARL only as the model. The results were very similar to those shown in Figs. 3(*a*,*b*). Again the correct solution was clear in both functions. The rotation function gave a clear solution, ten times the σ level of the map and about twice as high as the next peak, using 1795 reflections for the data and 1703 for the model. The R-factor search also gave a single correct solution with an R value of 33·1% based on 1999 reflections. The next lowest R value was 37·8% with an average

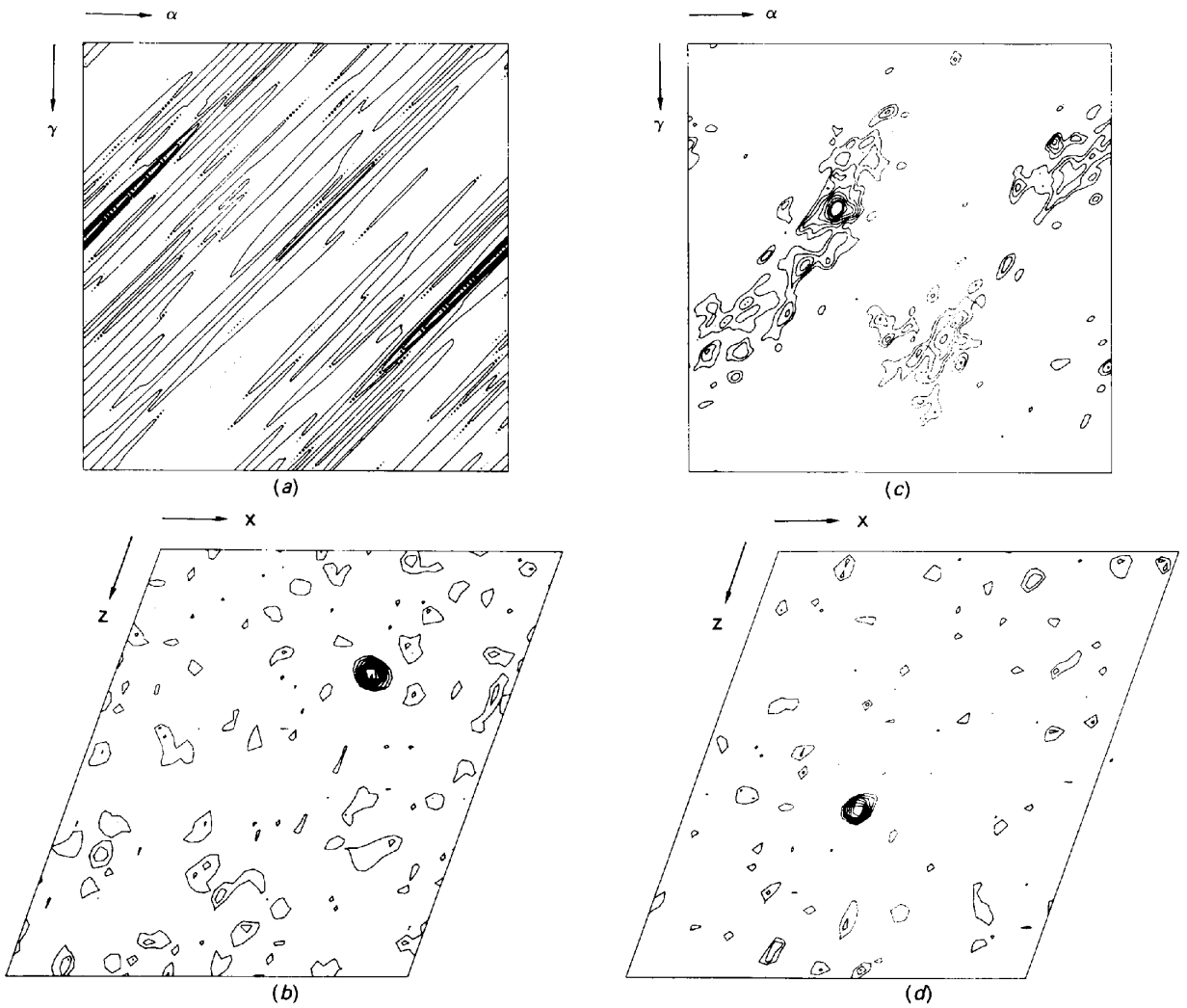


Fig. 3. Results of molecular replacement. (*a*) Section $\beta = 51°$ of the rotation map with subtilisin Carlsberg–eglin-C as a model using the resolution range between 3 and 10 Å. (*b*) Section $y = 0$ of the corresponding R-search map where the function plotted is the number of standard deviations of the point below the average R factor for the whole asymmetric unit of the synthesis. (*c*), (*d*) The comparable rotation-function and translation-function solution using only eglin-C with the 'difference' amplitudes as described in the text.

value of 39·5%. The orientation and position for the proteinase molecule were very close to those found for the complex above.

The SBCARL chain alone was refined against the SBMEP data, using restrained least-squares minimization. The R factor, using data between 10·0 and 2·5 Å resolution, was 30·7% after six cycles of minimization.

At this point the calculated structure factors for the SBCARL model were scaled, to allow for the fact that SBCARL makes up about 80% of the complex, and subtracted from the observed structure factors to give a residual set, assumed to arise primarily from the eglin-C. The rotation and translation functions were computed against these residual structure factors to establish the orientation and position of the eglin-C independently. A straightforward solution was obtained for both functions at the first attempt, and the results are shown in Figs. 3(*c*,*d*). The rotation function gave an immediate solution, seven times the σ level and two times higher than the next peak, using 1662 'observed' and 1651 calculated data. The R-factor search also gave a simple solution with $R = 45{\cdot}0\%$ below an average value of 48·8%, with the next lowest value being 46·9%. As is discussed below, the eglin-C is in a somewhat different orientation in relation to SBMEP compared to SBCARL, with a rotation of the molecule of about 16° about the binding residues as a hinge: this explains the difficulty in using the complete SBCARL–eglin-C as a model for the SBMEP–eglin-C complex.

The eglin-C and SBCARL molecules were then independently rotated and translated into the SBMEP–eglin-C cell and used as the starting model for the refinement. Such an approach, using subunits or domains of proteins as the search molecules and subtracting out known parts, may prove generally useful in molecular-replacement studies.

Structure refinement

The structure was refined using the restrained least-squares minimization procedure of Hendrickson & Konnert (Konnert & Hendrickson, 1980), incorporating the fast Fourier algorithm (Agarwal, 1978) as developed by Eleanor Dodson (*e.g.* Baker & Dodson, 1980). Throughout the refinement the X-ray terms were assigned unit weights relative to one another. All reflections were included: no σ cutoff was applied. The weighting scheme used for the stereochemical restraints is shown in Table 2. The program *FRODO* (Jones, 1978) running on an Evans and Sutherland PS330 interactive graphics unit was used for all molecular graphics operations. These included inspection of parts of the model with both $(3F_o - 2F_c)$, $(F_o - F_c)$ Fourier syntheses and 'omit' maps at appropriate steps in the refinement, and introduction of water molecules into the model.

Table 2. *Target values (σ) and standard deviations for the restrained parameters in the refinement*

The weight corresponds to $1/\sigma^2$.

		σ	Standard and deviations	No. of parameters
Distances (Å)				
Bond lengths	(1–2 neighbours)	0·020	0·012	2535
Bond angles	(1–3 neighbours)	0·040	0·039	3467
Dihedral angles	(1–4 neighbours)	0·050	0·044	927
Planar groups		0·020	0·009	433
Chiral volumes (Å^3)		0·150	0·126	397
Non-bonded contacts (Å)				
Single torsion contacts		0·500	0·180	821
Multiple torsion contacts		0·500	0·300	1190
Possible hydrogen-bonding contacts		0·500	0·190	227
Torsion angles (°)				
Peptide plane (ω)		2·0	1·5	345
Staggered (c_1)		15·0	17·4	364
Orthonormal (c_2)		20·0	32·3	37

A total of 89 cycles of least-squares minimization were carried out, divided up into 15 steps. An overview of the results is shown in Table 3, which gives the geometrical and thermal parameters, R factor and number of atoms in the model after each step. The procedure followed in each step of the refinement, and the data and model included therein, are now described.

Step 1. Cycles 1 to 6. The starting model consisted of the molecules of SBCARL and eglin-C independently placed in the SBMEP–eglin-C unit cell. The sequence of eglin-C is already known (Seemüller, Eulitz, Fritz & Strobel, 1980; Knecht, Seemüller, Liersch, Fritz, Braun & Chang, 1983). Only residues Lys8*I* to His68*I* of eglin-C were included in the initial model, as the terminal residues had been observed to be sometimes disordered in previous analyses. No solvent molecules or calcium ions were included. The side chains of the enzyme were left as those of SBCARL until the rebuild before step 4. The positional parameters only were refined using data between 10·0 and 2·5 Å. The initial R factor of 39·31% fell to 25·25%. The value for the matrix parameter weighting the X-ray data against the geometrical restraints was 2·0.

Step 2. Cycles 7 to 12. The resolution was extended to include data in the range 10·0 to 2·2 Å. The other parameters were as for step 1. The R factor fell from an initial value of 29·06 to 26·62%.

Step 3. Cycles 13 to 20. The resolution was extended again to take the full set of data between 10·0 and 2·0 Å. The same resolution range was used in all following steps. The positional parameters were varied in cycles 13 to 18, and only the atomic B values in cycles 19 and 20. The other parameters were as for step 1. The R factor dropped from 29·14 to 25·11%.

Table 3. *Progress of the refinement*

Step	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Cycle numbers	1–6	7–12	13–20	21–28	29–36	37–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89
No. of reflections	8924	13029	17212	17212	17212	17212	17212	17212	17212	17212	17212	17212	17212	17212	17212
No. of atoms	2430	2430	2430	2444	2501	2596	2631	2654	2662	2705	2705	2776	2788	2800	2795
No. of amino-acid residues	335	335	335	336	336	336	336	336	338	338	338	339	339	338	339
No. of solvent molecules		–	–		57	142	177	195	202	234	234	297	309	339	312
Minimum resolution (Å)	10·0	10·0	10·0	10·0	10·0	10·0	10·0	10·0	10·0	10·0	10·0	10·0	10·0	10·0	10·0
Maximum resolution (Å)	2·5	2·2	2·0	2·0	2·0	2·0	2·0	2·0	2·0	2·0	2·0	2·0	2·0	2·0	2·0
R factor at start (%)	39·3	29·1	29·1	30·1	23·7	20·1	18·9	18·4	17·6	17·5	18·5	17·0	17·2	16·2	16·0
R factor at end (%)	25·3	26·6	25·1	23·8	20·6	18·3	17·9	17·6	16·7	16·0	16·2	15·2	15·0	15·3	15·1
X-ray/geometry weighting	2·0	2·0	2·0	1·5	1·5	0·8	0·6	0·6	0·6	0·6	0·5	0·5	0·5	0·25	0·25
Mean temperature factor, B(Å^2)															
Main-chain atoms	9·3	9·3	12·7	13·7	13·6	14·0	14·4	14·9	15·1	15·6	15·7	15·5	15·6	15·3	15·5
Side-chain atoms	12·9	12·9	16·7	17·6	17·9	19·0	20·0	21·1	21·2	22·0	22·2	21·5	21·9	21·6	22·2
Overall protein	10·9	10·9	14·5	15·5	15·5	16·3	17·0	17·7	17·9	18·5	18·6	18·3	18·4	18·2	18·5
Overall solvent	–	–	–		24·5	28·0	30·9	33·4	34·3	35·1	36·6	39·1	40·0	42·0	39·2
Total (protein + solvent)	10·9	10·9	14·5	15·5	15·7	16·9	17·9	18·9	19·1	19·9	20·1	20·5	20·8	21·0	20·8
Standard deviations from ideal values															
1–2 distances (bonds) (Å)	0·102	0·081	0·102	0·042	0·070	0·064	0·043	0·046	0·025	0·024	0·020	0·019	0·018	0·012	0·012
1–3 distances (angles) (Å)	0·168	0·155	0·171	0·081	0·123	0·105	0·092	0·077	0·065	0·062	0·051	0·052	0·047	0·040	0·039
Planar groups (Å)	0·055	0·047	0·028	0·025	0·036	0·038	0·032	0·021	0·018	0·017	0·016	0·015	0·013	0·010	0·009
Chiral volumes (Å^3)	0·555	0·510	0·600	0·293	0·433	0·360	0·330	0·285	0·230	0·224	0·182	0·174	0·152	0·128	0·126
Torsion angle (ω) (°) of peptide plane	8·6	7·3	11·8	3·5	4·6	3·8	3·7	3·1	2·6	2·5	2·3	2·2	2·1	1·6	1·5

Step 4. Cycles 21 to 28. The model was rebuilt using *FRODO* before this step. In particular the side chains of SBCARL were modified to those of SBMEP where appropriate and fitted to the density in the $(3F_o - 2F_c)$ and $(F_o - F_c)$ Fourier syntheses. The occupancies for some or all of the atoms of four side chains (Ser3, Arg45, Ser78 and Pro129) were set to zero as there was no significant density in which to place them. Positional parameters were varied for six cycles, followed by atomic B values for two cycles. The matrix parameter was reduced to 1·5. The R factor fell from 30·63 to 23·77%.

Step 5. Cycles 29 to 36. As in the previous step the positional parameters were varied in the first six cycles, followed by the B values in the last two. Other parameters were as for the previous step. 57 water molecules were added to the model. The initial R factor was 22·57% and fell to 20·62%. At this stage the geometry was rather poor, Table 3.

Step 6. Cycles 37 to 44. The matrix parameter was reduced to 0·8 to tighten up the restraints on the geometry. In the rebuild before this step the number of water molecules was increased to 142 and the atoms which had been given zero occupancy (step 4) were built into the density. Water molecules were mainly selected from the difference Fourier synthesis by an automatic procedure taking account of peak height, hydrogen-bond formation and non-bonded contacts (V. Lamzin, unpublished). The R factor started at 20·10% and dropped to 18·32%.

Step 7. Cycles 45 to 49. In this and all subsequent steps three cycles of positional refinement were followed by two of B values. The matrix value was decreased to 0·6 to further tighten the geometrical restraints. Two of the water molecules were now reassigned to be Ca^{2+} ions on the basis of very low B values and of coordination geometry, with appropriately short contacts in both cases. The number of water molecules was increased to 175 and several regions of protein chain were rebuilt. The R factor fell from 18·86 to 17·88%. All temperature factors which refined to values less than 3·0 were reset to 3·0 at the beginning of this and subsequent steps, in order to avoid problems with the atomic peaks being too sharp in the reciprocal Fourier transformation of the Agarwal procedure.

Step 8. Cycles 50 to 54. The water molecules were increased in number to 177. The conditions were essentially as for step 7. The R factor began at 18·40% and ended at 17·62%.

Step 9. Cycles 55 to 59. The number of waters was increased to 200 and parts of the chain rebuilt as in previous steps. R fell from 17·65 to 16·65%. The geometry at the end of this step was somewhat improved with respect to its target values, Table 3.

Step 10. Cycles 60 to 64. Two residues were added at the C terminus of the eglin-C molecule, Val69*I* and Gly70*I*, as there was clear density for them in the $(3F_o - 2F_c)$ and $(F_o - F_c)$ Fourier syntheses. There were now 232 water molecules in the model. The R factor fell from 17·46 to 16·02%.

Step 11. Cycles 65 to 69. Before this step the whole model was idealized using the *'REFI'* regularization procedure within *FRODO*. The matrix parameter was decreased to 0·5, to tighten the restraints. Five cycles of refinement were then performed as in step 10. R fell from 18·53 to 16·15%, with the geometry greatly improved.

Step 12. Cycles 70 to 74. The number of water molecules was increased to 295. Regions of the protein with high B values or poor geometry were inspected and adjusted as necessary. There was low density visible in both the $(3F_o - 2F_c)$ and $(F_o - F_c)$ Fourier syntheses at the N terminus of the eglin-C

model, Lys8*I*. This could be interpreted as residue Leu7*I*, which was introduced to the model. The *R* factor fell from 16·95 to 15·24%.

Step 13. Cycles 75 to 79. Water molecules with high *B* values, greater than 50, were inspected on the graphics station and either removed from the model or moved into nearby density. New water molecules were added, and the resulting overall number of waters was 307. Again poorly defined protein regions were rebuilt. The *R* factor started at 17·16% and dropped to 14·95%.

Step 14. Cycles 80 to 84. The model was inspected and rebuilt as for the previous step. The number of water molecules was now 337. The temperature factors for all the atoms of Leu7*I*, introduced in step 12, had refined to relatively high values, about 60, and the residue was left out for this step. The matrix parameter was further tightened to 0·25, and the *R* factor fell from 16·17 to 15·26%. The geometrical parameters were greatly improved.

Step 15. Cycles 85 to 89. The model was carefully inspected as before the last two steps. A few small adjustments to the protein were carried out, in particular a rotation by roughly 180° about the main-chain ψ angle of Pro58*I*. Low, but significant, density was again apparent for Leu7*I*, and this was reintroduced to the model. The water molecules were carefully inspected and those with *B* values greater than 70 or without good density were rejected. The *R* factor dropped from 15·79% to a final value of 15·05%.

Refinement was stopped at this point as it was felt that no significant improvement would result using the criteria described in the next section. All atoms have been assigned occupancies of 1·0. Many of the side chains and water molecules have high temperature factors as discussed below. No attempt was made to model alternative side-chain conformations at this resolution. The final model comprises residues 1–275 of SBMEP, residues 7*I*–70*I* of eglin-C, two Ca^{2+} ions and 312 water molecules.

Accuracy of the model

Some criteria for the quality of the final model are briefly presented here. The *R* factor is plotted as a function of resolution Fig. 4(*a*). The σ_A plot of Read (1986) is shown in Fig. 4(*b*) from which the root-mean-square error in the atomic coordinates can be estimated as 0·20 Å. This should be considered when comparing the model with those of SBCARL and SBPN below. The deviation from ideality of the stereochemical parameters together with their target values are presented in Table 2. The low value of the matrix parameter weighting the X-ray data against the stereochemical restraints resulted in the good geometry for the final model.

The statistics of the electron densities after the last step of the refinement are now summarized. The syntheses were calculated on approximately the absolute level, using the correct value of the cell volume, and 40 000 electrons for *F*(000) for the $(3F_o - 2F_c)$ map. The latter value does not include the disordered solvent contribution. The root-mean-square value of the density in the final difference map was 0·063 e Å^{-3} and the largest negative and positive features were −0·30 and 0·31 e Å^{-3}. In the final $(3F_o - 2F_c)$ synthesis the root-mean-square deviation from the mean density was 0·22 e Å^{-3}, the mean value 0·29 e Å^{-3}, the minimum −0·68 e Å^{-3} and the maximum 2·05 e Å^{-3}. The electron density corresponding to many of the water molecules added in the last steps of the refinement was only just significant above the 3σ level of the difference map.

Fig. 5 shows the Ramachandran plot (Ramakrishnan & Ramachandran, 1965) for the whole complex. The only non-glycine residues with values of conformational angles (φ,ψ) significantly outside the allowed regions are: Asp32 (−156, −144°) and Asn77 (−149, −147°). Aspartic acid 32 is a member of the catalytic triad and its conformation is strained by involvement of its side-chain carboxyl O*D*2 atom in strong hydrogen bonds with the N atom of the

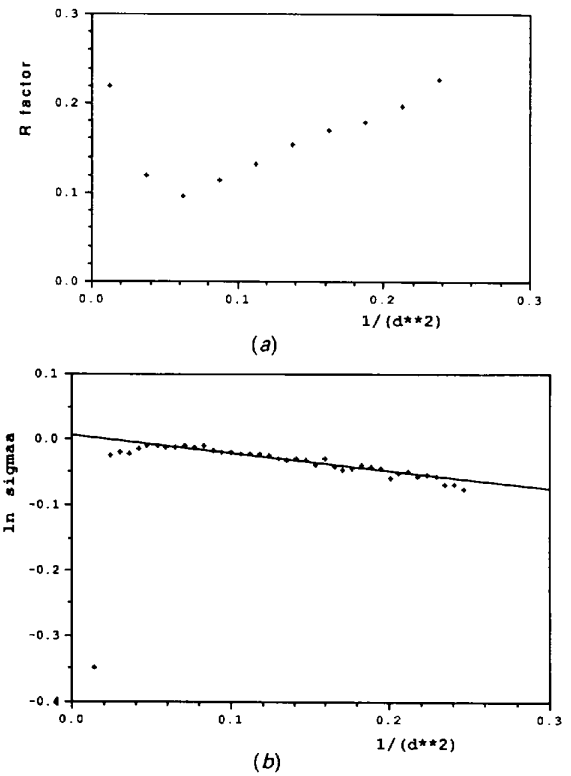


Fig. 4. (*a*) The final *R* factor as a function of resolution and (*b*) the σ_A plot for the final model. The root-mean-square error in the coordinates estimated from the latter is 0·20 Å.

next residue, Ser33, and the N*D*1 atom of the His64 imidazole ring. The O*D*1 atom of Asp32 forms a hydrogen bond to the buried water Wat452. The main-chain atoms of Asp32 are held in place by hydrogen bonds to the O atom of Val93 and the N atom of Asp95. The asparagine 77 amide oxygen is one of the atoms coordinating the calcium Ca1. These two residues have similar 'not-allowed' values in the other subtilisin structures, *e.g.* the SBCARL–eglin complex (McPhalen & James, 1988; Bode *et al.*, 1987).

Two regions of the $(3F_o - 2F_c)$ Fourier synthesis computed with phases from the final model are shown in Fig. 6: the areas around the tightly and the weakly bound calcium ions, sites Ca1 and Ca2, respectively.

In the final model of SBMEP all atoms have been assigned unit occupancy. The average *B* values for the various parts of the model are listed in Table 6. A small number of regions show substantial thermal motion or static disorder which is reflected in very weak density in the Fourier syntheses and high values of the corresponding temperature factors. Residues with at least one atom having a *B* value greater than 70 at the end of refinement are: Lys12, Gln19, Asn25, Arg45, Asn56, Gln59, Ser98, Gln103, Lys141, Val143, Ser158, Ser161, Asn184, Gln185, Ser188, Gln206, His238, Thr240, Thr248, Tyr256, Gln271, Gln275; and for the inhibitor: Leu7*I*, Lys8*I*, Asp19*I*, Gln20*I*, Arg22*I*, Glu23*I*, Tyr32*I*, Asn33*I*, Tyr49*I*, Asn60*I*. None lie in the 'functional' regions such as the active site or calcium ion binding sites. Gln103, which changes to Thr in SBCARL, lies in the *S*1 site, which is not very substrate specific in the subtilisins. Of these 32 residues, however, 19 have just one atom with *B* above the limit. Only eight of them have three or more atoms with *B* above 70.

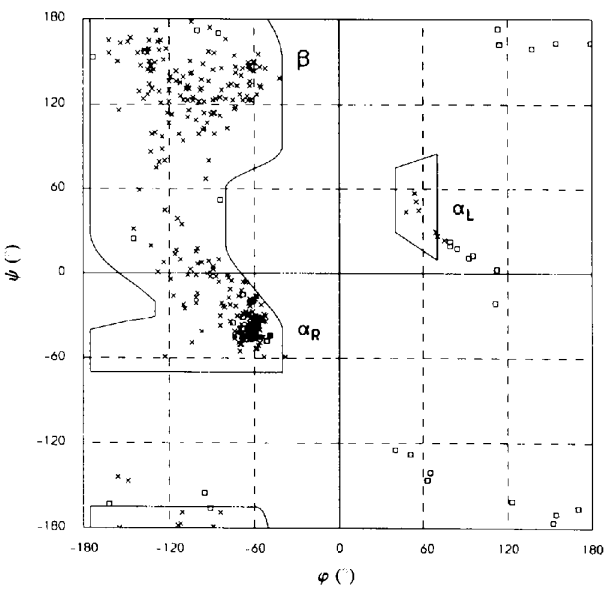


Fig. 5. The Ramachandran plot. The allowed regions for (φ,ψ) angles are marked for non-glycine residues in secondary structural elements: α_R, α_L and β for right- and left-handed helix and β-sheet, respectively. Squares correspond to glycines, crosses to other residues.

The surface area of each residue accessible to solvent (Kabsch & Sander, 1983) in SBMEP, excluding the eglin-C atoms from the computation, and inversely for eglin-C excluding the protein, are plotted as a function of residue number in Fig. 7(*a*). The variation of the average *B* value for the main chain and the side chain with residue number of SBMEP and eglin-C is shown in Fig. 7(*b*). The surface-area plot is repeated for SBMEP and for eglin-C in Fig. 7(*c*), with the calculation including all surrounding protein atoms from the other member of the complex or symmetry-equivalent neighbours. The total solvent-accessible surface area of 'native' SBMEP molecule extracted from the crystal complex is 10 073 Å^2. This is reduced to 9241 Å by introduction of the eglin-C, and further to 7872 Å^2 by taking into account all crystal contacts. The last value, and Fig. 7(*c*), reflects the true exposure of the SBMEP in the crystal. Careful inspection of Figs. 7(*b*) and 7(*c*) reveals that those residues having high temperature factors, especially for the side chains, mostly lie on the surface with large exposure to solvent. The eight residues with several atoms having *B* factors greater than 70 (Asn25, Arg45, Gln59, Gln185, Leu7*I*, Lys8*I*, Glu23*I*, Tyr49*I*) all lie on the surface with a large area accessible to solvent of 78, 107, 130, 65, 80, 109, 130 and 123 Å^2 respectively, calculated as for Fig. 7(*c*).

Homology in the *Bacilli* subtilisins

The sequences for the six *Bacilli* subtilisins SBMEP, SBI168, SBAS, SBPN, SBDY and SBCARL are aligned in Fig. 8, and the corresponding amino-acid compositions are shown in Table 4. They are numbered according to that of SBPN as is conventional, and are given in full to clear up some confusions from the published sequence data and the coordinates in the Brookhaven Protein Data Bank. The sequences given for BI168 (Stahl & Ferrari, 1984), SBAS (Yoshimoto *et al.*, 1988), SBPN (Wells, Ferrari, Henner, Estell & Chen, 1983; Vasantha, Thompson, Rhodes, Banner, Nagle & Filpula, 1984) and SBCARL (Jacobs, Eliasson, Uhlen & Flock, 1985) are those for the corresponding genes. These differ somewhat from those reported for direct sequencing of the protein for SBAS (Kurihara *et al.*, 1972), SBPN (Markland & Smith, 1967) and SBCARL (Smith *et al.*, 1968); the direct sequence has not been published for SBI168. [The situation is

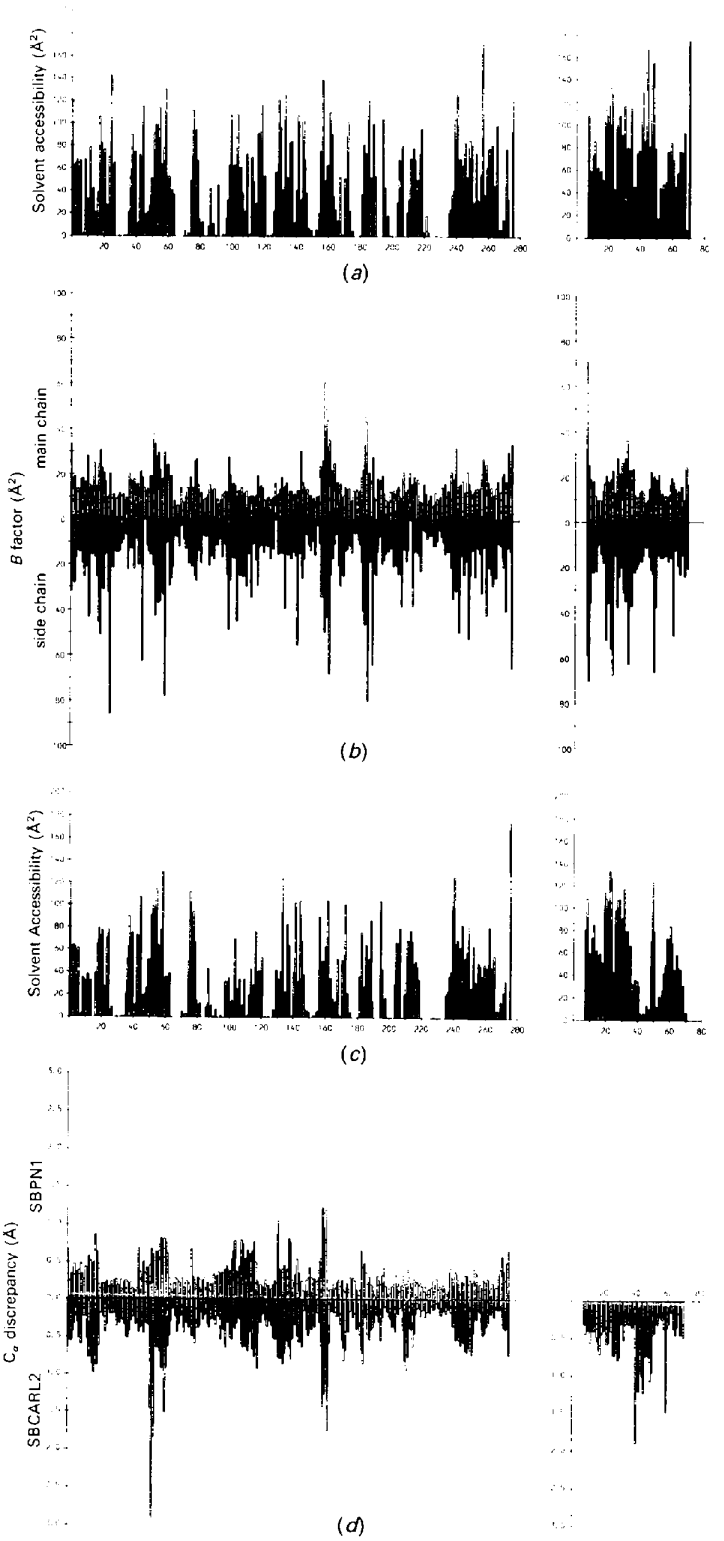
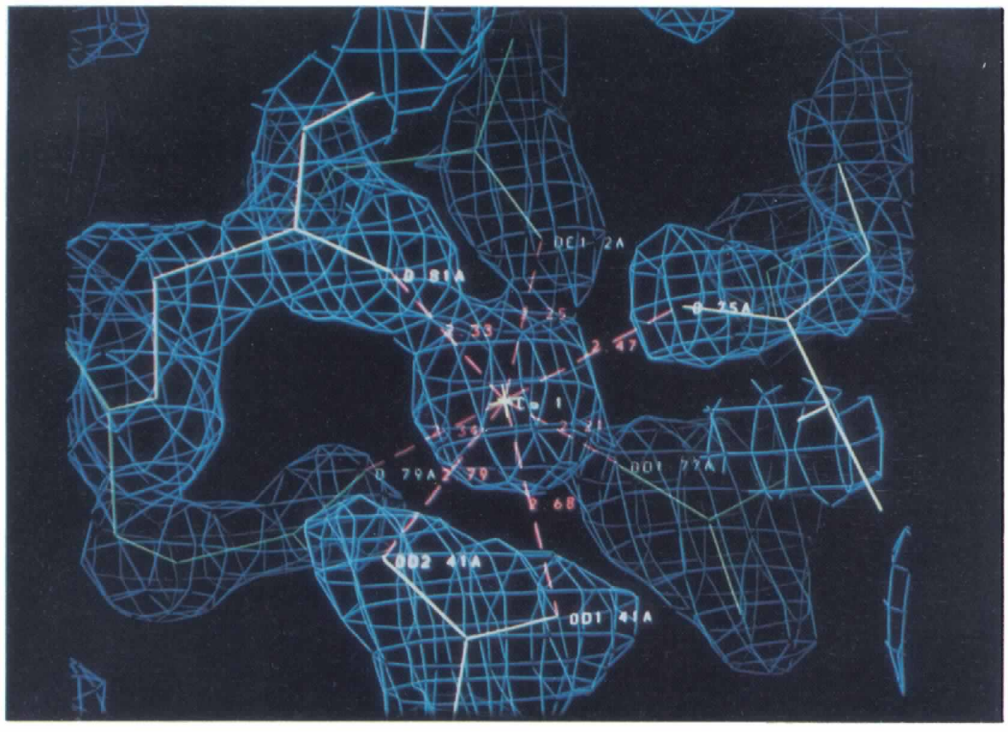
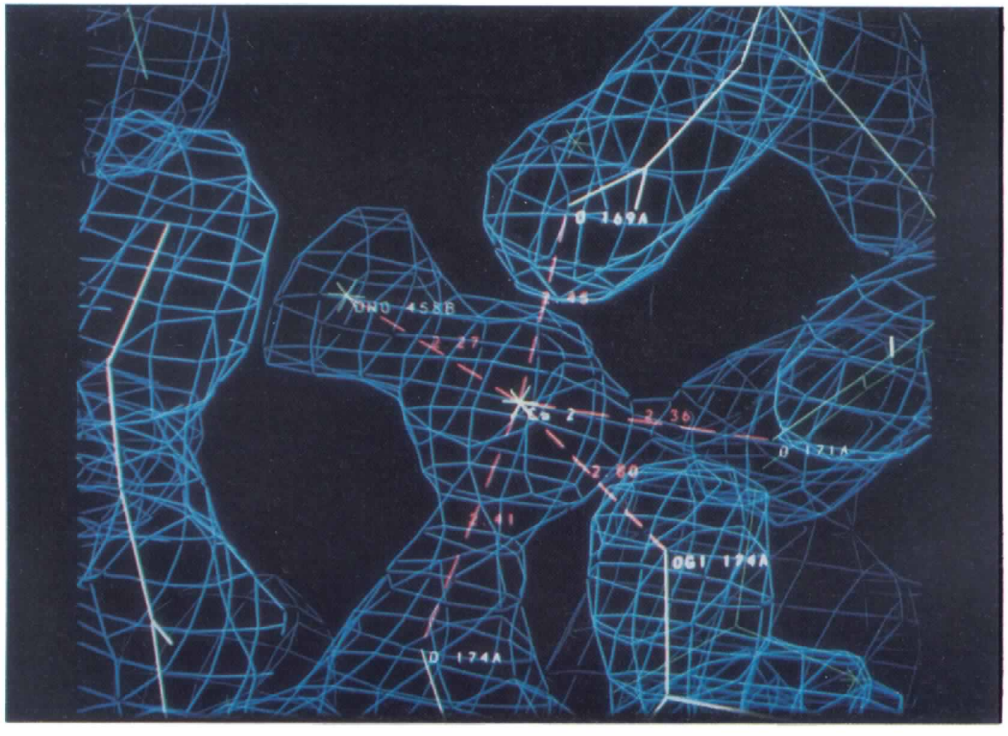


Fig. 7. (*a*) The accessible surface area for 'native' SBMEP and 'native' eglin-C, each in the absence of the other and omitting contacts with symmetry-related molecules in the crystal, as a function of the residue number. (*b*) The average *B* value for the main- and for the side-chain atoms of individual residues for SBMEP and eglin-C. (*c*) The accessible area for SBMEP and eglin-C, accounting for all contacts within the crystal. (*d*) The deviation between the refined C_α positions of SBPN overlapped on SBMEP, and SBCARL on SBMEP: the eglin molecules in the SBMEP and SBCARL complexes were overlapped independently of the enzymes.



(*a*)



(*b*)

Fig. 6. Two regions of the final $(3F_o-2F_c)$ electron density map: (*a*) the environment of the strongly bound calcium ion, Ca1, and (*b*) the environment of the Ca2 ion. The atoms are labelled by their name, residue number and chain identifier (*A* for enzyme, *B* for waters).

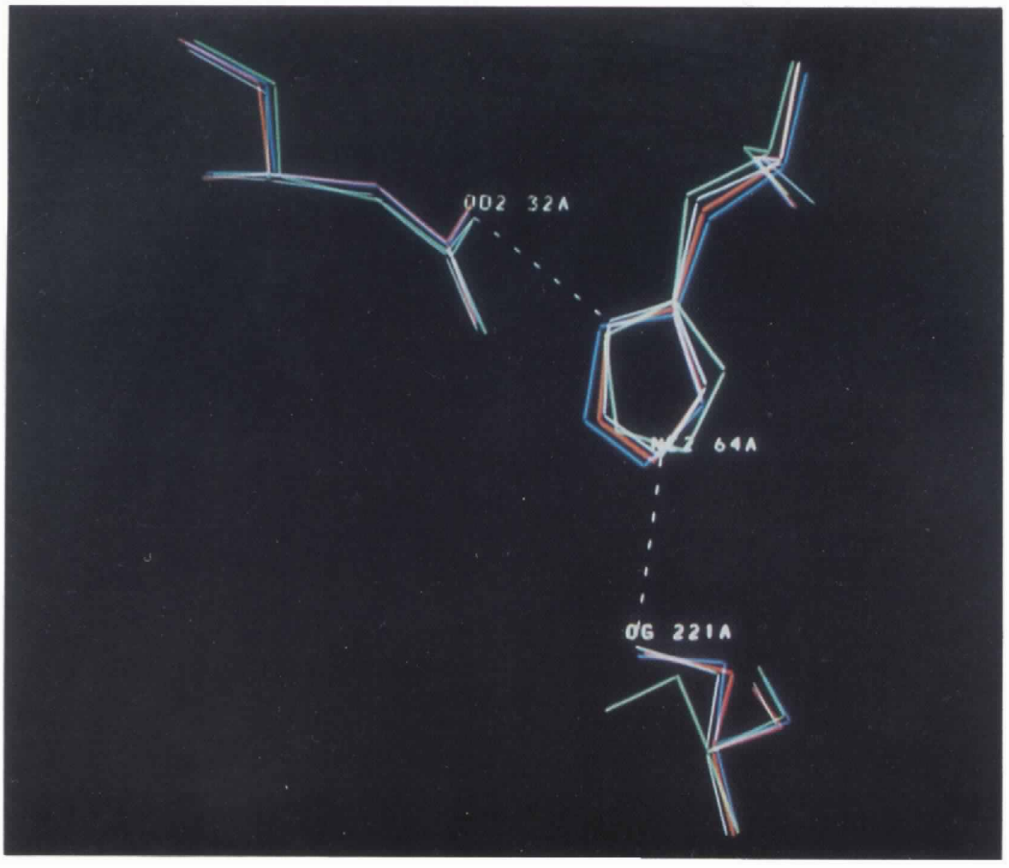


Fig. 12. Superposition of the catalytic triad of residues for SBMEP (white), SBPN1 (red), SBPN2 (green) and SBCARL (blue).

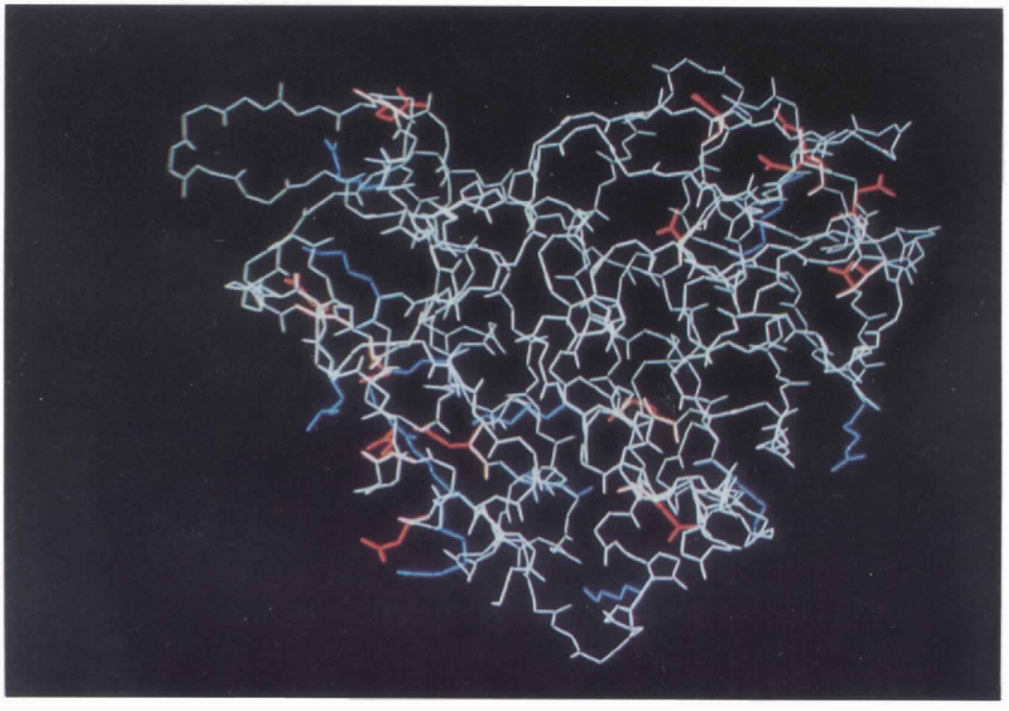


Fig. 13. The distribution of charged residues on the SBMEP molecule. The side chains of arginine and lysine residues are in blue, and those of glutamic and aspartic acids in red.

further confused by the extra four errors introduced in quoting the original protein sequence of SBAS by Yoshimoto *et al.* (1988)]. The gene sequences are generally believed to be more accurate, and these are used in Fig. 8 and in all discussions. Recent X-ray studies of SBPN (*e.g.* Bott *et al.*, 1988) use the published gene sequence (Wells *et al.*, 1983). However the coordinates for the two high-resolution studies of the SBCARL–eglin-C complex are based on the original protein sequence (Smith *et al.*, 1968) (probably the only available at the time of these studies) and hence differ from the gene sequence at five positions. It would perhaps be of interest to reinspect the three-dimensional structures in terms of the gene sequence for this protein. For SBDY only the protein sequence has been determined and this is used as published (Nedkov *et al.*, 1983).

For SBMEP again only the direct protein sequence is available (Svendsen *et al.*, 1986). This sequence was used in the present X-ray analysis with one change. During the refinement of the X-ray structure unequivocal evidence emerged from the electron density indicating the true identity of two residues as Ala88-Ser89 rather than the inverted dipeptide. Thus they were changed to Ala88-Ser89 in Fig. 8, in our final coordinate set and in all discussions. This order also agrees with the gene sequence data for the closely related SBI168 and SBAS.

The six sequences are highly homologous. There are no deletions or insertions between SBMEP, SBI168, SBAS and SBPN, and only a single deletion at the same position, 56, is required in aligning these four with the closely related pair SBDY and SBCARL. The number of differences between the sequences is shown in Table 5. The three proteins SBMEP, SBI168 and SBAS form a very closely related subgroup, with sequence identity approaching 100% and only two differences between the first two proteins. In addition, all three of these come from mesophilic *Bacilli*. Unfortunately, with the

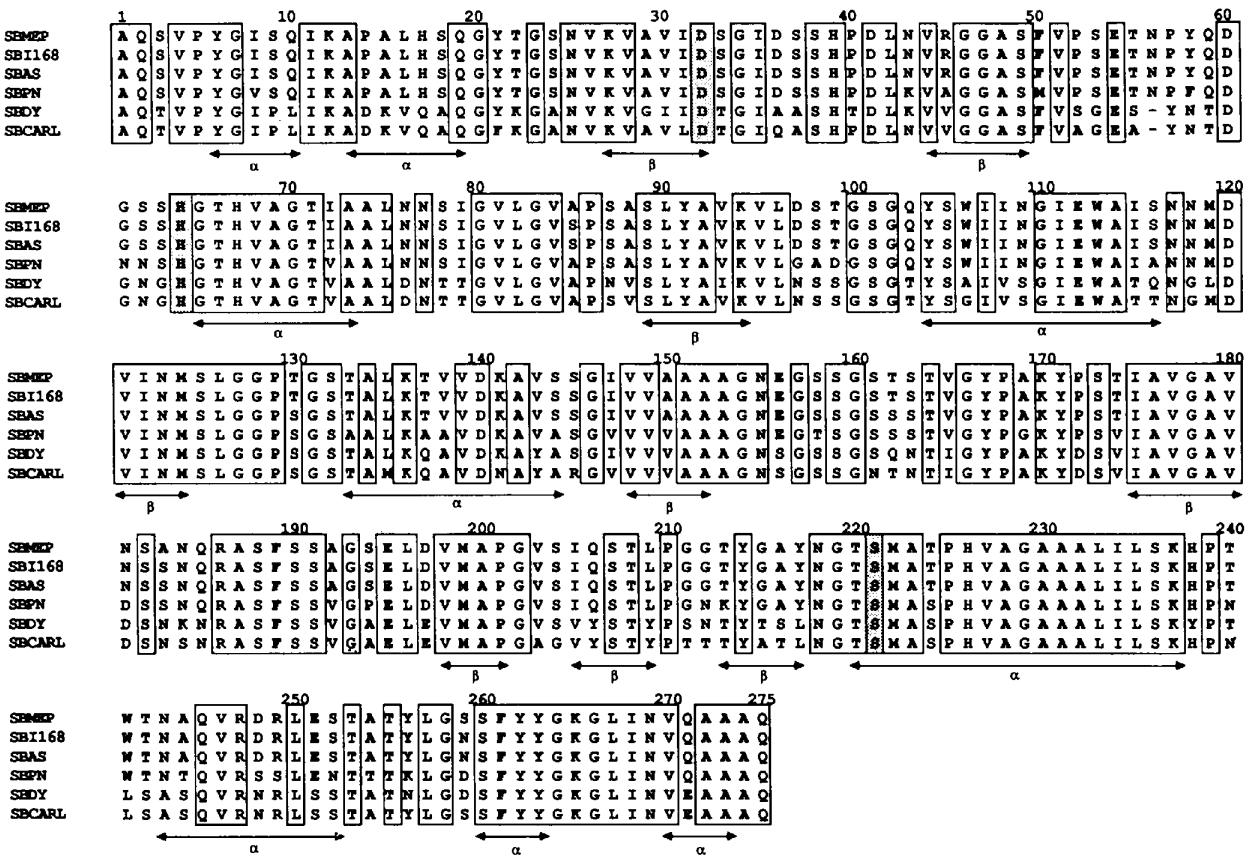


Fig. 8. The sequences of those members of the *Bacillus* subtilisin family given in Table 1. The sequence of SBMEP is amended from the published protein sequence by the inversion of the order of a pair of residues at positions 88 and 89 based on the X-ray analysis, as described in the text. The sequences of SBAS, SBPN and SBCARL are all from the published gene sequences rather than the original protein sequence determinations. This means that the SBCARL sequence differs in five positions from that used in previously published X-ray analyses. The sequence of SBDY is taken from the published protein sequence. Residues identical in all sequences are boxed, and the members of the catalytic triad are shaded.

Tables 4. *Amino-acid composition*

The amino-acid composition deduced from the sequences of the *Bacilli* subtilisin proteinases shown in Fig. 8. The nomenclature for the species is given in Table 1. The overall charge for the protein chain is also shown.

	SBMEP	SBI168	SBAS	SBPN	SBDY	SBCARL
Ala	36	34	34	37	37	40
Arg	4	4	4	2	3	4
Asn	15	16	16	18	17	18
Asp	9	9	9	10	10	9
Cys	0	0	0	0	0	0
Gln	10	10	10	10	8	7
Glu	5	5	5	5	5	5
Gly	33	33	33	33	34	35
His	6	6	6	6	4	5
Ile	16	16	16	13	14	10
Leu	15	15	15	15	17	16
Lys	8	8	8	11	12	9
Met	4	4	4	5	3	5
Phe	3	3	3	3	3	4
Pro	13	13	13	14	9	10
Ser	39	40	42	37	35	32
Thr	19	19	17	13	19	20
Trp	3	3	3	3	1	1
Tyr	12	12	12	10	14	13
Val	25	25	25	30	29	31
Asp + Glu	14	14	14	15	15	14
Arg + Lys	12	12	12	13	15	13
Charge	−2	−2	−2	−2	0	−1

Table 5. *Sequence homology among the subtilisin proteinases*

The number of substituted residues between the sequences of the *Bacilli* subtilisin proteinases as given in Table 1. The numbers in brackets are the deletions and insertions required to align the sequences.

	SBMEP	SBI168	SBAS	SBPN	SBDY	SBCARL
SBMEP	–	3(–)	5( )	38 (–)	81 (1)	80 (1)
SBI168		–	2(–)	38 ( )	82 (1)	82 (1)
SBBAS			–	36 (–)	81 (1)	82 (1)
SBPN				–	81 (1)	83 (1)
SBDY					–	31 (0)
SBCARL						–

Table 6. *Temperature factors* (Å^2) *for the refined structure*

	Average *B*	No. of atoms
Overall	20·8	2792
SBMEP		
Total	17·7	1948
Main chain	15·2	1100
Side chains	21·0	848
Eglin-C		
Total	21·4	530
Main chain	16·9	256
Side chains	25·6	274
Solvent	39·2	312
Calcium ions	17·8	2

exception of the gene sequence (Yoshimoto *et al.*, 1988) SBAS appears not to have been studied for many years now, and few data have been reported on the properties of SBI168. In this situation SBMEP must be taken as the representative member of this mesostable subgroup. SBPN differs by about 40 substitutions, and the SBDY/SBCARL pair by about 80 substitutions from this subgroup. The SBDY/SBCARL pair differ from one another at only 31 positions.

The present X-ray analysis of SBMEP means that a structure is now available for all three subfamilies of *Bacilli* subtilisins: SBMEP, SBPN and SBCARL. The sequence and compositional differences will be futher discussed with relation to the stucture and function of the enzymes in some of the following sections.

Description of the overall structure

The overall fold of the molecule is shown schematically in Fig. 9. The molecule resembles a hemisphere with a crevice on the flat surface making up the active site. SBMEP is an α/β protein and its folding follows closely that described originally for SBPN (Wright *et al.*, 1969), and for SBCARL (McPhalen & James, 1988; Bode *et al.*, 1987), and is not further discussed in detail here. The hydrogen-bonding scheme for the backbone of SBMEP is drawn schematically in Fig. 10. The core of the molecule is an eight-stranded parallel β-pleated sheet. There are nine α-helices distributed on either side of the sheet, and a second smaller anti-parallel β-sheet with only two strands on the surface of the molecule.

In discussing the binding sites of subtilisins it is customary to use the nomenclature of Schechter & Berger (1967). In brief, the substrate amino-acid residues are numbered from the scissile peptide bond as *P*1, *P*2 *etc.* towards the N terminus, and as *P*1′, *P*2′ *etc.* towards the C terminus. The corresponding binding sites on the surface of the enzyme are label-

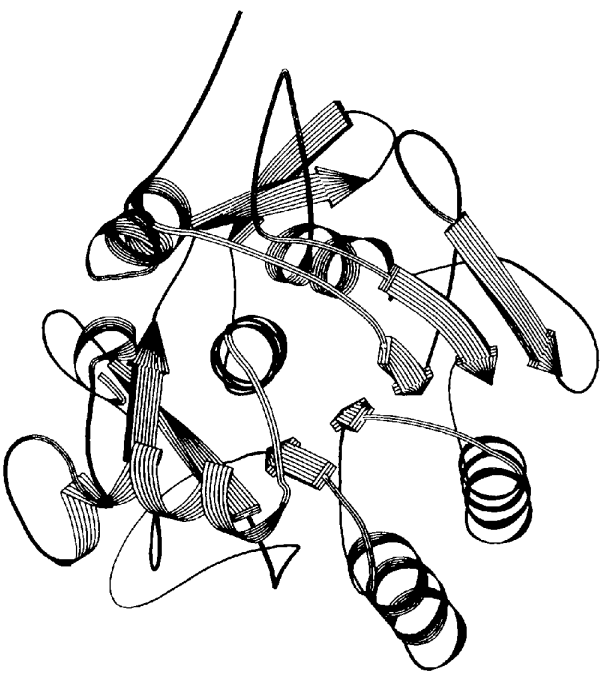


Fig. 9. A ribbon diagram (Priestle, 1988) showing the overall fold of the subtilisin molecule, α-helices are shown as cylindrical coils and β-strands as arrows.

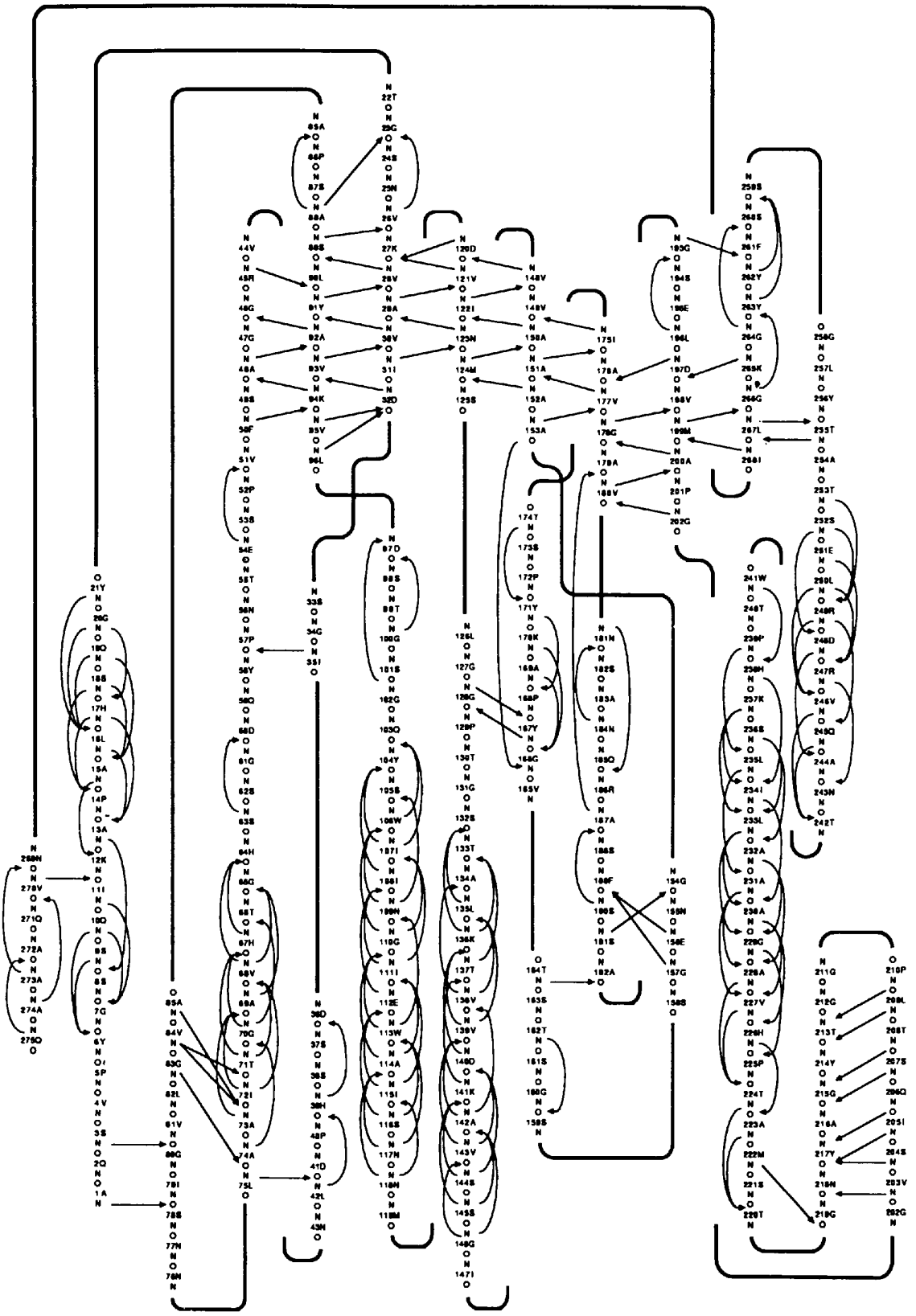


Fig. 10. The hydrogen-bonding scheme involving main-chain atoms for SBMEP.

Table 7. *Structural homology among the subtilisin proteinases*

The deviation (Å) in C_α positions for the superimposed models of SBMEP, SBPN and SBCARL. The superposition involved all C_α atoms excepting residues 52–59 in the loop surrounding the deletion at position 55 in SBCARL. The upper number is the mean (not r.m.s.) deviation and the lower the maximum. SBPN1 coordinates are taken from McPhalen & James (1988), SBPN2 from Bott *et al.* (1988), SBCARL1 from McPhalen & James (1988), SBCARL2 from Bode *et al.* (1987) and SBCARL3 from Neidhart & Petsko (1988).

	SBPN1	SBPN2	SBCARL1	SBCARL2	SBCARL3
SBMEP	0·30	0·28	0·36	0·38	0·45
	1·26	1·75	1·74	1·77	2·25
SBPN1		0·24	0·40	0·41	0·43
		1·14	2·06	2·44	1·53
SBPN2			0·38	0·40	0·44
			1·77	2·16	1·63
SBCARL1				0·11	0·37
				0·41	3·04
SBCARL2					0·39
					3·43

led as *S*1, *S*2 *etc.* and as *S*1′, *S*2′ *etc.* respectively. This convention is adhered to throughout the paper.

The refined coordinates of SBMEP, SBPN and SBCARL were superimposed using least-squares minimization of the deviation between the C_α atoms. The results of the comparison are shown in Table 7. The structures superimposed on the presently determined SBMEP coordinates from the eglin complex are as follows. SBPN1 are the SBPN coordinates from the complex with chymotrypsin inhibitor (McPhalen & James, 1988) and SBPN2 are for the PMSF-inhibited enzyme (Bott *et al.*, 1988). There are three SBCARL coordinate sets. SBCARL1 is from the complex with eglin-C (McPhalen & James, 1988), SBCARL2 from the same complex studied by Bode *et al.* (1987), and finally SBCARL3 is for the native enzyme (Neidhardt & Petsko, 1988). The loop of residues from position 52 to 59 which contains the deletion in SBCARL and SBDY was omitted from the superposition: all the remaining 266 C_α atoms were included. The surface area accessible to the solvent for SBMEP isolated from eglin and its crystalline environment is 10 073 Å^2. The corresponding values for SBPN and SBCARL are 10 071 and 9940 Å^2 respectively.

Several conclusions can be drawn from the comparisons.

(1) The base line for the comparison is the deviation between the two independent determinations of the structure of the SBCARL–eglin-C complex. The mean value of the deviation is 0·11 Å and the maximum 0·41 Å. This should be compared with the mean expected error in the coordinates estimated as described by Read (1986) of 0·20 Å.

(2) The SBCARL3 set (Neidhardt & Petsko, 1988) was derived from SBPN as a starting model. It shows a mean deviation from the refined models of the SBCARL–eglin-C complexes almost as large as that from the refined SBPN structures (*i.e.* the starting model), and a maximum deviation which is indeed greater. This illustrates the difficulty in obtaining accurate structures with certainty at medium resolution. The SBCARL3 coordinates are excluded from the rest of the discussion.

(3) The overall structures of SBMEP and the two structures of SBPN (one non-covalently complexed with a peptide inhibitor and one covalently inhibited with PMSF) are almost as similar as those of the two SBPN structures themselves.

(4) The differences between the superimposed C_α positions of SBMEP and of SBPN and SBCARL are plotted as a function of residue number in Fig. 7(*d*). The eglin molecules in the SBMEP and SBCARL complexes were superimposed independently of the enzymes and this comparison is included in the figure. The residues of the 'deletion loop', 52–59, were included in the superposition used to generate the figure, other than the deleted residue itself. There are no major differences between the tertiary structures and clearly the overall folds of the three proteins are very similar. There is a definite correlation between the deviations of SBPN1 and SBCARL2 from SBMEP. The only residues for which the C_α positions differ by more than 1 Å between SBMEP and SBPN1 are 130 (Thr in SBMEP, Ser in SBPN and SBCARL), and residues 159–161 (Ser-Gly-Ser in SBMEP and SBPN, Ser-Gly-Asn in SBCARL). The comparable deviations between SBMEP and SBCARL2 are at positions 52–54 and 61, just before and after the deleted residue 55, residues 159–160 and 162 (Thr in all three proteins). All lie on the surface, residue 130 in a loop between a helix and a sheet, residues 159–162 form a β-turn. Each of these regions has high temperature factors in the SBMEP structure, Fig. 7(*b*). Residue 130 is involved in contacts with eglin as discussed below.

Eglin

High-resolution X-ray structures are now known for eglin-C complexed to SBCARL (McPhalen & James, 1988; Bode *et al.*, 1987), two complexes of eglin-C with thermitase (Gros, Betzel *et al.*, 1989; Gros, Fujinaga *et al.*, 1989) and the present complex with SBMEP. The binding mode of eglin-C to subtilisins has been discussed in detail in the above mentioned publications and was found to be comparable in the present SBMEP complex. In brief, eglin-C has a compact hydrophobic core made up of side chains from both parallel β-strands and the α-helix. The binding loop lies in an extended conformation between two sharp turns Pro38*I*-Ser41*I* and Arg48*I*-Arg51*I* and projects away like a handle from the body of the molecule (Bode *et al.*, 1987). The connections between the central part of the binding loop

and the body of the molecule are relatively weak. In the present analysis, as in the previous studies of the SBCARL and thermitase–eglin-C complexes, the residues of the binding loop are generally in well defined electron density with low temperature factors.

In the X-ray analyses of the SBCARL–eglin-C complex (McPhalen & James, 1988; Bode *et al.*, 1987) there was no electron density for residues Thr1*I* to Leu7*I* and the final model only included residues Lys8*I* to Gly70*I*. It was postulated that the N-terminal peptide was cleaved from the inhibitor by residual amounts of active SBCARL in the crystallization solution. It was suggested that the most probable cleavage point was between residues Glu6*I* and Leu7*I*: however no density was seen for Leu7*I*. In the refined structures of the two thermitase–eglin-C complexes the inhibitor was again only modelled from residue Lys8*I* onwards. In the present analysis clear, albeit weak, density was visible for Leu7*I* in $(3F_o - 2F_c)$ or $(F_o - F_c)$ Fourier syntheses. However the refined *B* values for the atoms of Leu7*I* are high, suggesting only partial occupancy or, indeed, large thermal vibration. McPhalen & James (1988) also include residue Leu7*I* in their model for eglin, with rather high *B* values around 50.

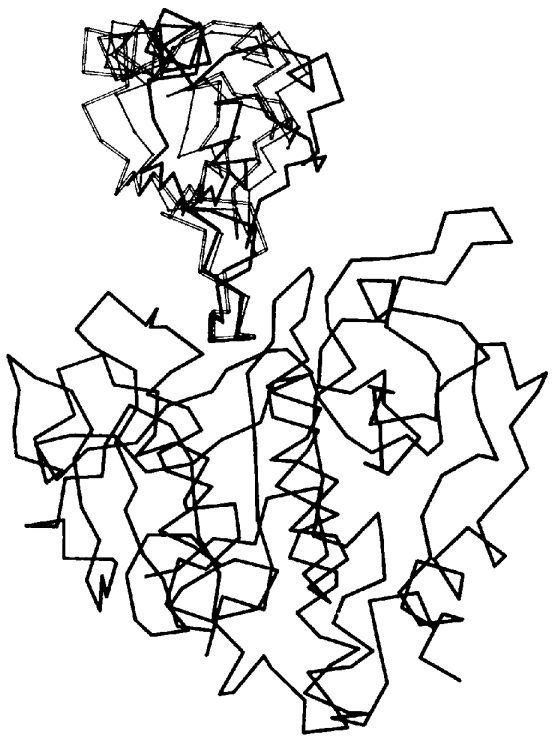


Fig. 11. The C_α backbone of the SBMEP–eglin-C complex showing the position of the inhibitor in the active site (thick lines). The eglin-C moieties from the complexes with SBCARL (open lines) and with thermitase (thin lines) are also shown after superposition of the subtilisin molecules of the latter complexes onto SBMEP.

We have compared the structures from the SBMEP complex, the type-II thermitase complex (Gros, Betzel *et al.*, 1989) and the SBCARL complex (McPhalen & James, 1988) using least-squares superposition of the C_α coordinates. Firstly, the overlap was carried out superimposing only the subtilisin molecules. The structure of the complex with SBMEP is shown in Fig. 11, with the relative positions of the eglin-C in the other two complexes superimposed. In all three complexes the eglin-C adopts a different orientation with respect to the subtilisin.

The eglin molecules were then superimposed using their C_α positions only and this revealed the following differences in the eglin-C orientation in the three pairs of complexes: SBCARL to SBMEP 16°, thermitase to SBMEP 11° and thermitase to SBCARL 10°, with the difference expressed as a single rotation. The resulting overlaps give the following statistics. For the eglin from SBMEP superimposed on that from SBCARL the mean C_α discrepancy is 0·53 Å and the maximum 1·90 Å. For the eglin of thermitase complex overlapped on SBMEP eglin the values are 0·33 and 2·41 Å respectively. The internal conformation of the eglin in each of the complexes is basically the same. Almost all the discrepancies greater than 1·9 Å lie within the binding loop, residues 40*I* to 50*I*. This confirms the motion of the loop relative to the relatively rigid core of the eglin-C in complex formation.

Closer inspection of the superimposed eglin-C molecules in SBMEP and SBCARL revealed some minor differences in conformation. There are two places in the chain where the peptide planes are rotated by about 180° in each case retaining the trans configuration: Glu39*I*-Gly40*I* have (φ,ψ) conformation angles of (−54°, −34°) and (−84°, 52°) in the SBMEP complex and (−56°, 139°) and (94°, −18°) in the SBCARL complex; Pro58*I*-Gly59*I* have (φ,ψ) angles of (−60°, −28°) and (−75°, −35°) in SBMEP and (−65°, 143°) and (87°, −37°) in SBCARL, respectively. These peptides lie in turns on the surface of the eglin-C molecules, and in the present structure are in well defined electron density. Several of the longer side chains lying on the surface of eglin adopt different conformations in the two complexes: Arg22*I*, Glu23*I*, His28*I*, Tyr49*I*, Asn61*I* and His68*I*. In the SBMEP complex the electron density gives unequivocal positions for all of these excepting Tyr49*I*.

In the present structure, as in the other complexes, the eglin core supports the conformation of the binding loop by an extensive network of salt bridges and hydrogen bonds. The salt bridges connect N*E* of Arg48*I* with O*D*1 of Asp46*I* (2·82 Å), O*D*2 of

Table 8. *Number of interatomic contacts less than* 4 Å *between SBMEP and SBCARL* (*Bode et al.*, 1987) *and eglin-C*

MEP	CARL	Tyr 35I	Phe 36I	Leu 37I	Pro 38I	Glu 39I	Gly 40I *P6*	Ser 41I *P5*	Pro 42I *P4*	Val 43I *P3*	Thr 44I *P2*	Leu 45I *P1*	Asp 46I *P1′*	Leu 47I *P2′*	Arg 48I *P3′*	Tyr 49I	Asn 50I	Arg 51I	Val 52I	Arg 53I	Total
Ser33	Thr										0,1										0,1
Ser62	Asn														3,8						3,8
Ser63	Gly														2,0						2,0
His65											10,9	1,1	6,5		0,2						17,17
Leu96											1,1										1,1
Thr99	Ser	0,3																		0,1	0,4
Gly100										3,3	7,6									0,3	10,12
Ser101				1,1					3,3	1,1											5,5
Gly102								1,0	6,4												7,4
Gln103	Thr								4,0												4,0
Tyr104							5,3		6,1												11,4
Ile107									2,2												2,2
Ser125											0,1	1,1									1,2
Leu126										3,3		3,3									6,6
Gly127								1,0	3,3	7,6		1,2									12,11
Gly128							0,2	1,1													1,3
Pro129	Ala					3,0	0,4														3,4
Thr130	Ser					9,0	3,6														12,6
Ala152												2,2									2,2
Gly154												1,2									1,2
Asn155												6,6	1,3	1,0							8,9
Phe189		1,0												6,3							7,3
Leu209	Tyr														0,3						0,3
Tyr217	Leu														2,1						2,1
Asn218													3,3	4,4		3,0					10,7
Gly219												2,2									2,2
Thr220												1,3									1,3
Ser221												10,11	2,2								12,13
Met222													1,0								1,0
Total		1,3		1,1		12,0	8,15	3,1	24,13	14,13	18,18	28,33	13,13	11,7	7,14	3,0				0,4	143,135

Asp46*I* with NH2 or Arg51*I* (3·12 Å) and the latter atom with the C-terminal carboxyl O*E* of Gly70*I* (2·81 Å). Hydrogen bonds connect the following atoms: NH1 of Arg53 with O of Thr44*I*, NH2 of Arg53 with O*G*1 of the same Thr44*I*, and the conserved water molecule Wat566 with O*G*1 of Thr44*I*, OD2 of Asp46*I* and NH1 or Arg51*I* (268, 2·83 and 3·03 Å, respectively).

Inspection of the superpositions shows that only part of the binding loop of eglin-C, residues 43*I* to 47*I*, adopt the same position with respect to subtilisin in all the complexes. The bulk of the eglin-C moiety extends away from the binding loop and is connected to it through two 'hinges' on the ends of the binding loop.

It has been suggested that the different relative orientation of eglin and subtilisin in the complexes results from the intermolecular packing contacts in the crystal (Gros, Betzel *et al.*, 1989). However comparison of eglin-C–SBMEP with eglin-C–SBCARL allows some rationalization of the difference in orientation in these complexes in terms of the eglin–subtilisin interface. Among the proteinase residues making a contact with the inhibitor of less than 4 Å, Table 8, there are eight substitutions between SBCARL and SBMEP: Thr33 by Ser33, Asn62 by Ser62, Gly63 by Ser63, Ser99 by Thr99, Thr103 by Gln103, Ser130 by Thr130, Tyr209 by Leu209 and Leu217 by Tyr217. For residue 103 only main-chain atoms are involved in intermolecular van der Waals contacts, and these are not influenced by the substitution. Substituting Ser130 in the SBCARL complex by Thr130 results in unacceptably close contacts, less than 3·0 Å, between this side chain and main chain of eglin-C residues Glu39*I* and Gly40*I*. Substitution of SBCARL Ser156 by SBMEP Glu156 allows the formation of a hydrogen bond between the carboxyl group and the N*E*2 atom of His68*I*. Residue Asn62 is not in contact with eglin-C in the SBCARL complex, but the corresponding SBMEP Ser62 is in weak (about 3·5 Å) hydrogen-bonding contact with Arg48*I*. In conclusion there are several contacts in the eglin–subtilisin interface which are affected by substitutions between SBCARL and SBMEP, and the eglin can rotate about its hinges to accommodate these changes while maintaining the binding loop in the active site in a favourable position.

In addition to the above differences in the eglin-C binding site, there is the question of residue 129 in SBCARL. This is present as Ala129 in the crystal models, but is Pro129 according to the gene sequence, as it is in all the other *Bacilli* subtilisins.

As in other complexes, the O*G*1 atom of Ser221 is in the correct position for nucleophilic attack on the scissile peptide bond with a distance 2·85 Å from the Leu45*I* C atom. This is in spite of the fact that the target distance in the restrained refinement was 3·4 Å for this pair of atoms. No significant distortion of the Leu45*I*-Asp46*I* peptide geometry is observed in the present analysis.

Table 9. *Active sites of the subtilisin proteinases*

The deviation (Å) in atomic positions for the superimposed active sites of SBMEP, SBPN and SBCARL: the upper number is the mean (not r.m.s.) and the lower the maximum deviation. All atoms in the residues were used in the superposition. The top right triangle gives the results for the three catalytic residues only (Asp32, His64 and Ser221). The lower left triangle also includes those residues interacting with substrate sites *P*5 to *P*2′ in the eglin complexes and conserved in all the sequences. The coordinates are as for Table 7, with SBCARL3 excluded.

	SBMEP	SBPN1	SBPN2	SBCARL1	SBCARL2
SBMEP		0·16	0·27	0·22	0·19
		0·29	1·02	0·42	0·36
SBPN1	0·43		0·30	0·13	0·12
	2·86		1·00	0·24	0·23
SBPN2	0·46	0·39		0·32	0·31
	2·83	2·24		0·94	0·90
SBCARL1	0·44	0·37	0·44		0·09
	3·24	2·93	3·66		0·16
SBCARL2	0·32	0·43	0·48	0·23	
	1·82	2·55	2·88	2·44	

Catalytic and specificity sites

The refined coordinates of the three catalytic residues (Asp32, His64 and Ser221) of SBMEP, SBCARL and SBPN were superimposed independently of the rest of the molecule, and the results are shown in the top right-hand triangle of Table 9. As might be expected the active site is highly conserved between all three enzymes with mean discrepancies of less than 0·35 Å in the five structures. The catalytic triad of residues is shown superimposed in Fig. 12. Several conclusions can be reached.

(1) The SBCARL1 and SBCARL2 pair again provide a base line for the comparison. The two independent determinations of the same structure result in coordinates for the catalytic residues which superimpose with an average difference of only 0·09 Å. This is smaller than the expected errors estimated from the σ_A plot of Read (1986).

(2) The SBPN1 coordinates for the chymotrypsin inhibitor complex of SBPN are closely similar to those of SBMEP, with a mean difference of only 0·16 Å and a maximum difference of 0·29 Å.

(3) In contrast the SBPN2 coordinates, for the PMSF inhibited SBPN structure, have maximum deviations of around 0·9–1·0 Å from each of the other four structures. The maximum differences can be seen from Fig. 12 to be the positions of the C*B* and O*G* atom of the catalytic serine, and the N*E*2 of His64 in contact with the serine. The O*G* atom can no longer form a hydrogen bond to the N*E*2 of His64 in the PMSF inhibited structure, and this readily explains the differences observed. It is of note that the differences in proteinase conformation on forming the covalent PMSF complex are small and largely restricted to these two residues.

In the bottom left triangle of Table 9 the results of superimposing the atoms making up the catalytic and substrate binding sites are shown. All the residues listed as making a contact less than 4 Å with sites *P*6 through *P*3′ of eglin-C in Table 8, and conserved in all the sequences in Fig. 8, were included in the comparison. These are His64, Leu96, Gly100, Ser101, Gly102, Tyr104, Ile107, Ser125, Leu126, Gly127, Gly128, Ala152, Gly154, Asn155, Phe189, Asn218, Gly219, Thr220 and Ser221. Again it is clear from the low values of the mean deviations that the active and substrate binding sites are highly conserved. Residues rejected from the comparison are described in the previous section on eglin-C.

In the structure of native SBMEP, 'extracted' from that of its complex with eglin-C, the catalytic triad of residues have surface areas accessible to solvent of 0 (Asp32), 37 (His64) and 18 Å^2 (Ser221). Thus Asp32 is completely buried as is found in other subtilisins. This relatively hydrophobic environment is believed to make it raise the p*K* value of the adjacent histidine while holding it in the conformation required for catalysis as in chymotrypsin (Fersht & Sperling, 1973). The recent model for the mechanism of serine proteinases involves the stabilization of the charge on the His64 side chain in the transition state (Warshel, Naray-Szabo, Sussman & Hwang, 1989). Furthermore in the eglin-C inhibited complex the accessible areas drop to zero for His64 and Ser221. With the binding of inhibitor, which can be viewed as a substrate analogue, all of the catalytic triad lie in a non-aqueous local environment.

The charged residues of SBMEP

The number of charged residues in SBMEP, SBPN and SBCARL is shown in Table 4. The number of charged groups is very similar in all six *Bacilli* proteinases. Their spatial distribution in SBMEP is shown in Fig. 13. As usual, the charged residues are almost all located on the surface of the molecule. As with SBCARL and SBPN the distribution is distinctly asymmetric with the basic side chains tending to lie on the opposite face of the molecule from the active site. The accessible surface area of the charged residues is given in Table 10 and is discussed in relation to those involved in salt bridges or Ca^{2+} binding and, in a later section, to chemical properties.

Ca^{2+} *sites*

There are two Ca^{2+} ion binding sites in each of the *Bacilli* subtilisin molecules. The first ion, Ca1, is tightly bound and cannot be completely removed with EDTA under non-denaturing conditions, whilst the second, Ca2, is more loosely bound and can be removed with EDTA (Voordouw *et al.*, 1976; Pantoliano *et al.*, 1988). The effective overall negative charge of the subtilisins given in Table 4 is changed by +4 by the binding of the two calcium

Table 10. *Surface area accessible to solvent (SA in Å²) for the residues of SBMEP with charged side chains; involvement in salt bridges (SB) and, for the carboxyl groups, the number in Ca^{2+} binding, is also indicated*

Positively charged groups			Negatively charged groups			
Residue	SA	SB	Residue	SA	SB	Ca
Arg45	116	–	Asp32	0	–	
Arg186	51	–	Asp36	22		–
Arg247	54	2	Asp41	0	–	2
Arg249	80	1	Asp60	11	–	
Lys12	80	–	Asp97	32	–	–
Lys27	65	–	Asp120	53		–
Lys94	1	1	Asp140	41	1	–
Lys136	83	1	Asp197	18	1	
Lys141	108	–	Asp248	85	1	–
Lys170	49	1	Glu54	39	1	–
Lys237	39	–	Glu112	71	–	–
Lys265	99		Glu156	140		
			Glu195	43	1	–
			Glu251	30	1	

Table 11. *Calcium binding ligands*

The table gives the ligand distances (Å) for the Ca^{2+} ions in SBMEP, and in SBPN1 (McPhalen & James, 1988) and SBCARL2 (Bode *et al.*, 1987) for comparison.

SBMEP			SBPN1			SBCARL2		
Ca1 (high-affinity site)								
Gln2	*OE1*	2·24	Gln2	*OE1*	2·37	Gln2	*OE1*	2·40
Asp41	*OD1*	2·68	Asp41	*OD1*	2·37	Asp41	*OD1*	2·47
Asp41	*OD2*	2·79	Asp41	*OD2*	2·60	Asp41	*OD2*	2·61
Leu75	O	2·46	Leu75	O	2·33	Leu75	O	2·38
Asn77	*OD1*	2·21	Asn77	*OD1*	2·46	Asn77	*OD1*	2·39
Ile79	O	2·33	Ile79	O	2·29	Thr79	O	2·36
Val81	O	2·33	Val81	O	2·29	Val81	O	2·41
Ca2 (low-affinity site)								
Ala169	O	2·45	Gly169	O	2·77	Ala169	O	2·39
Tyr171	O	2·36	Tyr171	O	2·96	Tyr171	O	2·48
Thr174	O	2·41	Val174	O	2·74	Val174	O	2·33
Thr174	*OG1*	2·78	Asp197	*OD2*	2·82	Wat476	O	2·42
Wat458	O	2·27	Glu195	O	3·03	Wat402	O	2·46
			Wat365	O	3·12			
			Wat369	O	3·02			

ions, providing both sites are fully occupied. Both Ca^{2+} ions were located in the SBMEP crystal structure. The chelating groups and the ligand to Ca^{2+} distances are given in Table 11. There are similar Ca^{2+} binding sites at equivalent positions in SBCARL and SBPN and the data for these sites are included in the table for comparison.

The Ca1 site is well defined in SBMEP with seven ligands in an approximately octahedral arrangement, Fig. 6(*a*). Asp41, the only negatively charged group in SBMEP in a first chelation shell of Ca^{2+}, has both of its OD atoms occupying one of the octahedral positions. In SBPN and SBCARL the ligands for the Ca1 ion are essentially identical, and the geometries of the sites are closely similar. The single difference lies in the substitution of Ile79 in SBMEP and SBPN by Thr79 in SBCARL; residue 79 coordinates through its carbonyl oxygen. The variation in bond lengths observed reflects the estimated errors in atomic positions of about 0·2 Å.

There is much more variation in the coordination of the lower affinity site, Ca2. In SBMEP there are only five ligands, four protein atoms and a water molecule, observed in the first coordination shell, Table 11. Three of the protein ligands are the carbonyl oxygens of Ala169, Tyr171 and Thr174. The fourth protein ligand is the side-chain *OG1* atom of Thr174. The three carbonyl atoms are around 2·4 Å from the calcium ion, a typical Ca^{2+} to oxygen distance; the *OG1* to Ca^{2+} distance is longer, 2·78 Å. The chelated water molecule is 2·27 Å from the Ca^{2+}, also a 'good' ligand distance. This water is firmly held in place by hydrogen bonds to the carbonyl oxygen of Glu195 and the charged *OD1* atom of Asp197. The coordination is roughly trigonal bipyramidal, with the three carbonyl oxygen atoms in equatorial positions. The region around Ca2 in SBMEP is shown in Fig. 6(*b*).

Bode *et al.* (1987) and McPhalen & James (1988) describe the binding of Ca2 to SBCARL in its eglin-C complex, with almost identical results, Table 11. There are five ligands for Ca2 and the coordination and position of Ca2 is similar to those in SBMEP. There are three protein ligands coordinating to Ca2: the carbonyl oxygens of Ala169, Tyr171 and Val174 as in SBMEP. The fourth protein ligand in SBMEP, the *OG1* atom, is missing as Thr174 is substituted by Val174 in SBCARL. To compensate, there are two water ligands for Ca2 in SBCARL, compared to one in SBMEP. Bode *et al.* (1987) show that both these waters form bridges to negatively charged carboxyl oxygens in SBCARL, Asp172 and Glu197, strengthening the Ca^{2+} binding. In SBMEP Asp172 is replaced by Pro172, and Glu197 by Asp197. Pro172 is unable to form a bridged charge interaction with Ca1, and in the SBMEP complex only one liganded water bridging Ca1 and Asp 197 is observed.

Pantoliano *et al.* (1988) show that in SBPN the weakly binding Ca2 site is not occupied at low calcium concentrations, and required about 40 m*M* $CaCl_2$ in the crystallization solution to saturate the site. Furthermore in the absence of Ca^{2+} there is an adjacent site, 2·9 Å away, which can be occupied by a monovalent cation such as sodium or potassium, or water. In the Ca2 site the ligands for the metal are listed in Table 11. Very similar chelation is reported by McPhalen & James (1988) for the SBPN complex with a chymotrypsin inhibitor. It is immediately clear that the situation is different from that in SBMEP and SBCARL. The Ca2 sites of the latter pair lie close to one another in the three-dimensional structure, whilst the SBPN site is significantly displaced, by about 1·3 Å. The coordination of the Ca2 in SBPN is also quite different. The metal is surrounded by seven ligands, Table 11, five from the protein and two water molecules. In addition to the 'standard' three carbonyl groups, Ca2 is bound to a further carbonyl, Glu195, and to a charged atom,

*OD*2 of Asp197. Two waters constitute the remaining ligands. However the average metal to ligand distance is 2·9 Å and the shortest 2·74 Å. This casts some doubt on the identity of the metal ion: the distances and coordination number are more typical of a monovalent alkali metal.

In summary, in all three proteinases there exists a weak and a strong binding site for calcium ions. The strong site, Ca1, is conserved in all three with almost identical ligands, and is fully occupied in all the crystal analyses. The weaker site, Ca2, is closely similar in position in the SBMEP and the SBCARL eglin stuctures, and significantly occupied in both. In addition the metal–ligand distances in these complexes are about 2·4 Å, as for Ca1. In contrast, the Ca2 site reported for SBPN is different in character, and may not be occupied by calcium. The crystallization conditions reported for SBCARL–eglin-C did not include calcium (McPhalen, Schnebli *et al.*, 1985; Bode *et al.*, 1986). In the present study 5 m*M* $CaCl_2$ was present in the crystallization solution. Calcium does not appear to have been used in the crystallization buffer for SBPN.

The chelating ligands for both of the calcium ions are insufficient to fully neutralize their formal +2 charge. Ca1 has a single negatively charged ligand: the carboxyl side group of Asp41. There is no indication from the protein sequence determination of deamidation of either Gln2 or Asn77. Ca2 has no charged ligands from the protein. It binds a single water, with a relatively short metal–ligand distance (2·27 Å): this could be hydroxyl ion.

The Ca^{2+} ions in proteins can be replaced in solution by Tb^{3+} or other lanthanide ions with a similar radius. However attempts to make such a substitution with native SBMEP were unsuccessful, reflecting the tight binding of Ca^{2+} to the enzyme. Substitution with Tb^{3+} has been achieved by pretreating the SBMEP with EDTA. The binding was followed by spectral changes at 245 and 300 nm, typical for the ionization of phenolic groups. The titration curve showed the presence of two nonequivalent Tb^{3+} binding sites, probably corresponding to the two calcium sites. However, taking into account that the calcium cannot be totally removed from the strongest metal binding site in subtilisins by EDTA, a second possibility is that one Tb^{3+} ion is bound to a third metal site. A structure analysis of the Tb^{3+} complex is planned, to clarify this point.

Salt bridges

We define a salt bridge to occur if the distance between the oppositely charged atoms is less than 3·6 Å. In addition only one bridge is counted between any pair of charged groups, even if more than one atom pair satisfies the criterion. Between a total of 13 positively and 15 negatively charged residues in SBMEP, including the chain termini, there are six salt bridges, Table 12. Arg247 is involved in two of the six. Thus 11 of the charged residues are involved in such bridges.

Table 12. *Intramolecular contacts between the charged groups of SBMEP, SBPN and SBCARL*

Salt bridges to symmetry-related molecules or to eglin are not listed here. The N terminus and C terminus are not involved in salt bridges.

Residue 1	Atom 1	Residue 2	Atom 2	Distance (Å)
SBMEP				
Lys94	*NZ*	Glu54	*OE*2	2·78
Lys136	*NZ*	Asp140	*OD*2	2·64
Lys170	*NZ*	Glu195	*OE*1	2·96
Arg247	NH2	Asp197	*OD*2	3·55
Arg247	NH2	Asp197	OD1	3·30
Arg247	N*E*	Glu251	*OE*1	2·83
Arg247	NH2	Glu251	*OE*1	3·44
Arg249	NH1	Asp248	*OD*1	3·07
SBPN1				
Lys141	*NZ*	Glu112	*OE*1	3·23
Lys141	*NZ*	Glu112	*OE*2	2·84
Lys136	*NZ*	Asp140	*OD*2	2·80
Arg247	NH2	Asp197	*OD*1	3·31
Arg247	NH2	Asp197	*OD*2	3·12
Arg247	N*E*	Glu251	*OE*1	2·96
Arg247	NH2	Glu251	*OE*1	3·32
Lys265	*NZ*	Glu251	*OE*2	2·95
SBCARL1				
Lys15	*NZ*	Glu271	*OE*1	3·40
Lys94	*NZ*	Glu54	*OE*1	2·81
Lys136	*NZ*	Asp140	*OD*2	2·87
Lys170	*NZ*	Glu195	*OE*2	3·23
Arg247	N*E*	Glu197	*OE*1	2·88
Arg247	N*E*	Glu197	*OE*2	3·58
Arg247	NH1	Glu197	*OE*2	2·90
SBCARL2				
Lys94	*NZ*	Glu54	*OE*2	2·87
Lys170	*NZ*	Glu195	*OE*1	3·57
Arg247	N*E*	Glu197	*OE*1	2·79
Arg247	NH2	Glu197	*OE*1	3·53
Arg247	N*E*	Glu197	*OE*2	3·46
Arg247	NH2	Glu197	*OE*2	2·69

The salt bridges of SBPN and the two structure determinations of SBCARL–eglin-C (McPhalen & James, 1988; Bode *et al.*, 1987) are also listed in Table 12. The problem of drawing firm conclusions about the conservation or loss of salt bridges between the different proteins is illustrated by the results of the two independent structure determinations of the SBCARL–eglin-C complexes. Using the present criteria there are five bridges in the stucture of McPhalen & James and only three in that of Bode *et al.*

One of the salt bridges which is 'missing' from the Bode structure is that between Lys136 and Asp140. This salt bridge is present in SBMEP and SBPN. The *B* values reported for the atoms of Lys136 rise to about 25–30 $Å^2$ at the end of the side chain, for both SBCARL structures, which is about two times higher than the average. In the SBMEP model the *B* values of the end of lysine side chain are about 15 $Å^2$, below the average. In the Bode model, while the major part of the residue lies in the same conformation as in that of McPhalen, the NZ atom points in the oppo-

site direction and this results in a distance between the N*Z* of Lys136 and the O*D*2 of Asp140 of 6·11 Å. Thus, in spite of the fact that the expected mean accuracy of the coordinates is about 0·2 Å, a small change in the model building can lead to different interpretations.

The SBMEP bridges are to some extent conserved in SBPN and SBCARL. Of the six bridges for SBMEP listed in Table 12, the residues involved in the first four are completely conserved in all six *Bacilli* subtilisin sequences in Fig. 8. Furthermore these four bridges themselves are generally conserved. The Lys94-Glu54 bridge is found in both refined structures of SBCARL but missing from SBPN where the closest charged contact is 4·6 Å. The absence of the Lys136-Asp140 bridge in Bode *et al.* (1987) is mentioned above. The third bridge between Lys170-Glu195 is present in both SBCARL structures, but missing from SBPN, where the closest contact is 4·0 Å. Thus in each of these the 'missing' bridges have charge–charge contacts somewhat longer than our limit of 3·6 Å. The Arg47-Asp197 is present in all four refined structures, with varying numbers of contacts between the charged atoms. One of the charged residues is substituted by an uncharged one in either SBPN (Ser249) or SBCARL (Ser251) in the last two salt bridges given for SBMEP.

Thus the number of salt bridges remains difficult to quantify absolutely, and their energy contribution even more difficult to evaluate. However from the fact that the number of bridges appears to be similar in all three subtilisin structures it may be concluded that they do not contribute substantially to the difference in resistance of the protein to thermal or otherwise induced denaturation. The difficulty in classifying and quantifying the salt bridges present in the structure is illustrated by Arg247 in SBMEP and Glu251 in SBCARL, which appear to form 'bifurcated' bridges, their charge being dissipated over two oppositely charged residues.

In summary there are five positively and six negatively charged groups involved in salt bridges, one negative group directly chelating Ca1 and a second one binding to Ca2 through a water molecule. This leaves eight positive and seven negative unpaired charges on the surface of the subtilisin molecule.

Physicochemical properties of SBMEP

Among the *Bacilli* subtilisins the physicochemical properties of SBMEP are somewhat different from the others. In making this point it should be remembered that SBAS and SBI168, the subtilisins most similar to SBMEP, have not been subjected to extensive studies. A number of these properties are discussed below. The first three are 'global' properties of the whole molecule, the rest are more characteristic of local regions of SBMEP.

Table 13. *Atomic temperature factors* (Å^2) *averaged for the total protein, and for the side- and the main-chain atoms independently, for the subtilisin structures quoted in Table* 9

	SBMEP	SBPN1	SBPN2	SBCARL1	SBCARL2
Main chain	15·8	10·3	9·2	9·3	9·2
Side chain	21·0	13·6	11·1	13·1	11·9
Total protein	17·7	11·7	10·1	10·9	10·3

(1) SBMEP has low solubility in water (a maximum of 10 mg ml^{-1} can only be attained with 1 *M* NaCl) in contrast to SBPN and SBCARL which are more soluble (> 50 mg ml^{-1}). This reflects a roughly fivefold difference in equilibrium constant between the aqueous and crystalline environment for the two proteins, *i.e.* a $\Delta\Delta G$ difference of about 4 kJ mol^{-1}. It is extremely difficult to identify the factor(s) responsible for such a small energy change, given the very large numbers of 'weak' interactions present in the protein; see the hydrogen-bonding scheme for example. Possible factors include the overall charge, differences in amino-acid composition *etc.* The complexity of identifying the residues responsible is illustrated by the results on sickle-cell haemoglobin (Wishner, Ward, Lattman & Love, 1975) where a single substitution greatly affects the solubility with gruesome side affects.

(2) The optimum growth temperature of *B. mesentericus* is 310 K and the melting point (T_m) for the thermal denaturation of SBMEP inhibited by phenyl methyl sulfonyl fluoride has been estimated as 316 K from preliminary fluorescence data (Genov, unpublished results). Thus SBMEP is a mesostable protein. Again, the difference in stability is small: an extra 2 kJ mol^{-1} is sufficient to increase the T_m from 316 K by about 30 K. Finding the residues responsible is difficult with so many changes between the sequences. As was found for the HU DNA binding proteins from *Bacillus* (Wilson, Vorgias, Tanaka, White & Kimura, 1990) the substitutions are almost exclusively located on the surface of the enzymes, with few occurring in the core, and they are restricted to the non-functional parts of the structure, away from the active site. However there is a pronounced increase in the average atomic temperature factors of SBMEP compared to the other subtilisins, Table 13. The others all have mean values for the protein atoms in the range 10–12 Å^2 whereas the corresponding value for SBMEP is 17·7 Å^2. This is observed for both the main and side-chain atoms. Thus at the ambient temperatures under which data were collected for these enzymes, SBMEP shows significantly increased thermal motion. It should be noted that

the data for SBMEP were (*a*) collected on film, (*b*) using a synchrotron source and (*c*) are of limited resolution compared to the other structures discussed. However our experience with a number of structures suggests that these differences do not explain the lower *B* values.

(3) SBMEP is the only subtilisin which is known to be sensitive to denaturation by 8*M* urea, being completely denatured after 24 h of exposure to this reagent (Genov & Shopora, 1978). The native enzymes SBPN, SBCARL and SBDY are not markedly perturbed under these conditions (Shopova, Boteva & Genov, 1984). This susceptibility to the urea denaturation parallels the reduced thermostability of SBMEP: the difference between the free energy of the folded and denatured state is smaller than for SBPN and SBCARL, and this is reflected in a reduced resistance to denaturation under various physical conditions.

(4) It is surprising to note that the SBMEP and SBAS have been reported to differ significantly in kinetic properties. In the hydrolysis of the synthetic substrate tosyl-L-arginine methyl ester by SBMEP, V_{max} has been observed to be pH dependent in the region 6–8 and to follow a sigmoidal curve (Genov & Shopova, 1974*b*). In contrast V_{max} for the hydrolysis of the same ester by SBAS has been reported to be pH independent in the same pH range (Markland & Smith, 1971). Given the near identity of the structures, this is difficult to rationalize. There are no changes apparent in the catalytic or substrate binding sites.

Relation to spectroscopic and chemical properties

In this section we attempt to explain some of the observed spectroscopic properties and chemical and photochemical reactivity data about functional groups in SBMEP on the basis of the X-ray model. The properties are discussed by residue type, as they are usually dependent on a particular side chain or set of related side chains. This discussion is not comprehensive but rather tries to rationalize those results where the solution and crystal studies can be seen to give complementary information on the enzyme.

Tryptophan

SBMEP contains three tryptophan residues at positions 106, 113 and 241. The behaviour of their indole side chains was studied by spectral and photochemical methods (Genov & Jori, 1973; Ricchelli, Jori, Shopova, Boteva & Genov, 1981; Ricchelli, Jori, Filippi, Boteva, Shopova & Genov, 1982). The native and PMSF inhibited enzyme exhibit an emission spectrum with a maximum at 342 nm typical for tryptophan residues in a rather polar environment. The activation energy for the thermal deactivation of the excited indole rings (38 kJ mol^{-1}) (Jori & Genov, 1973) as well as the Stern–Volmer constant K_{sv} for the quenching of the tryptophan fluorescence with acrylamide (9·0 M^{-1}) demonstrated that the intrinsic emission of SBMEP is dominated by partially 'exposed' tryptophan residues (Ricchelli *et al.*, 1981). In the X-ray model the overall surface accessible to solvent (Kabsch & Sander, 1983) of Trp106, Trp113 and Trp241 is 37, 48 and 49 $Å^2$ respectively. The *B* values for these residues are 14·4, 10·6 and 13·7 $Å^2$ (main-chain average) and 24·5, 10·8 and 28·9 $Å^2$ (side-chain average) respectively. All three residues lie close to the surface of the protein, with a similar surface area exposed to solvent, with the side chains of Trp106 and Trp241 showing substantially greater mobility. From this simplistic analysis all three appear equally likely to contribute to the fluoresence.

Cs^+-quenching experiments showed heterogeneity of the tryptophan emission: only 18% of it was accessible to the quenching action of these ions. In native SBMEP the ionic quencher Cs^+ will preferentially interact with carboxyl groups on the surface. The probable explanation of the fluorescence data is that the emission of one particular relatively exposed indole ring is specifically quenched. The K_{sv} value obtained with Cs^+ is higher in comparison with the respective K_{sv} for the model compound Ac-Trp-NH_2 and was thought to reflect an electrostatic attraction between Cs^+ and a negative charge located near to this particular tryptophan (Boteva, Dimov, Genov, Fittkau & Peters, 1988). Inspection of the X-ray model showed that the carboxylic group of the C-terminal residue Gln275 is about 4 Å from Trp241. There is no charged group in the micro-environment of the other indole rings. It can be concluded that Trp241, which is the most exposed tryptophan in SBMEP, interacts with Cs^+ ions and contributes about 18% of the total indole emission.

The pH dependence of the fluorescence of SBMEP, inhibited with phenyl methyl sulfonyl fluoride, after excitation at 295 nm showed quenching of the tryptophan emission with a p*K* of 6·7 (Genov, Shopova & Boteva, 1981). The most likely explanation is that the fluorescence of a particular tryptophan residue is quenched by a protonated imidazole group situated nearby. There are three tryptophan and six histidine residues in SBMEP. Examination of the X-ray model showed that the only candidates for a histidine-tryptophan pair close enough to be effective are the residues His238 and Trp241. The two side-chain rings are almost parallel and the distance between them in the crystal is close to 3 Å. An imidazolium cation is capable of forming

complexes with indole and probably the observed quenching occurs through such a charge-transfer complex formation.

Photooxidation studies of SBMEP in the presence of proflavin indicated that the indole side chains possess different levels of photoreactivity at neutral pH (Genov & Jori, 1973). One of them was oxidized with a rate constant equal to that for the free tryptophan in water at the same conditions. The second indole group was classified as 'slowly reactive' and the third one as 'unreactive'. The photochemical data probably reflect the different levels of exposure of the indole groups. Sequence experiments with the photooxidized SBMEP revealed that the photoreactive tryptophan is Trp106. The solvent accessible areas from the X-ray structure provide no explanation for this observation: all three residues have comparable surface areas (see above) and indeed that of Trp106 is the smallest. However Trp106 and Trp241 both have relatively high *B* values (see above) and this makes them the more likely candidates of the two photoreactive residues. The particular region of the residue which is exposed, the chemical environment or perhaps the mobility must confer the susceptibility to photooxidation on Trp106 rather than merely its overall accessibility to solvent.

Tyrosine

Spectrophotometric titration of SBMEP revealed the presence of three groups of tyrosine residues (Genov, 1975). Six residues titrate normally with a pK_{app} of 10·0 almost equal to the value for free tyrosine in aqueous solution. These residues must be exposed on the surface of protein and should be fully accessible to solvent. 4–5 tyrosine residues titrate with an abnormally high pK_{app} of 12·0 which suggests that they are buried to some extent in the interior of the protein molecule or participate in hydrogen bonds, or a combination of the two factors. Finally 1–2 residues possess a $pK_{app} > 13{\cdot}0$ and they can be titrated only after the alkaline denaturation of the protein. The results for the chemical reactivity of the tyrosines in SBMEP towards tetranitromethane (Genov, Karadjova & Velcheva, 1972) were completely in line with the spectrophotometric data and confirmed this classification. In the crystal structure six tyrosines, at positions 21, 58, 91, 167, 256 and 262, are exposed on the surface and are completely accessible to solvent. Three of the other tyrosines, positions 214, 217 and 263, can be classified as partially 'exposed' and the remaining three tyrosines, positions 6, 104 and 171, participate in hydrogen bonds with the O atoms of the carbonyl groups at residues in positions 203, 130 and 132, respectively. Thus the X-ray model readily explains the titration behaviour and the chemical reactivity of the tyrosine residues in SBMEP.

Lysine

Seven of the eight lysine residues in SBMEP can be chemically modified using maleic anhydride (Genov & Shopova, 1974*a*). The same result was obtained with glutaric anhydride. Sequence data showed that the unreactive residue is Lys94 (Svendsen, Genov & Idakieva, 1983). The accessible surface areas of the lysines are given in Table 10. All lysine residues except Lys94 are highly exposed on the surface of the protein. In contrast Lys94 is completely buried and is furthermore involved in a salt bridge to Glu54. The shielding by the neighbouring residues may thus rationalize the lack of reactivity of Lys94 towards maleic and glutaric anhydrides.

Arginine

The reaction between free arginine and glyoxal in aqueous solution is usually more than 80% efficient. In SBMEP, about 70% of the guanidine groups react, suggesting that only three out of the four arginine residues are modified with glyoxal (Idakieva & Genov, 1981). The solvent-accessible surface areas of the four arginines are shown in Table 10. Arg45 and Arg249 are both highly exposed to solvent, with accessible surface areas about twice that of Arg186 and Arg247. Hence the accessibilty alone is not sufficient to explain the lack of reactivity. However Table 12 also shows that Arg247 is involved in two salt-bridge interactions with Asp197 and Glu251. It is therefore likely that it is Arg247, with its reduced accessible surface area and salt-bridge network, which shows reduced activity towards glyoxal.

Carboxylic residues

SBMEP contains 15 carboxyl groups, including that of the C-terminal Gln275. The chemical reactivity of these groups was investigated by coupling glycinamide to the reactive carboxyl groups using *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide. Only seven residues reacted. Circular dichroic spectra showed no signficant changes in the three-dimensional folding of the SBMEP after this reaction. Using denatured SBMEP all carboxyl groups reacted, showing their greater accessibility to the reagent. The low reactivity of eight carboxylic groups can be explained by their location in the three-dimensional structure. The surface area accessible to solvent is shown in Table 10, together with the number of salt bridges formed and the interactions of carboxylic acid residues with calcium ions. Two residues are buried below the surface of the protein

with no accessible surface area, Asp32 and Asp41; the latter residue is in addition involved in chelating the strongly bound calcium ion, Ca1. This pair is presumably unreactive. A further six side chains are involved in salt-bridge formation: Asp140, Asp197, Asp248, Glu54, Glu195 and Glu251. These are also likely to have reduced reactivity, leaving Asp36, Asp60, Asp97, Asp120, Glu112, Glu195 and the C terminus as the reactive groups. Sequence analysis of peptides from glycinamide-treated protein would be required to test this hypothesis.

Methionine

SBMEP contains four methionine residues. The photooxidation of these has been studied and they can be classified according to their photoreactivity. Only two of them are photooxidized at neutral pH with a rate constant of $0{\cdot}40 \times 10^{-3}\ s^{-1}$, significantly lower than that found for free methionine in aqueous solution (Genov & Jori, 1973). Thus two residues are slowly reactive and two unreactive. In contrast, all four methionines can be photooxidized in the denatured protein. The decrease in reactivity in the native protein is not easily explained from the X-ray model. The accessible surface area for the four methiones, Met119, Met124, Met199 and Met222, is 6, 1, 0, 4 $Å^2$ respectively. Thus all four residues are almost completely buried in the native structure. Hence on the basis of solvent accessibility alone the most likely candidates for the slowly oxidized residues are Met119 and Met222, with at least some surface area exposed. However the *B* values for the four residues are 14·4, 7·8, 12·7 and 7·0 $Å^2$ (main-chain atom average) and 9·1, 17·5, 12·5 and 8·4 $Å^2$ (side-chain average) respectively, suggesting that the apparently completely buried Met124 and Met199 in fact have more mobile side chains.

Concluding remarks

The results present the first three-dimensional structure of a mesostable subtilisin. SBMEP was stabilised by complexation with the inhibitor eglin-C and crystallized as the 1:1 complex. The structure was solved by molecular replacement, using independent models for the proteinase and inhibitor, both refined at high resolution. The inhibitor position was defined using the residual structure-factor amplitudes after subtraction of the initially determined proteinase contribution. This procedure, involving consequential positioning of the various subunits or domains of complex structures, may be of general use.

The overall fold of SBMEP is very similar to that of the homologous SBCARL and SBPN. The catalytic sites are highly conserved between the three proteins. The first calcium ion binding site is also highly conserved. The second is less so. In SBPN Ca2 appears not be occupied by calcium in the structures described by McPhalen & James (1988) and Bott *et al.* (1988). In the SBCARL–eglin complex the Ca2 lies at an analogous position to that in SBMEP, but the coordination is slightly different, with a second carboxyl group chelating through a bridging water molecule replacing the weakly bound *OG*1 of a threonine residue in SBMEP. The extra, albeit weak, charged interaction may well enhance the stability of the SBCARL molecule.

The difference in inhibitor binding in comparison to the eglin-C complexes with other subtilisins can be explained partly by substitutions of some residues in contact with eglin-C, as well as merely by changes in crystal environment.

Several of the physicochemical measurements on SBMEP in solution, involving spectroscopic properties or reactivity of particular residues, could be directly related to their position in the three-dimensional structure.

Given the substantial number of amino-acid substitutions between SBMEP, SBCARL and SBPN it remains difficult to identify which substitutions are responsible for the differences in the three 'global' physicochemical properties of SBMEP compared to the other two subtilisins, namely the lower thermostability, susceptibility to urea denaturation and lower solubility of SBMEP. The amino-acid composition of the different *Bacillus* subtilisins, Table 4, does not indicate any clearly identifiable differences between the mesophiles and thermophiles. High-resolution structure determinations alone are necessary, but not sufficient for this. Site-directed mutagenesis of individual residues will be required to identify the contribution of the various substitutions to these properties. The higher thermal motion, reflected in the increased atomic *B* values for SBMEP compared to the other subtilisins, correlates with its lower thermostability. The coordinates have been deposited in the Brookhaven Protein Data Bank.*

We thank Professor Michael James for providing the subtilisin Carlsberg–eglin-C coordinates, and Dr Richard Bott for the subtilisin BPN′ coordinates, prior to their release through the Brookhaven Protein Data Bank.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1MEE, R1MEESF), and are available in machine-readable form from the Protein Data Bank at Brookhaven. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37046 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

References

AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

BAKER, E. N. & DODSON, E. J. (1980). *Acta Cryst.* A**36**, 559–572.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MAYER, E. F. JR, BRYCE, M. D., RODGERS, J. R., KENNARD, O., SIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BETZEL, CH., PAL, G. P. & SAENGER, W. (1988). *Eur. J. Biochem.* **178**, 155–171.

BETZEL, CH., TEPLYAKOV, A. V., HARUTYUNYAN, E. H., SAENGER, W. & WILSON, K. S. (1990). *Protein Eng.* **3**, 161–172.

BODE, W., PAPAMOKOS, E. & MUSIL, D. (1987). *Eur. J. Biochem.* **166**, 673–692.

BODE, W., PAPAMOKOS, E., MUSIL, D., SEEMÜLLER, U. & FRITZ, H. (1986). *EMBO J.* **5**, 813–818.

BOTEVA, R., DIMOV, I., GENOV, N., FITTKAU, S. & PETERS, K. (1988). *J. Photochem. Photobiol.* **1**, 461–468.

BOTT, R., ULTSCH, M., KOSSIAKOFF, A., GRAYCAR, T., KATZ, B. & POWER, S. (1988). *J. Biol. Chem.* **263**, 7895–7906.

DAUTER, Z., BETZEL, CH., HÖHNE, W.-E., INGELMAN, M. & WILSON, K. S. (1988). *FEBS Lett.* **236**, 171–178.

DRENTH, J., HOL, W. G. J., JANSONIUS, J. & KOEKOEK, R. (1972). *Eur. J. Biochem.* **26**, 177–181.

FERSHT, A. R. & SPERLING, J. (1973). *J. Mol. Biol.* **74**, 137–149.

GENOV, N. (1975). *Int. J. Pept. Protein Res.* **7**, 325–332.

GENOV, N. & JORI, G. (1973). *Int. J. Pept. Protein Res.* **11**, 185–189.

GENOV, N., KARADJOVA, M. & VELCHEVA, P. (1972). *Commun. Dept Chem. Bulg. Acad. Sci.* **4**, 343–348.

GENOV, N. & SHOPOVA, M. (1974*a*). *Int. J. Pept. Protein Res.* **6**, 149–154.

GENOV, N. & SHOPOVA, M. (1974*b*). *Int. J. Pept. Protein Res.* **6**, 347–351.

GENOV, N. & SHOPOVA, M. (1978). *Int. J. Pept. Protein Res.* **11**, 185–189.

GENOV, N., SHOPOVA, M. & BOTEVA, R. (1981). *Commun. Dept Chem. Bulg. Acad Sci.* **14**, 100–111.

GROS, P., BETZEL, CH., DAUTER, Z., WILSON, K. S. & HOL, W. G. J. (1989). *J. Mol. Biol.* **210**, 347–367.

GROS, P., FUJINAGA, M., DIJKSTRA, B. W., KALK, K. H. & HOL, W. G. J. (1989). *Acta Cryst.* B**45**, 488–499.

HIRONO, S., AKAGAWA, H., MITSUI, Y. & IITAKA, Y. (1984). *J. Mol. Biol.* **178**, 389–413.

IDAKIEVA, K. & GENOV, N. (1981). *Commun. FECS 1st Int. Conf. Chem. Biotechnol. Nat. Prod.* **3**, 372–376.

JACOBS, M., ELIASSON, M., UHLEN, M. & FLOCK, J. (1985). *Nucleic Acids Res.* **13**, 8913–8926.

JANY, K. D. & MAYER, B. (1985). *Hoppe-Seyler's Z. Physiol. Chem.* **366**, 485–492.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

JORI, G. & GENOV, N. (1973). *Int. J. Pept. Protein Res.* **5**, 171–177.

KABSCH, W. & SANDER, C. (1983). *Biopolymers*, **22**, 2577–2637.

KARADJOVA, M., BAKURDJIEVA, A. & VELCHEVA, P. (1970). *C. R. Acad. Bulg. Sci.* **23**, 431–434.

KNECHT, R., SEEMÜLLER, U., LIERSCH, M., FRITZ, H., BRAUN, D. G. & CHANG, J.-Y. (1983). *Anal. Biochem.* **130**, 65–71.

KONNERT, J. H. & HENDRICKSON, W. A. (1980). *Acta Cryst.* A**36**, 344–350.

KURIHARA, M., MARKLAND, F. S. & SMITH, E. L. (1972). *J. Biol. Chem.* **247**, 5619–5631.

LESLIE, A. G. W., BRICK, P. & WONACOTT, A. J. (1986). *CCP4 Newsl.* **18**, 33–39.

MCPHALEN, C. A. & JAMES, M. N. G. (1988). *Biochemistry*, **27**, 6582–6598.

MCPHALEN, C. A., SCHNEBLI, H.-P. & JAMES, M. N. G. (1985). *FEBS Lett.* **188**, 55–58.

MCPHALEN, C. A., SVENDSEN, I., JONASSEN, I. & JAMES, M. N. G. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 7242–7246.

MARKLAND, F. & SMITH, E. (1967). *J. Biol. Chem.* **242**, 5198–5211.

MARKLAND, F. & SMITH, E. (1971). *The Enzymes*, 3rd ed., Vol. 3, edited by P. BOYER, pp. 562–609. New York: Academic Press.

MATTHEWS, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

MELHOUN, B., BAUDYS, M., KOSTKA, V., HAUSDORF, G., FRÖMMEL, C. & HÖHNE, W. E. (1985). *FEBS Lett.* **183**, 195–200.

NEDKOV, P., OBERTHÜR, W. & BRAUNITZER, O. (1983). *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 1537–1540.

NEIDHART, D. J. & PETSKO, G. A. (1988). *Protein Eng.* **2**, 271–276.

PANTOLIANO, M. W., LADNER, R. C., BRYAN, P. N., ROLLENCE, M. L., WOOD, J. F. & POULOS, T. L. (1987). *Biochemistry*, **26**, 2077–2082.

PANTOLIANO, M. W., WHITLOW, M., WOOD, J. F., ROLLENCE, M. L., FINZEL, B. C., GILLILAND, G. L., POULOS, T. L. & BRYAN, P. N. (1988). *Biochemistry*, **27**, 8311–8317.

PRIESTLE, J. (1988). *J. Appl. Cryst.* **21**, 572–576.

RAMAKRISHNAN, C. & RAMACHANDRAN, G. N. (1965). *Biophys. J.* **5**, 909–933.

READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

RICCHELLI, F., JORI, G., FILIPPI, B., BOTEVA, R., SHOPOVA, M. & GENOV, N. (1982). *Biochem. J.* **207**, 201–205.

RICCHELLI, F., JORI, G., SHOPOVA, M., BOTEVA, R. & GENOV, N. (1981). *Int. J. Pept. Protein Res.* **17**, 330–337.

RUSSELL, A. J. & FERSHT, A. R. (1987). *Nature (London)*, **328**, 496–500.

RUSSELL, A. J., THOMAS, P. G. & FERSHT, A. R. (1987). *J. Mol. Biol.* **193**, 803–813.

SCHECHTER, I. & BERGER, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.

SEEMÜLLER, U., EULITZ, M., FRITZ, H. & STROBEL, A. (1980). *Hoppe-Seyler's Z. Physiol. Chem.* **361**, 1841–1846.

SHOPOVA, M., BOTEVA, R. & GENOV, N. (1984). *J. Mol. Struct.* **115**, 295–298.

SMITH, G. L., DE LANGE, R. J., EVANS, W. H., LANDON, M. & MARKLAND, F. S. (1968). *J. Biol. Chem.* **243**, 2184–2191.

STAHL, M. L. & FERRARI, E. (1984). *J. Bacteriol.* **158**, 411–418.

SVENDSEN, I., GENOV, N. & IDAKIEVA, K. (1983). *Carlsberg Res. Commun.* **48**, 583–591.

SVENDSEN, I., GENOV, N. & IDAKIEVA, K. (1986). *FEBS Lett.* **196**, 228–232.

TEPLYAKOV, A., KURANOVA, I. P., HARUTYUNYAN, E. H., VAINSHTEIN, B. K., FRÖMMEL, C., HÖHNE, W. E. & WILSON, K. S. (1990). *J. Mol. Biol.* **214**, 261–279.

TERADA, I., KWON, S.-T., MIYATA, Y., MATSUZAWA, H. & OHTA, T. (1990). *J. Biol. Chem.* **265**, 6576–6581.

VASANTHA, N., THOMPSON, L. D., RHODES, C., BANNER, C., NAGLE, J. & FILPULA, D. (1984). *J. Bacteriol.* **159**, 811–819.

VOORDOUW, G., MILO, C. & ROCHE, R. S. (1976). *Biochemistry*, **15**, 3716–3724.

WARSHEL, A., NARAY-SZABO, G., SUSSMAN, F. & HWANG, J.-K. (1989). *Biochemistry*, **28**, 3629–3637.

WELLS, J. A. & ESTELL, D. A. (1988). *Trends Biochem. Sci.* **13**, 291–297.

WELLS, J. A., FERRARI, E., HENNER, D. J., ESTELL, D. A. & CHEN, E. Y. (1983). *Nucleic Acids Res.* **11**, 7911–7925.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

WILSON, K. S., VORGIAS, C. E., TANAKA, I., WHITE, S. W. & KIMURA, M. (1990). *Protein Eng.* **4**, 11–22.

WISHNER, B. C., WARD, K. B., LATTMAN, E. E. & LOVE, W. E. (1975). *J. Mol. Biol.* **98**, 179–194.

WRIGHT, C. S., ALDEN, R. A. & KRAUT, J. (1969). *Nature (London)*, **221**, 235–242.

YOSHIMOTO, T., OYAMA, H., HONDA, T., TONE, H., TAKESHITA, T., KAMIYAMA, T. & TSURU, D. (1988). *J. Biochem.* **103**, 1060–1065.