The EMBO Journal vol.9 no.6 pp.1939 - 1947, 1990

The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction

Milton T.Stubbs, Bernd Laber, Wolfram Bode, Robert Huber, Roman Jerala[1], Brigita Lenarčič[1] and Vito Turk[1]

Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, FRG and [1]Department of Biochemistry, Jožef Stefan Institute, Jamova 39, YU-61000 Ljubljana, Yugoslavia

Communicated by R.Huber

A stoichiometric complex of human stefin B and carboxymethylated papain has been crystallized in a trigonal crystal form. Data to 2.37 Å resolution were collected using the area detector diffractometer FAST. The crystal structure of the complex has been solved by Patterson search techniques using papain as search model. Starting from the structure of chicken cystatin, the stefin structure was elucidated through cycles of model building and crystallographic refinement. The current crystallographic R factor is 0.19. Like cystatin, the stefin molecule consists of a five stranded β-sheet wrapped around a five turn α-helix, but with an additional carboxy terminal strand running along the convex side of the sheet. Topological equivalence of stefin and cystatin reveal the previous sequence alignment to be incorrect in part, through deletion of the intermediate helix. The conserved residues form a tripartite wedge, which slots into the papain active site as proposed through consideration of the tertiary structures of the individual components (Bode *et al.*, 1988). The main interactions are provided by the amino terminal 'trunk' (occupying the 'unprimed' subsites of the enzyme), and by the first hairpin loop, containing the highly conserved QVVAG sequence, with minor contributions from the second hairpin loop. The carboxyl terminus of stefin provides an additional interaction region with respect to cystatin. The interaction is dominated by hydrophobic contacts. Inhibition by the cysteine proteinase inhibitors is fundamentally different to that observed for the serine proteinase inhibitors.

Key words: crystal structure/cysteine proteinase inhibitor/cystatin superfamily/proteinase–inhibitor complex/stefin

Introduction

Human stefin B [also called cystatin B; see Barrett *et al.* (1986) for nomenclature] is a tight and reversibly binding inhibitor of the cysteine proteinases (Green *et al.*, 1984). It was originally discovered as an inhibitor of cathepsin B in a variety of tissues (Lenney *et al.*, 1979) and has been isolated from spleen (Järvinen and Rinne, 1982; Lenarčič *et al.*, 1986) and liver (Green *et al.*, 1984). Its most likely role is the regulation of proteolysis in the cytoplasm, due in part to possible rupture of lysosomes. It could also play a part in the control of intracellular and extracellular breakdown. It has been sequenced (Ritonja *et al.*, 1985; Lenarčič *et al.*, 1986) and has been cloned by two groups independently (Jerala *et al.*, 1988; Thiele *et al.*, 1988).

The inhibitor is a member of the stefin family, the characteristics of which are M_r ~11 000, no disulphide bonds and no carbohydrate. The stefin family itself forms a subgroup of the cystatin superfamily, whose other families are the cystatins (M_r ~13 000, two disulphides) and the kininogens, large plasma proteins whose amino terminal 'heavy' chains are composed of three glycosylated cystatin-like domains [see Barrett *et al.* (1986) for a comprehensive review].

The stefins and cystatins bind with 1:1 stoichiometry and high affinity to various members of the papain superfamily from both plants and mammalian tissues. For cystatin, the most widely studied of the group, the amino terminal residues prior to Gly9I–Ala10I [chicken cystatin numbering is used throughout (Table I and below); the suffix 'I' serves to distinguish inhibitor residues from enzyme residues] have been shown to be important for papain inhibition (Abrahamson *et al.*, 1987; Machleidt *et al.*, 1989). This does not seem to be the case for stefins, however; amino terminal truncation of oryzacystatin (Abe *et al.*, 1988) or stefin B (Thiele *et al.*, 1990) appears to have little or no effect on the inhibition of papain.

In addition, the highly conserved segment 53I–57I, which in most members of the superfamily reads Gln-Val-Val-Ala-Gly (QVVAG), has been implicated in binding to papain (Ohkubo *et al.*, 1984; Turk *et al.*, 1985; Teno *et al.*, 1987; Machleidt *et al.*, 1989). Some doubt has been cast on this by Nikawa *et al.* (1989), however, who have probed exchanges in this region such as KVVAG and QVTAG, with little loss in inhibition, whilst similar results have been obtained by Jerala *et al.* (1990) on substitution of Val54I by Ala, Ile or Phe.

The X-ray crystal structure of chicken egg white cystatin (Bode *et al.*, 1988) showed that the conserved QVVAG region is located in a β-hairpin loop, flanked on one side by the amino terminal segment Gly9I–Ala10I and on the other by a second hairpin loop, the latter containing the doublet Pro103I–Trp104I, also rather conserved in the cystatin superfamily. In the truncated cystatin species examined, Gly9I and both hairpin loops form a contiguous 'wedge'. For cystatin forms with elongated amino termini, residues prior to Gly9I are disordered, and can extend freely into solution (Bode *et al.*, 1990).

This wedge was found to be quite complementary to the active site groove of papain. Docking experiments suggested that the two hairpin loops of the cystatin could make extensive interactions with the 'primed' sites of the enzyme and the amino terminus of the 'native' elongated form could interact with the 'unprimed' sites, running along the surface of the enzyme like a trunk (Bode *et al.*, 1988, 1989; Bode, 1989; Machleidt *et al.*, 1989). In particular, the amino terminal segment preceding Ala10I of cystatin was proposed to be directed towards the substrate site S1 in the hypothetical

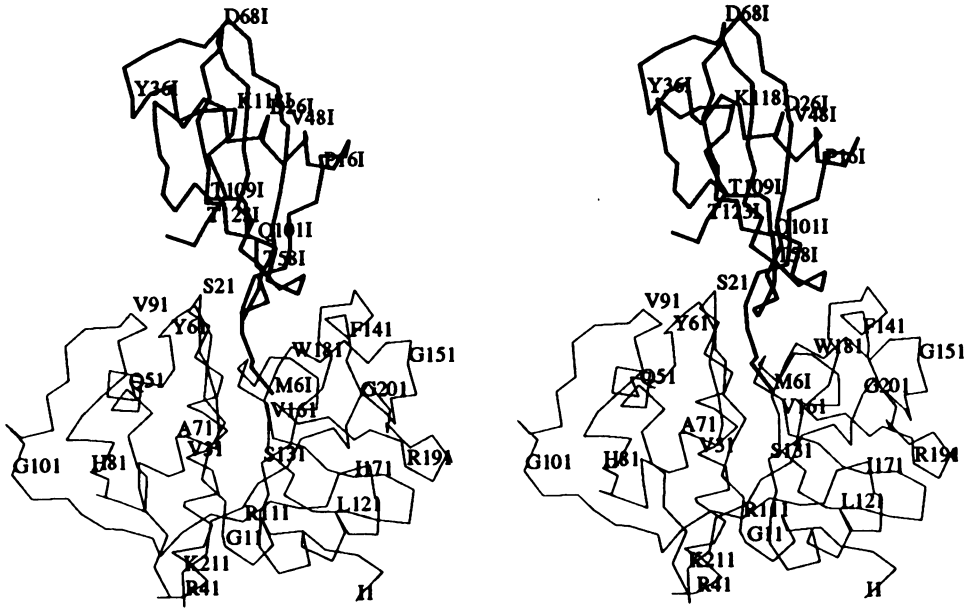

Fig. 1. α-Carbon trace of the complex formed by human stefin B (bold connections) and papain (thin connections). Every tenth residue is labelled. View is from the 'front', looking along the active site cleft with the amino terminal 'trunk' nearest.

complex with papain. This in turn would allow residues Leu7I, Leu8I and Gly9I to occupy the S3, S2 and S1 subsites, respectively, through formation of favourable main chain and side chain intermolecular interactions at S2 and tight turn formation between Leu7I and Ala10I. According to this docking model, Gly9I–Ala10I should be in an inappropriate conformation and too great a distance to be attacked by the reactive site Cys25. This mode of interaction—very different to that observed for serine proteinase inhibitors with serine proteinases (Huber and Bode, 1978; Bode *et al.*, 1990)—can explain among other things why complex formation can still occur when the reactive site thiol group of cysteine proteinases is blocked by small bound substituents (Anastasi *et al.*, 1983; Björk and Ylinjärvi, 1989).

Such a model requires experimental support. We have succeeded in crystallizing a 1:1 stoichiometric complex of recombinant human stefin B and carboxymethylated papain. These crystals have allowed X-ray data collection to 2.37 Å and the structure of the complex has been solved and refined to this resolution. As will be presented, the complex shows that the proposed docking model is in essence correct. It also yields the correct alignment of the stefins and the cystatins. The detailed structure and experimental details will be presented elsewhere.

Results

Overall structure

Figure 1 shows an α-carbon representation of the stefin B–carboxymethylated papain complex. The inhibitor indeed binds via its wedge shaped edge to the active site cleft of the enzyme. The main chain of papain can be traced in its entirety from Ile1 to Asn212. Comparison of the enzyme portion with the search structure (Priestle *et al.*, 1984; Drenth *et al.*, 1971; entry 1PPD from the Protein Data Bank, Bernstein *et al.*, 1977) and with oxidized papain (Kamphuis *et al.*, 1984, entry 9PAP) gives a root mean square deviation of ~0.35 Å for C^{α}s, whilst the two native papain structures show a deviation of ~0.22 Å. Deviations seem to be caused mainly by crystal packing; however, residues Arg58–Gly66, Ser70 and Gly138–Leu143 show shifts that most probably result from complex formation. In addition, the side chain of Trp69 is rearranged and of course the carboxymethyl group is attached to Cys25 (denoted Cym25 in the following).

The stefin molecule [numbered according to residues topologically equivalent in chicken cystatin (Bode *et al.*, 1988; Table I] can be traced unambiguously from the amino terminal Met6I to the carboxy terminal Phe125I, although two loop regions Asp68I–Glu92I (two residues, due to nomenclature) and Lys118I–His119I are somewhat poorly defined by density. The observed density is in full agreement with the cDNA sequence.

Stefin structure

The stefin component is shown in detail in Figure 2 and the secondary structure schematically in Figure 3. As in chicken cystatin, the main body of the molecule consists of a five-stranded antiparallel β-sheet, wrapped around a central five-turn α-helix, but with an additional extension of the C-terminus running along the 'back' of the sheet. Despite the topological similarities, rather surprising spatial discrepancies occur between human stefin B in complex and free chicken cystatin. Residues of the β-sheet, namely 10I–12I, 50I–62I, 94I–101I and 110I–111I can be superimposed as shown in Figure 4 with a r.m.s. deviation of 0.45 Å for C^{α}s; this can be reduced to 0.39 Å if section 55I–57I is not included. Using this alignment, the following characteristic features are observed as compared to cystatin (Bode *et al.*, 1988): (i) residues Met7I–Ala10I form a rather open type II turn, necessitating a glycine residue at the third position; the carbonyl group of Gly9I is in hydrogen bond contact with the carboxamide group of Gln53I; (ii) an extended strand A runs in an approximate poly-L-proline II conformation, with a wide β-bulge at Ala13I–Thr14I; this

Table I. Sequential alignment of human stefin B with chicken egg white cystatin according to topological equivalence in tertiary structure

	–	–	–	–	–	1	2	3	4	5	6	7	8	9	10
Stefin	–	–	–	–	–	M	M	S[a]	G	A	P	S	A	T	Q
Cystatin	S	E	D	R	S	R	L	L	G	A	P	V	P	V	D
	1I	2I	3I	4I	5I	6I	7I	8I	9I	10I	11I	12I	13I	14I	15I
	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Stefin	P	A	T	A	E	T	Q	H	I	A	D	Q	V	R	S
Cystatin	E	N	D	E	G	L	Q	R	A	L	Q	F	A	M	A
	16I	17I	18I	19I	20I	21I	22I	23I	24I	25I	26I	27I	28I	29I	30I
	26	27	28	29	30	31	32	33	34	35	–	–	36	37	38
Stefin	Q	L	E	E	K	Y	N	K	K	F	–	–	P	V	F
Cystatin	E	Y	N	R	A	S	N	D	K	Y	S	S	R	V	V
	31I	32I	33I	34I	35I	36I	37I	38I	39I	40I	41I	42I	43I	44I	45I
	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
Stefin	K	A	V	S	F	K	S	Q	V	V	A	G	T	N	Y
Cystatin	R	V	I	S	A	K	R	Q	L	V	S	G	I	K	Y
	46I	47I	48I	49I	50I	51I	52I	53I	54I	55I	56I	57I	58I	59I	60I
	54	55	56	57	58	59	60	61	–	–	–	–	–	–	–
Stefin	F	I	K	V	H	V	G	D	–	–	–	–	–	–	–
Cystatin	I	L	Q	V	E	I	G	R	T	T	C	P	K	S	S
	61I	62I	63I	64I	65I	66I	67I	68I	69I	70I	71I	72I	73I	74I	75I
	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Stefin	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Cystatin	G	D	L	Q	S	C	E	F	H	D	E	P	E	M	A
	76I	77I	78I	79I	80I	81I	82I	83I	84I	85I	86I	87I	88I	89I	90I
	–	62	64	64	65	66	67	68	69	70	71	72	73	74	75
Stefin	–	E	D	F	V	H	L	R	V	F	Q	S	L	P	H
Cystatin	K	Y	T	T	C	T	F	V	V	Y	S	I	–	P	W
	91I	92I	93I	94I	95I	96I	97I	98I	99I	100I	101I	102I	102AI	103I	104I
	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
Stefin	E	N	K	P	L	T	L	S	N	Y	Q	T	N	K	A
Cystatin	L	–	N	Q	I	K	L	L	E	S	K	C	–	Q	–
	105I	105AI	106I	107I	108I	109I	110I	111I	112I	113I	114I	115I	115AI	116I	117I
	91	92	93	94	95	96	97	98							
Stefin	K	H	D	E	L	T	Y	F							
Cystatin	–	–	–	–	–	–	–	–							
	118I	119I	120I	121I	122I	123I	124I	125I							

The latter numbering is used throughout this paper.
[a]This residue is cysteine in naturally occurring human stefin B.

forms a slightly longer regular structure in stefin than in cystatin; (iii) helix I is shifted ~1/2 turn, and displays a different orientation with respect to the β-sheet; in addition, the helix is kinked at residue Val28I in stefin; (iv) the loop joining helix I to strand B is shorter in stefin, resulting in a shorter strand B (deletion of two residues); (v) residues Val55I–Gly57I of the 'first hairpin loop' exhibit (surprisingly) a different conformation to that observed in cystatin; indeed, Val55I is in a slightly strained conformation (ϕ = −117, ψ = −146); (vi) the partially flexible helix II of cystatin is completely absent, through a deletion of 23 residues. Instead, the chain forms a five-residue open β-hairpin loop from Val66I to Asp93I; (vii) strand D is similar from residue 93I on and runs into a turn topologically equivalent to the 'second hairpin loop'. The internal organization of this loop, however, is rather different to that of cystatin; it is longer by two residues, and Leu102AI–Asn105AI form 1.5 turns of 3_{10} helix. The loop makes a wide excursion so that regular ladder formation only occurs from Phe100I on in strand E; (viii) the peptide chain of stefin extends beyond the carboxy terminus of cystatin by nine residues. Ala117I–Asp120I form a type I tight turn, from which the chain continues in an extended manner resting against the convex side of the sheet (opposite to the helix), terminating with Phe125I which covers a hydrophobic patch on the sheet exterior. Contacts to the rest of the molecule are made exclusively through side chain interactions.

The interface between stefin and papain

The interface region between papain and stefin is rather tightly packed (Figure 5). The surface of the wedge of stefin

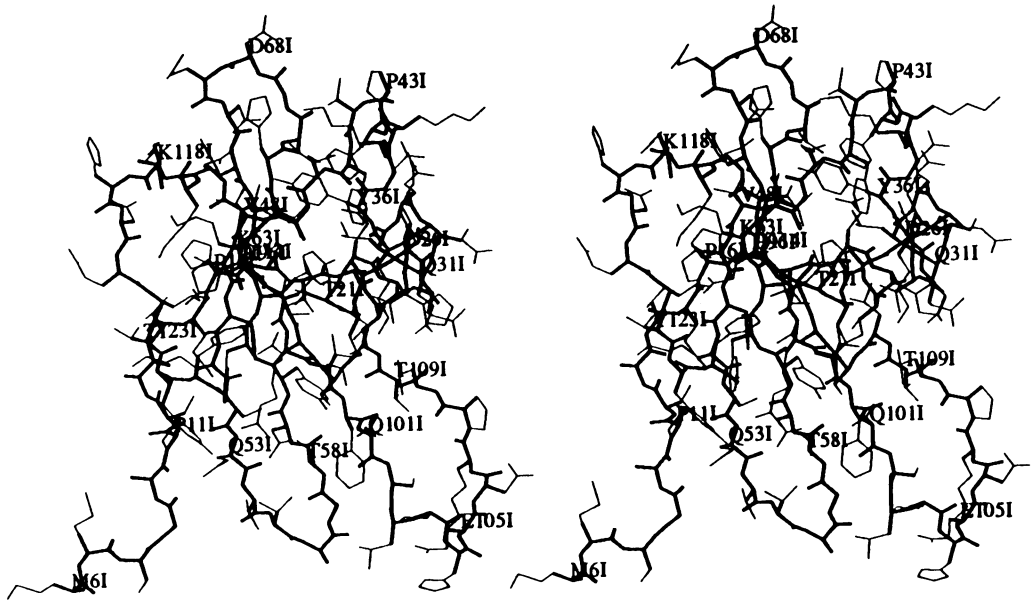

Fig. 2. Structure of human stefin B with the polypeptide backbone represented by thick connections. Every fifth residue is labelled. Orientation as if from the right hand side of Figure 1, with (reading from left to right along the bottom) the amino terminus, the QVVAG region and the second loop.

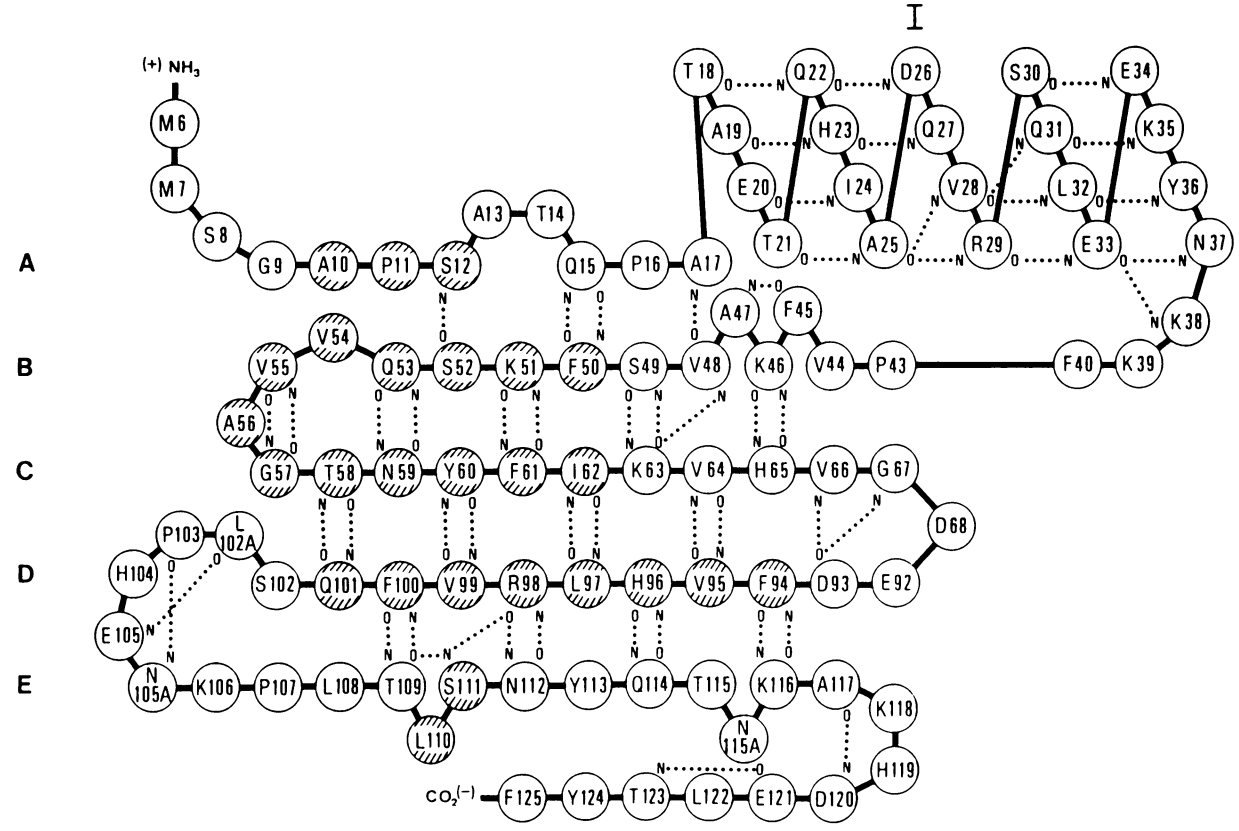

Fig. 3. Schematic representation of the polypeptide fold arrangement of human stefin B. Main chain hydrogen bonds displayed by dashed lines were selected according to definitions given by Kabsch and Sander (1983), using energy cut-off values of below −0.7 kcal/mol. Highlighted are residues structurally equivalent to those in chicken egg white cystatin.

is complementary to the active site groove of the papain molecule. Analysis of the accessible surface area of the protein components separated from and engaged in the complex show that a total area of 1117 Å^2 (from a total of 6907 Å^2) of stefin becomes buried in the complex, along with 977 Å^2 (of 10 612 Å^2) of papain, using a probe radius of 1.4 Å. This results in a total accessible surface area of 15 438 Å^2, and a total buried surface area of 2094 Å^2, or ~10% of the surface of the separated components. It is of interest what proportions of this interaction are made by the various interaction segments: 413 Å^2 (of 691 Å^2 exposed before) of the trunk become buried; 322 Å^2 (of 352 Å^2) of

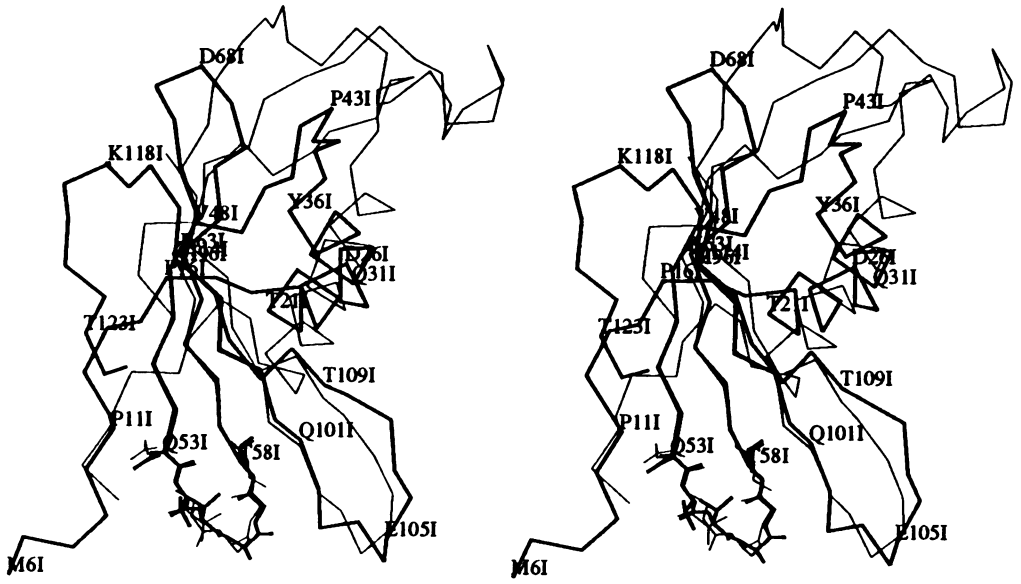

Fig. 4. Overlay of the α-carbon structure of chicken egg white cystatin (Bode *et al.*, 1988; thin connections) on that of human stefin B (bold connections). All atoms of the first hairpin loops of each are shown. Orientation as in Figure 2.

Table II. Matrix of intermolecular contacts <4 Å

Papain	Stefin																
	Trunk						1st loop						2nd loop				
	M6	M7	S8	G9	A10	P11	Q53	V54	V55	A56	N59	R98	L102A	P103	H104	Y124	Sum
N18										1							1
Q19									2	3							5
G20									3	7(1)							10(1)
S21									7		3(1)	1					11(1)
C22									3								3
G23									2								2
C25[a]			4[a]	4[a]				3(2)[a]									11(2)[a]
W26																	1
Y61		1															1
C63							2(1)									6(1)	8(2)
N64				2	2	2										11	17
G65			2	4													6
G66		3	5(2)														8(2)
Y67		8															8
P68	1																1
W69	2																2
V133			1														1
A136								1									1
A137								2									2
L143														2			2
D158			2	3(1)				1									6(1)
A160			3														3
W177									4(1)	3			4				11(1)
G180															7		7
Sum	3	12	18 (2)	13 (1)	2	2	2 (1)	7 (2)	21 (1)	14 (1)	3 (1)	1	4	2	7	17 (1)	
				50 (3)					47 (6)			1		13		17 (1)	128 (10)

Figures in parentheses indicate number of possible hydrogen bonds in total contacts.
[a]Note: all but one (S8I C^{β}–C25 S^{γ}) of the contacts made to C25 are to the carboxymethyl group.

the first hairpin loop; 205 Å^2 (of 389 Å^2) of the second hairpin loop; and 83 Å^2 (of 135 Å^2) of the carboxy terminus region become inaccessible.

A total of 128 intermolecular atom–atom interactions <4 Å occur in this region (Table II). The following segments of the stefin wedge contribute: Met6I–Pro11I,

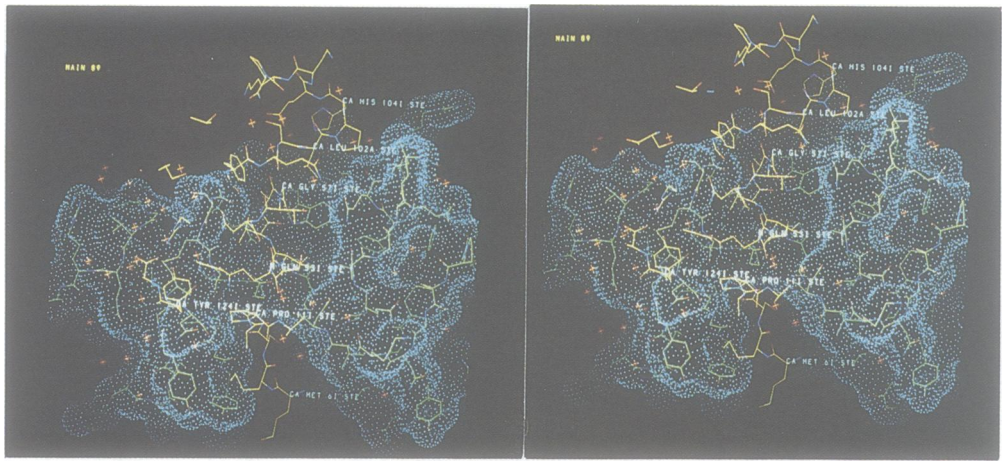

Fig. 5. View of the active site cleft of papain (green), overlaid with a corresponding Connolly dot surface (blue), and interacting residues from the trunk, the first and second hairpin loops and the carboxyl terminus of human stefin B (orange). View is from the 'top' of the complex, with the 'trunk' pointing straight down the page.

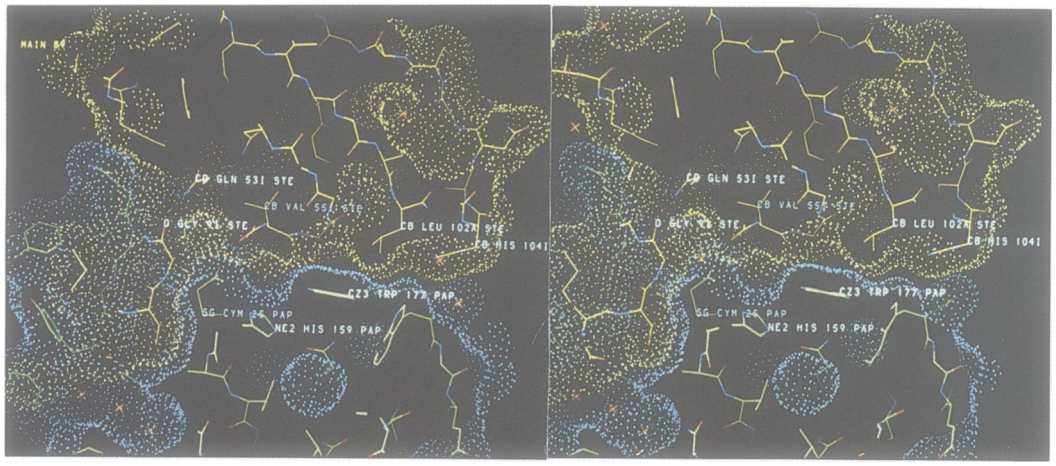

Fig. 6. View of the interaction between human stefin B (yellow on orange) and papain (blue on green), showing the space above the carboxymethylated Cym25. Orientation as in Figure 2.

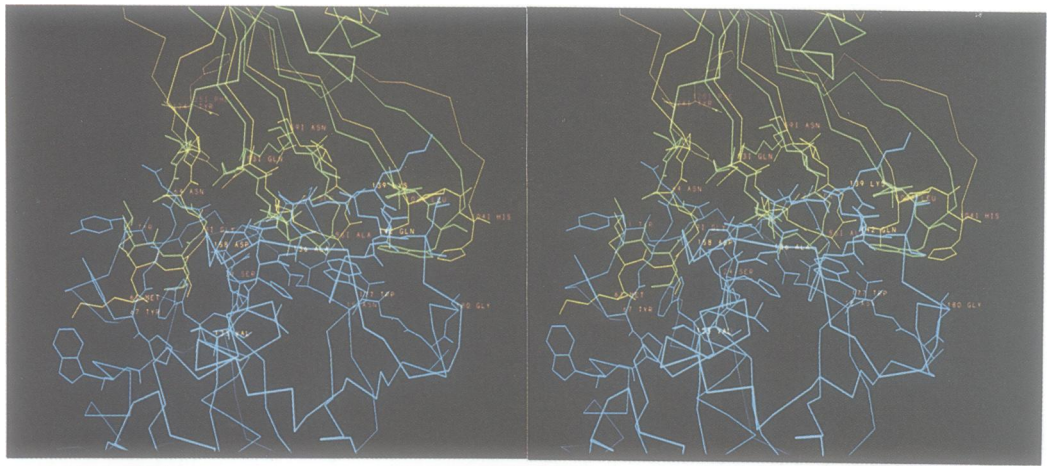

Fig. 8. Comparison of the proposed docking model of chicken egg white cystatin (green) with papain (blue) (Bode *et al.*, 1988) with the experimental complex of human stefin B (orange) with papain, after optimal alignment of papain residues.

Gln53I–Asn59I, Gln101I–His104I and Tyr124I–Phe125, although not all residues are in direct contact (see below). Particularly large numbers of contacts (>10) are made by the amino terminal residues Met7I, Ser8I (formally P3 and P2), residues Val55I, Ala56I of the first hairpin loop, and Tyr124I near the carboxy terminus.

All residues from the base and from both sides of the papain active site cleft are involved in complex interactions:

Table III. Polar contacts at the interface between stefin B and papain

Ser8I O				Gly66 N
Ser8I N				Gly66 O
Gly9I N				Asp158 O
Gln53I $N^{\epsilon 2}$				Cys63 O
Val54I O				Cym25 O^{1}
Val54I O				Cym25 O^{2}
Val55I O				Trp177 $N^{\epsilon 1}$
Ala56I N				Gly20 O
Asn59I $N^{\delta 2}$				Ser21 O
Try124I O^{η}				Cys63 O
Met7I O		423 OH		Asp158 O
Gly9I N		423 OH		Asp158 O
Ala10I N		423 OH		Asp158 O
Ala10I O	381 OH		387 OH	Cym25 O^{1}
Ala10I O	381 OH		387 OH	Asp158 O
Ala10I O	381 OH		391 OH	Asp158 $O^{\delta 1}$
Gln53I $O^{\epsilon 1}$		357 OH		Ser21 O
Gln53I $O^{\epsilon 1}$		357 OH		Gly23 N
Gln53I $O^{\epsilon 1}$		357 OH		Cys63 O
Val54I N	381 OH		387 OH	Cym25 O^{1}
Val54I N	381 OH		387 OH	Asp158 O
Val54I N	381 OH		391 OH	Asp158 $O^{\delta 1}$
Val54I O		387 OH		Cym25 O^{1}
Val54I O		387 OH		Asp158 O
Thr58I N		346 OH		Gln142 $O^{\epsilon 1}$
Thr58I $O^{\gamma 1}$		346 OH		Gln142 $O^{\epsilon 1}$
Gln101I $N^{\epsilon 2}$		346 OH		Gln142 $O^{\epsilon 1}$
Gln101I O		346 OH		Gln142 $O^{\epsilon 1}$
Ser102I O		346 OH		Gln142 $O^{\epsilon 1}$
His104I N		333 OH		Gln142 O
His104I N	333 OH		336 OH	Tyr144 O
His1041 $N^{\epsilon 2}$		322 OH		Trp177 O
Tyr124I O^{η}		357 OH		Ser21 O
Tyr124I O^{η}		357 OH		Gly23 N
Tyr124I O^{η}		357 OH		Cys63 O
Tyr124I O^{η}	525 OH		331 OH	Ser 21 O^{γ}
Tyr124I O^{η}	525 OH		331 OH	Gly90 O
Phe126I OH	349 OH		351 OH	Gln92 N

Direct contacts are listed first, followed by solvent mediated interactions.

Asn18–Trp26, Tyr61–Trp69, Val133–Leu143, Asp158–Ala160 and Trp177–Gly180. The contacts are more or less evenly distributed among these residues; however, greater numbers of contacts are observed for Gly20, Ser21, Cym25 (although all but one contact are through the carboxymethyl group), Asn64 and Trp177.

Of particular interest is that the side chain of Cym25 is enclosed by stefin residues Ser8I–Gly9I and Gln53I–Val54I (Figure 6). In conjunction with certain papain residues, these completely encapsulate the reactive site thiol group; yet even in the presence of the carboxymethyl group, the side chain is not tightly packed.

It is clear that the steric fit between hydrophobic parts of both interfaces dominates the interaction; polar contacts between the enzyme and inhibitor seem to be relatively unimportant. Only four main chain–main chain hydrogen bonds exist: three are made by the N-terminal trunk (Ser8I O–Gly66 N; Ser8I N–Gly66 O; Gly9I N–Asp158 O) and a fourth by the first hairpin loop (Ala56I N–Gly20 O). A further four hydrogen bond interactions are made between side chain and main chain atoms (Gln53I $N^{\epsilon 2}$–Cys63 O; Val55I O–Trp177 $N^{\epsilon 1}$; Asn59I $N^{\delta 2}$–Ser21 O; Tyr124I O^{η}–Cys63 O). An additional two hydrogen bonds are possible between the carboxylate group of Cym25 and Val54I O, if the carboxyl group is protonated (possible at pH 5.0). It is indeed remarkable that the carbonyl group of Val54I points towards both carboxyl groups of Cym25. Polar contacts at the interface are listed in Table III.

In contrast to the small number of direct hydrogen bonds, there are many polar interactions mediated via solvent molecules (Figure 7; Table III). A total of 13 solvent molecules bridge polar groups of stefin and papain, with bridge lengths of one or two solvent molecules. There are a total of 17 hydrogen bond contacts from stefin made to bridging molecules, terminating in 14 such contacts with papain (one of which is to Cym25 O^{1}). Two solvent molecules act as donor/acceptor to more than four groups; presumably they are statistically ordered. These water mediated contacts appear to play a particular role in the interaction between the C-terminus of stefin and papain.

Discussion

The experimentally determined structure of the complex formed between human stefin B and carboxymethylated papain has been described. The presented tertiary structure of stefin confirms the familial relationship between the stefins and the cystatins. In particular, conserved residues of stefin occur in equivalent positions to those in cystatin. Despite the obvious topological similarities, however, there are some pertinent differences. The shift in the helix, for example, would be impossible to predict, even when using modelling by homology. The published sequence alignment (Barrett *et al.*, 1986) has been shown to be incorrect for the 37 carboxy terminal residues. This is largely due to a deletion of 23 residues with respect to the cystatin sequence, source of a high penalty in most alignment algorithms. This results in a stefin molecule that is extended (in structural terms) by nine residues.

The presented structure shows that the proposed docking model (Bode *et al.*, 1988, 1989; Bode, 1989; Machleidt *et al.*, 1989) is in essence correct (Figure 8). In particular, the tripartite nature of the interaction is confirmed: the amino terminal 'trunk', the QVVAG loop and the second hairpin loop are all involved. The intra- and intermolecular hydrogen bonds of the 'trunk' have been correctly predicted. The near formation of a type II turn from Met7I to Ala10I, the geometry of which requires a glycyl residue in the third position, explains the absolute conservation of Gly9I in all members of the cystatin superfamily. The proposed interaction of Trp104I in cystatin is substituted by two residues in stefin, Leu102AI and His104I. The stefin molecule shows an additional binding site, due to the extended carboxy terminus. This, along with the involvement of solvent molecules, was impossible to predict.

It is quite clear that dimer formation due to the oxidation of Cys8I (in our case Ser8I) would lead to inactivation of the stefin molecule due to steric hindrance, as observed (Lenarčič *et al.*, 1986). The mutant Cys8I→Ser8I was in fact chosen for this reason (Jerala *et al.*, 1989).

Of particular interest are the differences observed in the peptide orientations of Val55I and Ala56I in the first hairpin loop. Of course, this could be an inherent structural

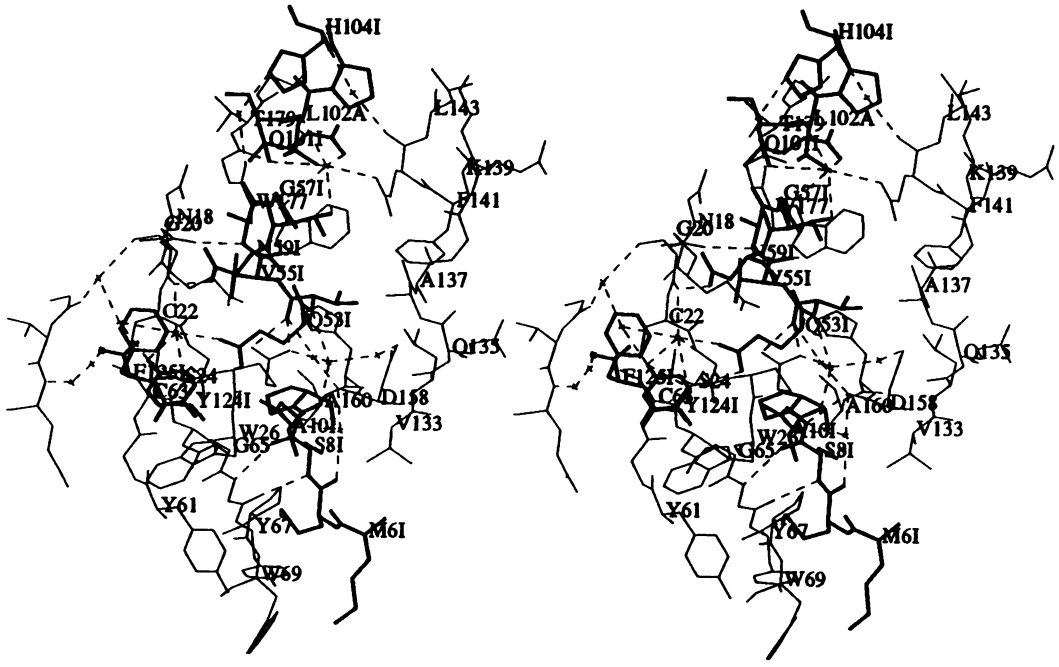

Fig. 7. Polar contacts at the papain–stefin interface: papain residues indicated by thin lines, stefin residues by bold lines, and possible hydrogen bonds by dashed lines. Of note is the relatively large number of solvent mediated contacts. Orientation as in Figure 5.

difference between stefin and cystatin (the latter sequence reads here QLSVG). This seems somewhat unlikely, however; the B-values for these residues in the cystatin structure indicate a degree of flexibility in this region (Bode *et al.*, 1988), while in contrast, B-values for these residues in the complex are rather low. In addition, Val55I is in a slightly strained conformation; it is hard to envisage such a conformation being preserved in the free inhibitor. Naturally, this raises the question as to whether this deformation is driven by complexation. It is of note that a hydrogen bond is formed between Ala56I N and Gly20 O, and that the carbonyl group of Val55I is within hydrogen bonding distance of Trp177 $N^{\epsilon 1}$, unpredictable on the basis of the free cystatin structure. Should this reorientation be complex dependent, then Gly57I would make an ideal hinge.

What of the relative importances of the trunk and the QVVAG region? It is clear from this structural analysis that they both make valuable contributions to the interaction; they are each rather well buried in the complex. Data have been presented (Abrahamson *et al.*, 1987; Machleidt *et al.*, 1989) indicating a loss in inhibition on amino terminal truncation of chicken cystatin; corresponding data for members of the stefin family (Abe *et al.*, 1988; Thiele *et al.*, 1990) show no such effect. It is not immediately clear why this is so. It may be that the interaction region of stefin forms better contacts than that of cystatin—the QVVAG region is deeply buried, and additional contacts are provided for by the carboxyl terminus. It should not be forgotten that the first hairpin loop of chicken cystatin has the sequence QLVSG; it is possible that less favourable contacts could be made in the two loop regions of cystatin that may need compensation by an amino terminal extension.

Exchanges of amino acids in the first hairpin loop should be quite easily explained (Nikawa *et al.*, 1989; Jerala *et al.*, 1990). The substitution Gln53I→Lys53I would result in the loss of a part of the connection between the trunk and the first loop and would therefore have only a minor effect. The exchanges Val54I→Ala54I, Ile54I or Phe54I, or Val55I→Thr55I each allow the same basic interactions to occur, and so once again should show only small differences. Such studies do not detract from the importance of the first hairpin loop in the interaction of stefins (and cystatins) with papain.

Clearly, not all questions concerning the interaction of cysteine proteinase inhibitors with their cognate enzymes can be solved by crystallographic methods. It is to be hoped, however, that such investigations lead to a firm structural foundation, on which physical, chemical and biochemical data can be analysed and hypotheses tested.

Materials and methods

Papain (2 × crystallized suspension) was purchased from Sigma. Papain (100% active) was prepared by covalent chromatography on a Hg–agarose column (Sluyterman and Wijdenes, 1970). Assay of papain activity and active site titration was performed according to Barrett *et al.* (1982). Fractions containing the active enzyme were pooled and desalted on Sephadex G-25 (PD-10) columns equilibrated with 50 mM phosphate buffer pH 6.0, 1 mM DTT. Carboxymethylation was performed according to Chaiken and Smith (1969) by adding 1/20 volume of 100 mM sodium chloroacetate in 50 mM phosphate buffer pH 6.0 to the enzyme solution. After 10 min at 37°C unreacted chloroacetate was removed by gel filtration on Sephadex G-25 (PD-10) columns equilibrated with 2 mM phosphate pH 7.0. The resulting carboxymethyl-papain (CM-papain) solution (residual activity ~0.5%) was concentrated by ultrafiltration (Amicon PM 10 membrane).

Recombinant human stefin B (Jerala *et al.*, 1988), differing from natural stefin B by a single amino acid exchange Cys8I→Ser8I (Jerala *et al.*, 1989), was precipitated by the addition of solid $(NH_4)_2SO_4$ to 0.56 g/ml. The precipitate was collected by ultracentrifugation, resuspended in a minimal amount of 2 mM phosphate buffer pH 7.0 and dialysed overnight against 1000 volumes of this buffer.

Crystallization experiments were performed at room temperature using the sitting drop vapour diffusion method. A stoichiometric complex between stefin and CM-papain was formed by adding both components using a 1.3-fold molar excess of inhibitor. This complex was crystallized by vapour diffusion against 10% PEG 1550 (Serva)/0.75 M phosphate pH 5.0. Plate-like needles and crystals appeared in the same droplet after ~10 days. Two crystal space groups were observed: tetragonal $P4_22_12$ with cell constants

$a = b = 66.0$ Å, $c = 186.5$ Å, $\alpha = \beta = \gamma = 90°$, with a resolution of ~2.6 Å; and trigonal $P3_121$ (as determined later, see below) with cell constants $a = b = 67.0$ Å, $c = 169.3$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$, diffracting to better than 2.3 Å. Each crystal form contains one complex in the asymmetric unit; the trigonal form showed better diffraction characteristics and was thus used for all further analyses.

X-ray intensity data were collected from a single trigonal crystal (mounted in 22.7% PEG/0.65 M phosphate pH 5.0) using a FAST television area detector diffractometer (Enraf-Nonius). A complete 2.37 Å data set was obtained by rotating the crystal for 90° about a^* and for 130° about an axis inclined by 40° to a^* in the plane containing a^* and c^*. Measurements (82 366), evaluated on-line using the program MADNES (Messerschmidt and Pflugrath, 1987) were scaled and corrected for absorption effects (Huber and Kopfmann, 1969; Messerschmidt *et al.*, manuscript submitted for publication) and merged and loaded by means of PROTEIN (Steigemann, 1974) yielding 16 084 unique reflections, comprising 89% of all data expected to 2.37 Å resolution (with 27% of all reflections expected in the last shell from 2.44 to 2.37 Å). R_{merge} (defined as $\Sigma|I - <I>|/\Sigma|I|$) for the individual measurements is 0.102 and for Friedel pairs after separate averaging 0.052.

The orientation of the papain component in the crystals was determined using Patterson search techniques (Huber, 1965) as implemented in PROTEIN, using intensity data between 8.0 and 3.5 Å resolution. The coordinates used were those of Priestle *et al.* (1984) and Drenth *et al.* (1971) as deposited in the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977). The correlation between the model self vectors and observed Patterson maxima showed a peak in Euler angle space of 3.2 arbitrary units, with next highest 2.8 and mean value 2.0. Relating the oriented molecule relative to the crystallographic axes was performed using the translation function of Crowther and Blow (1967) and programs written by E.E.Lattmann, modified by J.Deisenhofer and R.Huber. The highest peaks of the translation function were found on the five Harker sections and were self consistent, indicating clearly the assignment of space group $P3_121$. The values of the correlation for these highest peaks ranged from 847 to 1167 arbitrary units in the various Harker sections, whilst the second highest peaks were between 357 and 509. The R value (defined as $\Sigma(F_{obs} - F_{calc})/\Sigma F_{obs}$) of the correctly orientated model was 0.425.

A 3 Å $2F_{obs} - F_{calc}$ Fourier map was calculated using Sim weighted model phases. The map showed density for the papain molecule and in addition density belonging to the inhibitor. The putative complex of chicken cystatin with papain (Bode *et al.*, 1988) was aligned with the papain portion of the new cell; using the Fourier space correlation functions of PROTEIN, the cystatin moiety was adjusted as a rigid body to the observed density. Model building of the stefin component of the complex began by taking sections of the cystatin molecule and fitting them to their corresponding density if present, inserting residues as polyalanine using the PSFRODO version (Pflugrath *et al.*, 1984) of FRODO (Jones, 1978) on a PS390 interactive display system (Evans and Sutherland). The partial structure was refined using EREF (Jack and Levitt, 1978); cycles of model building and refinement allowed the gradual completion of the chain and the fitting of the known sequence to the density; the resolution was increased stepwise to the final resolution of 2.37 Å. The final model contains all main chain atoms and in addition 139 solvent molecules at stereochemically reasonable positions have been inserted. No correction for bulk solvent was made. The current R value for 16 033 data from 8.0 to 2.37 Å is 0.19. The total energy of the model is −1971 kcal/mol. The r.m.s. deviations from ideal values for bond lengths and bond angles are 0.014 Å and 2.1°, respectively.

Acknowledgements

We thank Dušan Turk for the use of his programme MAIN for display of the molecular surfaces. This work has been supported by the Research Council of Slovenia (grant C1-0515-106 to V.T.) and by the Sonderforschungsbereich 207 of the Ludwig-Maximilians-Universität, München (grants H-1 and H-2).

References

Abe,K., Emori,Y., Kondo,H., Arai,S. and Suzuki,K. (1988) *J. Biol. Chem.*, **263**, 7655–7659.

Abrahamson,M,. Ritonja,A., Brown,M.A., Grubb,A., Machleidt,W. and Barrett,A.J. (1987) *J. Biol. Chem.*, **262**, 9688–9694.

Anastasi,A., Brown,M.A., Kembhavi,A.A., Nicklin,M.J.H., Sayers,C.A., Sunter,D.C. and Barrett,A.J. (1983) *Biochem. J.*, **211**, 129–138.

Barrett,A.J., Kembhavi,A.A., Brown,M.A., Kirschke,H., Night,C.G., Tanai,M. and Hanada,K. (1982) *Biochem. J.*, **201**, 189–198.

Barrett,A.J., Rawlings,N.D, Davies,M.E., Machleidt,W., Salvesen,G. and Turk,V. (1986) In Barrett,A.J. and Salvesen,G. (eds), *Proteinase Inhibitors*. Elsevier, Amsterdam, pp, 515–569.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F, Jr, Brice,M.D, Rodgers,J.R, Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Björk,I. and Ylinjärvi,K. (1989) *Biochem. J.*, **260**, 61–68.

Bode,W. (1989) In Fritz,H., Schmidt,I. and Dietze,G. (eds), *The Kallikrein–Kinin System in Health and Disease*. Limbach-Verlag, Braunschweig, pp. 227–235.

Bode,W., Engh,R., Musil,D., Thiele,U., Huber,R., Karshikov,A., Brzin,J. and Turk,V. (1988) *EMBO J.*, **7**, 2593–2599.

Bode,W., Musil,D., Engh,R., Huber,R., Brzin,J., Kos,J. and Turk,V. (1989) In Katanuma,N. and Kominami,E. (eds), *Intracellular Proteolysis: Mechanisms and Regulations*. Japan Scientific Societies Press, Tokyo, pp. 297–304.

Bode,W., Engh,R., Musil,D., Laber,B., Stubbs,M., Huber,R. and Turk,V. (1990) *IInd Brdo Research Conference, Biol. Chem. Hoppe-Seyler*, in press.

Chaiken,I.M. and Smith,E.L. (1969) *J. Biol. Chem.*, **244**, 5095–5099.

Crowther,R.A. and Blow,D.M. (1967) *Acta Crystallogr. Sect. A*, **23**, 544–548.

Drenth,J., Jansonius,J.N., Koekoek,R. and Wolthers,B.G. (1971) *Adv. Prot. Chem.*, **25**, 79–115.

Green,G.D.J., Kembhavi,A.A., Davies,M.E. and Barrett,A.J. (1984) *Biochem. J.*, **218**, 939–946.

Huber,R. (1965) *Acta Crystallogr. Sect. A*, **19**, 353–356.

Huber,R. and Bode,W. (1978) *Acc. Chem. Res.*, **11**, 114–122.

Huber,R. and Kopfmann,G. (1969) *Acta Crystallogr. Sect. A*, **25**, 143–152.

Jack,A. and Levitt,M. (1978) *Acta Crystallogr. Sect. A*, **34**, 931–935.

Järvinen,M. and Rinné,A. (1982) *Biochim. Biophys. Acta*, **708**, 210–217.

Jerala,R., Trstenjak,M., Lenarčič,B. and Turk,V. (1988) *FEBS Lett.*, **239**, 41–44.

Jerala,R., Trstenjak,M., Kroon-Zitko,L., Lenarčič,B. and Turk,V. (1989) In Auerswald,E., Fritz,H. and Turk,V. (eds) *Proteinases and their Inhibitors: Recent Developments*. Kernforschungsanlage Jülich, pp. 55–58.

Jerala,R., Trstenjak-Prebanda,M., Kroon-Žitko,L., Lenarčič,B. and Turk,V. (1990) *IInd Brdo Research Conference, Biol. Chem. Hoppe-Syler*, in press.

Jones,A. (1978) *J. Appl. Crystallogr.*, **11**, 268–272.

Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.

Kamphuis,I.G., Kalk,K.H., Swarte,M.B.A. and Drenth,J. (1984) *J. Mol. Biol.*, **179**, 233–256.

Lenarčič,B., Ritonja,A., Sali,A., Kotnik,M., Turk,V. and Machleidt,W. (1986) In Turk,V. (ed.), *Cysteine Proteinases and their Inhibitors*. Walter de Gruyter, Berlin, pp. 473–487.

Lenney,J.F., Tolan,J.R., Sugai,W.J. and Lee,A.G. (1979) *Eur. J. Biochem.*, **101**, 153–161.

Machleidt,W., Laber,B., Assfald-Machleidt,I., Esterl,A., Wiegand,G., Kos,J., Turk,V. and Bode,W. (1989) *FEBS Lett.*, **243**, 234–238.

Messerschmidt,A. and Pflugrath,J. (1987) *J. Appl. Crystallogr.*, **20**, 306–315.

Nikawa,T., Towatari,T., Ike,Y. and Katanuma,N. (1989) *FEBS Lett.*, **255**, 309–314.

Ohkubo,I., Kurachi,K., Takasawa,T., Shiokawa,H. and Sasaki,M. (1984) *Biochemistry*, **23**, 5691–5697.

Pflugrath,J.W., Saper,M.A. and Quiocho,F.A. (1984) In Hall,S. and Ashiaka,T. (eds), *Methods and Application in Crystallographic Computing*. Clarendon Press, Oxford, p. 407.

Priestle,J.P., Ford,G.C., Glor,M., Mehler,E.L., Smit,J.D.G., Thaller,C. and Jansonius,J.N. (1984) *Acta Crystallogr. Sect. A*, **40** (Suppl.), Hamburg Congress Abstracts C-17.

Ritonja,A., Machleidt,W. and Barrett,A.J. (1985) *Biochem. Biophys. Res. Commun.*, **131**, 1187–1192.

Sluyterman,L.A.A. and Wijdenes,J. (1970) *Biochim. Biophys. Acta*, **200**, 593–595.

Steigemann,W. (1974) Ph.D. Thesis, Technische Universität, München.

Teno,N., Tsuboi,S., Itoh,N., Okamoto,H. and Okada,Y. (1987) *Biochem. Biophys. Res. Commun.*, **143**, 749–752.

Thiele,U., Auerswald,E.-A., Gebhard,W., Assfalg-Machleidt,I., Popovič,T. and Machleidt,W. (1988) *Biol. Chem. Hoppe-Seyler*, **369**, 1167–1178.

Thiele,U., Assfalg-Machleidt,I., Machleidt,W. and Auerswald,E.-A. (1990) *IInd Brdo Research Conference, Biol. Chem. Hoppe-Seyler*, in press.

Turk,V., Brzin,J., Lenarčič,B., Locnikar,P., Popovič,T., Ritonja,A., Babnik,J., Bode,W. and Machleidt,W. (1985) In Khairallah,E. and Bonds,J. (eds), *Intracellular Protein Catabolism V*. Alan R.Liss, New York, pp. 91–103.

Received on February 12, 1990; revised on March 9, 1990