The EMBO Journal vol.10 no.9 pp.2321 – 2330, 1991

The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity

D.Musil, D.Zucic, D.Turk, R.A.Engh, I.Mayr, R.Huber, T.Popovic[1], V.Turk[1], T.Towatari[2], N.Katunuma[2] and W.Bode

Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG, [1]Department of Biochemistry, Jozef Stefan Institut, Jamova 39, YU-61000 Ljubljana, Yugoslavia and [2]Department of Enzyme Chemistry, Institute for Enzyme Research, Tokushima University, Tokushima 770, Japan

Communicated by R.Huber

From the lysosomal cysteine proteinase cathepsin B, isolated from human liver in its two-chain form, monoclinic crystals were obtained which contain two molecules per asymmetric unit. The molecular structure was solved by a combination of Patterson search and heavy atom replacement methods (simultaneously with rat cathepsin B) and refined to a crystallographic R value of 0.164 using X-ray data to 2.15 Å resolution. The overall folding pattern of cathepsin B and the arrangement of the active site residues are similar to the related cysteine proteinases papain, actinidin and calotropin DI. 166 α-carbon atoms out of 248 defined cathepsin B residues are topologically equivalent (with an r.m.s. deviation of 1.04 Å) with α-carbon atoms of papain. However, several large insertion loops are accommodated on the molecular surface and modify its properties. The disulphide connectivities recently determined for bovine cathepsin B by chemical means were shown to be correct. Some of the primed subsites are occluded by a novel insertion loop, which seems to favour binding of peptide substrates with two residues carboxy-terminal to the scissile peptide bond; two histidine residues (His110 and His111) in this 'occluding loop' provide positively charged anchors for the C-terminal carboxylate group of such polypeptide substrates. These structural features explain the well-known dipeptidyl carboxypeptidase activity of cathepsin B. The other subsites adjacent to the reactive site Cys29 are relatively similar to papain; Glu245 in the S2 subsite favours basic P2-side chains. The above mentioned histidine residues, but also the buried Glu171 might represent the group with a pK_a of ~5.5 near the active site, which governs endo- and exopeptidase activity. The 'occluding loop' does not allow cystatin-like protein inhibitors to bind to cathepsin B as they do to papain, consistent with the reduced affinity of these protein inhibitors for cathepsin B compared with the related plant enzymes.

Key words: cathepsin B/crystal structure/cysteine proteinase/exopeptidase specificity

Introduction

The lysosomal cysteine proteinases play an important role in intracellular protein degradation (see Barrett and Kirschke, 1981; Brocklehurst *et al.*, 1987; Barrett *et al.*, 1988). Of these proteinases, cathepsin B is the most abundant and the most thoroughly studied. Besides its involvement in intracellular protein turnover, it has been implicated in tumour metastasis (see Sloane, 1990) and in other disease states (see Mort *et al.*, 1984), and might also be involved in extra-lysosomal processing of protein precursors (see Chan *et al.*, 1986).

Cathepsin B exhibits optimal activity in slightly acidic media and is irreversibly inactivated at alkaline pH values (Willenbrock and Brocklehurst, 1985a). Cathepsin B acts as an endopeptidase with relatively broad specificity and a slight preference for basic residues or phenylalanine at P2 (using the nomenclature of Schechter and Berger, 1967). Bulky side chains at P1 are disfavoured (Shaw and Kettner, 1981; Shaw *et al.*, 1983). A remarkable feature of cathepsin B is its distinct peptidyl dipeptidase activity (Aronson and Barrett, 1978; Bond and Barrett, 1980; Takahashi *et al.*, 1986; Polgar and Csoma, 1987) to liberate dipeptides sequentially from the carboxy terminus. Cathepsin B is inhibited by typical cysteine proteinase protein inhibitors such as cystatins and stefins; among the related cysteine proteinases, however, cathepsin B is lowest ranking in affinity (Nicklin and Barrett, 1984; Popovic *et al.*, 1988; Thiele *et al.*, 1990). Cathepsin B is readily inhibited by typical thiol blocking reagents, such as epoxy succinyl derivatives (Hashida *et al.*, 1980; Barrett *et al.*, 1982), peptidyl diazomethyl ketones and sulphonium salts (see Shaw, 1990).

The complete amino acid sequences of rat (Takio *et al.*, 1983), human (Ritonja *et al.*, 1985) and bovine cathepsin B (Meloun *et al.*, 1988) and the partial sequence of the porcine species (Takahashi *et al.*, 1984) have been communicated. According to the nucleotide sequences (Chan *et al.*, 1986; Fong *et al.*, 1986; Ferrara *et al.*, 1990), cathepsin B from man, rat or mouse is synthesized as a 339 amino acid polypeptide chain, which is processed to the mature single-chain molecule of 254 amino acid residues (Nishimura and Kato, 1987; Kominami *et al.*, 1988). In mammalian tissues, most of the active cathepsin B is found as a two-chain molecule consisting of 47 (or 49) and 205 (or 204) residues polypeptide chains (light and heavy chain) covalently cross-linked by a disulphide bridge.

The cathepsin B sequence indicated a close structural homology with the plant proteinase papain (Takio *et al.*, 1983). Comparisons with the sequences of the related cathepsins L and H and of papain and actinidin resulted in alignment proposals for cathepsin B (Takio *et al.*, 1983; Kamphuis *et al.*, 1985). Based on the three-dimensional structures of papain (Kamphuis *et al.*, 1984) and actinidin (Baker, 1980) the common structural features as well as sites of insertions and deletions were made more precise (Kamphuis *et al.*, 1985; Baker and Drenth, 1987).

Cathepsin B is, however, considerably larger than papain or actinidin, and the accommodation of some of the longer

polypeptide insertions and its spatial proximity to the active site residues remained unclear. Human cathepsin B possesses 14 cysteine residues (Ritonja *et al.*, 1985), with Cys29 representing the active site cysteine, and 12 other cysteine residues presumed to be involved in disulphide bridges. Very recently, the seven disulphide connectivities in the related bovine enzyme (with 16 cysteine residues) have been determined (Baudys *et al.*, 1990) resulting in a somewhat novel peptide alignment and a revised three-dimensional model for cathepsin B.

However, a clear understanding of its specificity and of its catalytic properties requires the availability of an experimental structure as provided by X-ray crystallography. Spindle-like crystals from rat liver cathepsin B were described a decade ago by two of us (Towatari *et al.*, 1979), and very recently monoclinic crystals of recombinant rat cathepsin B (Lee *et al.*, 1990) were reported, but no structure has been communicated so far. We have grown suitable crystals of human liver cathepsin B and determined its structure. In this publication, we present for the first time the folding of cathepsin B and those structural features which contribute essentially to its unique specificity, mainly based on the refined structure of the human enzyme. In a separate paper (D.Zucic *et al.*, in preparation) we will also present analysis and structure of the closely related rat enzyme at pH 5, describe the rat and the human cathepsin B structures in more detail and compare them with the related cysteine proteinases.

Results and discussion

Overall structure of cathepsin B

Cathepsin B is roughly disc shaped with a diameter of 50 Å and a thickness of 30 Å and a marked incision representing the active site cleft (on 'top' of the molecule, Figure 1). The polypeptide chain, similar to that of papain and actinidin (Baker and Drenth, 1987) and calotropin DI (Heinemann *et al.*, 1982), is folded into two distinct domains which interact with one another through an extended polar interface which opens to the V-shaped active site cleft. The left-hand 'L' domain (Figure 1) (to adopt the nomenclature introduced for papain) is formed by the amino-terminal half of the polypeptide chain (except about the first 10 residues) up to about Tyr148 (cathepsin B single chain nomenclature) and by the last (about four) carboxy-terminal residues. The right-hand 'R' domain (Figure 1) comprises the first 10 residues and the carboxy-terminal half of the polypeptide chain (except for about the last four residues). Both the amino- and the carboxy-terminal ends act as 'straps' which assist in clamping both domains together with the segment around Tyr148–Tyr150 and various non-covalent inter-domain contacts. The distinct secondary structure elements of cathepsin B defined according to hydrogen bonding are delineated in Figure 2.

Human cathepsin B possesses six disulphide bridges which are confined to the amino-terminal half of the molecule. The disulphide connectivities shown in Figure 2 are in full agreement with those recently proposed (Baudys *et al.*, 1990) for bovine cathepsin B (which possesses, however, an additional seventh disulphide at positions 148–252, i.e. at the inter-domainal interface). Two cysteine residues of cathepsin B, namely Cys29 and Cys240, are unpaired. Cys29 is topologically equivalent to the reactive cysteine of all other cysteine proteinases (Cys25 in papain); Cys240 is located close to the R-domain surface and is unique to cathepsin B.

The crystalline cathepsin B is of the double-chain form, i.e. it consists of one light and one heavy chain comprising residues Leu1–Asn47 and Val50–Asp254, respectively. The peptide chain is well defined by electron density from Leu1 onwards; there is, however, no appropriate electron density for the last two residues (Thr253 and Asp254), and both intermediate terminal residues left upon excision of the dipeptide Ala48–His49 (Chan *et al.*, 1986), Asn47 and Val50, are likewise conformationally undefined. This non-structured cleavage site is located on the surface spatially close to the likewise undefined carboxy-terminal end of Thr253–Asp254.

Comparison with papain and actinidin

166 amino acid residues of cathepsin B are topologically equivalent as defined by the program OVRLAP (Rossmann and Argos, 1975) with selection parameters $E_1 = E_2 = 2.0$) to papain (Kamphuis *et al.*, 1984; Stubbs *et al.*, 1990) with an optimal r.m.s. deviation of 1.04 Å for α-carbon atoms. Figure 3 shows the α-carbon chain of cathepsin B together with that of papain after optimal superposition of these 166 α-carbons. The helical segments of the L domain and the barrel-forming strands of the R domain superimpose most closely; neither domain is significantly shifted or rotated relative to the other. The active site residues, in particular the reactive Cys29 (25 in papain) involved in covalent intermediate formation, are among those showing the

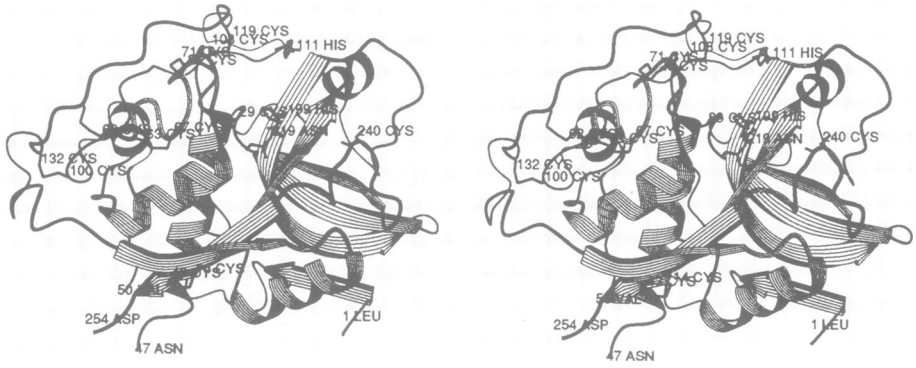

Fig. 1. Ribbon-plot of the human cathepsin B polypeptide chain; β-strands, helices and turns are represented by (twisted) arrows, helical ribbons and ropes (Priestle, 1988, modified by A.Karshikoff). The L and the R domains are on the left- and right-hand side, respectively. The view is along the active site cleft (in the upper half) with the 'occluding loop' at the back. Disulphide connections and some residues particularly addressed in the text are also given with full structure.

smallest deviations. 59 topologically equivalent residues are identical residues; thus the amino acid homology amounts to 24% (based on 248 residues). Polypeptide segments with identical or similar residues are confined to more central parts, whereas most of the deletions and insertions occur on the molecular surface (Figures 2 and 3).

Table I shows the alignment of these 166 amino acid residues of human cathepsin B and of papain according to topological equivalence. Thirteen segments deviate strongly from papain. Cathepsin B is considerably larger [254 residues, compared with 212 (papain) and 220 (actinidin)], and most of the large structural deviations are due to residue insertions. We therefore use the cathepsin B numbers; for comparison the corresponding papain numbers are added in brackets to the cathepsin B residue number.

Among the non-equivalent residues, 88 cathepsin B

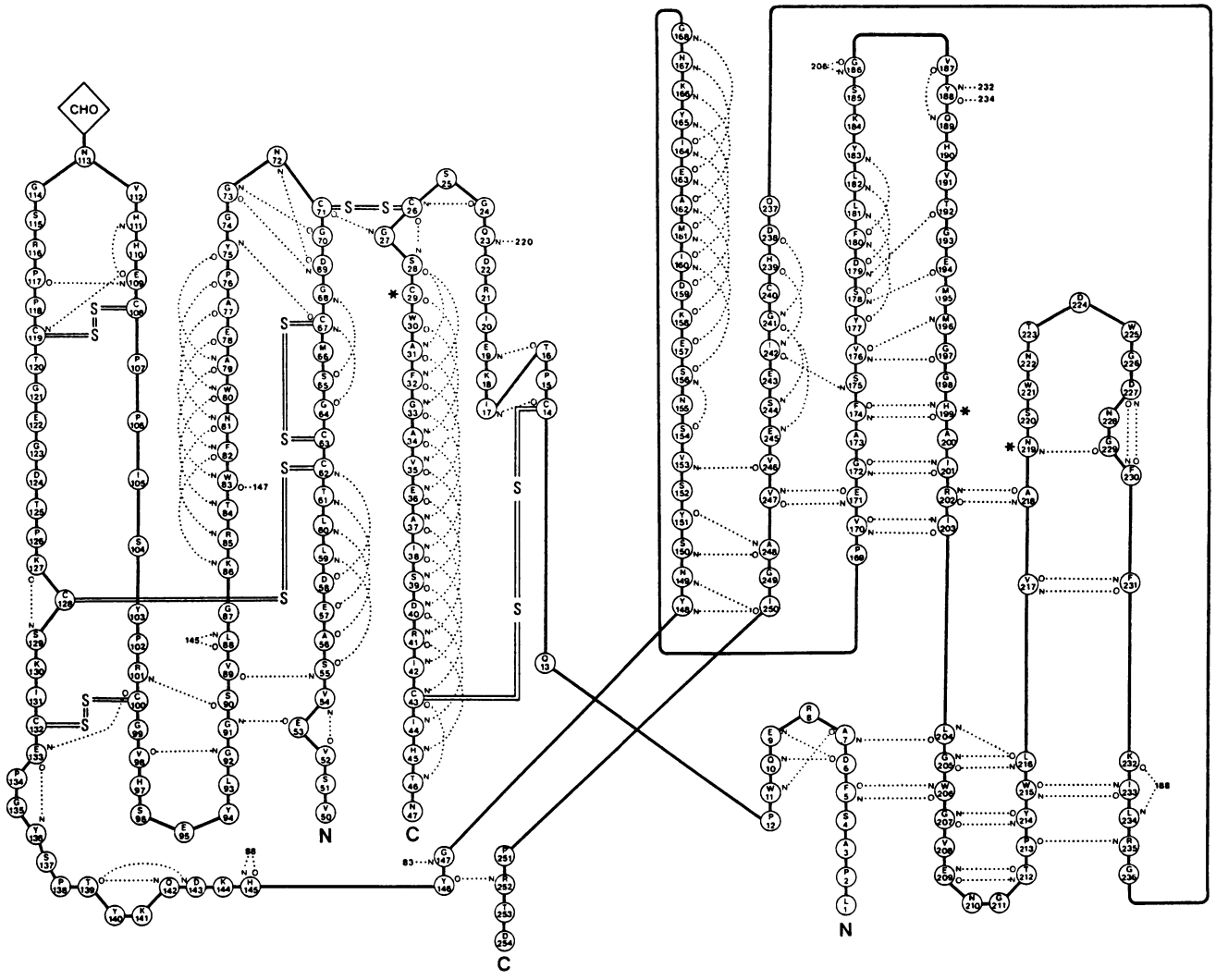

Fig. 2. Secondary structure of human cathepsin B (molecule 1) with segments making up the L and R domain on the left and on the right-hand side, respectively. Inter-main-chain hydrogen bonds are selected according to criteria given by Kabsch and Sander (1983) (using E < −0.7 kcal/mol). The disulphide connectivities are indicated, and the residues of the active site 'triad' (*) and the glycoslyated Asn113 are labelled.

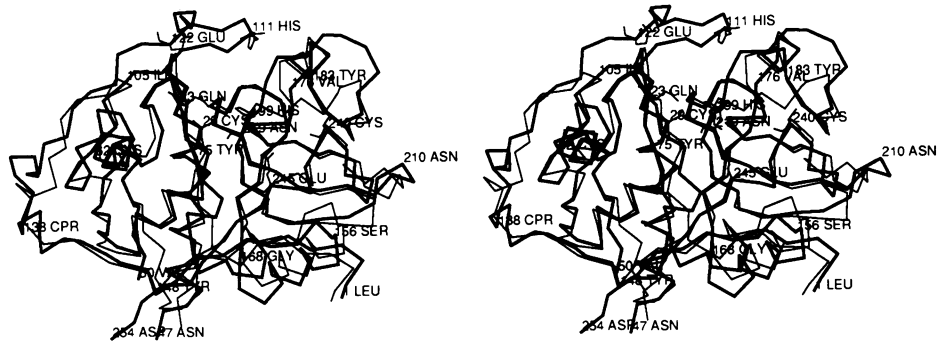

Fig. 3. CA structure of human cathepsin B (thick connections) superimposed with papain (thin connections) (Kamphuis *et al.*, 1984). The view is identical to that in Figure 1. The active site residues and some other residues are given with full structure.

Table I. Alignment of cathepsin B and papain; non-equivalent papain residues are given in lower case letters

Cath B	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
	L	P	A	S	F	D	A	R	E	Q	W	P	Q	C	P	T	I	K	E	I	R	D	Q	G	S	C	G	S	C	W	A	F	G	A	V	E	A	I	S	D
Papain	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15					16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
	I	P	E	Y	V	D	W	R	Q	k	g	a	v	t	p					V	K	N	Q	G	S	C	G	S	C	W	A	F	S	A	V	V	T	E	E	G
Cath B	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
	R	I	C	I	H	T	N	A	H	V	S	V	E	V	S	A	E	D	L	L	T	C	C	G	S	M	C	G	D	G	C	N	G	G	Y	P	A	E	A	W
Papain	37	38	39	40	41	42	43	44			45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60			61	62	63	64	65	66	67	68	69	70	71	72
	I	I	K	I	R	T	g	n			L	N	Q	Y	S	E	Q	E	L	L	D	C	D	R	r	s			Y	G	C	N	G	G	Y	P	W	S	A	L
Cath B	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
	N	F	W	T	R	K	G	L	V	S	G	G	L	Y	E	S	H	V	G	C	R	P	Y	S	I	P	P	C	E	H	H	V	N	G	S	R	P	P	C	T
Papain	73	74	75	76	77	78	79	80	81	82	83	84								85	86	87	88	89	90	91	92													
	Q	L	V	A	Q	Y	G	I	H	Y	r	n								T	Y	P	Y	E	g	v	q													
Cath B	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154		155	156	157	158	159
	G	E	G	D	T	P	K	C	S	K	I	C	E	P	G	Y	S	P	T	Y	K	Q	D	K	H	Y	G	Y	N	S	Y	S	V	S	–	N	S	E	K	D
Papain						93	94	95	96	97	98	99	100	101	102	103	104								105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
						R	Y	C	R	s	r	e	k	g	p	y	a								A	K	T	D	G	V	R	Q	V	Q	p	y	N	Q	G	A
Cath B	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175		176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198
	I	M	A	E	I	Y	K	N	G	P	V	E	G	A	F	S	–	V	Y	S	D	F	L	L	Y	K	S	G	V	Y	Q	H	V	T	G	E	M	M	G	G
Papain	121	122	123	134	125	126	127		128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	
	L	L	Y	S	I	A	N		Q	P	V	S	V	V	L	Q	a	a	G	K	D	F	Q	L	Y	R	G	G	I	F	V	G	p	c	g	n	k	v	d	
Cath B	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235			
	H	A	I	R	I	L	G	W	G	V	E	N	G	T	P	Y	W	L	V	A	N	S	W	N	T	D	W	G	D	N	G	F	F	K	I	L	R	–	–	–
Papain	159	160	161	162	163	164	165	166	167	168	169					170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194
	H	A	V	A	A	V	G	Y	G	p	n					Y	I	L	I	K	N	S	W	G	T	G	W	G	E	N	G	Y	I	R	I	K	R	g	t	g
Cath B		236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254																				
	–	G	Q	D	H	C	G	I	E	S	E	V	V	A	G	I	P	R	T	D																				
Papain	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212																						
	n	s	y	g	V	C	G	L	Y	T	S	S	F	Y	P	v	k	n																						

residues replace 46 residues of papain. Five of the non-equivalent segments in cathepsin B are only one or two residues longer than in papain; the segment between residues 190 and 199 where both chains are of almost equal length (eight cathepsin B residues versus seven papain residues) illustrates nicely that small differences in length can be paralleled by large spatial deviations extending over a larger molecular area (see below). At four other non-equivalent segments, however, four or seven additional residues (see Table I) are inserted in the cathepsin B chain leading to quite different structures extending over longer adjacent peptide segments. By far the largest insertion is observed between residues 104 and 126 (Figure 3), where three residues in papain (and actinidin) correspond to 21 cathepsin B residues which are arranged to form a completely novel surface structure most probably with significant implications for its specificity (see below).

Most of the insertions and deletions had been predicted correctly (Kamphuis *et al.*, 1985). However, the insertions of cathepsin B segments 10–19 (Trp11 has no counterpart in papain, for example), 91–99 and 130–144 (Figure 3) were incorrect; the observed equivalence of segment 145–154 was likewise not predicted. The full length of the large insertion between Ser104 and Pro126 was not foreseen, and Cys119 was erroneously expected to be equivalent to Cys95 of papain. Very recently the possible projection of this loop into the active site cleft was proposed on the basis of the experimentally determined disulphide connection Cys108–119 (Baudys *et al.*, 1990).

Two of the six disulphide bridges of cathepsin B [26–71(22–63) and 62–128(56–95)] are topologically similar to equivalent connections in papain and actinidin. The other four are specific to cathepsin B and cross-connect polypeptide segments of either slightly (14–43, 63–67 and 100–132) or completely different architecture (108–119). The third disulphide bridge of papain (153–200) likewise finds no counterpart in cathepsin B. Cys240(200) is topologically equivalent, but His190–Thr192 of cathepsin B takes a very different course unsuitable for a disulphide connection to this segment.

The L domain of cathepsin B

The central building elements of the L domain are three helices A, B and C (Figure 1), which are packed approximately perpendicularly, and which extend from Ser28 to His45, from Ser55 to Cys63, and from Tyr75 to Arg85 respectively, with regular α-helical conformations (see Figure 2). These helices as well as the entering, the connecting and the exiting strands' up to Ser90, have topological equivalents in the plant enzymes. The dipeptide Ala48–His49 is excised in the course of processing, creating new termini which remain close to each other; in the single-chain cathepsin B precursor, this segment is probably located on the surface, accessible to proteinases.

The adjacent chain segment from Ser90 to His145 of cathepsin B is much longer than the equivalent strands in papain and actinidin (see Figures 2 and 3 and Table I) and is organized in three wide hairpin loops of novel architecture. The loops contact one another through a few side chains and thus create a completely new surface patch. The first of them, Ser90–Arg101, is a twisted, irregular β-hairpin loop, which is an extension of the shorter turns in papain and actinidin.

About five residues of the next inter-loop segment (Cys100–Ser104) are similarly arranged in papain. At Ile105 the cathepsin B chain, however, changes direction and forms a wide loop of novel structure located on 'top' of the molecule (Figures 1 and 3) at the 'primed site' end of the active site cleft, which it partially 'occludes' (see below). 20 residues of cathepsin B between Ile105 and Pro126 have no equivalence in papain or actinidin where a short connection of only two residues exists. Nevertheless, the underlying structures (Asp22–Ser25 and Trp221–Trp-225), although partially buried in cathepsin B are structurally and sequentially similar to those in papain where they are exposed.

The cathepsin B segment crosses over itself at Cys108 and

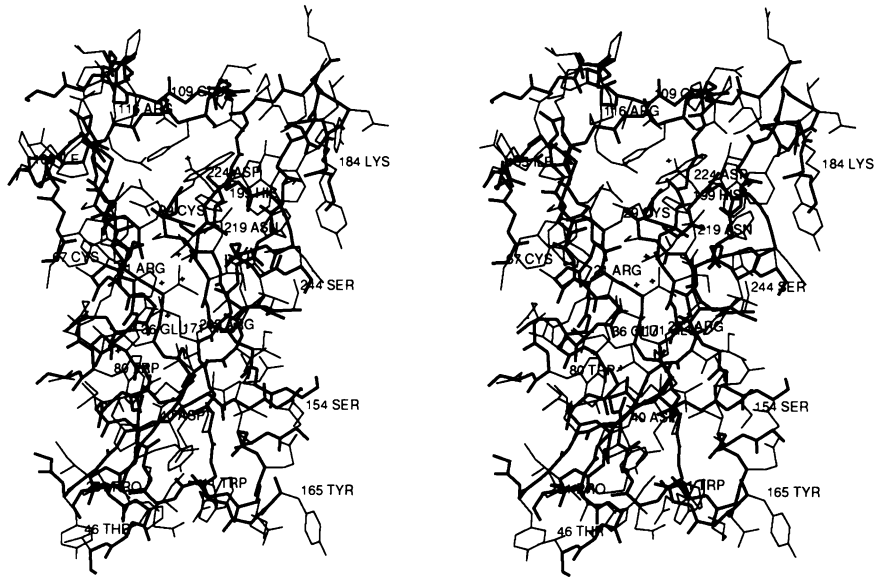

Fig. 4. Inter-domainal interface of cathepsin B. The main chain atoms are given in thick connections, and crosses mark internal solvent molecules. Most of the charged residues addressed (in particular Glu171 in 'front', Arg21, Glu36, Arg202 and Asp40 in the 'centre', and Glu109, Asp224, Arg116 on 'top') in the text are labelled. View as in Figure 1 and 3.

Cys119, and a novel disulphide bridge between both cysteine residues creates a covalently closed circular structure, the 'occluding loop'. Both flanking peptides leading to and away from this occluding loop contain Pro-Pro-Cys sequences; both segments, Pro106-Pro107-Cys108 as well as Pro117-Pro118-Cys119, exhibit quite regular polyproline II conformations and may confer stability to this loop, along with three inter-main chain hydrogen bonds located close to the disulphide bridge (Figure 2). The side chain of Glu109 (see Figure 4) projects towards the centre and interacts through hydrogen bonds with the amide nitrogens of Ser115 and Arg116; also, it makes electrostatic/hydrogen bond interactions with the adjacent side chains of Asp224 and Arg116 (see below). The side chains of His110 and His111, which are implicated in exopeptidase activity (see below), and Val112 act together with a buried water molecule to separate the loop from the supporting, highly conserved segment 220–230 (Figure 4). All other side chains of this loop, in particular those of Asn113 and Ser115, project away from the molecular surface. The Asn113 side chain, which is glycosylated in cathepsin B (Takio *et al.*, 1983; Ritonja *et al.*, 1985), is relatively flexible but still well defined by electron density; in neither molecule, however, does the density indicate the position of an asparagine-linked sugar, although the inter-molecular space in the crystal could accommodate it. This result is in accord with the finding of only minor amounts of sugar in human cathepsin B (see Barrett and Kirschke, 1981) and in cathepsin B from other species (see Takahashi *et al.*, 1984).

The cathepsin B segment Lys130–Lys144 forms the third L domain hairpin loop of novel architecture (the analogous but shorter segments of the plant cysteine proteinases also deviate here from one another despite their identical lengths). This loop segment is clamped to its support through a novel disulphide bridge (110–132) and through internal salt bridges made by Asp58 (topologically equivalent to a charge-uncompensated, buried Glu52 in actinidin) with Lys144 and Arg101. Pro138 is in *cis* conformation. From His145 onwards, the cathepsin B chain runs (as in the plant cysteine proteinases) in an extended conformation to the R domain, antiparallel to the carboxy terminus (Figure 2) and making a prominent bend at Tyr148.

The L domain of cathepsin B possesses a quite hydrophobic core centred between the three constitutive helices, which provide most of the 18 hydrophobic residues of cathepsin B. Several of them have identical counterparts in the papain structure; Trp80 and Trp83 replace, however, smaller hydrophobic residues. The hydrophobic core is separated from another hydrophobic satellite (Trp11, Ile17 and Ile44, located in the 'back' of the L domain, Figures 1 and 3) by an inter-domainal polar layer formed by three internal but charged residues (Arg21, Glu36 and Asp40) and water molecules of the interface channel (see Figure 4 and below). The guanidyl group of Arg21 is not hydrogen-bonded to a carboxylate, but is surrounded by a few internal solvent molecules and is close to the likewise unbalanced side chain of Glu36. Three charged residues (Asp22, Glu36 and Asp40) project into the polar interface made with the R domain (see below).

The R domain of cathepsin B

The carboxy-terminal half of the polypeptide chain possesses six extended strands which are (in the sequential order 1, 6, 2, 3, 4, 5; see Figures 1 and 2) aligned antiparallel and form an extended β-pleated sheet. This sheet is highly twisted and bent into a barrel-like structure which is, however, not closed (see Figure 2). Similar to the plant enzymes, this barrel structure is sealed on both ends by a relatively regular α-helical segment, Ser156–Gly168 (at the 'bottom' of the domain, Figure 1) and by a wide loop structure consisting of a helical segment (Tyr177–Tyr183) of irregular 3_{10}-helical conformation and a turn-strand-pair up to Thr192 (on 'top' of the domain, Figure 1). At Phe5/Ala7 the amino-terminal segment of cathepsin B crosses at the fourth strand (at Leu204/Gly205) and functions as a third strand in the sheet before Val170–Asp179 takes this position (see Figure 2).

Most of the barrel-forming strands are topologically similar to those in the plant enzymes, in spite of some quite different residues (e.g. segment 148–156, Table I). Also

the large multiple-turn loop which (at Ser178) emerges from the third strand exhibits a remarkably similar shape. From Val191 up to Gly198 the cathepsin B chain, although longer by only one residue, loops over a different part of the molecular surface than does the equivalent segment 151–158 in papain. The His190 side chain is located in cathepsin B at about the site occupied in papain by Cys153 (which is involved in a disulphide bridge formed with Cys200). In cathepsin B there is, of course, also no equivalent residue for the preceding *cis* Pro152 found in the plant enzymes.

Loop segment Asn219–Gly229, which connects the fifth and the sixth barrel strands (see Figures 1 and 3), extends towards the L domain as in the plant enzymes, but in cathepsin B forms the support for the unique 'occluding loop' Cys108–Cys119 ('behind' the active site cleft, Figures 1 and 5). Asn222 exhibits (as in actinidin) a strained main chain conformation (+66°/−175°). The second part of this loop is arranged in multiple adjacent turns, with segment Trp225–Phe230 folded into a short left-handed helix.

Segment Arg235–Val246, which connects the sixth and the second barrel strands (Figures 1 and 2), is organized on the surface as a multiple-turn structure, partially of a left-handed 3_{10}-helical conformation. In contrast to the buried Cys200 side chain of papain (where it is disulphide-linked with Cys153) the topologically equivalent Cys240 side chain is unpaired and positioned in a slight depression framed by the side chains of five aromatic residues. Although its thiol group is effectively removed from the surface, both mercury compounds bind with high occupancy.

The terminal β-strand (number 2 in the sheet) traverses to the L domain at about Gly249. A disulphide bridge between residues 148 and 252 as observed in bovine cathepsin B (Baudys *et al.*, 1990) is also in agreement with the structure of human (and rat) cathepsin B. The two terminal residues (Thr253 and Asp254) are exposed to bulk solvent and are disordered in human cathepsin B.

The whole R domain barrel of cathepsin B encloses a large and completely hydrophobic core, formed by about 19 hydrophobic residues. Noteworthy is the large portion of aromatic residues in cathepsin B, some of which are identical with plant enzymes residues. Almost all of the charged residues are located on the surface; three charged residues (Glu171, Arg202 and Asp224) extend into the interface (see below).

The inter-domainal interface of cathepsin B

Both domains are in contact, enclosing a water-filled channel. In the upper half of the molecule (Figure 1) both domains diverge, leaving the V-shaped active site cleft, which is blocked at the 'rear' (Figure 1) by the 'occluding loop'. Only a few hydrophobic residues are in direct contact across the interface (Figure 4).

Most of the interface contacts are made by polar and charged residues. The side chain of Glu171 (in 'front', Figure 4) is completely buried in a quite hydrophobic, relatively conserved environment, but lacks an adjacent, positively charged neighbour. One of its carboxylate oxygens is engaged in a hydrogen bond with Trp80 NE; the second oxygen is in a favourable position (see Figure 6) to form (if protonated) a hydrogen bond with Trp30 O (which itself is significantly rotated out of the helix axis direction and therefore only engaged in a quite weak intra-helix hydrogen bond of 3.6 Å with Ala34 N); the acceptance (if deprotonated) of a hydrogen bond from Ala34 N (3.1 Å, see Figure 6) would be probable only after considerable rotation of the amide group 33–34 out of the helix axis. This in turn might affect the nearby active site residues (carboxylate171–Cys29 SG 7.3 Å). It is noteworthy that the same structure is observed in rat cathepsin B at pH 5; these observations indicate that Glu171 may have a higher pK_a (>5).

A few other charged side chains project with their side chains into the channel which is filled with six hydrogen bond-connected internal solvent molecules. Within this channel, Asp40 forms a salt bridge with Arg202, and Glu36 projects into it, sandwiched between the side chains of Arg21 and Arg202 (Figure 4). The channel has a shape and polarity as in the plant enzymes; however, most of the charged residues involved are not topologically equivalent (e.g. Glu171, Arg202, Glu36, Ala218 and Ser39 of cathepsin B (Figure 4) are replaced by Ser131, Val/Ala162, Ala/Val32, Lys174 and Glu35 in papain/actinidin).

The inter-domainal channel is (on 'top', Figure 4) 'sealed' by the exposed loop 220–230, in the centre of which another

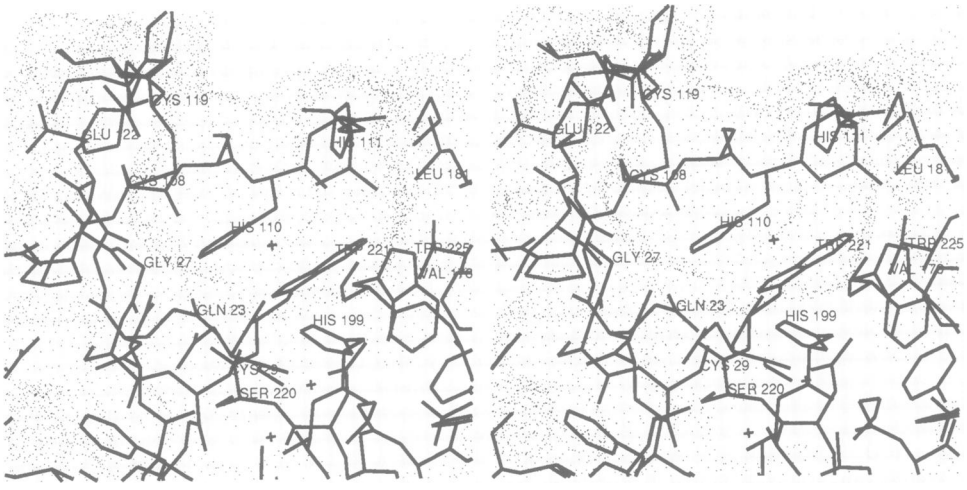

Fig. 5. Part of the active site cleft and the 'occluding loop' of human cathepsin B, displayed together with the equivalent Connolly dot surface (made with MAIN). View as in Figures 1, 3 and 4.

isolated water molecule is encaptured. The unique 'occluding loop' 108–119 floats on this support via another buried water molecule and a few side chain interactions. Asp22 can form a buried salt pair with His110 (via NE2). Asp224 makes a hydrogen bonded salt-bridge with Arg116; one of its oxygens is furthermore in hydrogen bond distance (2.8 Å) to one of the carboxylate oxygens of Glu109 which should be protonated to allow this interaction (Figure 4). Again, this geometry remains the same at pH 5 as judged from the rat cathepsin B structure.

The active site cleft of cathepsin B

The active site cleft of cathepsin B located on 'top' of the interface region is, compared with papain, blocked in the 'rear' (Figures 3 and 5) by the characteristic 'occluding loop'. From this loop the adjacent imidazole side chains of His110 (salt-bridged with Asp22, see above) and His111 (hydrogen bonded only to solvent molecules) project towards the active site residues; segment Cys119–Glu122 exiting from this loop forms a higher rim on the left-hand side (Figures 3 and 5). Similar to the plant enzymes, the remainder of the cleft is formed by the curled segment Gly24–Trp30 and by the extended strand Asn72(64)–Pro-75(67) on the left-hand side, and by segments Phe-180–Leu182 and Met196–His199 on the right-hand side. The latter segment up to Gly198, however, is in a quite different conformation; Gly198 is topologically equivalent with Asp158, a non-essential residue in papain; it only influences the pK_as of the thiolate/imidazolium couple (Menard *et al.*, 1990).

The floor of the cleft is, as in the related plant enzymes, mainly made by the side chains of Trp221(177), Gln23(19), His199(159), Ala200 and Ala173(133). The side chain of the reactive Cys29(25) is (as well as the adjacent residues, see Figure 6) well defined by electron density. The two peaks ('ears') in the difference electron density on either side (Figure 6) presumably indicate the partial presence of two oxygens of a tetrahedral sulphinic acid group. One of the oxygens would be within hydrogen bonding distance of Cys29 N and Gln23 ND1 thus occupying the 'oxyanion hole' (Drenth *et al.*, 1976) of cathepsin B; the second oxygen is placed close to Trp30(26) N and to Ala200(160) N. The SG atom of Cys29 [presumed to be a thiolate of pK_a 3.4 (Willenbrock and Brocklehurst, 1985b)] is unfavourably positioned to form a hydrogen bond with the imidazolyl ND1 atom of the catalytic His199(159) (distance 3.6 Å). The

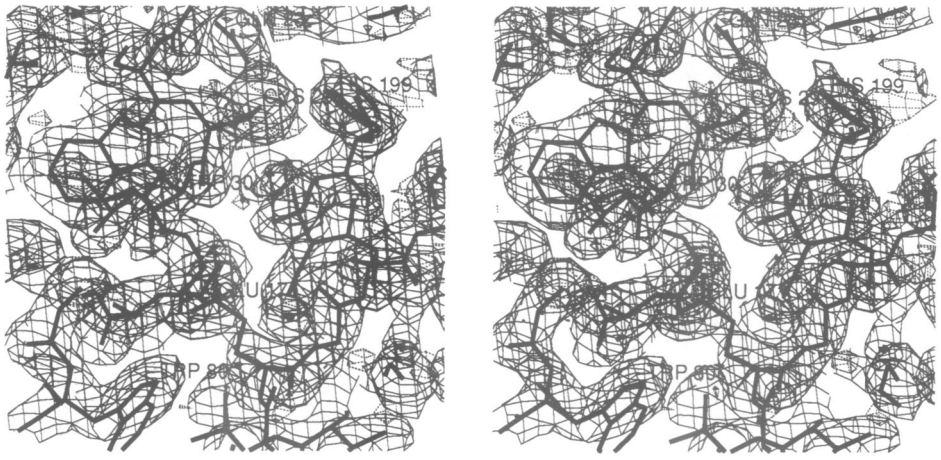

Fig. 6. Part of the structure of human cathepsin B (molecule 1) around the reactive Cys29(25) superimposed by the final 2 $F_{obs}-F_{calc}$ electron density and the positive difference density (contouring levels are 1.0 and 2.0 σ); the residual positive density at Cys29 SG indicates the additional (partial) presence of two oxygens. Also visible is Glu171 and surrounding possible hydrogen-bonding partners. View as in Figures 1, 3–5.

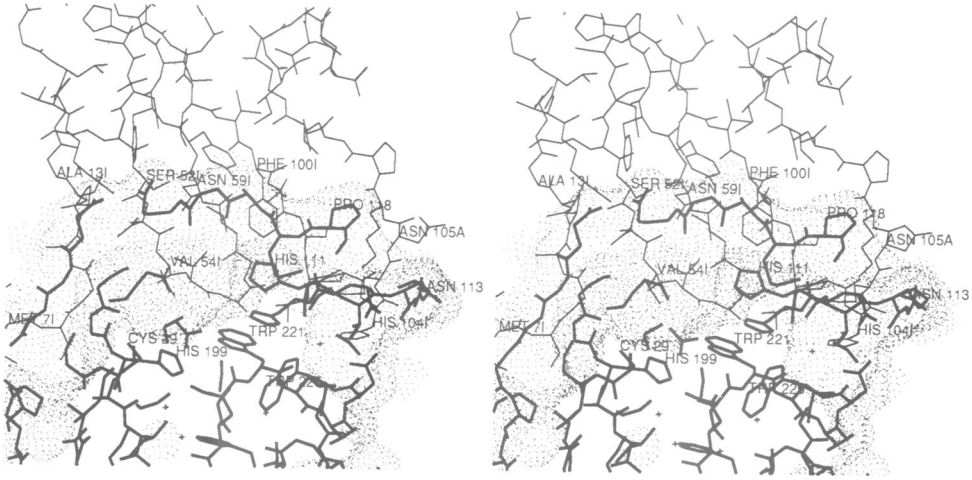

Fig. 7. Binding region and Connolly surface of cathepsin B (thick connections) together with the binding 'edge' of stefin B arranged as it binds to papain (Stubbs *et al.*, 1990). The view is perpendicular to Figures 1, 3–6 and across the active site cleft, with the 'occluding loop' on the right-hand site. Stefin B would collide with cathepsin B (mainly through its 'second hairpin loop' and the 'occluding loop', respectively).

imidazole NE2 forms a much shorter (2.6 Å) hydrogen bond with Asn219(175), which is (as in the plant enzymes) shielded by the side chain of Trp221(177); this indole moiety is, however, partially covered by the novel His111 in cathepsin B.

The presumed S2 subsite (in 'front', Figure 4) with Tyr75(67), Pro76(68), Ala77(69), Ala173(133) and Ala-200(160) is constructed similarly to that in papain/actinidin. Due to the bent chain at Gly198 and because of lack of a topologically equivalent residue for papain residue Val157, this S2 groove is larger, however. Gln245 (Ser/Met205 in the plant enzymes) provides an electrostatic anchor for positively charged P2 side chains.

Probable interaction of peptidyl substrates with cathepsin B

A peptide substrate will probably bind in the manner proposed by Drenth *et al.* (1976) for papain, with its scissile peptide bound to the active site Cys29 thiolate. The P1 carbonyl group may hydrogen bond to Cys29 N and Gln23 NE2, with the amide nitrogen of the P1′ residue close to ND1 of the imidazole of His199. The preceding main chain can be fixed, as in papain, by three hydrogen bonds with the carbonyl and the amide group of Gly74 and the carbonyl of Gly198. The P2 side chain can extend along the cleft floor (Ala200 and Ala173). The carboxylate group of Glu245 is well placed to make a salt bridge with the guanidyl group of an arginine (see Figure 4). The side chain of the P1 residue must project out of the cleft, in the direction of the intruding side chain of Glu122 (Figure 5), resulting in electrostatic side chain interactions, in agreement with the acceptance of arginines at P1. The following P1′ side chain should also project out of the cleft; because of a more hydrophobic surrounding at the bottom and the presence of two carbonyl groups (120 and 121) at the rim, both small hydrophobic as well as large polar (non-acidic) residues seem to be acceptable.

The exact position and conformation of the P2′ residue is more difficult to predict; it probably depends, however, on the length of the chain in the carboxy-terminal direction. In a substrate terminating with P2′, the carbonyl group of P1′ could be directed towards NE2 of Trp221. The P2′ side chain would find a quite hydrophobic shallow depression, surrounding the side chains of His199, Ala176, Phe180, Leu181 and the benzene part of Trp221 (Figure 5); the terminal carboxylate group would necessarily be near the exposed imidazole ring of His111 (and not too far from the adjacent His110 NE2) and make electrostatic interactions (in the case of protonated imidazolium groups), which would assist in fixing the substrate with respect to the active site. This interaction accounts well for the observed dipeptidyl exopeptidase activity of cathepsin B and would imply a positively charged imidazolium group(s). The pK_a values of His110 and His111 can be estimated from their environment to be higher and lower, respectively, than normal values for histidines. Thus His111 might represent the 'acidic' group which seems to control the exopeptidase activity of cathepsin B with a pK_a of ~5.5 (Polgar and Csoma, 1987). The strained substrate conformation presumably necessary at P1′ and P2′, and the hydrophobic nature of the presumed S2′ subsite could account for the 'inactivity' of cathepsin B for substrates with proline at P2′/P1′ and arginine at P2′ (Marks *et al.*, 1986).

The arrangement of the P2′ residue (and probably also of the P1′ residue) of a longer bound substrate must be different from this 'exopeptidase interaction' (and also different from the still hypothetical interaction in papain, due to the occluding loop of cathepsin B). In particular, the positively charged histidine residue(s) (not present in papain) might interfere with this 'endopeptidase interaction'. This may explain the increase in endopeptidase catalytic activity with pH governed by a group to be deprotonated at about pH 5.5 (Willenbrock and Brocklehurst, 1985b; Polgar and Csoma, 1987; Pohl *et al.*, 1987). Neutral histidines might also be preferred by 2-naphthylamide substrates.

Two of us have suggested that Arg202 might be responsible for exopeptidase activity (Takio *et al.*, 1983). Polgar and Csoma (1987) recently proposed a more elaborate structural model according to which an electrostatic cluster of three charged groups (Glu171, Arg202 and Glu245), presumed from modelling with papain to be close to one another in cathepsin B, should exist above pH 5.5; protonation of the acidic groups below pH 5.5 would release Arg202 from this cluster which would subsequently approach and affect the thiolate of Cys29. The side chains of the three residues are, however, neither in direct contact with one another nor located (except Glu245 which is at the extreme end of the S2 pocket) at the surface of the cleft. The guanidyl group of the buried residue Arg202 projects into the water-filled channel (Figure 4) and seems to be well integrated in an extended hydrogen bond network with Asp40 and internal solvent molecules. Also the adjacent internal Glu36, although lacking a hydrogen bonded compensating charge, appears likewise to be relatively insensitive to protonation/deprotonation. Glu171 is, however, an interesting candidate due to its completely buried carboxylate group (Figures 4 and 6) lacking not only any compensating charge, but (except after some conformational change, see above) also lacking a second hydrogen bond donor; deprotonation of Glu171 (with 7.3 Å the second closest ionizable group to Cys29 SG) with increasing pH values could well affect the structure (and thus activity) around the active site. All other ionizing groups are ≥ 10 Å from Cys29 SG. Recombinant methods should further clarify the role of several of these residues, in particular of His111 and Glu171.

Probable interaction with protein inhibitors

Hypothetical models of cathepsin B complexes formed with the low molecular weight protein inhibitors stefin B and chicken egg white cystatin were made by superimposing the papain components of the experimental papain–stefin B complex (Stubbs *et al.*, 1990) and the papain–cystatin docking model (Bode *et al.*, 1988), respectively. These docking experiments revealed, however, that the 'second hairpin loops' (Bode *et al.*, 1988) [in particular their turn parts at Leu102I-Pro103I-His104I-Glu105I (stefin B), see Figure 7, and Pro103I–Trp104I (cystatin)] would collide with segment His111-Val112-Asn113 of the cathepsin B occluding loop. Most of the sterical hindrance could, however, be released by some tilting and simultaneous rotation of either inhibitor relative to the active site cleft. Preliminary experiments indicate, however, that the fit will not be as good as for papain. Thus, the partial occlusion of the active site cleft of cathepsin B easily accounts for the reduction in affinity for roughly three to four orders of magnitude for the interaction with chicken cystatin and stefin

Table II. Data collection statistics

	Derivative		
	NATI	HGAC	MEHG
Number of measurements	32145	32529	30701
Number of independent reflections	18640	18900	17143
Resolution (Å)	2.15	2.28	2.37
R_{merge} after scaling and absorption correction	0.0861	0.0963	0.1237
R_{merge} of Friedel pairs after averaging	0.0504	0.0600	0.0569
Measured/possible reflections			
to 2.5 Å	0.872	0.923	0.901
to 2.15 Å	0.691		

Table III. Final model of human cathepsin B

Number of active protein atoms	3771
Number of inactive protein atoms	105
Number of active solvent atoms	283
R.m.s. standard deviation from target values	
bond lengths:	0.015 Å
bond angles:	2.565°
Total internal energy	−2363 kcal/mol
Resolution range	8−2.15 Å
Crystallographic R value	0.164
Mean temperature factor	16.1 Å^2
Lower temperature factor limit	3.0 Å^2

B, but less for stefin A (Popovic *et al.*, 1988; Thiele *et al.*, 1990); the considerable affinity dependence on the length of the amino-terminal 'trunk' of cystatin (Bode *et al.*, 1988, 1989) stresses the importance of the intermolecular interaction in the S2 subsite (Thiele *et al.*, 1990).

Materials and methods

Crystallization

Cathepsin B used for crystallization was isolated from human liver in a two-chain form according to procedures published previously (Zvonar *et al.*, 1979). Native crystals of dimensions up to 0.6 × 0.3 × 0.2 mm^3 were obtained at 20°C by hanging drop vapour diffusion against 1.4−1.5 M sodium phosphate at pH 3.5−4.0 and 1.1−1.25 M ammonium sulphate at pH 3.5−4.5. These crystals belong to the monoclinic space group $P2_1$ and have the cell constants a = 86.23 Å, b = 34.16 Å, c = 85.56 Å, $\alpha = \gamma = 90°$, β = 102.9°. The asymmetric unit contains two molecules corresponding to a volume per unit molecular mass of 2.2 Å^3/Da. Native crystals were transferred to 2.25 M sodium phosphate, pH 3.75, before data collection.

Two mercury derivatives, isomorphous to native crystals, were obtained in 2.0 M ammonium sulfate at pH 5.0 by treating native crystals with 1 mM dithiothreitol and subsequently (after dialyzation against thiol-free mother liquor) with 1 mM mercury acetate for 16 days at 20°C (HGAC) or with 6 mM methyl mercury chloride for 4 days (MEHG, see Table II).

Data collection and processing

X-ray diffraction data for all crystals were collected using the FAST television area detector diffractometer (Enraf-Nonius). The data were evaluated on-line with the program MADNES (Messerschmidt and Pflugrath, 1987), corrected for absorption effects using the program ABSCORR (Messerschmidt *et al.*, 1990) and loaded and scaled by means of PROTEIN (Steigemann, 1974). The statistics of the native and the derivative data are given in Table II.

Heavy atom refinement

Two independent positions for each mercury compound were found (similar positions for both derivatives; Table II) from difference Patterson maps. They were refined (with a mean figure-of-merit of 0.70 for data to 4.0 Å resolution) and phases were derived using PROTEIN. Later it turned out that both sites represented equivalent sites (Cys240) on each independent molecule, related by a local symmetry axis; the mercury atoms are covalently bound to the sulphur atoms.

Patterson search

Patterson searches for cathepsin B were done using a truncated model (comprising 867 out of 1631 atoms) constructed on a graphic system (PS390 Evans and Sutherland) by visual comparison of the refined models of papain (Kamphuis *et al.*, 1984) and actinidin (Baker, 1980) under consideration of the amino acid sequences (Ritonja *et al.*, 1985) and the proposed alignments (Kamphuis *et al.*, 1985).

A rotation search in Patterson space (PROTEIN) using a model Patterson map (8000 highest peaks from 15 to 3.5 Å) and a crystal Patterson map (calculated with intensity data from 8 to 3.5 Å) yielded two highest peaks of comparable heights (1st peak: 45.4; 2nd peak: 42.4; 3rd peak: 38.8; mean value 9.4; standard deviation 6.9). A self rotation search in Patterson space done with crystal data from 15.0 to 3.0 Å showed that the two independent molecules are related by an almost 90° rotation around an axis parallel to the crystallographic y-axis, in agreement with the results obtained in the cross rotation search. The height of these two peaks was ~85% of the maximal correlation of 0° rotation.

Translation searches using programs written by E.E.Lattman (modified by J.D.Deisenhofer and R.Huber) as well as using the correlation function implemented in XPLOR (Brünger, 1990) provided a consistent solution for only one of both molecules (molecule A). This particular structure was used to derive native phases and to calculate difference Fourier maps for both derivatives. Two sites were found consistent with the results obtained from heavy atom refinement, thus corroborating the translation solution and defining a common origin. The approximate position of the second independent protein molecule was determined by (i) rotating molecule A according to the results obtained from the rotation searches and (ii) shifting it according to the difference vectors between the two independent heavy atom positions. The correct position was then determined with a fine search using XPLOR's correlation function. This peak, with a correlation value of 0.18, disappeared into background noise when a grid size of ≥0.5 Å was used. Positions and orientations of both molecules were subsequently further refined by TRAREF (Huber and Schneider, 1985).

Model building and refinement

A 3 Å $2F_{obs}-F_{calc}$ Fourier map calculated with combined phases obtained from the partial model and form isomorphous replacement (Hendrickson and Lattman, 1970) was improved by averaging over the two independent molecules using the program MAIN (D.Turk). This averaged Fourier map displayed not only most parts of the conserved polypeptide segments, in particular the helices and the β-sheet, but in addition showed electron density for several omitted side chains, insertion loops and a novel disulphide bridge (Cys14−Cys43). These newly visible structural parts were modelled against the Fourier map on the graphic system using the PSFRODO version (Pflugrath *et al.*, 1984) of FRODO (Jones, 1978). The resulting model was refined using a combined crystallographic and energy refinement procedure with the program EREF (Jack and Levitt, 1978). Phases from this model were again combined with the MIR phases, the resulting Fourier map averaged, and the current model inspected and modelled etc.

At a certain stage (R = 0.323, resolution from 8 to 2.5 Å, 3077 active atoms corresponding to 84% of all protein atoms), however, the density representing large insertions (in particular the 'occluding loop', see above) remained broken and inappropriate for tracing the chain with certainty. At this time, crystals of rat liver cathepsin B became available and native data were collected. The current model of human cathepsin B was used to determine the position and orientation of the three independent molecules of rat cathepsin B (a more detailed description of its structure solution and structure will be published elsewhere, Zucic *et al.*, 1991). The Fourier map of rat cathepsin B, obtained by cyclic 3-fold averaging (MAIN, D.Turk), clearly displayed this structural part; this helped to complete model building of human cathepsin B. After several cycles of energy refinement (EREF) and further model building (FRODO) the crystallographic R value dropped to about 0.27 using 2.3 Å data; further cycles of rebuilding and refinement using simulated annealing (XPLOR) reduced the R value to 0.23. At this stage, water molecules were progressively added and individual B values refined. The final model refined by EREF has a R value of 0.164 (Table III).

Acknowledgements

We thank Dr J.Kos and U.Baumann for support. The financial support of the SFB207 of the LM Universität München (grants H-1 and H-2), of the Fonds der Chemischen Industrie, of the Research Council of Slovenia (grant

C1-0515-106) and of the Bundesministerium für Forschung und Technologie, BEO 21-0319208A, are gratefully acknowledged.

References

Aronson,N.N. and Barrett,A.J. (1978) *Biochem. J.*, **171**, 759–765.

Baker,E.N. (1980) *J. Mol. Biol.*, **141**, 441–484.

Baker,E.N. and Drenth,J. (1987) In Jurnak,F.A. and McPherson,A. (eds) *Biological Macromolecules and Assemblies: Vol. 3 Active Sites of Enzymes.* John Wiley and Sons, New York, pp. 314–367.

Barrett,A.J. and Kirschke,H. (1981) *Methods Enzymol.*, **80**, 535–561.

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

Barrett,A.J., Buttle,D.J. and Mason,R.W. (1988) *ISI Atlas for Science: Biochemistry*, pp. 256–260.

Baudys,M., Meloun,B., Gan-Erdene,T., Pohl,J. and Kostka,V. (1990) *Biol. Chem. Hoppe-Seyler*, **371**, 485–491.

Bode,W., Engh,R., Musil,D., Thiele,V., 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 Katunuma,N. and Kominami,E. (eds), *Intracellular Proteolysis: Mechanisms and Regulations.* Japan Scientific Societies Press, Tokyo, pp. 297–304.

Bond,J. and Barrett,A.J. (1980) *Biochem. J.*, **189**, 17–25.

Brocklehurst,K., Willenbrock,F. and Salih,E. (1987): Neuberger,A. and Brocklehurst,K. (eds), *Hydrolytic Enzymes.* Elsevier, Amsterdam, pp. 39–158.

Brünger,A.T. (1990) *Acta Crystallogr.*, **A46**, 46–57.

Chan,S.J., Segundo,B.S., McCormick,M.B. and Steiner,D.F. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7721–7725.

Drenth,J., Kalk,K.H. and Swen,H.M. (1976) *Biochemistry*, **15**, 3731–3738.

Ferrara,M., Wojcik,F., Rhaissi,H., Mordier,S., Roux,M.-P. and Béchet,D. (1990) *FEBS Lett.*, **273**, 195–199.

Fong,D., Calhoun,D.H., Hsieh,W.T., Lee,B. and Wells,R.D. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2909–2913.

Hashida,S., Towatari,T., Kominami,E. and Katunuma,N. (1980) *J. Biochem.*, **88**, 1805–1811.

Heinemann,U., Pal,G.-P., Hilgenfeld,R. and Saenger,W. (1982) *J. Mol. Biol.*, **161**, 591–606.

Hendrickson,W.A. and Lattman,E.E. (1970) *Acta Crystallogr.*, **B26**, 136–143.

Huber,R. and Schneider,M. (1985) *J. Appl. Crystallogr.*, **18**, 165–169.

Jack,A. and Levitt,M. (1978) *Acta Crystallogr.*, **A34**, 931–935.

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,B.A. and Drenth,J. (1984) *J. Mol. Biol.*, **179**, 233–256.

Kamphuis,I.G., Drenth,J. and Baker,E.N. (1985) *J. Mol. Biol.*, **185**, 317–329.

Kominami,E., Tsukuhara,T., Hara,K. and Katunuma,N. (1988) *FEBS Lett.*, **231**, 225–228.

Lee,X., Ahmed,F.R., Hirama,T., Huber,C.P., Rose,D.R., To,R., Hasnain,S., Tam,A. and Mort,J.S. (1990) *J. Biol. Chem.*, **265**, 5950–5951.

Marks,N., Berg,M.J. and Benuck,M. (1986) *Arch. Biochem. Biophys.*, **249**, 489–499.

Meloun,B., Baudys,M., Pohl,J., Pavlik,M. and Kostka,V. (1988) *J. Biol. Chem.*, **263**, 9089–9093.

Menard,H., Khouri,H.E., Plouffe,C., Dupres,R., Ripoll,D., Vernet,T., Tessier,D.C., Laliberté,F., Thomas,D.Y. and Storer,A.C. (1990) *Biochemistry*, **29**, 6706–6713.

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

Messerschmidt,A., Schneider,M. and Huber,R. (1990) *J. Appl. Crystallogr.*, **23**, 436–439.

Mort,J.S., Recklies,A.D. and Poole,A.R. (1984) *Arthritis Rheum.*, **27**, 509–515.

Nicklin,M.J.H. and Barrett,A.J. (1984) *Biochem. J.*, **223**, 245–253.

Nishimura,Y. and Kato,K. (1987) *Biochem. Biophys. Res. Commun.*, **148**, 254–259.

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

Pohl,J., Davinic,S., Bláha,I., Strop,P. and Kostka,V. (1987) *Anal. Biochem.*, **165**, 96–101.

Polgar,L. and Csoma,C. (1987) *J. Biol. Chem.*, **262**, 14448–14453.

Popovic,T., Brzin,J., Kos,J., Lenarcic,B., Machleidt,W., Ritonja,A., Hanada,K. and Turk,V. (1988) *Biol. Chem. Hoppe-Seyler*, **371**, Suppl., pp. 175–183.

Priestle,J.P. (1988) *J. Appl. Crystallogr.*, **21**, 572–576.

Ritonja,A., Popovic,T., Turk,V., Wiedenmann,K. and Machleidt,W. (1985) *FEBS Lett.*, **181**, 169–172.

Rossmann,M.G. and Argos,P. (1975) *J. Biol. Chem.*, **250**, 7525–7532.

Schechter,I. and Berger,A. (1967) *Biochem. Biophys. Res. Commun.*, **27**, 157–162.

Shaw,E. (1990) In Meister,A. (ed.) *Advances in Enzymology and Related Areas of Molecular Biology.* John Wiley and Sons, New York, Vol. 63, pp. 271–347.

Shaw,E. and Kettner,C. (1981) *Acta Biol. Med. Germ.*, **40**, 1503–1511.

Shaw,E., Wikstrom,P. and Ruscica,J. (1983) *Arch. Biochem. Biophys.*, **222**, 424–429.

Sloane,B.F. (1990) *Semin. Cancer Biol.*, **1**, 137–152.

Steigemann,W. (1974) Dissertation TU, München.

Stubbs,M.T., Laber,B., Bode,W., Huber,R., Jerala,R., Lenarcic,B. and Turk,V. (1990) *EMBO J.*, **9**, 1939–1947.

Takahashi,T., Dehdarani,A.H., Schmidt,P.G. and Tang,J. (1984) *J. Biol. Chem.*, **259**, 9874–9882.

Takahashi,T., Dehdarani,A.H., Yonezawa,S. and Tang,J. (1986) *J. Biol. Chem.*, **261**, 9375–9381.

Takio,K., Towatari,T., Katunuma,N., Teller,D.C. and Titani,K. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3666–3670.

Thiele,U., Assfalg-Machleidt,I., Machleidt,W. and Auerswald,E.-A. (1990) *Biol. Chem. Hoppe-Seyler*, **371**, Suppl., 125–136.

Towatari,T., Kawabata,Y. and Katunuma,N. (1979) *Eur. J. Biochem.*, **102**, 279–289.

Willenbrock,F. and Brocklehurst,K. (1985a) *Biochem. J.*, **227**, 511–519.

Willenbrock,F. and Brocklehurst,K. (1985b) *Biochem. J.*, **227**, 521–528.

Zvonar,T., Kregar,I. and Turk,V. (1979) *Croat. Chem. Acta*, **52**, 411–416.

Received on April 29, 1991