J. Biochem. 127, 635–643 (2000)

Substrate Specificity of Bovine Cathepsin B and Its Inhibition by CA074, Based on Crystal Structure Refinement of the Complex

Atsushi Yamamoto,* Koji Tomoo,*,[1] Tadaoki Hara,* Mitsuo Murata,† Kunihiro Kitamura,† and Toshimasa Ishida*,[1]

Department of Physical Chemistry, Osaka University of Pharmaceutical Sciences, 4-20-1 Nasahara, Takatsuki, Osaka 569-1094; and †Research Center, Taisho Pharmaceutical Co., Ltd., 1-403 Yoshino-cho, Ohmiya, Saitama 330-0031

Received December 3, 1999; accepted January 27, 2000

The crystal structure of the bovine spleen cathepsin B (BSCB)–CA074 complex was refined to R = 0.152 using X-ray diffraction data up to 2.18 Å resolution. BSCB is characterized by an extra Cys148–Cys252 disulfide bridge, as compared with rat and human CBs. Although the crystal structures of these enzymes showed similar overall folding, a difference was observed in the occluding loop, a structural element specific only to CB. Comparison of the torsion angles indicated the different flexibilities of their loop structures. The oxirane C6 atom of CA074 was covalently bonded to the Cys29 S^{γ} atom (C3-S^{γ}=1.81 Å), where the S-configuration was transformed to the R-form. Concerning the oxirane carbon atom that participates in the covalent bonding with the Cys residue, an acceptable rule has been proposed. The substrate specificities at the Sn (n = 1–3) and Sn' (n=1 and 2) subsites of CB, together with the interaction features as to CA074, have been discussed in comparison with the crystal structure of the papain-CA028 (a CA074-related inhibitor) complex.

Key words: cathepsin B, CA074, X-ray crystal structure, tertiary structure, substrate specificity, binding mode.

Lysosomal cathepsins are papain-like cysteine proteases that are abundant in mammalian cells and are responsible for intracellular proteolysis (*1*). At present, cathepsins B, H, L, S, K, and D have been identified as members of the cathepsin family. Among them, cathepsin B (CB, EC 3.4.22.1) is one of the most abundant proteinases in mammalian cells. The abnormal activity of CB has been implicated in muscular dystrophy (*2*), bone resorption (*3*), pulmonary emphysema (*4*), and tumor metastasis (*5*). Therefore, potent low-molecular-weight inhibitors specific for CB could be useful as therapeutic drugs for these diseases, in addition to their application as powerful probes for clarification of the physiological functions of this enzyme.

Previously, we designed a CB-selective inhibitor, CA074 [*N*-(L-3-*trans*-propyl-carbamoyloxirane-2-carbonyl)-L-isoleucyl-L-proline], based on the tertiary structure of CB (*6*), which was predicted from the crystal structure of the papain-E64c complex (*7*). CA074 was proven to be a CB-specific inhibitor *in vitro* (*8*) as well as *in vivo* (*9*). At present, this covalent-type inhibitor is being widely used as a probe to clarify the biological functions of CB (*10–12*).

The actual binding mode of CA074 as to CB at the atomic level was elucidated by X-ray crystal structure analysis of the bovine spleen cathepsin B (BSCB)–CA074 complex (*13*). Based on further structural refinement of the complex, this paper reports the structural features of BSCB and its substrate specificity revealed on comparison of the crystal structures of the present complex and the papain-CA028 one. The chemical structures of the covalent-type inhibitors used in this study are shown in Fig. 1. The numbering for the amino acid sequence of BSCB is shown in Fig. 2.

EXPERIMENTAL PROCEDURES

Preparation, Crystallization, and Data Collection—The purification of crude BSCB (Sigma) and the preparation of its complex with CA074 were reported previously (*13*, *14*); no enzymatic activity of the complex was detected with the method of Barrett (*15*). Bipyramidal complex crystals were obtained from a complex solution of 20 mg/ml concentration by vapor diffusion at 20°C using the hanging-drop method, where a droplet of 10 μl initial volume was equilibrated against a reservoir solution of 50 mM sodium citrate buffer (pH 3.5) containing 2 M sodium phosphate. Potential heavy-atom [$Hg(AcO)_2$ and K_2PtCl_6] derivatives of the complex crystals were prepared by the direct addition of the heavy-atom solutions to small drops including freshly grown crystals (*13*).

X-ray diffraction data up to 2.18 Å resolution were collected with an imaging plate detector (R-AXIS IIc) with a Rigaku rotating-anode X-ray source using Ni-filtered Cu Kα radiation at 15°C. Although the details of the data collection statistics for the native and heavy-atom derivative crystals were reported in Ref. *13*, the cell constants should

[1] To whom correspondence should be addressed.

Abbreviations: CB, cathepsin B; CA074, [*N*-(L-3-*trans*-propylcarbamoyloxirane-2-carbonyl)isoleucyl-proline; BSCB, bovine spleen cathepsin B; HLCB, human liver cathepsin B; RCB, rat cathepsin B; RMSD, root-mean-square deviation; Z-Arg-Ser(OBzl), benzyloxycarbonyl-L-arginyl-(*O*-benzyl)-L-serine; Bz-Arg-βNA, *N*-benzyloxy-L-arginyl-β-naphthylamide.

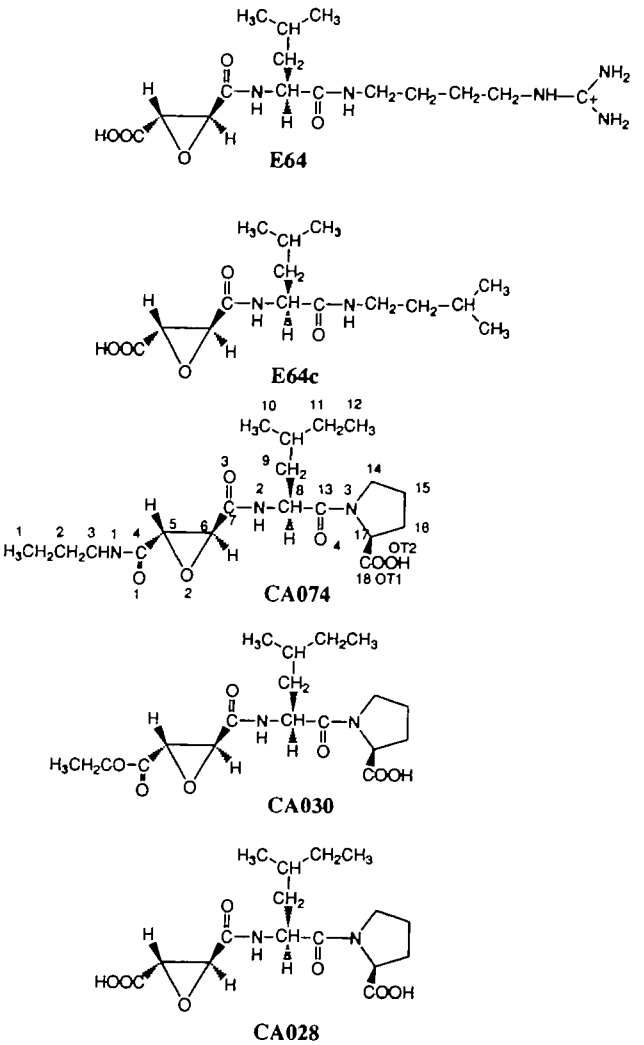


Fig. 1. **Chemical structures of E64, E64c, CA074, CA030, and CA028.** The atomic numbering used for CA074 is also given.

be revised to $a = b = 73.06$ (*10*) Å, $c = 141.79$ (*20*) Å.

Structure Solution and Refinement—The structure was determined using the phase information on two heavy-atom derivatives, where their atomic positions were identified using difference Patterson and difference Fourier maps, and refined with the program PHASES (*16*). The initial multiple isomorphous replacement map showed clear electron density for the helix domains of the protein. The map was further improved by means of solvent flattening techniques using the program PHASES, making it possible to trace the entire structure, except for the "occluding loop". The program FRODO (*17*) was used to build the protein. The sequence phase of the protein was established by the alignment of electron densities corresponding to the Trp, Tyr, and Phe aromatic rings, and to the disulfide bridges. Finally, 251 residues, excluding Gly48 and Arg49, were identified.

Crystallographic refinement and map creation were performed using X-PLOR (*18*); each stage of the refinement cycles consisted of (i) simulated annealing and conventional conjugate-gradient least squares, (ii) a conformational check using the Ramachandran plot, and (iii) manual rebuilding on the electron density map using the FRODO program. After several cycles of refinement of BSCB, the difference maps ($|2F_o-F_c|$ and $|F_o-F_c|$) were of sufficient quality to build the entire CA074 structure. Further manual rebuilding, refinement of CA074, and addition of clearly identifiable water molecules improved the *R*-factor. The final cycles of refinement reduced the *R*- and free *R*-factors to 15.2 and to 18.5%, respectively, a total of 2,091 non-H atoms (BSCB, CA074, and 156 water molecules) being included for the refinement using 17,671 independent reflections from 7.0 to 2.18 Å resolution. Gly48 and Arg49 were not visible in the electron density map and therefore were deleted from the refinement. The atomic coordinates have been deposited in the Brookhaven Protein Data Bank (ID code: 1QDQ).

RESULTS AND DISCUSSION

Overall Structure—The overall structure of the BSCB-CA074 complex is shown in Fig. 3a. BSCB is an α+β-type protein of four α-helices (residues 29–45, 56–62, 76–85, and 157–167) and seven β-strands (residues 5–7, 146–153, 170–178, 194–204, 212–219, 229–235, and 246–252), and consists of a light chain (residues 1–47) and a heavy chain (residues 50–253) (Fig. 2). Although the light and heavy chains of BSCB are linked by the sequence Gly48-Arg49 in the single-chain form (*19*), the dipeptide sequence gave no appropriate electron density. This is consistent with the observation that these two residues are accessible to cleavage and dipeptide excision leading to the production of the two chains during processing (*19*); moreover, no or unclear observation of this dipeptide sequence was reported for the X-ray crystal structures of HLCB (*20*) and its complex with the CA030 inhibitor (*21*), and recombinant RCB (*22*) and its complex with the *Z*-benzyloxycarbonyl-Arg-Ser (OBzl)-chloromethylketone inhibitor (*22*).

It has been reported that the 129th residue is either Asn or Ser, and this difference is explained by the genetic variety within one species in different batches of BSCB (*19*). The present results clearly produced the electron density of Ser129. Asn113 has been thought to be the glycosylated residue in CB. However, the electron density indicated no such structure, although the map was relatively poor; similarly, the glycosylated structure has not been assigned in the crystal structures of HLCB and RCB.

When the crystal structure of BSCB is compared with those of HLCB and RCB, a strong similarity is observed in the conformations of their backbone chains. The structures of HLCB and RCB can be superimposed on that of BSCB by means of a least squares fit between the corresponding Cα atoms, and the RMSDs for Cα, main-chain, and side-chain atoms lie in the ranges of 0.29–0.33, 0.30–0.35, and 0.64–0.73Å, respectively. Similar to that of papain, the tertiary structure of BSCB can be divided into two domains (Fig. 3b), designated as the L- (residues 13–147 and 251–253) and R- (residues 1–12 and 148–250) domains. An extended polar active site cleft is formed at the interface between these two domains; the Cys29 residue, the active center of CB, is positioned at the middle of this active site.

The regions (residues 45–53, 119–124, and 209–213) located at the surface of the molecule assumed relatively large temperature factors of 20–65 Å^2, indicating high conformational flexibility. Among them, residues 119–124 be-

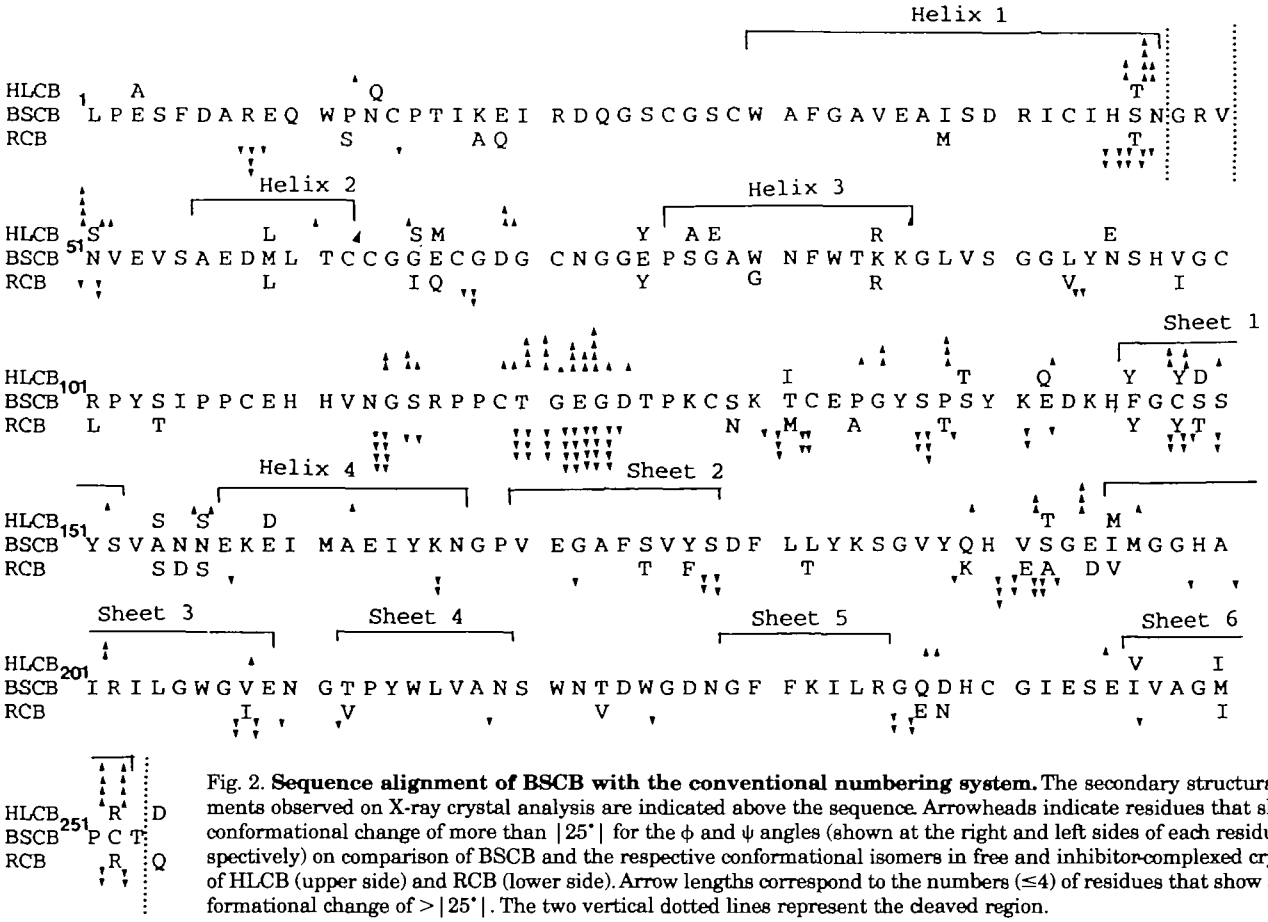


Fig. 2. **Sequence alignment of BSCB with the conventional numbering system.** The secondary structural elements observed on X-ray crystal analysis are indicated above the sequence. Arrowheads indicate residues that show a conformational change of more than |25°| for the ϕ and ψ angles (shown at the right and left sides of each residue, respectively) on comparison of BSCB and the respective conformational isomers in free and inhibitor-complexed crystals of HLCB (upper side) and RCB (lower side). Arrow lengths correspond to the numbers (≤4) of residues that show a conformational change of >|25°|. The two vertical dotted lines represent the cleaved region.

long to the region called the "occluding loop" (residues 105–125); this loop is a structural feature of CB because it is missing in other lysosomal enzymes and has been proposed to be important for the peptidyldipeptidase activity of CB (*23*).

Conformation around Disulfide Bonds—BSCB has seven disulfide bridges at Cys14–Cys43, Cys26–Cys71, Cys62–Cys128, Cys63–Cys67, Cys100–Cys132, Cys108–Cys119, and Cys148–Cys252 (*24*). The seventh disulfide bridge is the structural feature of BSCB, because the CBs from chicken (*25*), mouse (*26*), man (*27*), and rat (*28*) all lack this bridge. This extra disulfide bridge in BSCB appears to be related to its enzymatic activity, because human and rat CBs are about three times less active than bovine CB against Bz-Arg-βNA, a low-molecular-weight substrate analogue (*15*).

The conformations around the disulfide bridges in BSCB are summarized in Table I, together with the averaged values of those in HLCB (free and inhibitor-complexed forms) and RCB (free and inhibitor-complexed forms). The conformation of the disulfide bridge can be classified according to the chirality of the torsion angle around the sulfur–sulfur bond (χ_3): right-handed ($\chi_3 = 99° \pm 9°$) or left-handed ($\chi_3 = -85° \pm 9°$), and is further subdivided into several distinct classes according to other torsion angles (*29*, *30*). The conformations around the Cys14–Cys43 and Cys62–Cys128 bonds belong to one of the most highly populated classes of left- and right-handed disulfide regions, respectively. Although the conformations of the other disulfide bonds differ significantly from the standard value at only one dihedral angle, they all appear to be energetically stable, because the values of χ_1 and χ_1' are all in one of the three preferable regions (~±60° and ~180°), and the χ_2 and χ_2' values are expected to be the most variable (*30*).

The Cys148–Cys252 bridge only formed in BSCB assumes a geometry similar to that of Cys62–Cys128, although the value of χ_1 (= −90°) is rather rare for disulfide bridges in proteins. No notable difference is observed between the torsion angles in BSCB and the others, implying conformational similarity among these enzymes. Thus, it can be concluded that the respective disulfide bridges in BSCB assume reasonable geometries and the extra disulfide bond of Cys148–Cys252 does not impose any specific constraint on the molecular conformation and is smoothly formed during the entire folding of the backbone chain.

Conformational Flexibility of the Occluding Loop—As stated earlier, the occluding loop exhibited relatively large thermal parameters, probably due to the lack of any defined secondary structure. The flexibility of this loop becomes much clearer when the crystal structure of BSCB is compared with those of HLCB (free and inhibitor-complexed forms) and RCB (free and inhibitor-complexed forms). The degree of conformational variation among them is shown in Fig. 2. The large variations at sequences 45–47/51–52 and 251–253 clearly result from conformational instabilities due to excision of sequence 49–50 or to the C-ter-

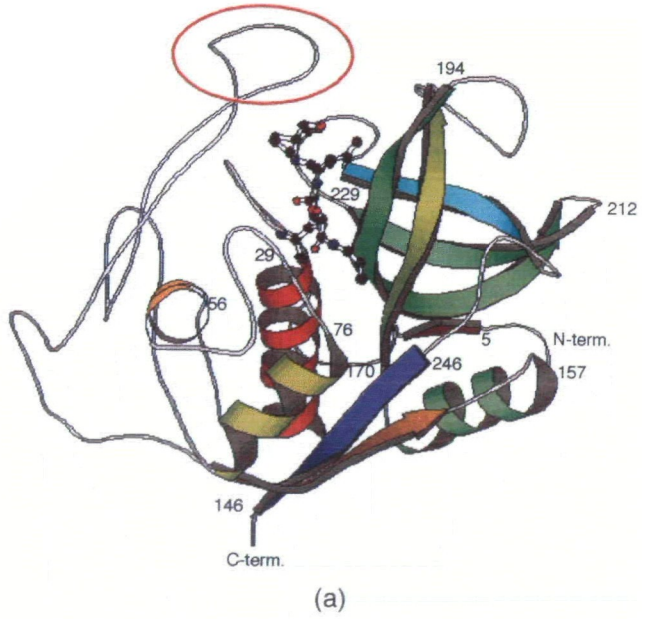
minal end, respectively. Although relatively large differences in the corresponding torsion angles are observed between the nonidentical amino acid residues of the respective CBs, the number of successive residues, each of which shows such a large difference, is at most three or less (≤3). In contrast, the backbone and side chains of the occluding loop, particularly the Cys119-Thr120-Gly121-Glu122-Gly-123-Asp124 sequence, show continuous and significant variation (Fig. 2), despite there being the same conserved sequence in these enzymes. Because some of the ϕ, ψ, ω, and χ angles in Cys119–Asp124 assume values which are in the energetically unfavorable or metastable regions of the Ramachandran plot, it may be assumed that this loop sequence plays an important role in eliminating the constraints imposed by the tertiary disposition of the structurally rigid secondary α-helical and/or β-sheet structures.

Comparison of the crystal structures of CBs so far analyzed has suggested that the observed molecular conformation reflects the inherent structural feature of the enzyme (Fig. 4). Therefore, the flexibility of the occluding loop appears to be related to a biological event. As one biological function of the occluding loop, the peptidyldipeptidase activity, that catalyzes the removal of C-terminal dipeptides due to the blocking of the primed terminus of the active site, has been proposed (*23*). On the other hand, the X-ray crystal structure of human procathepsin B (*31*) suggests the function of fixing the proregion at the primed site by raising the loop structure. These festures appear to be based on the conformational flexibility of the occluding loop.

CA074 Binding Site—The interaction scheme for CA074 at the BSCB active site is shown in Fig. 4; the binding mode was quite the same as that in the case of the HLCB-CA030 complex (*21*). Significant atomic interactions and torsion angles of CA074 are listed in Tables II and III, respectively.

The electron density map clearly showed covalent bond formation between the oxirane C6 atom of CA074 and the Cys-29 S^{γ} atom (C6-S^{γ} = 1.81 Å), a free O2H group thus being formed at the C5 atom. The CA074 molecule in its complexed state takes on an energetically stable conformation, as judged from the respective reasonable torsion angles. The C-terminal Pro carboxyl group was treated as an anionic state, based on the pK_a (=1.95) of proline and the C–O bond lengths. CA074 was bound to the S′ and S subsites of BSCB *via* a substrate-like interaction, where the Pro, Ile, carbamoyl, and propyl groups correspond to P2′, P1′, P1, and P2 sites, respectively. This binding mode is in contrast with those commonly observed in the papain–E64 and related complexes, in which the peptide moiety of the inhibitor is located at the S subsite (*7*, *32*, *33*). Previously we reported (*6*) that the Ile-Pro sequence, particularly the existence of a Pro residue at the C-terminus, is necessary for the specific inhibition for CB, suggesting that the binding mode of CA074 at the active site is primarily dominated by intimate P2′–S2′ and P1′–S1′ interactions. Thus, it is obvious that the double hydrogen bonding between the carboxyl two oxygens of the Pro C-terminus and the imidazole nitrogens of His110 and His111 of the occluding loop are important for CB-specific inhibition of CA074. The importance of such double hydrogen bonding has already been reported by Turk *et al.* (*21*).

Oxirane Carbon Atom Participating in Covalent Bonding with a Cys Residue—The oxirane C6 atom of CA074 (Fig. 1) was covalently bonded to the Cys 29 S^{γ} atom *via* a sub-



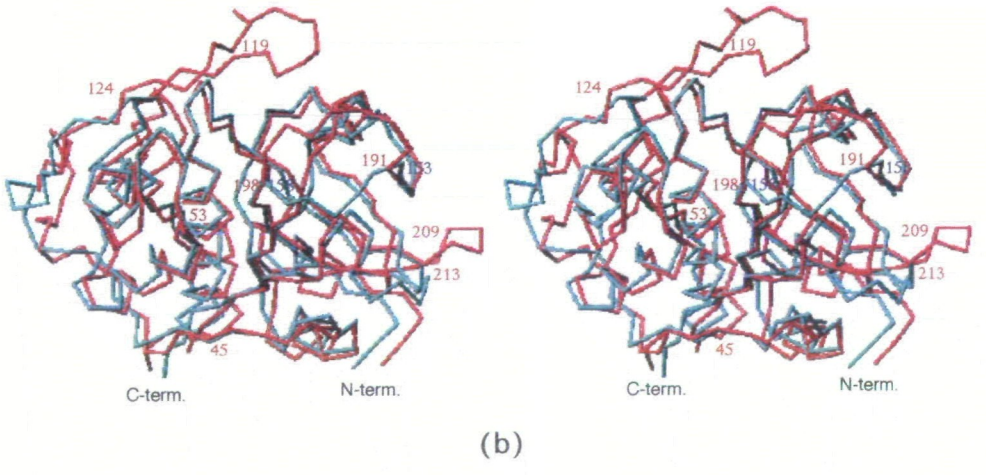


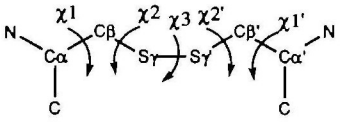
Fig. 3. **(a) Overall view of the BSCB-CA074 tertiary complex, and (b) stereoscopic superposition of the Cα-chain foldings of BSCB (red) and papain (blue).** In (a), the α-helices and β-strands are shown as spirals and waved sheets, respectively. CA074 is shown as a ball-and-stick model. The occluding loop is enclosed with a red circle. Some sequence numbers are given in (a) and (b). The red- and blue-colored numbers in (b) correspond to BSCB and papain, respectively.

strate-like interaction. This atomic position is different from that commonly observed in the papain-CA028 (*32*), -E64 (*33*), and -E64c (*7*) complexes, where a covalent bond is formed at the oxirane C5 atom. This is clearly due to the difference in substrate specificity between CB and papain. Molecular dynamic simulations for the BSCB-CA074 (*34*) and CB-E64 (*35*) complexes showed the energetic advantage of the Ile-Pro residues for binding to the S1′ and S2′

TABLE I. **Comparison of the disulfide torsion angles of BSCB with the averaged angles in HLCB and RCB.**[a]

Disulfide bridge		BSCB	Average[b]	Disulfide bridge		BSCB	Average[b]
Cys14–Cys43	χ^1	–66	–67 (7)	Cys26–Cys71	χ^1	171	168 (3)
	χ^2	–64	–63 (4)		χ^2	83	92 (2)
	χ^3	–87	–87 (5)		χ^3	–96	–97 (6)
	$\chi^{2'}$	–101	–102 (4)		$\chi^{2'}$	–51	–53 (8)
	$\chi^{1'}$	–179	–175 (4)		$\chi^{1'}$	–80	–81 (8)
Cys62–Cys128	χ^1	–69	–64 (6)	Cys63–Cys67	χ^1	179	–179 (3)
	χ^2	–61	–69 (8)		χ^2	–172	–175 (5)
	χ^3	100	98 (5)		χ^3	81	81 (6)
	$\chi^{2'}$	60	68 (7)		$\chi^{2'}$	96	94 (7)
	$\chi^{1'}$	170	172 (7)		$\chi^{1'}$	–68	–72 (4)
Cys100–Cys132	χ^1	173	178 (4)	Cys108–Cys119	χ^1	52	50 (4)
	χ^2	87	79 (7)		χ^2	88	90 (6)
	χ^3	–85	–92 (6)		χ^3	94	95 (5)
	$\chi^{2'}$	–50	–48 (6)		$\chi^{2'}$	–64	–56 (6)
	$\chi^{1'}$	–68	–62 (4)		$\chi^{1'}$	–57	–57 (5)
Cys148–Cys252	χ^1	–90					
	χ^2	60					
	χ^3	106					
	$\chi^{2'}$	–65					
	$\chi^{1'}$	–62					

[a]The torsion angles in a disulfide group are defined in the diagram below.



χ1:N-Cα-Cβ-Sγ, χ2:Cα-Cβ-Sγ-Sγ'; χ3:Cβ-Sγ-Sγ'-Cβ'; χ2':Cα'-Cβ'-Sγ'-Sγ, χ1':N-Cα'-Cβ'-Sγ'

[b]This value is the averaged torsion angle and its estimated standard deviation (in parentheses) of the respective conformational isomers in free and inhibitor-complexed HLCB and RCB crystals.

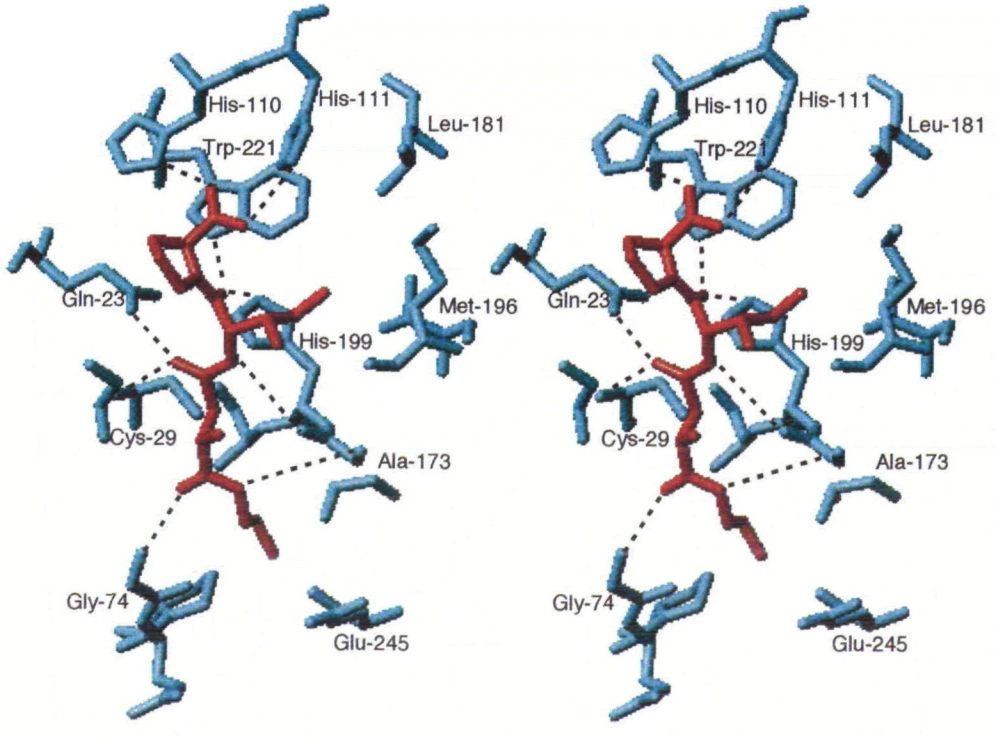


Fig. 4. **Stereoscopic view of the mode of binding of CA074 (red) to the active site (blue) of BSCB.** The atomic pairs given in Table II are indicated by dotted lines.

subsites over the S1 and S2 ones. Thus, it could be considered that the C6 atom is subjected to covalent bond formation with the Cys S^{γ} atom. In the binding to papain, on the other hand, the orientation of the Ile-Pro residues as to the S1 and S2 subsites is energetically more favorable than that as to the S1′ and S2′ ones, resulting in covalent bond formation at the C5 atom, as shown for the papain-CA028 complex (*32*). Thus, regarding the atomic position of the oxirane carbon which participates in the covalent bond formation with the Cys S^{γ} atom, an acceptable rule could be proposed, *i.e.*, the nucleophilic attack of the Cys S^{γ} atom occurs at the oxirane carbon atom which possesses the P′ substituent. The inhibitors shown in Fig. 1 have carbonyl or carboxyl O atoms at both the C5 and C6 atoms of the oxirane ring. These O atoms could be located at the "oxyanion hole" of the enzyme, which is formed by the Gln23 $N^{\epsilon 2}H$ and Cys29 NH groups in BSCB. Thus, the covalent bond formation at either the C5 or C6 atom depends on which side chain of the inhibitor could be located at the S′ subsites. In the case of BSCB, the carbonyl group which is linked to the Ile-Pro residue is bound *via* two interactions, *i.e.*, hydrogen bonding of O3···$N^{\epsilon 2}H$ (Gln23) = 2.82 Å and an electrostatic interaction of O3 NH (Cys29) = 3.48 Å. After the binding of the O3 atom at the oxyanion hole, the Cys29 S^{γ} atom would attack the carbon atom of the oxirane ring, consequently leading to covalent bond formation through electrostatic attraction between the $S^{\gamma}H$ group and the oxirane carbon atom which possesses the P′ substituent (S^{γ} = –0.234 e.u. and C3 = +0.234 e.u.).

On the other hand, since the active cleavage sites of papain and CB are similarly long and narrow (Fig. 3), the structural requisite for the inhibitor/substrate should be that it is flat in order to be able to bind to the active site smoothly. This is the main reason why the E64-like inhibitor should be (5*S*, 6*S*) or (5*R*, 6*R*) to exhibit inhibitory activity against cysteine proteases (*36*); such a configuration makes the inhibitor planar and accessible for binding to the active site.

Configuration Change of the Oxirane Carbon on Covalent Bond Formation with a Cys Residue—Although we reported previously that the covalent-bonded C6 atom took on an almost planar *S* configuration (*13*), further refinement of the complex structure showed clearly that the configuration should be revised to the *R*-form. Since we only used the (5*S*, 6*S*)-form among the four possible enantiomers of CA074 in this work, the $S \rightarrow R$ conversion of the C6 atom must have occurred; the $S \rightarrow R$ conversion of the C5 atom is due to the change in the ranking order of substituents. This (5*S*, 6*S*) → (5*R*, 6*R*) chirality change coincides with the case of the binding of E64c to papain (*7*), where the C5*S* configuration of the inhibitor is converted to the *R*-form through covalent bond formation with the Cys25 S^{γ} atom of papain. Based on (i) the superiority of the oxirane (5*S*, 6*S*)-form to the (5*R*, 6*R*)-form regarding inhibitory activity (*37*) and (ii) the MD simulation result for the papain-E64c complex (*38*), the energetic advantage of the *S* configuration for the interatomic interaction between the inhibitor and enzyme is suggested, and the (5*S*, 6*S*) → (5*R*, 6*R*) conversion through the covalent bond formation appears to be a general feature of the covalent-type inhibitor of cysteine protease.

On the other hand, the $S \rightarrow R$ configuration change of the oxirane carbon atom possessing the P′ substituent implies that the nucleophilic attack by the Cys S^{γ} H of its carbon atom [C6 for CA074–CB and C5 for E64c–papain] should occur from the O1 side of the oxirane ring plane.

Structural Feature of the Binding Pocket—To clarify the outline of the substrate-binding pocket of BSCB and to further characterize its structural specificity, it would be useful to compare the present complex structure with those of other related complexes. Recently, our group analyzed the X-ray crystal structure of the papain-CA028 complex (*32*) to determine at the atomic level how the Ile-Pro sequence is oriented in papain and why this sequence does not lead to specificity for papain. Since CA028 is a CA074-related inhibitor (Fig. 1) and exhibits inhibitory activity against papain to some extent, the structural features of the S*n*′ and S*n* subsites in BSCB could be easily characterized by comparison with the papain-CA028 complex. A stereoscopic comparison is shown in Fig. 5, where the topologically equivalent amino acid residues in BSCB and papain are superimposed on each other with reference to the literature (*20*).

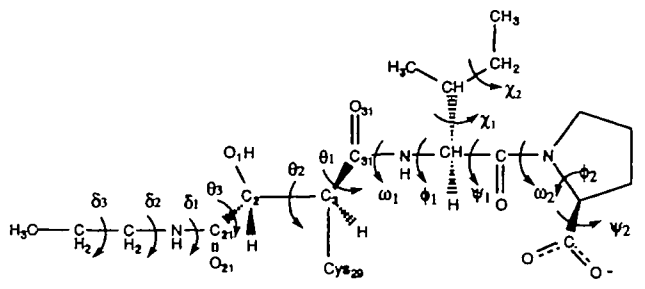
Sn′ (n = 1 and 2) Subsite—The S2′ subsite of BSCB is mainly composed of Gln23, Gly24, His110, and His111 of the occluding loop, Gly121, and Trp221. The occluding loop is located so as to block the orientation of the substrate at

TABLE II. **Interatomic distances (less than 3.5 Å) between the functional groups of the CA074 and BSCB active sites.**

CA074	BSCB	Distance (Å)
OT1	$N^{\delta 1}H$ (His110)	2.69
OT2	$N^{\epsilon 2}H$ (His111)	3.24
O4	$N^{\epsilon 1}H$ (Trp221)	3.12
O4	$N^{\epsilon 2}H$ (His199)	3.32
N2H	O=C (Gly198)	3.39
O3	$N^{\epsilon 2}H$ (Gln23)	2.82
O3	NH (Cys29)	3.48
O1	NH (Gly74)	3.22
N1H	N (Gly198)	3.39

TABLE III. **Torsion angles (°) of CA074.**

Torsion angle	Angle (°)
ψ_2	172
ϕ_2	–83
ω_2	158
ω_1	–177
ϕ_1	–48
ψ_1	143
χ_1	54
χ_2	158
θ_1	93
θ_2	173
θ_3	80
δ_1	177
δ_2	91
δ_3	160



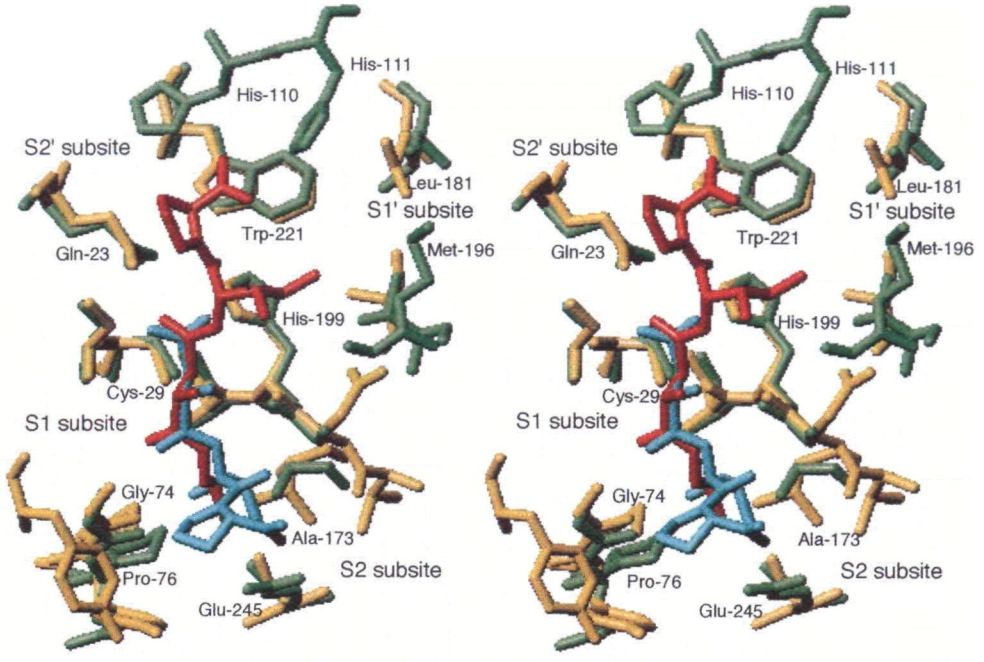


Fig. 5. **Stereoscopic view of the S*n*′ and S*n* (*n* = 1–2) subsites of BSCB (green).** The corresponding residues of papain (yellow) are superimposed for comparison. CA074 (red) or CA028 (blue), which is complexed with BSCB or papain, respectively, is also shown.

the S*n*′ ($n \geq 3$) subsites. The hydrogen bond pairing between the carboxylic two oxygens in Pro and the imidazole nitrogens of His110 and His111 is an important factor for the CB specificity of CA074. The remaining residues play roles in maintaining the Pro pyrrolidine ring through hydrophobic interactions (Gln23 and Trp221). This subsite is most likely responsible for the substrate specificity of CB, because in the cases of other cysteine proteases such as papain, the S2′ subsite which extends from the active site has not been very clearly determined to impose any specificity on the substrate and is largely open to solvent attack.

The S1′ subsite is composed of the hydrophobic core of Val176, Leu181, Met196, His199, and Trp221 in the R-domain, and stably holds the Ile side chain of CA074 through (Ile)C=O· $N^{\epsilon 1}$H(Trp221) hydrogen bonding, (Ile)C=O· $N^{\epsilon 2}$H(Gln23) (=3.56 Å), (Ile)C=O $N^{\epsilon 2}$H(His199) and (Ile)NH·O=C(Gly198) electrostatic interactions, and the hydrophobic interaction of the Ile side chain with the S1′ subsite (Leu176 and Met196). The compactness and hydrophobicity of the CB S1′ subsite are almost the same as those of papain (the corresponding residues are Ala136, Gln142, Val157, and Trp177). However, because the side chain of Met196 is located such that it interrupts the pocket, the CB S1′ subsite is significantly narrow compared with that of papain.

In addition to the fixation of the Ile side chain through hydrogen bond formation, Trp221 appears to contribute markedly to stabilization of the S1′ subsite through edge-to-face and face-to-face aromatic interactions with His111 and His199, respectively; the respective averaged center-to-center distances are 3.55 and 3.77 Å. It is important to note that the mutual spatial orientations among these residues are essentially maintained throughout the MD simulation in the free and inhibitor-complexed BSCBs (*34*), suggesting the importance of these residues in the formation of the S1′ subsite core.

On the other hand, the Gln 23$N^{\epsilon 2}$H and Cys29 NH groups form an "oxyanion hole", which has been thought to stabilize the reaction intermediates; these residues are located on the opposite side from the hydrophobic S1′ pocket across the inhibitor and produce a hydrophilic condition. In papain, the Gln19 and Cys25 residues cause the same condition.

Sn (n = 1 – 3) Subsite—The S1 subsite corresponds to the region surrounding a free O2H at the C5 atom, produced through the covalent bond formation of the inhibitor with Cys29, and is composed of Gly27, Gly74, and Gly198. The NH of Gly74 forms a hydrogen bond with the carbonyl O1 of CA074, stabilizing the binding with the inhibitor. Since such an interaction is commonly observed in papain-inhibitor complexes, it could be assumed that Gly74 plays a role by creating a kind of polar acceptor site, which is important for stably fixation of the substrate or inhibitor at the Sn subsite. NH of Gly198 forms a short contact with N1H of CA074. Although no specific interaction with the enzyme could be observed for the newly produced O2H group, a water solvent is located close to the group (3.50 Å), probably participating in hydrogen bond formation.

The S2 subsite, that consists of Pro76, Ala173, Ala200, and Glu245, forms a hydrophobic pocket, the space of which would be wide enough to accommodate a P2 side chain such as Phe. The propyl chain of CA074 is located at this site and is stabilized through hydrophobic interactions with the backbone chains of Pro76, Ala173, and Ala200. Compared with the corresponding S2 subsite of papain (composed of Pro68, Val133, Ala160, and Ser205, respectively), BSCB rather widely extends from the active site of cleavage toward Glu245 and is open to solvent attack. This is mostly due to the fact that the His191–Gly198 sequence of BSCB takes on a different spatial orientation from the corresponding Pro153–Asp158 sequence of papain (see Fig. 3b).

On the other hand, the S3 subsite has not been so well demonstrated to impose any specificity for the substrate or

inhibitor. This is in contrast with the case of papain, in which the Tyr61, Gly66, and Tyr67 residues tightly form a compact hydrophobic space to accommodate the P3 moiety of an inhibitor or substrate.

At present, CA074 is being widely used as a selective inhibitor to characterize the biological function of CB. However, it lacks the satisfactory conditions required for therapeutic application because of its irreversible inhibition. Thus, the most appealing alternative now is to develop a noncovalent-type reversible inhibitor against CB, because it could be expected to selectively control the enzymatic activity for only a certain period of time. The present structural features observed for the BSCB-CA074 complex provide a detailed picture of the inhibitor–enzyme interaction and would be useful for designing noncovalent-type inhibitors with high specificity and potent activity.

REFERENCES

1. Kirschke, H. and Barrett, A.J. (1987) Chemistry of lysosomal proteases in *Lysosomes: Their Role in Protein Breakdown* (Glaumann, H. and Ballard, F.J., eds.) pp. 193–283, Academic Press, London
2. Katunuma, N. and Kominami, E. (1987) Abnormal expression of lysosomal cysteine proteinases in muscle wasting diseases. *Rev. Physiol. Biochem. Pharmacol.* **108**, 1–20
3. Dalaisse, J.M., Eeckhout, Y., and Vaes, G. (1984) *In vivo* and *in vitro* evidence for the involvement of cysteine proteinases in bone resorption. *Biochem. Biophys. Res. Commun.* **125**, 441–447
4. Johnson, D. and Travis, J. (1977) Inactivation of human alpha 1-proteinase inhibitor by thiol proteinases. *Biochem. J.* **163**, 639–641
5. Sloane, B.F., Roxhin, J., Robinson, D., and Honn, K.V. (1990) Role for cathepsin B and cystatins in tumor growth and progression. *Biol. Chem. Hoppe-Seyler* **371** (Suppl.), 193–198
6. Sumiya, S., Yoneda, T., Kitamura, K., Murata, M., Yokoo, C., Tamai, M., Yamamoto, A., Inoue, M., and Ishida, T. (1992) Molecular design of potent inhibitor specific for cathepsin B based on the tertiary structure prediction. *Chem. Pharm. Bull.* **40**, 299–303
7. Yamamoto, D., Matsumoto, K., Ohishi, H., Ishida, T., Inoue, M., Kitamura, K., and Mizuno, H. (1991) Refined X-ray structure of papain•E-64-c complex at 2.1 Å resolution. *J. Biol. Chem.* **266**, 14771–14777
8. Murata, M., Miyashita, S., Yokoo, C., Tamai, M., Hanada, K., Hatayama, K., Towatari, T., Nikawa, T., and Katunuma, N. (1991) Novel epoxysuccinyl peptides. Selective inhibitor of cathepsin B, *in vitro*. *FEBS Lett.* **280**, 307–310
9. Towatari, T., Nikawa, T., Murata, M., Yokoo, C., Tamai, M., Hanada, K., and Katunuma, N. (1991) Novel epoxysuccinyl peptides. A selective inhibitor of cathepsin B, *in vivo*. *FEBS Lett.* **280**, 311–315
10. Buttle, D.J., Murata, M., Knight, C.G., and Barrett, A.J. (1992) CA074 methyl ester: a proinhibitor for intracellular cathepsin B. *Arch. Biochem. Biophys.* **299**, 377–380
11. Ohshita, T., Nikawa, T., Towatari, T., and Katunuma, N. (1992) Effects of selective inhibition of cathepsin B and general inhibition of cysteine proteinases on lysosomal proteolysis in rat liver in vivo and in vitro. *Eur. J. Biochem.* **209**, 223–231
12. Werle, B., Ebert, W., Klein, W., and Spiess, E. (1995) Assessment of cathepsin L activity by use of the inhibitor CA-074 compared to cathepsin B activity in human lung tumor tissue. *Biol. Chem. Hoppe-Seyler* **376**, 157–164
13. Yamamoto, A., Hara, T., Tomoo, K., Ishida, T., Fujii, T., Hata, Y., Murata, M., and Kitamura, K. (1997) Binding mode of CA074, a specific irreversible inhibitor, to bovine cathepsin B as determined by X-ray crystal analysis of the complex. *J. Biochem.* **121**, 974–977
14. Yamamoto, A., Kaji, T., Tomoo, K., Ishida, T., Inoue, M., Murata, M., and Kitamura, K. (1992) Crystallization and preliminary X-ray study of the cathepsin B complexed with CA074, a selective inhibitor. *J. Mol. Biol.* **227**, 942–944
15. Barrett, A.J. (1977) Cathepsin B and other thiol proteinases in *Proteinnases in Mammalian Cells and Tissues* (Barrett, A.J., ed.) pp.181–208, North-Holland Publishing, Amsterdam
16. Furey, W. and Swaminathan, S. (1990) PHASES, a program package for the processing and analysis of diffraction data from macro-molecules. *Am. Crystallogr. Assoc. Meeting Abstr. Ser. 2* **18**, 73
17. Jones, T.A. (1978) A graphic model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268–272
18. Brunger, A.T. (1992) X-plor Manual Version 3 1, Yale University
19. Meloun, B., Baudys, M., Pohl, J., Pavlik, M., and Kostka, V. (1988) Amino acid sequence of bovine spleen cathepsin B. *J. Biol. Chem.* **263**, 9087–9093
20. Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., and Bode, W. (1991) The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.* **10**, 2321–2330
21. Turk, D., Podobnik, M., Popovic, T., Katunuma, N., Bode, W., Huber, R., and Turk, V. (1995) Crystal structure of cathepsin B inhibited with CA030 at 2.0 Å resolution: a basis for the design of specific epoxysuccinyl inhibitors. *Biochemistry* **34**, 4791–4797
22. Jia, Z., Hasnain, S., Hirama, T., Lee, X., Mort, J.S., To, R., and Huber, C.P. (1995) Crystal structures of recombinant rat cathepsin B and a cathepsin B-inhibitor complex. Implications for structure-based inhibitor design. *J. Biol. Chem.* **270**, 5527–5533
23. Illy, C., Quraishi, O., Wang, J., Purisima, E., Vernet, T., and Mort, J.S. (1997) Role of the occluding loop in cathepsin B activity. *J. Biol. Chem.* **272**, 1197–1202
24. Baudys, M., Meloun, B., Gan-Erdene, T., Pohl, J., and Kostka, V. (1990) Disulfide bridges of bovine spleen cathepsin B. *Biol. Chem. Hoppe-Seyler* **371**, 485–491
25. Dong, A.-S., Stransky, G.I., Whitaker, C.H., Jordan, S.E., Schlesinger, P.H., Edwards, J.C., and Blair, H.C. (1995) Avian cathepsin B cDNA: sequence and demonstration that mRNAs of two sizes are produced in cell types producing large quantities of the enzyme. *Biochem. Biophys. Acta* **1251**, 69–73
26. Chan, S.J., San Segundo, B.S., McCormick, M.B., and Steiner, D.F. (1986) Nucleotide and predicted amino acid sequences of cloned human and mouse preprocathepsin B cDNAs. *Proc. Natl. Acad. Sci. USA* **83**, 7721–7725
27. Ritonja, A., Popovic, T., Turk, V., Wiedenmann, K., and Machleidt, W. (1985) Amino acid sequence of human liver cathepsin B. *FEBS Lett.* **181**, 169–172
28. Takio, K., Towatari, T., Katunuma, N., Teller, D.C., and Titani, K. (1983) Homology of amino acid sequence of rat liver cathepsin B and H with that of papain. *Proc. Natl. Acad. Sci. USA* **80**, 3666–3670
29. Richardson, J.S. (1981) The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 223–231
30. Katz, B.A. and Kossiakoff, A. (1986) The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. *J. Biol. Chem.* **261**, 15480–15485
31. Turk, D., Podobnik, M., Kuhelj, R., Dolinar, M., and Turk, V. (1996) Crystal structures of human procathepsin B at 3.2 and 3.3 Angstroms resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide. *FEBS Lett.* **384**, 211–214
32. Matsumoto, K., Murata, M., Sumiya, S., Mizoue, K., Kitamura, K., and Ishida, T. (1998) X-Ray crystal structure of papain complexed with cathepsin B-specific covalent-type inhibitor: substrate specificity and inhibitory activity. *Biochim. Biophys. Acta* **1383**, 93–100
33. Varughese, K.I., Ahmed, F.R., Carey, P.R., Hasnain, S., Huber, C.P., and Storer, A.C. (1989) Crystal structure of a papain–E-64 complex. *Biochemistry* **28**, 1330–1332
34. Yamamoto, A., Tomoo, K., Miyagawa, H., Takaoka, Y., Sumiya,

S., Kitamura, K., and Ishida, T. (2000) Molecular dynamics simulations of bovine cathepsin B and its complex with CA074. *Chem. Pharm. Bull.* in press
35. Feng, M.H., Chan, S.L., Xiang, Y., Huber, C.P., and Lim, C. (1996) The binding mode of an E-64 analog to the active site of cathepsin B. *Protein Eng.* **9**, 977–986
36. Schaschke, N., Assfalg-Machleidt, I., Machleidt, W., Turk, D., and Moroder, L. (1997) E-64 analogues as inhibitors of cathepsin B. On the role of the absolute configuration of the epoxysuccinyl group. *Bioorg. Med. Chem.* **5**, 1789–1797
37. Tamai, M., Matsumoto, K., Omura, S., Koyama, I., Ozawa, S., and Hanada, K. (1986) In vitro and in vivo inhibition of cysteine proteinases by EST, a new analog of E-64. *J. Pharmacobio-Dyn.* **9**, 672–677
38. Yamamoto, D., Ohishi, H., Ishida, T., Inoue, M., Sumiya, S., and Kitamura, K. (1990) Molecular dynamics simulation of papain–E-64 (N-[N-(L-3-trans-carboxyoxirane-2-carbonyl)-L-leucyl]-agmatine) complex. *Chem. Pharm. Bull.* **38**, 2339–2343