Crystal Structure of Cathepsin B Inhibited with CA030 at 2.0-Å Resolution: A Basis for the Design of Specific Epoxysuccinyl Inhibitors[†,‡]

D. Turk,[*,§] M. Podobnik,[§] T. Popovič,[§] N. Katunuma,[∥] W. Bode,[⊥] R. Huber,[⊥] and V. Turk[§]

Department of Biochemistry and Molecular Biology, Jozef Stefan Institute, Jamova 39, 61111 Ljubljana, Slovenia, Tokushima Bunri University, Tokushima 770, Japan, and Max Planck Institute for Biochemistry, 85152 Martinsried, Germany

Received October 21, 1994; Revised Manuscript Received January 17, 1995[⊗]

ABSTRACT: Crystals of cysteine protease human cathepsin B inhibited with CA030 (ethyl ester of epoxysuccinyl-Ile-Pro-OH) [Murata, M., *et al.* (1991) *FEBS Lett. 280*, 307−310; Towatari, T., *et al.* (1991) *FEBS Lett. 280*, 311−315] were isomorphous to a previous published structure of cathepsin B [Musil, D., *et al.* (1991) *EMBO J. 10*, 2321−2330]. The crystal structure of the complex was refined at 2.0-Å resolution to an *R*-value of 0.194. CA030 is well-defined in the electron density. The Ile-Pro-OH part of CA030 mimics a substrate P1′ and P2′ residues. The structure thus reveals for the first time a substratelike interaction in the S1′ and S2′ sites of a papain-like cysteine protease. The CA030 ethyl ester group occupies the S2 site. The structure confirms the role of residues His 110 and His 111 as the receptors of a peptidic substrate C-terminal carboxylic group. The structure suggests that an epoxysuccinyl fragment can be used to extend binding into primed and nonprimed substrate binding sites of a papain-like cysteine protease.

Abnormal cathepsin B activities have been implicated in the development and progression of a variety of human diseases. The most extensively studied lysosomal cysteine protease, cathepsin B is capable of degrading components of the extracellular matrix, in such diseases as muscular dystrophy (Katunuma, 1987), rheumatoid arthritis [reviewed in Werb et al. (1989)], and tumor invasiveness [reviewed in Sloane (1990)]. With regard to the latter, there is increasing evidence that cathepsin B is implicated in the progression of tumors from the premalignant to the malignant state. A specific cathepsin B inhibitor is therefore highly desirable: blocking cathepsin B activity will help to clarify the role of the enzyme in living organisms and also serve as a drug prototype. A specific cathepsin B inhibitor should exploit the unique feature of cathepsin B, namely, its carboxydipeptidyl activity: cathepsin B is the only known cysteine protease that cleaves the last two residues from a peptide chain.

In 1978 a new type of irreversible cysteine protease inhibitor was isolated, identified as 1-[L-*N*-(*trans*-epoxysuccinyl)leucyl]amino-4-guanidinobutane and named E 64[1] (Figure 1) (Hanada *et al.*, 1978). E 64 has no real preference for a particular cysteine protease. It inhibits the plant enzymes papain and actinidin as well as the mammalian cathepsins B, L, H, and S (Barrett *et al.*, 1986). Crystal structures of the papain−E 64 (Varughese *et al.*, 1989) and papain−E 64c complexes (Yamamoto *et al.*, 1991; Kim *et al.*, 1992) have shown that E 64 and analogues bind along the protein surface in the reverse direction to the peptide chain of substrate analogues (Drenth *et al.*, 1976; Schröder *et al.*, 1993).

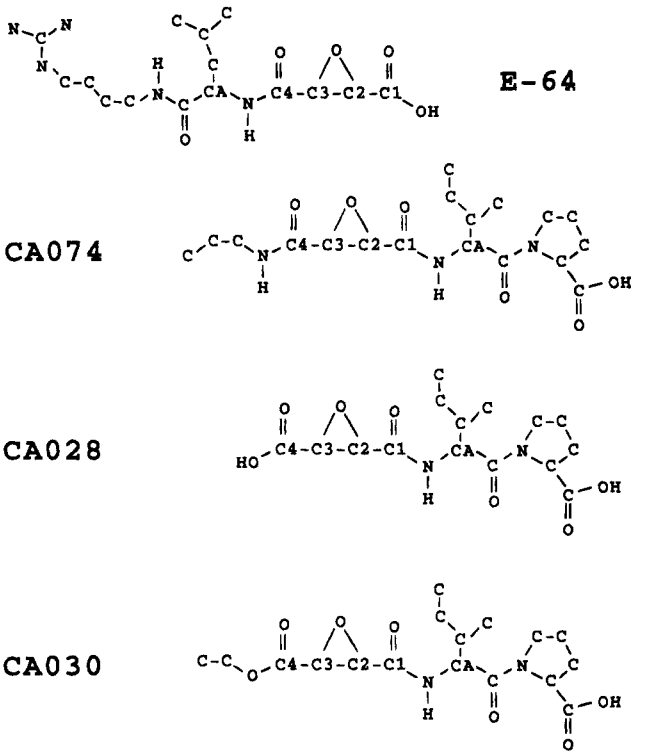

FIGURE 1: Topological schemes of E 64, 1-[L-*N*-(*trans*-epoxysuccinyl)leucyl]amino-4-guanidinobutane; CA028, epoxysuccinyl-L-Ile-L-Pro-OH; CA030, ethyl ester of epoxysuccinyl-L-Ile-L-Pro-OH; and CA074, propyl amide of epoxysuccinyl-L-Ile-L-Pro-OH.

[†] This work was supported by the Slovenian Ministry of Science and Technology. D. T. wants to thank to Kernforschungsbereich Jülich, BRD, for financial support for a part of his stay in Martinsried. Dedicated to Professor Harald Tschesche on the occasion of his 60th birthday.

[‡] The structure coordinates have been deposited in the Brookhaven Protein Data Bank and will be released 1 year after the paper publication as entry 1CSB.

* Author to whom correspondence should be addressed.

[§] Jozef Stefan Institute.

[∥] Tokushima Bunri University.

[⊥] Max Planck Institute for Biochemistry.

[⊗] Abstract published in *Advance ACS Abstracts,* March 15, 1995.

[1] Abbreviations: E 64, 1-[L-*N*-(*trans*-epoxysuccinyl)leucyl]amino-4-guanidinobutane; CA028, epoxysuccinyl-L-Ile-L-Pro-OH; CA030, ethyl ester of epoxysuccinyl-L-Ile-L-Pro-OH; CA074, propyl amide of epoxysuccinyl-L-Leu-L-Pro-OH; RMS, root mean square.

0006-2960/95/0434-4791$09.00/0 © 1995 American Chemical Society

A novel type of epoxysuccinyl peptides, analogues of E 64, were reported by Murata *et al.* (1991) and Towatari *et al.* (1991). They systematically modified both termini of E 64 in order to design a selective cathepsin B inhibitor. They succeeded by mimicking a carboxy-terminal dipeptide by replacing the terminal guanidinobutylamine of E 64 with a proline residue and terminating the chain with a free carboxylic group. They also systematically converted the proline carboxylic group to an amide, methyl ester, hydroxymethyl, and aldehyde and showed that the carboxylic group is responsible for selective binding of these E 64 analogues to cathepsin B. Besides, they also altered the other terminus of the epoxysuccinyls by converting the succinyl carboxy group to various ester and amide derivatives. The most successful ones, an ethyl ester and a propyl amide of epoxysuccinyl-Ile-Pro-OH, were named CA030 and CA074, respective (Figure 1). Their work showed that on the cathepsin B surface there must be a site involved in selective recognition of a free carboxylic group responsible for the cathepsin B dipeptidyl activity. Without a reliable three-dimensional structure of cathepsin B it was not possible to recognize the residue(s) involved in this interaction.

We determined the X-ray crystal structure of human cathepsin B (Musil *et al.*, 1991), which revealed that a large insertion loop (106−124, cathepsin B nomenclature) blocks the primed site of the active-site cleft and is thereby responsible for the cathepsin B exopeptidase activity. The role of a carboxylic group receptor for the C-terminus of a substrate was assigned to the two consecutive histidine residues of the insertion loop (His 110 and His 111). This proposal was later confirmed by others (Buttle *et al.*, 1992; Gour-Salin *et al.*, 1993), who also showed that a carboxylic group is essential for inhibition of cathepsin B with CA030-related derivatives. Although an attempt to determine the binding geometry of CA074 in complex with cathepsin B by means of X-ray crystallography has been communicated (Yamamoto *et al.*, 1992), no structure has been reported so far. The crystal structure of the complex between CA030 and human cathepsin B presented here is to our knowledge the first direct experimental evidence for these proposals.

Below we present the 2.0-Å resolution crystal structure of the complex between CA030 and human cathepsin B. We elucidate the contributions of various cathepsin B residues involved in substrate and inhibitor binding, and we try to address the specificity requirements of the S2, S1, S1′, and S2′ sites of cathepsin B. In addition, we offer possible structure-based explanations for the binding affinities of several cathepsin B inhibitors. This leads to suggestions regarding design of more specific inhibitors, which may be of great importance for *in vivo* studies of cysteine proteases and also as possible novel therapeutic agents.

EXPERIMENTAL PROCEDURES

Human kidney cathepsin B was isolated as described previously (Popvič *et al.*, 1988). A racemic mixture of *trans-S,S-* and *trans-R,R*-epoxysuccinyl-based CA030 was obtained from N. Katunuma. The complex of cathepsin B with CA030 was prepared by addition of 100 μL of a 1.0 mM CA030 (racemic mixture) solution in DMSO to 1.5 mg of cathepsin B dissolved in 1 mL of 0.02 M sodium acetate buffer, pH 5.2, containing 1 mM EDTA. The solution showed no cathepsin B activity when tested with substrate Bz-ArgNA according to the procedure described by Barrett (Barrett *et al.*, 1972).

Crystals were grown in a manner similar to that described previously (Musil *et al.*, 1991) at 21 °C by the hanging drop vapor diffusion method against 1.5 M ammonium sulfate buffered with 0.1 M sodium acetate at pH 5.0. The solution of inhibited cathepsin B was concentrated to about 5 mg/mL. Droplets were formed by mixing 4 μL of protein solution with 4 μL of the reservoir solution. The crystals are isomorphous with the human cathepsin B crystals (Musil *et al.*, 1991). They belong to the monoclinic space group $P2_1$ and have the cell dimensions $a = 86.51$ Å, $b = 34.15$ Å, and $c = 86.53$ Å and $\beta = 103.1°$. X-ray data were collected using a MAR image plate detector system with a Rigaku rotation anode and Cu Kα radiation. Data were evaluated with XDS (Kabsch, 1988). The total number of measurements is 51 421. These data were later processed with PROTEIN 3.1 (Steigemann, 1974). Only measurements with $I/\sigma > 2.0$ were accepted. The resulting data set contains 80% of all possible reflections in the range of 8.0−2.0-Å resolution. The last resolution shell between 2.05 and 2.00 Å contains 64% of all possible reflections. A total of 24 924 unique reflections were used in the final refinement procedure.

Since the crystals were isomorphous to the known structure (Musil *et al.*, 1991), the crystallographic refinement started from the original atomic coordinates (the asymmetric unit contains two cathepsin B molecules). All crystallographic refinement was done with X-PLOR version 3.1 (Brünger, 1992) on a CONVEX 3 computer using the PARHCSDX force field (Engh & Huber, 1991). During refinement the resolution of included data was gradually increased from 3.0 to 2.3 Å. At this stage a model of CA030 was built into the electron density and solvent molecules were for the first time added with the MAIN automatic procedure (Turk, 1992). A topology file for the missing hydroxysuccinyl and ethyl fragments was created using MAIN. The thiol ether bond length and angle energy terms for the active-site sulfur atom attached to the C2 atom of the epoxysuccinyl fragment were taken from the methionine residue. New solvent molecules were added and old ones were checked and if necessary removed after each refinement cycle within the MAIN automatic procedure. The procedure first checks already refined solvent molecules. It removes the ones with too-high temperature factors and too-low density in the $2F_{obs} - F_{calc}$ map. Then it searches for peaks in the $F_{obs} - F_{calc}$ map, removes the peaks with too-low density in the $2F_{obs} - F_{calc}$ map, keeps only the peaks that are in hydrogen-bonding distance to an oxygen or nitrogen atom, and removes the peaks related by the crystal symmetry operations.

The R-value of the refined model at 2.0-Å resolution is 0.194. The temperature factor refinement included bond ($b\sigma = 4.0$) and angle ($a\sigma = 6.0$) restraints. The minimal allowed atomic temperature factor value was set to 5.0 Å^2 (for some additional data see Table 1). The root mean square (RMS) deviations for all crystallographically "real" atoms from their ideal bond length and angle targets are 0.011 Å and 1.58°, respectively.

Furthermore, cathepsin B complexes with E 64 and with Ala-Arg-Arg-Leu-Trp-OH, a substrate with a reasonable binding geometry, were modeled. During energy minimization in addition to the usual energy terms (bond, angle, dihedral, improper dihedral, electrostatic, and van der Waals)

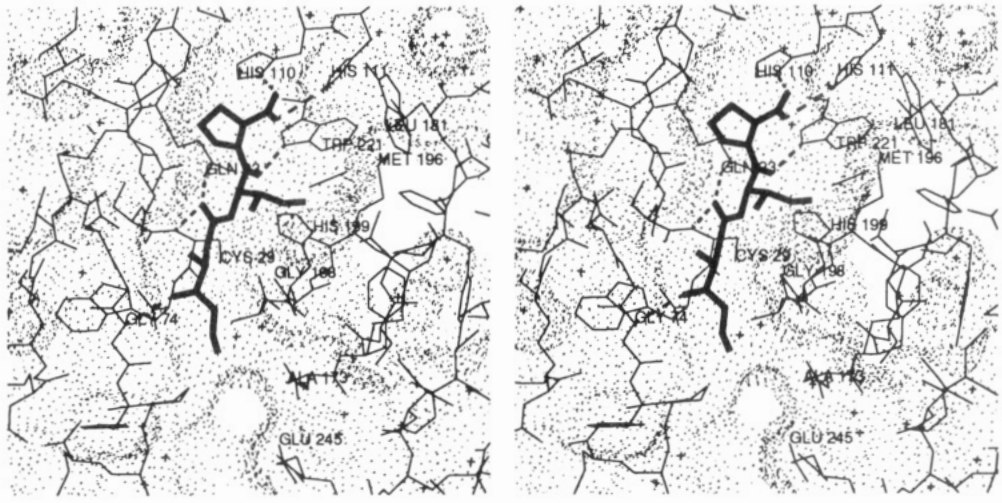

FIGURE 2: Stereo plot of CA030 (thick line) in its cathepsin B environment (thin line and points of cathepsin B Connolly surface) shown in a view along the active-site cleft. Relevant residues are marked and intramolecular hydrogen bonds are drawn with a broken line.

Table 1: Average Temperature Factors for Different Segments of Structure

segment[a]	N^b	B (Å^2)	σ^c (Å^2)
all	4386	13.6	10.6
cathepsin B 1	1909	11.2	7.4
cathepsin B 2	1883	12.0	7.3
inhibitor 1	26	20.7	4.3
inhibitor 2	26	24.2	5.1
active site 1	417	11.1	7.6
active site 2	406	12.8	7.0
solvent	542	26.7	17.7

[a] The marks 1 and 2 denote molecules 1 or 2 of cathepsin B and of their inhibitors. The active-site segments include all atoms of cathepsin B within a 10-Å distance range from any inhibitor atom. It merits to mention that 33 solvent molecules are extremely well defined; their temperature factors are equal to the minimal allowed value (5 Å^2). [b] Number of scattering atoms. [c] Standard deviation.

distance constraints were introduced to satisfy the hydrogen-bonding pattern observed in papain structures and to pull the P2 equipositioned residue into the hydrophobic environment. The distance constraints enforced proper hydrogen-bonding interactions and also pressed the model atoms against the cathepsin B surface to compensate the absence of solvent pressure, because the calculations were carried out *in vacuo*. E 64 was modeled into the cathepsin B environment lead by the papain structures (Varughese *et al.*, 1989; Yamamoto *et al.*, 1991). The succinyl fragment of E 64 was overlaid with corresponding parts of the CA030 structure. This E 64 model was then energy-minimized within the cathepsin B environment by fixing all cathepsin B atoms. The following distance contraints were used: The chosen target atoms for the guanidinium group of the agmatin residue were Asp 68 CG (Tyr 64) and Tyr 75 OH (71), the agmatin amide bond N atom was pulled toward the Gly 74 O (66), and the leucine side chain was pulled towards the Ala 173 CB. In this paper we use the cathepsin B and papain numbering; the numbers in parentheses refer to topologically equivalent papain residues, based on optimal superposition of both structures (Musil *et al.*, 1991).

The binding geometry of the substrate was constructed by overlaying the papain crystal structures of chloromethyl substrate analogue inhibitors (Drenth *et al.*, 1976) and the structure of CA030 in complex with cathepsin B presented here. The papain structure coordinates were first transformed to overlay the matching Cα atoms of the cathepsin B structure and displayed on the screen. The substrate model was then manually placed over the leading structures and energy-minimized while keeping all cathepsin B atoms fixed. The following distance constraints were used: The chosen target atoms for the guanidinium groups of the P2 and the P1 Arg residues of Glu 245 CD and Glu 122 CD, respectively, the O and N atoms of P2 Arg were pulled toward the N and O atoms of Gly 74 (66), P1 Arg O was restrained to the oxyanion hole formed by Gln 23 NE1 (19) and Cys 29 N (25), the P1′ O atom was pulled toward the NE1 Trp 211 (177) atom, and the OT2 and OT1 atoms of the P2′ residue were pulled toward the ND1 atom of His 110 and ND2 atom of His 111.

All interactive model building steps, energy calculations, and the automatic solvent positioning and verification were done using MAIN (Turk, 1992), running the program either on a Hewlett-Packard/9000, a Silicon Graphics Indigo 2, or a CONVEX 3 computer.

RESULTS AND DISCUSSION

CA030 Binding. The CA030 molecule is bound all along the active site cleft of cathepsin B. The Ile-Pro-OH part of the CA030 very probably mimics residues P1′ and P2′ of a substrate and thereby reveals the S1′ and S2′ sites of cathepsin B (Figures 2 and 3). Only the ethyl ester part of CA030 extends into the S2 site. The geometry of CA030 binding to cathepsin B is well-defined by electron density, allowing an unambiguous electron density map interpretation for the residues Ile-Pro (Figure 4). There are slight differences between the two molecules (designated 1 and 2) in the asymmetric unit. The only atom lacking appropriate electron density in inhibitor 1 contoured at a 0.8σ level is the C2 carbon (density not shown) bound to the cysteine 29 sulfur atom, while atoms of inhibitor 2 lie entirely within electron density contoured at the same level. The positioning of atoms of both bound inhibitor molecules is identical within the limits of the structure determination accuracy (see Table 2).

The Connolly surface of the left-rim side of the S2′ seems to be relatively smooth (Figure 2). The proline ring is

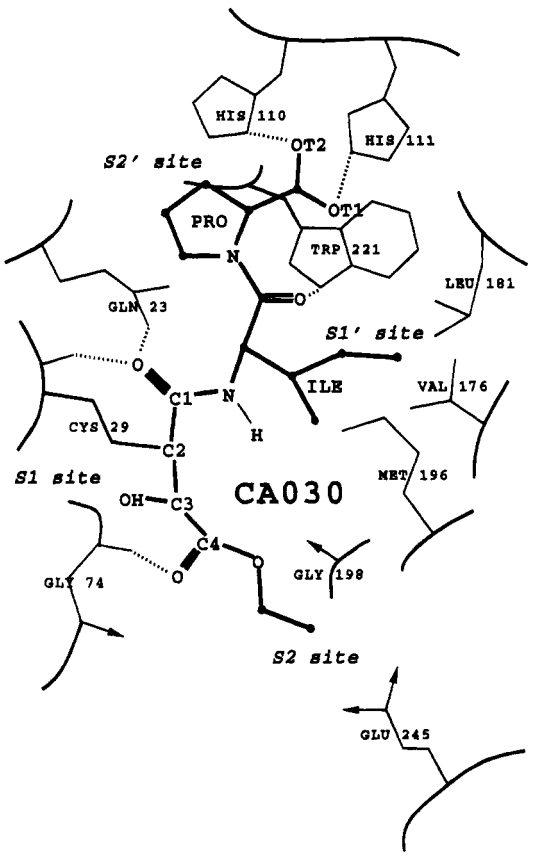

FIGURE 3: Schematic presentation of CA030 binding in the cathepsin B environment. CA030 is shown with thick lines, cathepsin B residues are shown with thin lines, cathepsin B carbonyl not involved in hydrogen-bond network with CA030, but important for inhibitor design, are presented with arrows.

Table 2: Relevant Distances between Atoms of Cathepsin B and CA030

		distance (Å)	
CA030	cathepsin B	molecule 1	molecule 2
O2(C4) EPO	N GLY 74	3.1	3.1
O (C1) EPO	NE2 GLN 23	2.3	2.8
O (C1) EPO	N CYS 29	3.1	3.1
N ILE	O GLY 198	4.0	4.0
O ILE	NE1 TRP 221	3.1	3.0
OT2 PRO	ND1 HIS 110	2.7	2.6
OT2 PRO	NE2 HIS 111	3.0	2.9
OT1 PRO	NE2 HIS 111	3.2	2.8
OT1 PRO	SOLV	3.0	

situated in the shallow S2′ groove formed by the L domain residues. Although the proline seems to fill the cleft properly, many other side chains might also fit. The CA030 carboxylic groups are oriented against the two histidine residues (His 110 and His 111). The oxygen atom (OT2) is pointing between His 110 and His 111; it is situated slightly closer to His 110 ND1 (2.6 Å) than to His 111 NE2 (2.9 Å). The binding environment of the other carboxylic oxygen atom (OT1) is slightly different in both molecules. The OT1 atom of inhibitor 1 is 3.2 Å away from His 111 NE2 and contacts a well-defined solvent molecule (at 3.0 Å distance), while the OT1 atom of inhibitor 2 is placed closer to His 111 NE2 (2.9 Å away) and lacks an equivalent solvent molecule within hydrogen-bonding distance.

The isoleucine peptide nitrogen is too far (4.0 Å) from the carbonyl oxygen of Gly 198 to form a good hydrogen bond, though its hydrogen atom would be properly oriented to form it. The isoleucine side chain is placed in a hydrophobic pocket surrounded by side chains of residues Val 176, Leu 181, Met 196, and Trp 221, making up the S1′ site (formed by the residues of the R domain only). The isoleucine carbonyl oxygen is oriented toward the NE1 atom of Trp 221 (3.1 Å away from it), forming a favorable hydrogen bond.

The terminal ethyl group of CA030 occupies part of the S2 binding site. The C4−O4 carbonyl (ester) group points toward the peptide-bond nitrogen of Gly 74 (Table 2). The active-site sulfur atom of Cys 29 is covalently bound to the C2 succinyl atom. The O bound to C1 carbonyl oxygen occupies the "oxyanion hole" of cathepsin B. It is located within a hydrogen-bond distance (2.3−2.8 Å) to the NE2 atom of Gln 19 and (3.1−3.2 Å) to the peptide-bond nitrogen of Cys 29. The electron density does not allow an unambiguous interpretation of the chiralities at the C2 and C3 atoms (see Figure 4). Under the assumption that only the *R,R* and the *S,S* forms of *trans*-epoxysuccinyl were present in the reaction solution (for chirality conversion see Figure 5), we tried to evaluate occupancies of both possible (*R,R* and *S,S*) enantiomers simultaneously; however, the refinement starting from different occupancy values did not converge.

Cathepsin B Molecules. The binding of CA030 changes slightly the orientation of the two cathepsin B molecules when compared to the original structure (Musil *et al.*, 1991). The root mean square (RMS) deviation of distances between the Cα atom positions for the first and second molecule of cathepsin B (Musil *et al.*, 1991) and of the CA030-inhibited cathepsin B are 0.41 and 0.61 Å, respectively (7 Cα atom pairs of the first pair of molecules and 22 Cα atom pairs of the second pair of molecules with interatomic distances longer than 1.0 Å, mostly located at chain termini, were omitted from the calculations). These values were calculated directly from the coordinates. After the superposition of the equivalent cathepsin B molecules was optimized, the RMS deviation of distances between equivalent Cα atom pairs decreases to 0.26 and 0.32 Å for each pair of equivalent molecules, respectively. These values are essentially equal to 0.30 Å, which is the Cα atom pairs RMS deviation of distances between both cathepsin B molecules of the complex presented here. This indicates that inhibitor slightly affects the orientation of both cathepsin B molecules as compared to the original structure and that all compared cathepsin B molecules are essentially equal. All residues lie within the allowed regions of the Ramachandran plot.

E 64 Analogues. Towatari *et al.* (1991) reported an increased selectivity of E 64 analogues for cathepsin B in the order HO-, EtO-, prNH-tES-Ile-Pro-OH, named CA028, CA030, and CA074, respectively (see Figure 1). Their affinity strength seems to be related to the possibility of forming a hydrogen bond with the amide proton of Gly 74 and the carbonyl oxygen of Gly 198 (Asp 158 in papain). The first compound (CA028) actually has a charged carboxylic group that should be repelled by the carbonyl oxygen of Gly 198 because of its negative charge. The second one (CA030, the one we used in this structure determination) carries no negative charge at this site but is also incapable of forming a hydrogen bond such as that formed by the CA074 amide proton. CA074 exhibits an additional advantage compared to CA030, namely, that its amide bond is

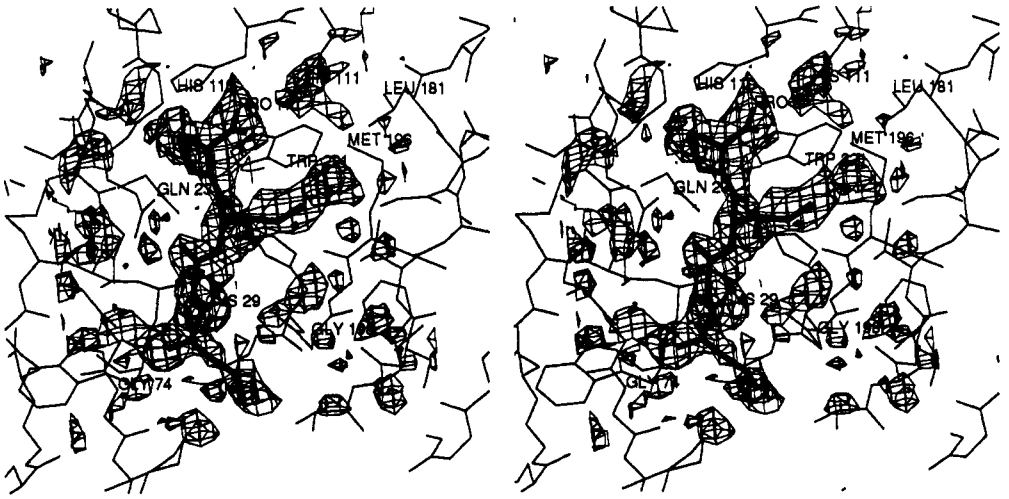

FIGURE 4: Omit map of the final CA030 model contoured at 1.85σ shown together with the CA030 model (thick line) and cathepsin B (thin line). Omit map by excluding CA030 model atoms and CB and SG of Cys 29 atoms from phasing after two consecutive passes of 50 steps of positional refinement followed by 20 steps of temperature factor refinement. (The omit map presentation was chosen in order to improve the clarity of the plot, although the adequate density for the inhibitor is visible in the first maps.)

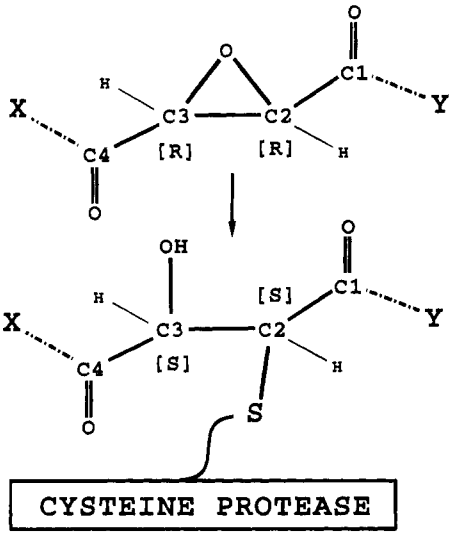

FIGURE 5: Scheme of a *trans*-epoxysuccinyl fragment before and after its reaction with a cysteine protease active site. Here are shown two chiral centers, C2 and C3, converted from *R,R* to *S,S* when reacting with an active-site cysteine. The C2 chirality is inverted because of the sulfur nucleophilic attack, while the C3 atom chirality changes only formally due to change of the substituent's ranking order.

less sensitive to hydrolysis *in vivo* than an ester bond; hydrolysis of the CA030 ester bond converts CA030 to CA028.

Modeling of E 64 against the cathepsin B surface shows that E 64 can bind to cathepsin B in a conformation very similar to that observed in papain. According to our model, the leucine residue occupies the S2 subsite and the positively charged E 64 guanidinium group is anchored between the carboxylic group of Asp 69 (Tyr 61) and Tyr 75 (Tyr 67) (Figure 6).

It must be mentioned that our model of E 64 differs from the complexes of E 64c with papain [Protein Data Bank entries 1PE6 (Yamamoto *et al.*, 1991) and 1PPP (Kim *et al.*, 1992)] in the positioning of the leucine residue for several reasons: The S2 site of cathepsin B is differently shaped than the papain S2 site. The Glu 245 of cathepsin B compared to papain Ser 205 is still too far to affect an E 64 leucine binding; however, the cathepsin B Ala 173 is smaller than equivalent papain residue Val 133 and the chain of cathepsin B from Gly 198 backward runs in the direction of the papain Asp 158 side chain. This allows placement of the E 64 leucine side chain deeper into the cathepsin B S2 site than the papain S2 site. Besides, the ϕ, ψ angles of the deposited E 64c−papain complexes (−43°, 67° and −151°, −40°) lie slightly outside the allowed Ramachandran plot regions and also outside the region of a β-strand conformation, whereas our model (−54°, 117°) lies inside the β-strand region of the Ramachandran plot.

The S2 pocket of cathepsin B is terminated by Glu 245, a unique residue favoring accommodation of positively charged substrate P2 residues such as arginines. Large hydrophobic residues such as phenylalanine are accepted as well; however, Hasnain *et al.* (1993) reported that Glu 245 to Gln and Ala mutants exhibit significantly weaker binding constants toward substrates with Arg at the P2 position. It is certainly worth trying to exploit this unique feature of cathepsin B when designing specific cathepsin B inhibitors.

An effort in this direction was recently documented by Gour-Salin *et al.* (1994a), who on the basis of *trans*-epoxysuccinyl E64 analogues varied the residue presumably occupying the S2 site in cysteine proteases. It was found that an arginine analog of E 64 (replacing leucine) is a less selective cathepsin B inhibitor than expected. Our modeling studies with E 64, based on the CA030−cathepsin B complex and papain complexes with E 64 (Varughese *et al.*, 1989; Yamamoto *et al.*, 1991), are consistent with these findings: A parallel β-strand hydrogen-bonding network of E 64, formed with cathepsin B residue Gly 74, includes the O4 atom of the epoxysuccinyl fragment and the N atom of the residue agmatin and cannot be superimposed with the antiparallel hydrogen-bonding network of a single substrate residue P2 (Figure 6). In fact, the E 64 leucine residue, which occupies the S2 binding site, does not form any hydrogen bond with the Gly 74 at all; the substrate P2 residue, however, forms two hydrogen bonds. This results in slightly varied positions of the Cα atoms of the E 64 leucine residue and of a substrate arginine P2 residue, respectively, and also in different orientations of the Cα−Cβ bonds, further affecting the more distant side-chain

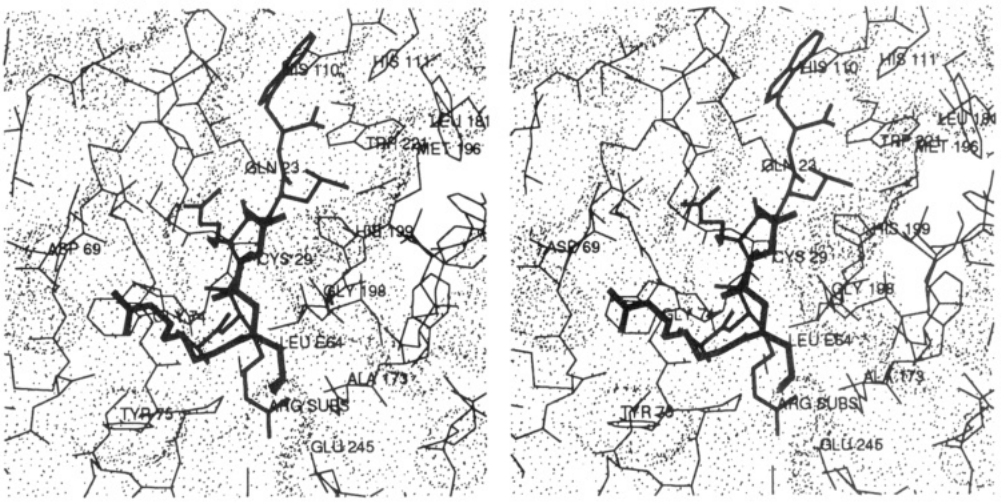

FIGURE 6: E 64 (thick line) and a substrate (Ala-Arg-Arg-Leu-Trp-OH) model (medium line) superimposed on cathepsin B (thin line) shown in a view along the active-site cleft.

atoms. The side chains of an arginine or a phenylalanine residue replacing the E 64 leucine would be too long for the cathepsin B S2 site and would collide with the side chain of cathepsin B Glu 245. The equivalent residue in papain is Ser 205, whose side chain is shorter by two atoms ($C\delta$, $O\epsilon$), eliminating this steric hindrance. In a very recent paper the importance of a different hydrogen-bonding network of a substrate and an E 64 analog was exposed in a cathepsin L, cathepsin S study (Gour-Salin *et al.*, 1994b).

Murata *et al.* (1991) reported a series of compounds in which they systematically replaced the tail attached to the CA030 and CA074 frameworks, where esters and amides of hydrophobic groups of different sizes are entering the cathepsin B S2 pocket. No substantial gain was achieved in binding affinities against cathepsin B by introducing larger hydrophobic groups. The selectivity, when compared to cathepsins H and L, was even worse. This loss of selectivity might be also explained as result of steric interference of larger hydrophobic groups with the side chain of Glu 245.

Substrate Binding. On the basis of our CA030−cathepsin B structure and of X-ray structures of substrate analogue chloromethyl ketone inhibitors in complex with papain (Drenth *et al.*, 1976), we modeled an Ala-Arg-Arg-Leu-Trp-OH peptide substrate bound to cathepsin B, assuming the Arg-Leu as the scissile peptide bond. The structures of papain in complex with substrate analogue chloromethyl ketone and aldehyde-based inhibitors (Drenth *et al.*, 1976, Schröder *et al.*, 1993) were taken to model the binding of the P1, P2, and P3 residues. The structure of the papain−stefin B complex (Stubbs *et al.*, 1990) showed that stefin B residue Ser 8 occupies the S2 site and that the chain from the Met 7 residue to the N-terminus follows a different route, thus leaving a further substrate chain trace amibiguous. The P2 residue of a substrate forms a double hydrogen-bond ladder with Gly 74 (66) in an antiparallel β-strand manner. E 64 and E 64c utilize the same glycine to form a short but parallel β-strand hydrogen-bonding pattern (Varughese *et al.*, 1989; Yamamoto *et al.*, 1991). It is notable that antiparallel substrate (P3) binding along a highly conserved Gly 216 residue (S3) occurs also in the trypsin-like serine protease family and that the high affinity of small peptide analogue thrombin and trypsin inhibitors seems to require formation of these two hydrogen bonds (Brandstetter *et al.*, 1992), although not always as antiparallel interactions (Pol *et al.*, 1994).

Selection Rules. The major difference of cathepsin B when compared to other cysteine proteases is the occluding loop (106−124) where the two histidines (His 110 and His 111) responsible for the carboxydipeptidase activity of cathepsin B are placed (Musil *et al.* 1991). Utilizing this particular structural feature should aid the design of selectivity of a cathepsin B inhibitor.

The residues [Gln 23 (19), Gly 74 (66), and Trp 221 (177)] involved in the hydrogen-bonding network of a bound substrate are conserved among the cysteine proteases of plant as well as animal origin (Jerala, 1993). Therefore, it is reasonable to assume that binding of a substrate or a substratelike inhibitor to a cysteine protease surface will be governed by these interactions. Epoxysuccinyl inhibitors extending into the nonprime site can in addition form one more hydrogen bond with the carbonyl oxygen atom of Gly 198 (Asp 158) (Yamamoto *et al.*, 1991). When some of these interactions are disrupted (protein hydrogen-bond donors and acceptors find no corresponding partners from the inhibitor side), the result is a lower affinity for the particular inhibitor.

Side chains of P1 and P2′ residues of a bound substrate (Figure 6) point upward, away from a cysteine protease surface, indicating that these positions are not very selective. The side chains of P3, P2, and P1′ residues, however, are placed in contact with the enzyme surface. This makes the S3, S2, and S1′ sites of a cysteine protease potential targets for the design of selective cysteine protease inhibitors.

Proline as a P2′ residue of an epoxysuccinyl-based inhibitor represents a special case. If bound to the primed site, its ring prevents rotation about the N−Cα bond and thereby helps to orient the subsequent carboxylic group against the His 110 and 111. If bound to the nonprimed site, however, proline cannot occupy the S3 site as the guanidinobutane group of E 64 can. Proline is the only amino acid residue lacking the peptide bond proton. Therefore it is unable to form the hydrogen bond with Gly 74(66) carbonyl group (thus could not adopt a required parallel β-strand conformation, and its CG, CD atoms would sterically interfere with the Gly 74 atoms). So we conclude that

proline can be a selective residue. It favors the S2′ site and would very unlikely bind into the nonprimed site.

Little data is available regarding the S1′ site specificity of cysteine proteases. The commercially available fluorogenic substrates have no naturally occurring amino acid residue at the P1′ position. Also, substrate analogue inhibitors have not explored variations in the primed side of the substrate binding. Menard *et al.* (1993) reported a substrate specificity study with a series of peptide chain substrates where the P1′ amino acid residue was systematically changed. These substrates included also a P3′ residue so that their results regarding the P1′ specificity should be taken with some precaution for designing cathepsin B inhibitors that terminate within the P2′ moiety. On the basis of our modeling study, also discussed above, the positioning of a P1′ residue should not be significantly affected by addition of a P3′ amino acid. Therefore we assume that the data presented by Menard *et al.* (1993) are a valid description of a P1′ residue specificity, although a good cathepsin B substrate should be terminated at P2′. Menard's study showed that actually any residue with a larger hydrophobic side chain (Leu, Phe, Tyr, Trp) could fit into the cathepsin B S1′ site and that smaller and hydrophilic or even charged side chains are not well accepted. If we add to the residues leucine and phenylalanine also isoleucine (that is present in CA030), these three residues then make a list of amino acid residues that bind well to the S1′ cathepsin B site.

Another, yet unexploited, selective feature of cathepsin B is the S2 site. A positive charge forming a salt bridge with Glu 245 should increase inhibitor selectivity. Since an arginine residue is too long, a more selective cathepsin B inhibitor still carrying a distal positive charge should have a shorter side chain. Our modeling studies show that even a lysine side chain still seems to be slightly too long.

How to Continue? The epoxysuccinyl fragment with its two carbonyl groups is a basis to which two peptide chains with two C-termini can be attached. A peptide chain attached to the C1−O1 carbonyl group thus follows a substrate chain direction (CA030), while a peptide chain attached to the C4−O4 carbonyl group runs in the direction opposite to it (E 64). An amino acid residue, when starting from the C4−O4 succinyl carbonyl, will occupy the S2 site and, when starting from the C1−O1 carbonyl, will occupy the S1′ site. The bidirectionality of the epoxysuccinyl fragment allows us to exploit binding in both directions simultaneously, in contrast to chloromethyl ketones and diazomethane-based inhibitors, which bind only in the direction toward the N-terminus of a substrate chain. An appropriate inhibitor binding into the S2′ site and continued through S1′, S1, and S2 sites to the S3 site or even further might enhance binding to such an extent that an epoxysuccinyl fragment, necessary for covalent bond formation with the active-site cysteine, might be avoided. An appealing challenge now is to design a noncovalently bound reversible inhibitor of cathepsin B that would be capable *in vivo* of selectively turning off cathepsin B activity for a certain, controlled period of time.

ACKNOWLEDGMENT

Drs. R. A. Engh and D. Lamba are gratefully acknowledged for critical reading of the manuscript.

REFERENCES

Barrett, A. J. (1972) *Anal. Biochem. 47*, 280−293.
Barrett, A. J. (1986) Introduction to proteinases, in *Protease Inhibitors* (Barrett, A. J., & Salvesen, G., Eds.) pp 3−22, Elsevier, Amsterdam.
Brandstetter, H., Turk, D., Hoeffken, H. W., Grosse, D., Stürzebecher, J., Martin, P. D., Edwards, B. F. P., & Bode, W. (1992) *J. Mol. Biol. 226*, 1085−1099.
Brünger, A. T. (1988) *J. Mol. Biol. 203*, 803−816.
Brünger, A. T. (1992) X-PLOR Version 3.1, Manual, Yale University Press, New Haven and London.
Buttle, D. J., Murata, M., Knight, C. G., & Barrett, A. J. (1992) *Arch. Biochem. Biophys. 299*, 377−380.
Drenth, J., Kalk, K. H., & Swen, H. M. (1976) *Biochemistry 15*, 3731−3738.
Engh, R. A., & Huber, R. (1991) *Acta Crystallogr., Sect. A47*, 392−400.
Gour-Salin, J. B., Lachance, P., Plouffe, C., Storer, A. C., & Menard, R. (1993) *J. Med. Chem. 36*, 720−725.
Gour-Salin, J. B., Lachance, P., Magny, M. C., Plouffe, C., Menard, R., & Storer, A. C. (1994a) *Biochem. J. 299*, 389−392.
Gour-Salin, J. B., Lachance, P., Bonneau, P., Storer, A. C., Kirschke, H., & Broemme, D. (1994b) *Bioorg. Chem. 22*, 227−241.
Hanada, K., Tamai, M., Yamagishi, M., Ohmura, S., Sawada, J., & Tanaka, I. (1978) *Agric. Biol. Chem. 42*, 523−528.
Hasnain, S., Hirama, T., Huber, C. P., Mason, P., & Mort, J. S. (1993) *J. Biol. Chem. 268*, 235−240.
Jerala, R. (1993) Doctoral Thesis, Ljubljana University, Ljubljana, Slovenia.
Kabsch, W. (1988) *J. Appl. Crystallogr. 21*, 67−71.
Katunuma, N., & Kominami, E. (1987) *Rev. Physiol. Biochem. Pharmacol. 108*, 1−20.
Kim, M., Yamamoto, D., Matsumoto, K., Inoue, M., Ishida, T., Mizuno, H., Sumiya, S., & Kitamura, K. (1992) *Biochem. J. 287*, 797−803.
Menard, R., Carmona, E., Plouffe, C., Brömme, D., Konishi, Y., Lefebvre, J., & Storer, A. C. (1993) *FEBS Lett. 328*, 107−110.
Murata, M., Miyashita, S., Yokoo, C., Tamai, M., Hanada, K., Hatayama, K., Towatari, T., Nikawa, T., & Katunuma, N. (1991) *FEBS Lett. 280*, 307−310.
Musil, D., Zucič, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovič, T., Turk, V., Towatari, T., Katunuma, N., & Bode, W. (1991) *EMBO J. 10*, 2321−2330.
Poll, T., Engh, R. A., Bode, W., Sucher, G., Leinert, H., & van der Saal, W. (1995) *J. Med. Chem.* (submitted for publication).
Popovič, T., Brzin, J., Kos, J., Lenarčič, B., Machleidt, W., Ritonja, A., Hanada, K., & Turk, V. (1988) *Biol. Chem. Hoppe-Seyler 369*, 175−183.
Schröder, E., Phillips, C., Garman, E., Harlos, K., & Crawford, C. (1993) *FEBS Lett. 315*, 38−42.
Sloane, B. F., Moin, K., Krepela, E., & Rhozhin, J. (1990) *Cancer Metastasis Rev. 9*, 333−352.
Steigemann, W. (1974) Thesis, University Munich, Munich, Germany.
Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarčič, B., & Turk, V. (1990) *EMBO J. 9*, 1939−1947.
Towatari, T., Nikawa, T., Murata, M., Yokoo, C., Tamai, M., Hanada, K., & Katunuma, N. (1991) *FEBS Lett. 280*, 311−315.
Turk, D. (1992) Ph.D. Thesis, Technische Universität München, München, Germany.
Varughese, K. I., Ahmed, F. R., Carey, P. R., Hasnain, S., Huber, C. P., & Storer, A. C. (1989) *Biochemistry 28*, 1330−1332.
Werb, Z. (1989) Proteinases and matrix degradation, in *Textbook of Rheumatology* (Keller, W. N., Harris, E. D., Ruddy, S., & Sledge, C. S., Eds.) pp 300−321, W. B. Saunder Co., Philadelphia, PA.
Yamamoto, A., Kaji, T., Tomoo, K., Ishida, T., Inoue, M., Murata, M., & Kitamura, K. (1992) *J. Mol. Biol. 227*, 942−944.
Yamamoto, D., Matsumoto, K., Ohishi, H., Ishida, T., Inoue, M., Kitamura, K., & Mizuno, H. (1991) *J. Biol. Chem. 266*, 14771−14777.

BI942466Y