FEBS 18361 FEBS Letters 407 (1997) 47–50

The crystal structure of human cathepsin L complexed with E-64

Akira Fujishima[a,*], Yumi Imai[a], Toshiyuki Nomura[b], Yukio Fujisawa[b], Yoshio Yamamoto[a], Tohru Sugawara[a]

[a] *Molecular Chemistry Laboratory, Pharmaceutical Research Division, Takeda Chemical Industries, Ltd., Juso-honmachi 2-17-85, Yodogawa-ku, Osaka 532, Japan*

[b] *Molecular Pharmacology Laboratory, Pharmaceutical Research Division, Takeda Chemical Industries, Ltd., Juso-honmachi 2-17-85, Yodogawa-ku, Osaka 532, Japan*

Received 2 January 1997; revised version received 18 February 1997

Abstract **We have determined the three dimensional structure of the complex of human cathepsin L and E-64, an irreversible inhibitor of cysteine proteases, at 2.5 Å resolution. The overall structure was similar to that of other known cysteine proteases and apparently identical to the mature region of procathepsin L. The electron density for E-64 is clearly visible except for the guanidinobutane moiety. From comparison of the active sites of cathepsin L and B, we found the following: (1) The S′ subsites of cathepsin L and B are totally different because of the 'occluding loop' lying on the end of the S′ subsites of cathepsin B. (2) The S_2 pocket of cathepsin L is shallow and narrow compared to that of cathepsin B. (3) The S_3 subsites of the two enzymes are more similar than the other subsites, but cathepsin L may accommodate a more bulky group at this site. Knowledge of the active site structure of cathepsin L should be helpful for the structure-based design of potent and specific inhibitors which are of therapeutic importance.**

Key words: Cathepsin L; E-64; X-ray crystal structure

1. Introduction

Cathepsin L is one of the major cysteine proteases expressed in animal tissues. Although the enzyme is well known as a lysosomal protease, there is increasing evidence that cathepsin L may play extracellular roles of physiological importance, such as bone resorption [1] and thyroid hormone liberation [2]. In addition, human cysteine proteases are implicated in the pathogenesis of several diseases including rheumatoid arthritis [3], osteoporosis, tumor metastasis [4] and periodontal diseases [5]. Therefore, potent and selective inhibitors of the cysteine proteases are expected to be good candidates for therapeutic agents.

Since both cathepsin B and L are abundant species among the mammalian cysteine proteases, comparative studies of the active sites in both cathepsin L and B are strongly desired to aid the design of selective inhibitors. Crystal structures of cathepsin B have been determined for the human and rat enzymes in the presence or absence of a synthetic inhibitor [6–8]. Regarding cathepsin L, the three-dimensional structure of the proenzyme has been published recently [9]. We describe here the crystal structure of mature cathepsin L focusing on its active site. This study will provide a reliable basis for understanding and predicting activities and specificities of inhibitors.

2. Materials and methods

Recombinant human procathepsin L was expressed in mouse myeloma cells, purified and processed to obtain the mature enzyme bound to E-64 as described by Nomura et al. [10]. Briefly, procathepsin L was secreted in culture media up to about 10 mg/l concentration and was purified to homogeneity by series of column chromatographies. The precursor was processed to obtain the mature cathepsin L. This process was initiated by changing the pH to 4.0 and terminated by adding E-64. The complex was further purified by gel filtration.

Crystallization was performed by hanging-drop vapor diffusion at 6°C. Prismatic crystals were grown from a solution containing 18% (w/w) polyethylene glycol 6000, 0.1 M Bis-Tris HCl pH 6.0, 0.1 M $CaCl_2$. They belong to a space group $P2_12_12_1$ with unit cell dimensions $a = 48.8$ Å, $b = 103.9$ Å, $c = 47.8$ Å. The asymmetric unit contains one molecule of the complex.

X-ray data were collected at room temperature on a Rigaku R-AXIS IIc detector mounted on an RU-200 generator. Intensity data up to 2.5 Å resolution were extracted and averaged by the program PROCESS (Rigaku) and used for structural analysis. The data set was 99.1% complete and the merging R factor was 5.7%.

The rotation and translation functions were calculated by the program package MERLOT [11] with data between 8.0 and 4.0 Å resolution. We initially used the papain structure (PDB ID code: 1ppp) as a search model but failed to obtain an interpretable translation function map for any high peaks on the rotation function maps. Use of a hypothetical model of cathepsin L, which had been built by standard homology modeling methods with a papain structure as the template, led to successful determination of the position of cathepsin L in the crystal. The crystallographic R value for this model after rigid-body refinement was 0.604.

X-PLOR [12] and QUANTA (Release 4.1, Molecular Simulations Inc.) were used for calculation of maps, revision of models and structure refinement. The parameter set of CHARMm (Release 23.1, Molecular Simulations Inc.) were used throughout the model building and structure refinement procedures. The R value of the current model is 0.175 and the free R value is 0.249 for a data set between 8.0 to 2.5 Å resolution. The RMS deviations from the target values are 0.017 Å for bond lengths, 2.95° for bond angles, and 1.87° for improper torsion angles. The current model contains 63 water molecules.

3. Results and discussion

3.1. Structure of cathepsin L

Most of the backbone structure became traceable during the refinement and manual model revision cycles, except for the nine residues from Phe-172 to Asn-180. (The sequence number of cathepsin L referred to in this report is based on the mature protein sequence which begins with Ala-114 of the preproprotein.) The electron density for these residues is not clear on a $|2F_o - F_c|$ map calculated with the current model suggesting the flexible and disordered nature of this region. A

*Corresponding author. Fax: (81) (6) 300 6306.
E-mail: fujishim@lab.takeda.co.jp

Abbreviations: E-64, 1-[L-*N*-(*trans*-epoxysuccinyl)leucyl]amino-4-guanidinobutane; RMS, root mean square; Bis-Tris, bis(2-hydroxyethyl)-iminotris(hydroxymethyl)methane

PII S0014-5793(97)00216-0

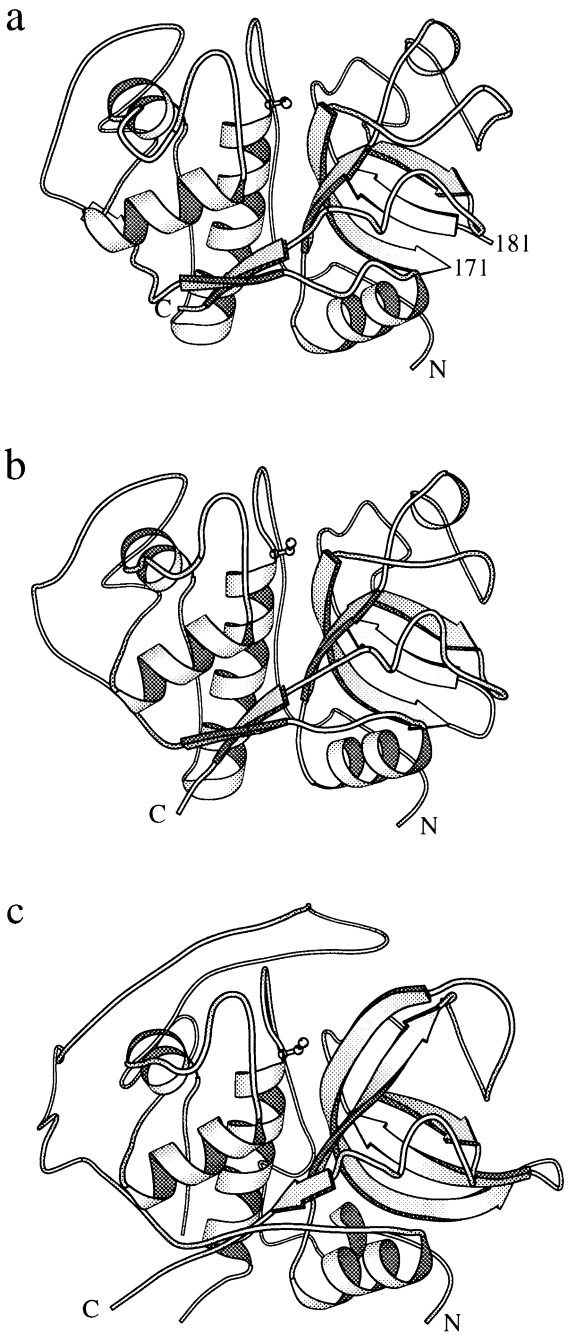

Fig. 1. Overall structures of human cathepsin L (a), papain (b) and human cathepsin B (c) drawn with MolScript [17]. N and C represent the amino and carboxyl termini, respectively. The catalytic cysteines are presented using the ball-and-stick model. The atomic coordinates of papain and cathepsin B were taken from the Brookhaven Protein Data Bank entries 9pap and 1huc, respectively.

similar disordered structure was also observed in the same region of procathepsin L [9].

As expected from the high sequence identity (41%), the overall structure of human cathepsin L is similar to that of papain (Fig. 1). Among the 220 residues of mature cathepsin L, 183 Cα atoms can be superimposed onto the corresponding Cα atoms of papain with an RMS deviation of 0.82 Å. The only major difference is an extended loop corresponding to the sequence from Phe-172 to Asn-179 of cathepsin L. Though we could not model these residues, as mentioned above, they are distant from the active site of this enzyme. Molecular packing in the crystal shows that the missing residues are also apart from the active sites of adjacent molecules. Therefore, we may safely assume that there are no structural perturbations around the active site due to the missing residues.

Coulombe et al. [9] have shown that the prosegment of procathepsin L consists of two components: the N-terminal globular domain and C-terminal extended structure. The former interacts with the loop from His-140 to Asp-155 of the mature domain, and the latter lies in the active site cleft. However, judging from the figures presented in the paper [9], there is no substantial difference in any part of the mature region between the precursor and mature form of cathepsin L. It seems that the tertiary structure formation of the mature cathepsin L is completed in the proenzyme, though precise studies based on the atomic coordinates remain to be performed.

3.2. Interactions between cathepsin L and E-64

There are several crystallographic studies which have revealed interactions between plant cysteine protease and E-64. Varughese et al. [13] determined the crystal structure of the papain-E-64 complex and showed the first molecular view of the E-64 adduct bound with cysteine protease. The structure of the actinidin-E-64 complex was also determined [14]. Yamamoto et al. [15] and Kim et al. [16] reported that there are two binding modes for E-64-c, a synthetic derivative of E-64, to complex with papain.

The E-64 adduct interacts with cathepsin L via almost the same manner as found in papain and actinidin crystals. The carboxyl group of E-64 is in the oxyanion hole which consists of the side chains of Gln-16, Trp-189, His-163 and the main chain of Cys-25. There are six hydrogen bonds between E-64 and cathepsin L. Three of them are found between the carboxyl group of E-64 and the residues forming the oxyanion hole. The other three seem to be important to fix the inhibitor in the active site cleft. The O5 and N10 atoms of E-64 form hydrogen bonds with Gly-68 like a parallel β-sheet (Fig. 3). The N6 atom of E-64 is likely to be involved in a weak (3.4 Å) hydrogen bond with the main chain carbonyl oxygen of Asp-162. These hydrogen bonds are commonly observed in papain and actinidin accommodating E-64 or E-64-c. This may be a natural consequence because those residues are conserved not only in the amino acid sequences but also in the three-dimensional structures of the three proteases. The 4-guanidinobutane moiety of E-64 is not well defined in the final omit map (Fig. 2), and the atomic temperature factors of the guanidino group are greater than 60 $Å^2$. These indicate that this moiety is not fixed in the cleft. The side chain of the leucine moiety is located in the S_2 pocket stretching down toward Met-70 and Ala-135. These residues, together with Leu-69, Met-161 and Gly-164, form the bottom of the hydrophobic S_2 pocket.

3.3. Comparison of active sites between cathepsin L and B

As there is no evidence for distinguishing each subsite in the S′ region of the cleft, we will refer to the S′ subsites as the S′ region as a whole. The S′ region of cathepsin L is similar to that of papain and is well described as a shallow depression rather than a cleft. The indole ring of Trp-189 is located at the center serving as a flat floor. The floor is surrounded by the side chains of Gln-21, Gly-20, Asn-18, Glu-192, Trp-193 and

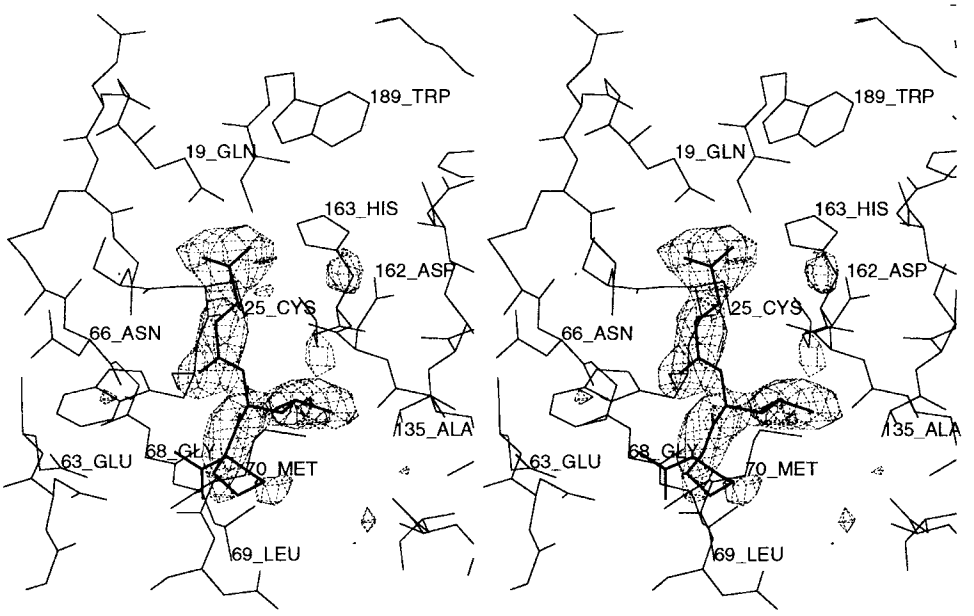

Fig. 2. Stereoview of an omit $|F_o-F_c|$ map showing the electron density of E-64. The map was phased by the final model without E-64 and contoured at the 2.5 σ level.

Leu-144. It is in the S′ region that the most remarkable difference between the active sites of cathepsin L and B resides. As described by Musil et al. [6], cathepsin B bears the 'occluding loop' lying on the equivalent position to the S′ region of cathepsin L. The loop corresponds to a long insertion sequence (106–124) in cathepsin B. This additional piece of structure not only accounts for the characteristic carboxydipeptidase activity of cathepsin B but also provides specific hooks (interaction points) for selective inhibitors. Actually, it has been revealed by X-ray analysis that a B selective inhibitor interacts with two histidine residues on the occluding loop [7]. Since the structural environments are totally different, we may expect that a compound which could fit in the S′ region of cathepsin L would not be a good inhibitor of cathepsin B and vice versa.

The S_2 subsite of cathepsin L is deep and hydrophobic. These characteristics should be common in cysteine proteases which prefer bulky and hydrophobic residues at the P_2 position of the substrate. However, there are several substitutions of amino acid residues and conformational changes of a loop structure which make the shape of S_2 pocket of cathepsin L different from that of cathepsin B. If we look down the active site cleft of cathepsin L, putting the primed subsites up and unprimed subsites down, like Fig. 4, Asp-162, Met-161, Asp-160 and Ala-214 form a continuous wall on the right hand side of the S_2 pocket. This wall makes the S_2 pocket of cathepsin L narrow and clear. In cathepsin B, there is no equivalent residue for the Met-161 of cathepsin L. Consequently, the space corresponding to that residue is opened as a part of the S_2 pocket. This is because the loop from Val-191 to Gly-198 in cathepsin B (from Pro-154 to Asp-162 in cathepsin L) runs through a different route and the peptide strand is kinked at Gly-198. It should be noted that the main chain carbonyl oxygen of Met-161 of cathepsin L is exposed to the solvent. Although E-64 has no interaction with this oxygen, it may be possible to utilize this hydrogen bond acceptor for designing cathepsin L selective inhibitors. At the bottom of the pocket, there is the side chain of Met-70 (Pro-76 for cathepsin B) lying across the cleft, which makes the S_2 pocket of cathepsin L shallower than that of cathepsin B.

By analogy with papain, the S_3 subsite of cathepsin L can be assigned to a region in the middle of the left wall of the cleft (Fig. 4). There is an extended structure of Gly-67 and Gly-68 at the center of this subsite, surrounded by the side chains of Asn-66, Glu-63 and Leu-69 as well as the carbonyl oxygen of Gly-61. The S_3 subsites of the two enzymes are more similar than the other regions in the active site cleft. However, the side chain of Tyr-75 in cathepsin B, which corresponds to Leu-69 in cathepsin L, makes the subsite narrower than that of cathepsin L. Therefore, cathepsin L may accommodate a more bulky group in this region.

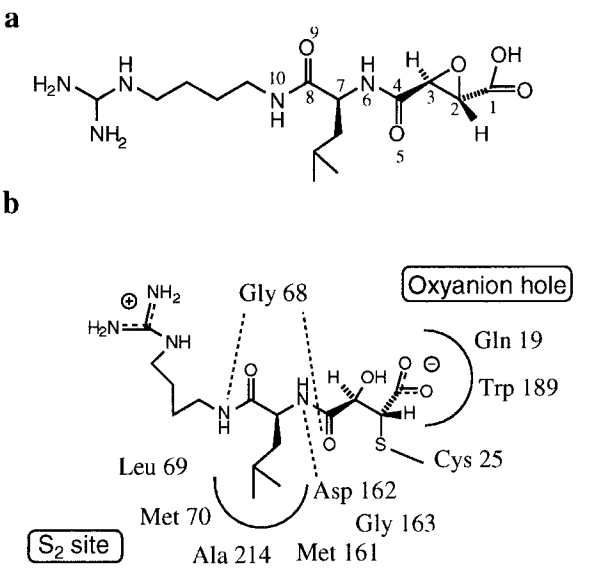

Fig. 3. (a) Chemical structure and atom numbering of E-64. (b) Schematic diagram showing the interactions of E-64 adduct and surrounding residues of cathepsin L. The guanidino group is not fixed in the crystal.

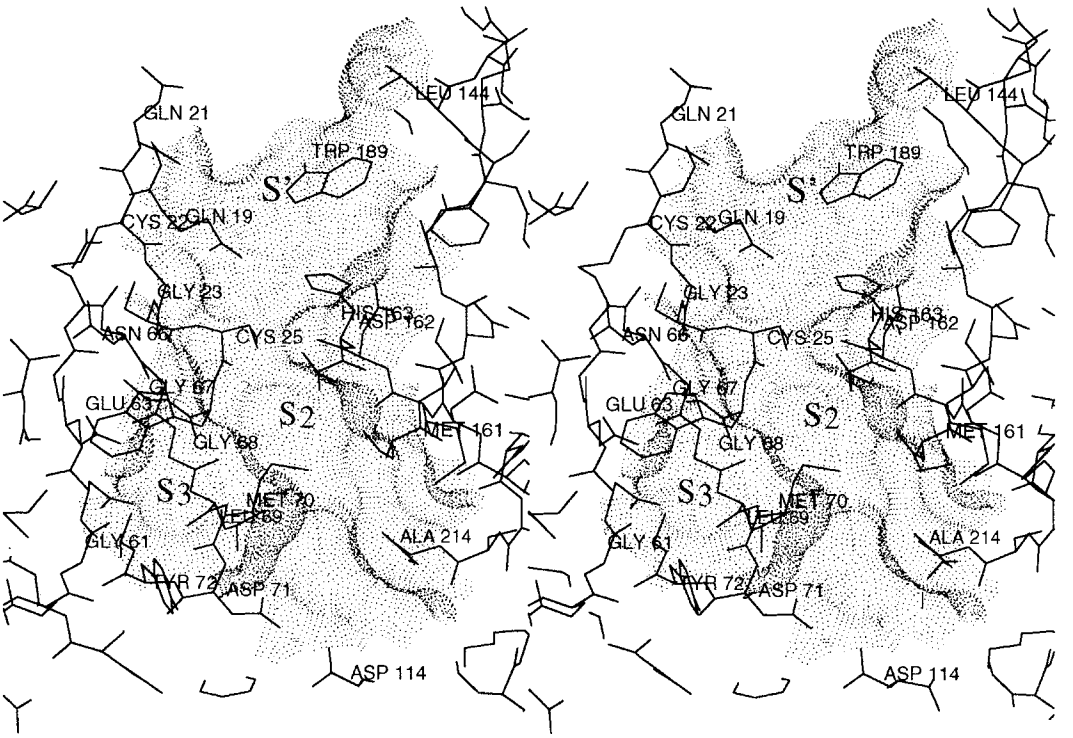

Fig. 4. Stereoview of the active site cleft of cathepsin L. E-64 atoms were omitted for simplicity. Dots represent the Connolly surface with solvent radius of 1.4 Å.

Acknowledgements: We gratefully acknowledge Dr. K. Meguro and Mr. T. Yasuma for valuable discussions and encouragement. Thanks are also due to Dr. D.G. Cork for reviewing the manuscript.

References

[1] H. Kakegawa, T. Nikawa, K. Tagami, H. Kamioka, K. Sumitani, T. Kawata, M. Drobnic-Kosorok, B. Lenarcic, V. Turk, N. Katunuma, FEBS Lett. 321 (1993) 247–250.
[2] K. Brix, P. Lemansky, V. Herzog, Endocrinology 137 (1996) 1963–1974.
[3] R.E. Esser, R.A. Angelo, M.D. Murphy, L.M. Watts, L.P. Thornburg, J.T. Palmer, J.W. Talhouk, R.E. Smith, Arthritis Rheum. 37 (1994) 236–247.
[4] C. Tomssen, M. Schmitt, L. Goretzki, P. Oppelt, L. Pache, P. Dettmar, F. Janicke, H. Graeff, Clin. Cancer Res. 1 (1995) 741–746.
[5] A. Trabandt, U. Muller-Ladner, J. Kriegsmann, R.E. Gay, S. Gay, Lab. Invest. 73 (1995) 205–212.
[6] D. Musil, D. Zucic, D. Turk, R.A. Engh, I. Mayer, R. Huber, T. Popovic, V. Turk, T. Towatari, N. Katunuma, W. Bode, EMBO J. 10 (1991) 2321–2330.
[7] D. Turk, M. Podovnik, T. Popovic, N. Katunuma, W. Bode, R. Huber, V. Turk, Biochemistry 34 (1995) 4791–4797.
[8] Z. Jia, S. Hasnain, T. Hirama, X. Lee, J.S. Mort, R. To, C.P. Huber, J. Biol. Chem. 270 (1995) 5527–5533.
[9] R. Coulombe, P. Grochulski, J. Sivaraman, R. Menard, J.S. Mort, M. Cygler, EMBO J. 15 (1996) 5492–5503.
[10] T. Nomura, A. Fujishima, Y. Fujisawa, Biochem. Biophys. Res. Commun. 228 (1996) 792–796.
[11] P.M.D. Fitzgerald, J. Appl. Crystallogr. 21 (1988) 273–278.
[12] Brunger, A.T. (1993) X-PLOR Version 3.1 Manual, Yale University Press, New Haven/London.
[13] K.I. Varughese, F.R. Ahmed, P.R. Carey, S. Hasnain, C.P. Huber, A.C. Storer, Biochemistry 28 (1989) 1330–1332.
[14] K.I. Varughese, Y. Su, D. Cromwell, S. Hasnain, N. Xoung, Biochemistry 31 (1992) 5172–5176.
[15] D. Yamamoto, K. Matsumoto, H. Ohishi, T. Ishida, M. Inoue, K. Kitamura, H. Mizuno, J. Biol. Chem. 266 (1991) 14771–14777.
[16] M.-J. Kim, D. Yamamoto, K. Matsumoto, M. Inoue, T. Ishida, H. Mizuno, S. Sumiya, K. Kitamura, Biochem. J. 287 (1992) 797–803.
[17] P.J. Kraulis, J. Appl. Crystallogr. 24 (1991) 946–950.