Crystal structures of human procathepsin B at 3.2 and 3.3 Å resolution reveal an interaction motif between a papain-like cysteine protease and its propeptide

Dušan Turk[a,b,*], Marjeta Podobnik[a], Robert Kuhelj[a], Marko Dolinar[a], Vito Turk[a]

[a]*Dept. of Biochem. and Mol. Biol., Jozef Stefan Institute, Jamova 39, 61111 Ljubljana, Slovenia*
[b]*Max-Planck-Institute for Biochemistry, 82152 Martinsried, Germany*

Received 14 March 1996

Abstract A wild-type human procathepsin B was expressed, crystallized in two crystal forms and its crystal structure determined at 3.2 and 3.3 Å resolution. The structure reveals that the propeptide folds on the cathepsin B surface, shielding the enzyme active site from exposure to solvent. The structure of the enzymatically active domains is virtually identical to that of the native enzyme [Musil et al. (1991) EMBO J. 10, 2321–2330]: the main difference is that the occluding loop residues are lifted above the body of the mature enzyme, supporting the propeptide structure.

Key words: Cathepsin B; Cysteine protease; Proenzyme; Crystal structure; Papain

1. Introduction

The lysosomal cysteine proteases, cathepsins B, H, L, S and C, are well characterized proteins with closely related amino acid sequences, belonging to the papain superfamily [1]. They have been implicated in many normal cellular processes such as protein degradation and turnover [2], antigen presentation [3,4], bone remodeling [5] and prohormone processing [6]. The two most extensively studied cysteine proteases, cathepsin B and cathepsin L, appear to be involved in a variety of disease states such as rheumatoid arthritis [7], Alzheimer's disease [8], tumor invasiveness [9] and osteoporosis [10].

Lysosomal enzymes are synthesized in normal cellular processes as glycosylated higher molecular weight precursors, which, during their maturation, undergo several processing steps by limited proteolysis [2]. Cathepsin B maturation includes removal of the N-terminal propeptide, the C-terminal extension and a dipeptide between residues 47 and 50 (mature enzyme numbering). The product is an enzymatically active molecule with two chains covalently cross-linked by a disulfide bridge [11,12].

The crystal structures of human and rat cathepsin B [13,14] have confirmed high structural similarity to plant enzyme papain [16] and structural reasons for cathepsin B peptidyl dipeptidase activity [13,15] have also been revealed. An understanding of the molecular basis for the activation mechanism [17–19] of cathepsin B requires knowledge of the 3-dimensional structure of the proenzyme. Crystal structures of several serine, aspartic and zinc proteases and their zymogens provided clues necessary for understanding of the structural bases of their activation mechanisms ([20–23]; only some papers are referenced here). The 3-dimensional structure of the human cathepsin B proenzyme reported here now sheds light on an as yet unresolved interaction between a cysteine protease and its propeptide.

2. Materials and methods

Recombinant human procathepsin B was expressed in *E. coli* in the form of inclusion bodies as described by Kuhelj et al. [24]. After folding, the protein was concentrated by a Centricon 10 concentrator (Amicon) to a concentration of 15 mg/ml. When compared to the procathepsin B cDNA sequence [11], the protein expressed possessed an amino acid substitution in the pro-part (Val-P9 to Leu), probably due to polymorphism [24]. The protein was expressed with an additional Met on the N-terminus and without the C-terminal hexapeptide extension. Two crystal forms were obtained in sitting drops at 20°C using the vapor diffusion technique. Hexagonal crystals (dimensions $0.3\times0.1\times0.1$ mm^3) were grown from 2 M ammonium sulphate, 700 mM Tris-HCl pH 7.1. They belong to space group $P6_522$ and have one molecule per asymmetric unit. Orthorhombic crystals (dimensions $0.2\times0.1\times0.1$ mm^3) were grown from 30% PEG 6000, 0.2 M ammonium sulphate, 0.1 M HEPES-NaOH, pH 7.3. They belong to space group $P2_12_12_1$ with one molecule per asymmetric unit. Diffraction data up to 3.3 Å resolution for hexagonal crystals and 3.2 Å for the orthorhombic ones were collected on an MAR imaging plate system using a Rigaku rotation anode with copper radiation as the X-ray source. All diffraction data were processed using the MOSFLM [25] and CCP4 package [26]. The completeness of data is 94.2% in the resolution range of 10.0–3.3 Å for the hexagonal crystal form and 88.5% in the resolution range of 10.0–3.2 Å for the orthorhombic crystal form.

The orientation and position of the procathepsin B molecules in both crystal forms were found with the help of the AMoRe suite of programs [27]. Human native cathepsin B structure [13] was used for the starting search model. The rotation and translation functions were computed using data between 15 and 4 Å resolution, and gave R values of 0.416 and 0.420 for the hexagonal and orthorhombic crystal forms, respectively.

Henceforth, the program MAIN running on an Silicon Graphics Indigo 2 [28] was used interactively for electron density map calculations, model building, structure refinement and electron density averaging procedures unless explicitly specified otherwise. The parameter set of Engh and Huber [29] was used throughout the model building and structure refinement procedures.

Visual inspection of the first electron density map showed clear electron density for the search model cathepsin B and revealed some regions where the propeptide chain was bound. It was also evident that the cathepsin B occluding loop residues (106–124) did no have adequate electron density. A model with occluding loop residues omitted was then crystallographically refined and the resulting map used for building segments of propeptide chain as a polyalanine model.

When the orthorhombic crystal form data became available, the electron density between the two crystal forms was averaged in several cycles. The molecular model was built and crystallographically refined

*Corresponding author. Fax: (386) (61) 273 594.
E-mail: dusan.turk@ijs.si

Abbreviations: RMS, root mean square; PEG, polyethylene glycol.

SSDI S0014-5793(96)00309-2

in both crystal forms simultaneously. Besides the usually applied energy terms, we also used non-crystallographic symmetry constraints between the models of the two crystal forms. These enabled us to obtain the unbroken density of the whole proenzyme chain, including the occluding loop in the orthorhombic form, and allowed us to complete the model building. Parts of the chain in the hexagonal form, however, were still not visible. The resulting crystallographic R values were 0.255 for orthorhombic and 0.265 for hexagonal data. Subsequent molecular dynamics procedures (at 500 K for 3000 steps) followed by annealing using X-PLOR [30] of the two crystal forms models brought the R values down to 0.229 and 0.230 for the orthorhombic and hexagonal crystal forms, respectively. After the models from both crystal forms were superimposed, it became evident that the annealing procedure kept most parts of the structures close to each other, only the propeptide residues in the region from Gly-P47 to Lys-P62 being significantly displaced (up to 2.0 Å). The overlap of superimposed electron density maps in this region became worse and further minimization attempts with MAIN to enforce similarity in the region by applying noncrystallographic symmetry constraints only raised the R value. We concluded therefore that the differences between the two crystal forms in this part of the propeptide chain are real.

The crystallographic R values of the final models are 0.2158 (5051 unique reflections in resolution range between 10.0 and 3.2 Å) and 0.2313 (4011 unique reflections in the resolution range between 10.0 and 3.3 Å) for the orthorhombic and hexagonal model, respectively. The backbone conformation of all residues is within the Ramachandran plot limits. The RMS deviations from target values are 0.0114 and 0.0119 Å for bond lengths and 1.77 and 1.83° for angles of the crystallographically weighted atoms for the orthorhombic and hexagonal models, respectively. No individual or group atomic B value refinement was applied to either of the structures.

3. Results and discussion

The crystal structures of the single chain procathepsin B comprise the propeptide sequence from Arg-P1 to Lys-P62 and the cathepsin B sequence from Leu-1 to Asp-254 (Fig. 1). The whole chain, apart from a few N-terminal residues, is well defined by the electron density maps of the orthorhombic crystal form. In addition to the N-terminal residues, the hexagonal structure lacks adequate electron density for the propeptide residues Glu-P59 and Asp-P60 and mature enzyme residues Asn-113, Gly-114 and Ser-115 from the occluding loop region. The following description of the structure refers to the orthorhombic model.

The structure of the procathepsin B mature enzyme domains is virtually identical to that of cathepsin B [13] (Fig. 1). The RMS deviation between the CA atoms of mature enzyme structure [13] and structurally equivalent CA atoms of the two independently refined proenzyme structures are 0.43 and 0.44 Å for the orthorhombic and hexagonal forms, respectively. The mature enzyme domains contact each other through an extended polar interface across a large surface in the middle of the enzyme. The structure is reminiscent of a book that, when closed, brings together the active site residues Cys-29 and His-199 at the bottom of a 'V'-shaped active site cleft situated on the top of the enzyme structure. (The mature enzyme domains are denoted throughout the text by R-and L-according to their right and left position in the standard view used in Fig. 1.)

The only substantial difference between the cathepsin B structure and the structure of enzymatic domains in the proenzyme is observed in the occluding loop region (residues between Ile-105 and Pro-126). The occluding loop in the cathepsin B structure blocks the active site cleft from the back, whereas in the proenzyme it is lifted away from the enzyme surface, building a wall that interacts extensively with the propeptide, thus supporting its structure (Fig. 2).

The propeptide chain embraces the mature cathepsin B domains (see Fig. 1). The propeptide starts as a disordered chain that runs into a quite regularly folded region with defined secondary structure elements (Leu-P9 to Cys-P42). The folded region begins as a short extended chain followed by a α-β structural motif (three-turn α-helix and associated short antiparallel β-structure). From there the next six residues run in an extended conformation with a kink in the middle. From the kinked region the propeptide chain folds down to the active site as a two-turn α-helix (Asp-P34 to Cys-P42) which sits on top of the active site. From there on the chain runs in an extended conformation towards the cathepsin B N-terminus, quite loosely packed on the enzyme surface.

Interactions between the propeptide chain and the mature

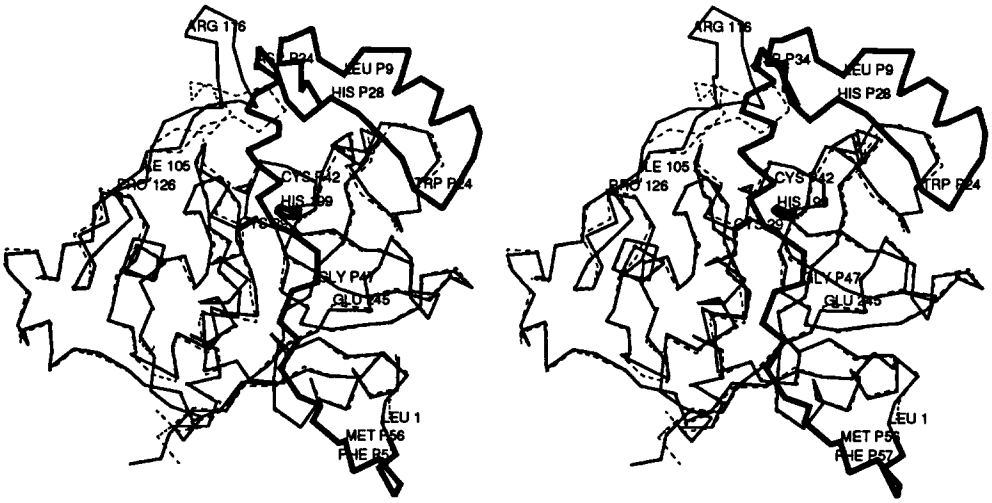

Fig. 1. CA plot of human procathepsin B structure superimposed on the human cathepsin B structure [13]. The proenzyme fold is shown by continuous lines: the propeptide chain is indicated by the bold line, the mature enzyme domains by thin lines and the active-site residues Cys-29 and His-199 are marked and their side chains are shown by bold lines. Proenzyme residues mentioned in the text are marked. The human cathepsin B structure is presented using broken lines.

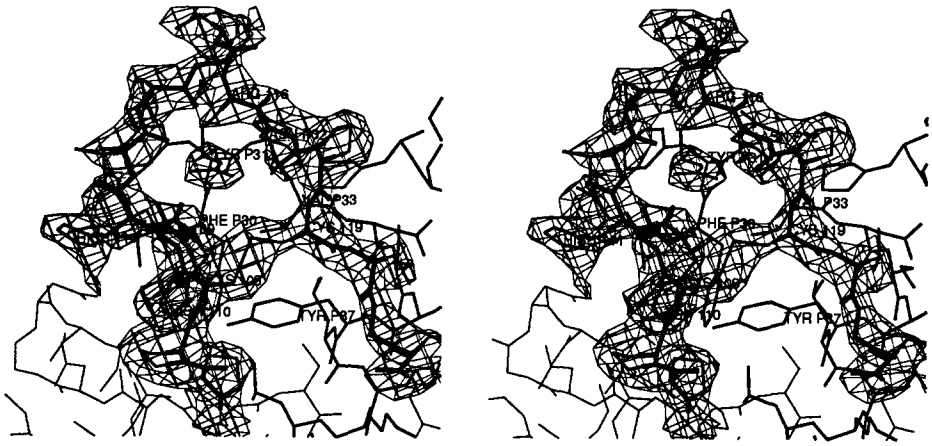

Fig. 2. Electron density map phased by the final model and contoured at the 1.0 σ level is shown for 2.3 Å around the occluding loop residues (only residues between Pro-106 and Cys-119 are seen). The occluding loop residues are drawn in bold, propeptide residues (in the back) with the line of intermediate thickness and the rest of mature cathepsin B domain is shown in thin line.

enzyme domains exhibit hydrophobic as well and hydrophilic character. No disulfide bridge or salt bridges are observed in the interface.

4. Propeptide folded region

The folded region of the propeptide sits on top of the Phe-174– His-199 loop, the α-β structural motif in its middle being anchored with the Trp-P24 indole ring (Fig. 3). This results in the propeptide shielding a substantial part of the cathepsin B hydrophobic surface from solvent. The location of the β part of the α-β structural motif on the top of the loop is further stabilized by an anti-parallel β-strand hydrogen bond arrangement.

The residues of the kinked region (from Asn-P29 to Asp-P34) interact with the cathepsin B occluding loop through a number of hydrogen bonds that involve main as well as side chain interactions. The side chains of the propeptide residues Tyr-P37 and Arg-P40 are squeezed between the occluding loop and the body of the mature enzyme. The Tyr-P37 ring is placed below the occluding loop disulfide bridge 108–119 and the Arg-P40 side chain is placed beneath Glu-122. These residues serve to lift the occluding loop away from the rest of the enzyme surface.

Removal of the propeptide thus results in substantial destabilization of the occluding loop conformation and it essentially falls down on the enzyme surface into the arrangement observed in the mature cathepsin B structure. The occluding loop of cathepsin B thus appears to be quite flexible and capable of adopting two quite different conformations in spite of an internal disulfide bridge and two pairs of sequential proline residues.

We assume that the propeptide N-terminal region (including residues from Ser-P5 to Cys-P42), similarly to the occluding loop, cannot keep its tertiary structure observed in the proenzyme after dissociation from the enzyme domain. The spatial arrangements of those parts with regular secondary structure are stabilized mostly through contacts with the surface of cathepsin B. The only exception is the α-β structural motif which is internally stabilized by two hydrogen bonds:

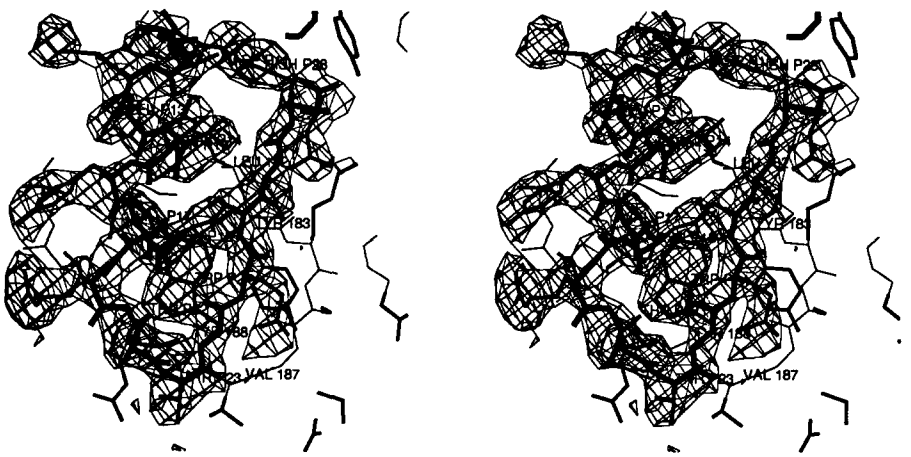

Fig. 3. Electron density map phased by the final model and contoured at the 1.0 σ level is shown for 2.3 Å around the α-β structural motif including residues from P9 to P42 and their surroundings. The Trp-P24 moiety is shown embedded in the middle of the hydrophobic surface of the cathepsin B Phe-174–His 199 loop. The propeptide main chain is drawn with the thickest lines and the side chains with lines of intermediate thickness. Thin lines represent the bonds of the cathepsin B mature form domains.

The first is formed between the Asn-P18 side chain OD1 atom and the main chain amide hydrogen of Ala-P26 and brings together the last helical turn and the middle of the β strand. The second one connects the carbonyl groups of Leu-P9 and Ser-P10, situated before the beginning of the helix, to the imidazole ring of His-P28, the last residue in the β region. Thus the substantial hydrophobic surface of the propeptide folded region that, in the proenzyme structure, spans the cathepsin B R-domain hydrophobic surface might, in solution, merge into a single hydrophobic core enclosed by the hydrophilic surface of the secondary structure elements which, in the proenzyme structure, are exposed to solvent. This reasoning finds support in kinetic measurements reported by Fox [31], where it was shown that inhibition of cathepsin B by its propeptide exhibits slow binding kinetics. Thus binding of cathepsin B propeptide to the enzyme would appear to involve structural rearrangement.

5. Propeptide linker peptide

After leaving the folded region, the propeptide chain runs in an extended chain along the active site cleft across the S1 and S2 binding subsites (Gly-P43–Leu-P46), making a turn across Glu-245 at Gly-P47–Pro-P52, and then continues quite loosely packed along the L-and R-domain interface towards the cathepsin B N-terminal residues (Gln-P53–Lys-P62). The chain is slightly lifted above the cathepsin B surface, with only a couple of main chain interactions, and is therefore rather flexible. The flexibility is indicated by the different chain placement in the two crystal structures. Deviation between some equivalent atoms exceeds 2 Å. Flexibility in this part is not surprising, since cleavage of peptide bonds in this region results in cathepsin B activation [17,18]. The scissile bonds are exposed to solvent in the proenzyme structure and thus available for cleavage by an approaching protease. Besides activation with other enzymes, including mature cathepsin B, it was also reported that procathepsin B, when introduced into an acidic environment, is capable of autocatalytic processing [18,19,32]. The reported autocleavage scissile bond is between the residues Met-P56 and Phe-P57 [33]. Obviously the acidic environment neutralizes some carboxylic group(s), which results in liberation of the bound propeptide and consequent exposure of the active site to solvent. However, since no salt bridge between propeptide and mature enzyme is observed in the proenzyme structure the mechanism of procathepsin B autoactivation remains to be answered.

Acknowledgements: We gratefully acknowledge the assistance of F. Schneider and A. Bergner in data collection, data processing and initial molecular replacement with the first hexagonal data set. Profs. W. Bode and R. Huber are gratefully acknowledged for their permanent interest and critical discussions. Thanks for critical reading of the manuscript go to Prof. R. Pain. Atomic coordinates of the two structures will be deposited in the Brookhaven Protein Data Bank. This work was supported by the Slovenian Ministry of Science and Technology and the European Community Grant CHRX-CT94-0535.

References

[1] Berti, P.J. and Storer, A.C. (1995) J. Mol. Biol. 246, 273–283.
[2] Bond, J.S. and Butler, P.E. (1987) Annu. Rev. Biochem. 56, 333–364.
[3] Roche, P.A. and Cresswell, P. (1991) Proc. Natl. Acad. Sci. USA 87, 1927–1933.
[4] Morton, P.A., Zacheis, M.L., Giacoletto, K.S., Manning, J.A. and Schwartz, B.D. (1995) J. Immunol. 154, 137–150.
[5] Tezuka, K., Tezuka, Y., Maejima, A., Sato, T., Nemoto, K., Kamioka, H., Hakeda, Y. and Kumegawa, M. (1994) J. Biol. Chem. 269, 1106–1109.
[6] Krieger, T.J. and Hook, V.Y.H. (1991) J. Biol. Chem. 266, 8376–8383.
[7] Buttle, D.J., Saklatvala, J. and Barrett, A.J. (1993) Agents Action Suppl. 39, 161–165.
[8] Pentaceska, S., Burke, S., Watson, S.J. and Devi, L. (1994) Neuroscience 59, 729–738.
[9] Sloane, B.F., Moin, K. and Lah, T.T. (1994) in: Biochem. Mol. Aspects Select. Cancers, vol. 2 (Pretlow, T.G. and Pretlow, T.P. eds.) pp. 411–460, Academic Press, San Diego.
[10] Delaisse, J.M., Pascale, L. and Vaes, G. (1991) Biochem. J. 279, 167–174.
[11] Chan, S.J., San Segundo, B., McCormick, M.B. and Steiner, D.F. (1986) Proc. Natl. Acad. Sci. USA 83, 7721–7725.
[12] Takio, K., Towatari, T., Katunuma, N., Teller, D.C. and Titani, K. (1983) Proc. Natl. Acad. Sci. USA 80, 3666–3670.
[13] Musil, D., Zučič, D., Turk, D., Engh, R.A., Mayr, I., Huber, R., Popovič, T., Turk, V., Towatari, T., Katunuma, N. and Bode, W. (1991) EMBO J. 10, 2321–2330.
[14] Jia, Z., Hasnain, S., Hirmama, T., Lee, X., Mort, J.S., To, R. and Huber, C.P. (1995) J. Biol. Chem. 270, 5627–5633.
[15] Turk, D., Podobnik, M., Popovič, T., Katunuma, N., Bode, W., Huber, R. and Turk, V. (1995) Biochemistry 34, 4791–4797.
[16] Drenth, I., Jansonius, J.N., Koekoek, R., Swen, H.M. and Wolthers, B.G. (1968) Nature 249, 54–57.
[17] Rowan, A.D., Mason, P., Mach, L. and Mort, J.S. (1992) J. Biol. Chem. 267, 15993–15999.
[18] Mach, L., Stüwe, K., Hagen, A., Ballaun, C. and Glössl, J. (1992) Biochem. J. 282, 577–582.
[19] Mach, L., Schwihla, H., Stüwe, K., Rowan, A.D., Mort, J.S. and Glössl, J. (1993) Biochem. J. 293, 437–442.
[20] Huber, R. and Bode, W. (1978) Acc. Chem. Res. 11, 114–122.
[21] James, M.N.G. and Sielecki, A.R. (1986) Nature 319, 33–38.
[22] Coll, M., Guasch, A., Aviles, F.X. and Huber, R. (1991) EMBO J. 10, 1–9.
[23] Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., Almog, O., Gilliland, G. and Gallagher, T. (1995) Biochemistry 34, 10310–10318.
[24] Kuhelj, R., Dolinar, M., Pungerčar, J. and Turk, V. (1995) Eur. J. Biochem. 229, 533–539.
[25] Leslie, A.G.W. (1994) Mosflm Users Guide, Mosflm version 5.23, MRC Laboratory of Molecular Biology, Cambridge, UK.
[26] CCP4 (1994) Acta Crystallogr. D50, 760–763.
[27] Navaza, J. (1994) Acta Crystallogr. A50, 157–163.
[28] Turk, D. (1992) Doctoral Thesis Technische Universität München.
[29] Engh, R.A. and Huber, R. (1991) Acta Crystallogr. A47, 392–400.
[30] Brünger, A.T. (1992) X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR, Yale University Press, New Haven, CT.
[31] Fox, T., De Miguel, E., Mort, J.S. and Storer, A.C. (1992) Biochemistry 31, 12571–12576.
[32] Mach, L., Mort, J.S. and Glössl, J. (1994) Maturation of human procathepsin B. J. Biol. Chem. 269, 13030–13035.
[33] Mach, L., Mort, J.S. and Glössl, J. (1994) J. Biol. Chem. 269, 13036–13040.