Acta Cryst. (1996). D52, 874–875

Crystallization of rat procathepsin B†

J. SIVARAMAN,[a,b,c] RENÉ COLOUMBE,[a,b,c] MARIE-CLAUDE MAGNY,[d] PATRIZIA MASON,[d] JOHN S. MORT[b,d] AND MIROSLAW CYGLER[a,b,c]* at [a]*Biotechnology Research Institute, National Research Council of Canada, 6100 Royalmount Avenue, Montreal, Quebec H4P 2R2, Canada,* [b]*Protein Engineering Network of Centers of Excellence,* [c]*Montreal Joint Centre for Structural Biology and* [d]*Shriners Hospital for Crippled Children and Department of Surgery, McGill University, Montreal, Quebec, Canada. E-mail: mirek.cygler@bri.nrc.ca*

(Received 29 November 1995; accepted 7 February 1996)

Abstract

Rat procathepsin B has been expressed in the yeast *Pichia pastoris*. To facilitate crystallization of the proform two mutations were introduced: Cys29Ser to avoid self-processing and Ser115Ala to eliminate an *N*-glycosylation site. The recombinant protein was purified and crystallized by vapor diffusion against mother liquor containing 100 m*M* KSCN, 100 m*M* phosphate buffer, pH 6.5 and polyethylene glycol (PEG) 3350 as a precipitating agent. Crystal size was increased by multiple macroseeding. At a 16% PEG concentration trigonal crystals were obtained, with the space group $P3_121$ and $a = 99.6$, $c = 141.4$ Å, $\gamma = 120°$. They diffract to 2.8 Å resolution using a rotating-anode source. At a concentration of 11% PEG, rod-shaped crystals were grown. They are monoclinic, space group $P2_1$, $a = 62.8$, $b = 67.9$, $c = 100.4$ Å, $\beta = 98.2°$ and diffract to approximately 3.5 Å.

1. Introduction

Proteases are synthesized remote from their site of action and it is vital for the biological system that their proteolytic activity be released only at their final destination (Knop, Schiffer, Rupp & Wolf, 1993). To accomplish this goal most proteases are synthesized as inactive precursors and require a processing step to attain activity. In most cases the activation of the proenzyme is achieved by proteolytic removal of the N-terminal fragment, the proregion. Additional proteolytic processing steps usually follow, often yielding multiple-chain forms of the active enzyme (Barrett & Kirschke, 1981). The size of the proregion, its mechanism of inhibition and the mechanism of processing leading to the activation of the enzyme differs between protease families. Crystallographic studies on proenzymes have been reported to date for serine (Freer, Kraut, Robertus, Wright & Xuong, 1970; Huber & Bode, 1978; Bryan *et al.*, 1995), aspartic (James & Sielecki, 1986; Moore, Sielecki, Chemaia, Tarasova & James, 1995) and metalloproteases (Coll, Guasch, Avilés & Huber, 1991; Becker *et al.*, 1995). These studies indicated the structural basis for the inactivation of these enzymes by their propeptides. No such information is yet available for the cysteine protease family.

Cysteine proteases represent a major component of the lysosomal proteolytic repertoire and play an important role in intracellular protein degradation (Knop *et al.*, 1993). Cathepsin B is a lysosomal cysteine protease and its sequence (Takio, Towatari, Katunuma, Teller & Titani, 1983) and structure (Musil *et al.*, 1991) clearly identifies it as a member of the large family of papain-like enzymes (Berti & Storer, 1995). As with other cathepsins, it is synthesized as a precursor on membrane-bound ribosomes, passes through the Golgi and is transported to the lysosomes where it is processed to the mature form in the acidic pH of the lysosomal environment (Bohley & Seglen, 1992). Cathepsin B is capable of autoprocessing *in vitro* in acidic pH (Mach, Mort & Glössl, 1994). It is not clear if this process is primarily an intra- or intermolecular event. Detailed sequence comparison of the proteases from the papain family have shown that two subclasses are clearly distinguishable. They differ in the length of the propeptide and there is little if any sequence homology in this region between the two subclasses. Additionally, there are specific regions of the mature proteins where the sequence differences are much higher than the rest of the chain (Karrer, Peiffer & DiTomas, 1993; Berti & Storer, 1995). Cathepsin B represents the subfamily with the shorter prosegment. It is characterized by an ~65 residue long proregion, as compared to that of the second class which, like cathepsin L for example, have a proregion of ~100 residues. The three-dimensional structures of human and rat cathepsin B, and their complexes with various inhibitors, have been reported (Musil *et al.*, 1991; Turk *et al.*, 1995; Jia *et al.*, 1995) but the structure of the proregion is unknown. We describe here the crystallization of an active site mutant of procathepsin B, which is incapable of self-processing.

2. Methods and results

The cDNA for rat proenzyme has been expressed as an α-factor fusion in the methylotrophic yeast *Pichia pastoris*. Although autoprocessing can be slowed down by maintaining the proenzyme preparation at a high pH, processing proceeds over the long periods required for crystallization trials. We have, therefore, constructed a mutant form of procathepsin B where the active-site Cys29 has been replaced by a serine. In addition, the *N*-glycosylation site occurring within the so-called occluding loop (Musil *et al.*, 1991) was deleted by mutating Ser115, within the Asn-*X*-Ser glycosylation concensus sequence, to an alanine. That leaves only one *N*-glycosylation site, Asn21 located within the Asn-Thr-Thr sequence in the proregion.

Pichia pastoris strain GS115 containing the integrated procathepsin B construct from the vector pPIC9 (Invitrogen) was cultured in buffered glycerol complex medium as recommended by the manufacturer. Cells from a 3 d 4 l initial culture were resuspended into 800 ml of buffered methanol complex medium for induction of the recombinant protein by methanol. After a further 3 d the medium was concentrated tenfold by ultrafiltration and applied, as two batches, onto Sephacryl S-200 HR column (5 × 70 cm) in

† NRCC publication No. 39916.

© 1996 International Union of Crystallography
Printed in Great Britain – all rights reserved

25 mM Tris-HCl, pH 7.5 containing 0.15 M NaCl at a flow rate of 2 ml min^{-1}. Fractions (14 ml) were analyzed by SDS-PAGE and the procathepsin B peak pooled and concentrated. This sample was of sufficient purity to warrant the crystallization trials. Yields of 36 and 80 mg were obtained for two different cultures.

The crystallization trials were conducted using the hanging-drop method and a wide range of conditions were tested using the factorial procedure (Jancarik & Kim, 1991). The protein concentration was near to 7.0 mg ml^{-1} and the drops were prepared by mixing 2 μl of the protein solution with 2 μl of the well solution. Although the initial screening did not result in obtaining crystals it indicated that the polyethylene glycol (PEG) was the most promising precipitating agent. In subsequent screening attempts we used different molecular weight PEG's, various buffers with pH near neutral and salts. Finally, small crystals appeared in the drops equilibrated against the wells containing PEG 3350 as a precipitant, $NaKPO_4$ buffer, pH 6.5 and KSCN. By systematically scanning around these preliminary conditions we have found that the morphology of the crystals obtained depended on the concentration of the precipitant. Changing the precipitant concentration by ~5% led to a different morphology. The two crystal morphologies corresponded to two crystal forms. The original crystals were too small for data collection but both forms were grown to larger sizes by multiple transfers of growing crystals to new drops with fresh protein (macroseeding).

For routine growth of procathepsin B crystals the protein was concentrated to 7.4 mg ml^{-1}. The crystals of form 1 were obtained from 100 mM KSCN, 100 mM phosphate buffer, pH 6.5 and 14–16% PEG 3350. They appeared after 5–7 d and had the shape of hexagonal plates. For macroseeding, the well contained the same buffer and salt but only 14% PEG 3350. The drop was prepared by mixing 3 μl of protein solution with 3 μl of well solution and was equilibrated with the well solution for 1 d prior to transfer of the seed. A single crystal was then washed thoroughly in the well solution and transferred to the drop. When the crystal stopped growing, usually after approximately 5–7 d, it was transferred to a fresh drop. After repeating this process three to four times the crystals grew to the maximum size of 0.5 × 0.4 × 0.25 mm. These crystals belong to the trigonal system, space group $P3_121$, with $a = 99.6$, $c = 141.4$ Å, $\gamma = 120^\circ$. With the assumption of two molecules in the asymmetric unit the $V_m = 2.85$ Å^3 Da^{-1} is within the expected range and corresponds to a solvent content of 56%. With a rotating-anode generator the diffraction of these crystals extends to approximately 2.8 Å.

At a lower PEG 3350 concentration (11%) different crystals appeared. These form 2 crystals were rod shaped and could not be grown to the same size as those of form 1. The macroseeding was carried out using the same technique as described for form 1, except that the PEG 3350 concentration in the well was kept at 11%. After three to four transfers to fresh drops the crystals reached their maximum size of 0.4 × 0.12 × 0.08 mm. Additional transfers did not increase their size any further. These crystals are monoclinic, likely $P2_1$, with $a = 62.8$, $b = 67.9$, $c = 100.4$ Å, $\beta = 98.2^\circ$ and appear to have also two molecules of procathepsin B in the asymmetric unit ($V_m = 3.00$ Å^3 Da^{-1}). These crystals diffract less well than form 1, approximately to 3.5 Å resolution.

We have collected a native data set from crystals of form 1 mounted in a thin-walled glass capillary. Data were collected to 2.8 Å resolution at room temperature on an R-AXIS IIC area detector mounted on a RU-300 rotating-anode generator with a 0.5 × 10 mm focal size and operating at 50 kV, 260 mA. The crystal-to-detector distance was 130 mm. The data were collected in 1.5° oscillation frames over a 170° oscillation range. The frames were processed with R-AXIS software resulting in 82 919 observations with $I > \sigma(I)$ that merged to 19 015 independent reflections (redundancy 4.4) with an $R_{merge} = 0.097$. Completeness of the data to 2.8 Å resolution was 91%. The last data shell, 2.9–2.8 Å, contained rather weak reflections, with the average $\langle I/\sigma(I)\rangle = 1.7$. The rotation function using the rat cathepsin B as a model gave two clear solutions corresponding to two molecules in the asymmetric unit. The translation function allowed us to distinguish between two possible space groups, $P3_121$ and $P3_221$, in favor of the former. This model forms only approximately 75% of the procathepsin B. We are now in the process of determining the structure of this proenzyme.

We thank Dr Andrew C. Storer and Dr Robert Ménard for helpful discussions and critical reading of the manuscript.

References

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

Becker, J. W., Marcy, A. I., Rokosz, L. L., Axel, M. G., Burbaum, J. J., Fitzgerald, P. M. D., Cameron, P. M., Esser, C. K., Hagmann, W. K., Hermes, J. D. & Springer, J. P. (1995). *Protein Sci.* **4**, 1966–1976.

Berti, P. J. & Storer, A. C. (1995). *J. Mol. Biol.* **246**, 273–283.

Bohley, P. & Seglen, P. O. (1992). *Experientia*, **48**, 151–157.

Bryan, P., Wang, L., Hoskins, J., Ruvinov, S., Strausberg, S., Alexander, P., Almog, O., Gilliland, G. & Gallagher, T. (1995). *Biochemistry*, **34**, 10310–10318.

Coll, M., Guasch, A., Avilés, F. X. & Huber, R. (1991). *EMBO J.* **10**, 1–9.

Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T. & Xuong, N. H. (1970). *Biochemistry*, **9**, 1997–2009.

Huber, R. & Bode, W. (1978). *Acc. Chem. Res.* **11**, 114–122.

James, M. N. G. & Sielecki, A. R. (1986). *Nature (London)*, **319**, 33–38.

Jancarik, J. & Kim, S. H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Jia, Z., Hasnain, S., Hirama, T., Lee, X., Mort, J. S., To, R. & Huber, C. P. (1995). *J. Biol. Chem.* **270**, 5527–5533.

Karrer, K. M., Peiffer, S. L. & DiTomas, M. E. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 3063–3067.

Knop, M., Schiffer, H. H., Rupp, S. & Wolf, D. H. (1993). *Curr. Opin. Cell Biol.* **5**, 990–996.

Mach, L., Mort, J. S. & Glössl, J. (1994). *J. Biol. Chem.* **269**, 13036–13040.

Moore, S. A., Sielecki, A. R., Chemaia, M. M., Tarasova, N. I. & James, M. N. G. (1995). *J. Mol. Biol.* **247**, 466–485.

Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N. & Bode, W. (1991). *EMBO J.* **10**, 2321–2330.

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

Turk, D., Podobnik, M., Popovic, T., Katunuma, N., Bode, W., Huber, R. & Turk, V. (1995). *Biochemistry*, **34**, 4791–4797.