Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

An unfolding/refolding step helps in the crystallization of a poorly soluble protein

Guido Kaulmann,[a,b] Gottfried J. Palm,[b,c] Klaus Schilling,[a] Rolf Hilgenfeld[b,d] and Bernd Wiederanders[a]*

[a]Institut für Biochemie I, Klinikum der Friedrich-Schiller-Universität Jena, Am Nonnenplan 2, D-07740 Jena, Germany, [b]Department of Structural Biology and Crystallography, Institute of Molecular Biotechnology, Beutenbergstrasse 11, D-07745 Jena, Germany, [c]Abteilung für Biochemie I, Institut für Chemie und Biochemie, Ernst-Moritz-Arndt-Universität, Soldmannstrasse 16, D-17489 Greifswald, Germany, and [d]Institute of Biochemistry, University of Lübeck, Ratzeburger Allee 160, D-23538 Lübeck, Germany

Correspondence e-mail: bwie@mti.uni-jena.de

Received 4 December 2002
Accepted 17 April 2003

Proteins that are unstable or poorly soluble often elude crystallization. Here, a novel strategy is presented that leads to the crystallization of the isolated N-terminal propeptide of human procathepsin S, a proteinase belonging to the cathepsin L-like endopeptidases of the clan CA1 cysteine peptidases. Being very hydrophobic, the propeptide is extremely poorly soluble in aqueous solvents at neutral pH. Solubility is much better at acidic pH, but the native structure is destroyed under these conditions. A novel approach to the crystallization of this poorly soluble protein is presented in which it is first unfolded in an acidic buffer (pH 4.5) and then mixed with a nearly neutral crystallization buffer (pH 6.75) in which the native conformation should form spontaneously. Crystals were grown at a high concentration of MES (1.14 *M*) with 10% 2-propanol as precipitant. They belong to a tetragonal space group, with unit-cell parameters $a = b = 151.1$, $c = 75.8$ Å. Diffraction data to a resolution of 3.5 Å were obtained.

1. Introduction

The crystallization of poorly soluble proteins is a great challenge. Characteristics such as high hydrophobicity affect and/or prevent their solubilization at the concentrations required for crystallization using several techniques such as sitting-drop vapour-diffusion, hanging-drop vapour-diffusion, batch and dialysis methods (McPherson, 1982, 1999) as well as the gel-acupuncture approach (García-Ruiz & Moreno, 1994). We are at present investigating the propeptide of cathepsin S, a member of the cathepsin L-like papain subfamily (Karrer *et al.*, 1993; Kirschke *et al.*, 1986, 1989; Wiederanders *et al.*, 1992). The 99-residue cathepsin S propeptide (MW 12 kDa) can be regarded as a foldase which promotes the folding process of the mature enzyme (Kreusch *et al.*, 2000; Schilling *et al.*, 2001; Pietschmann *et al.*, 2002). This function requires the tertiary structure of the isolated propeptide, which folds independently of the other parts of the molecule (Maubach *et al.*, 1997; Schilling *et al.*, 2001). The existence of such an ordered structure was first indicated by various spectroscopic and fluorometric investigations of recombinant cathepsin S propeptides (Maubach *et al.*, 1997), but the structure is unknown. These results raise questions as to (i) how the isolated propeptide is structured, (ii) how it interacts with the mature enzyme domains during the folding process and (iii) whether the propeptide spontaneously folds into its native structure upon a shift from acid to neutral pH.

In order to verify our hypothesis of an independently folding tertiary structure, we crystallized this poorly soluble (36 µg ml^{-1}; Maubach *et al.*, 1997) protein using the vapour-diffusion method in a hanging drop. With regard to its ability to fold spontaneously upon a shift from acidic to neutral pH, the protein was dissolved in an acidic buffer, in which the propeptide is unfolded or partially folded. This allows its use at high concentration without the formation of aggregates, precipitates or other phases before mixing with the nearly neutral crystallization buffer (MES with added 2-propanol), which enables the folding of the propeptide and the growth of crystals. We believe that this novel approach will be of general usefulness for proteins which elude crystallization owing to low solubility.

2. Materials and methods

2.1. Cloning and expression

The recombinant human cathepsin S propeptide was expressed in *Escherichia coli*, purified and refolded from inclusion bodies as described previously (Maubach *et al.*, 1997; Schilling *et al.*, 2001). The purified protein was concentrated to 7.5 mg ml^{-1} and stored in aliquots in 50 m*M* sodium acetate buffer pH 4.5 with added 15% acetonitrile.

2.2. Crystallization

In order to circumvent the problem of the achievable protein concentration being too low for crystallization, we dissolved the protein in 50 m*M* sodium acetate buffer pH 4.5 with added 15% acetonitrile. Under these condi-

tions, even the secondary structure of the propeptide is broken, as shown by CD spectroscopy (Maubach *et al.*, 1997). However, the solubility of the protein under these conditions is no longer a limiting factor for crystallization and sufficiently high propeptide concentrations can be achieved. In a first step, the protein (7.5 mg ml^{-1}) was subjected to Crystal Screen I and II crystallization buffers (Hampton Research) at temperatures of 277 and 293 K. After a year, many very small microcrystals were only observed in 0.1 *M* MES (2-morpholinoethanesulfonic acid) pH 6.5, 10% dioxane and 1.6 *M* ammonium sulfate at 293 K. Further screens were carried out using various concentrations of MES, dioxane, ammonium sulfate and protein at various pH values (above 6.0), but no improvement in crystal growth was observed. The subsequent use of 10% 2-propanol as precipitant in MES buffer at concentrations above 1 *M* and a pH above 6.7 led to reproducible growth of diffracting crystals. Finally, a 2 µl hanging drop containing the protein solution (7.5 mg ml^{-1} protein concentration) was equilibrated with the same amount of crystallization solution containing 1.14 *M* MES pH 6.75 and 10% 2-propanol. The mixture of the crystallization solution and protein solution in a 1:1 ratio gave a pH of ∼6.7. The hanging drops were incubated at 293 and 303 K.

2.3. Diffraction data collection

Crystals were flash-cooled in a 100 K nitrogen stream generated by an Oxford Cryosystems Cryostream cooling system (Oxford Cryosystems, Oxford, England). No further cryoprotectant was needed in addition to the 2-propanol in the mother liquor. A complete data set to 3.5 Å resolution was collected at a wavelength of 0.9101 Å using the EMBL synchrotron beamline X11 at Deutsches Elektronen-Synchrotron, Hamburg, Germany. Diffraction intensities were recorded using a MAR Research CCD detector (X-ray Research, Hamburg, Germany). Diffraction data were indexed, integrated and scaled using the *HKL* package (Otwinowski & Minor, 1997). *TRUNCATE* (Collaborative Computational Project, Number 4, 1994) was used to derive structure amplitudes from the measured intensities.

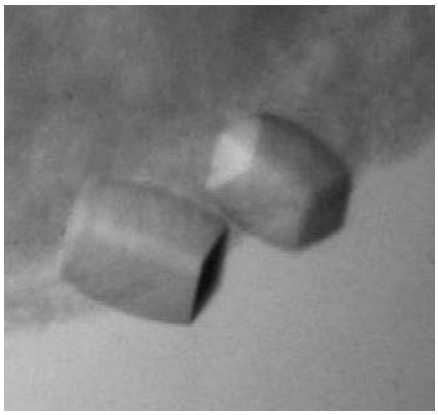

Figure 1
Tetragonal crystals of human cathepsin S propeptide. The crystals have dimensions of 0.25 × 0.15 × 0.10 mm.

3. Results

The first crystals were obtained after three months and grew to maximum dimensions of 20 × 20 × 10 µm at 293 K. At 303 K, the first crystals appeared after 3–5 d and reached maximum dimensions of 0.25 × 0.15 × 0.10 mm within 3–6 months (Fig. 1). Variation of the buffer concentration revealed a significant influence on the number of crystals grown. Above 1.0 *M* MES the appearance of crystals and crystal growth was observed, whereas below 1.0 *M* many small crystals were observed in only a few droplets. The largest crystals were obtained at a MES concentration of 1.14 *M* at 303 K. Variation of the 2-propanol concentration (±10%) yielded no improvement with respect to the number of crystals, velocity of crystal growth or crystal size; in the absence of 2-propanol no crystals were observed. The pH of MES plays a crucial role in the crystallization of the propeptide. As expected, no crystals were observed below pH 6.75, indicating the importance of the nearly neutral pH for the folding of the propeptide molecules (Maubach *et al.*, 1997) incorporated into crystals. Increasing the pH to 8.0 gave no improvement of the crystals. After mixing both buffer compounds in the droplet, formation of a precipitate could be observed, which can be explained by the extremely high supersaturation of about 100 (3.75 mg ml^{-1}/36 µg ml^{-1}) within the droplet. The amount of precipitate decreased during crystal growth. The crystals are tetragonal, with unit-cell parameters $a = b = 151.1$, $c = 75.8$ Å. The Laue class was determined as 4/*mmm*. The poor quality and low resolution of the collected data prevented the exact determination of systematic absences and identification of screw axes. Assuming the presence of eight molecules per asymmetric unit, the packing density is 2.3 Å^3 Da^{-1}, corresponding to approximately 45.7% solvent content in the crystal. This value is within the range observed for protein crystals (Matthews, 1968). Data-collection parameters are given in Table 1.

Table 1
Data-collection statistics.

Values for the outermost resolution shell (3.63–3.50 Å) are shown in parentheses.

Laue class	4/*mmm*
Space group	$P4_x2_y2$
Unit-cell parameters (Å)	$a = b = 151.1$, $c = 75.8$
Resolution range (Å)	30.0–3.50
No. of observations	74 063
No. of unique reflections	11482
$R_{\rm merge}$† (%)	8.8 (44.5)
$R_{\rm r.i.m.}$‡ (%)	9.7 (49.4)
$R_{\rm p.i.m.}$§ (%)	3.8 (19.0)
$I > 3\sigma(I)$ (%)	84.1 (60.4)
Completeness of data [cutoff $-3\sigma(I)$]	99.9 (100.0)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - I(hkl)|/\sum_{hkl}\sum_i I_i(hkl)$, where the index *hkl* sums over all reflections and *i* sums over all equivalent and symmetry-related reflections (Stout & Jensen, 1968). ‡ $R_{\rm r.i.m.}$ is the the redundancy-independent merging *R* factor (Weiss & Hilgenfeld, 1997), which is identical to the $R_{\rm meas}$ of Diederichs & Karplus (1997). $R_{\rm r.i.m.} = \sum_{hkl}[N/(N-1)]^{1/2}\sum_i |I_i(hkl) - I(hkl)|/\sum_{hkl}\sum_i I_i(hkl)$, with *N* being the number of times a given reflection has been observed. § $R_{\rm p.i.m.}$ is the precision-indicating merging *R* factor (Weiss & Hilgenfeld, 1997). $R_{\rm p.i.m.} = \sum_{hkl}[1/(N-1)]^{1/2}\sum_i |I_i(hkl) - I(hkl)|/\sum_{hkl}\sum_i I_i(hkl)$.

4. Discussion

For the first time, a poorly soluble propeptide of a lysosomal cysteine proteinase has been crystallized. Only one other example of a crystallized propeptide had been obtained previously, the so-called proregion of a bacterial serine proteinase called α-lytic protease (Sauter *et al.*, 1998). The strategy used for the crystallization of the cathepsin S propeptide consists of the combination of a conventional method such as vapour diffusion in a hanging drop with the use of unfolded cathepsin S propeptide dissolved at a suitable concentration in an acidic buffer. The successful crystallization prompted the assumption of protein folding during mixture with a crystallization solution, which allows as a first step a pH change from acidic to nearly neutral conditions within the droplet. This leads to the formation of precipitate (aggregation of partially or completely folded molecules) and a number of solubilized folded molecules within the droplet, which is limited because of their poor solubility. In a second step, the growth of crystals occurs after equilibration of the droplet with the reservoir buffer, in which 2-propanol, which has a higher vapour pressure than water, transfers from the reservoir into the droplet, acting as precipitant. The growth of crystals depends on the presence of a pool of folded and dissolved molecules in the droplet. This pool seems to

be replenished by the precipitate, which decreases in mass over time. The crystals incorporate folded molecules; the molecules removed from the solution are substituted by molecules from the precipitate. The observation of a higher crystal-growth rate on raising the temperature from 293 to 303 K may be interpreted in terms of an increased solubility of folded molecules. The growth of crystals at the expense of precipitate, which appeared to be crystalline, prompted speculation concerning 'Ostwald ripening' (Ostwald, 1897). This phenomenon cannot be excluded because the requirement of small crystals grown within the precipitate seems to be given for the growth of crystals at the expense of small crystals, as described by Giegé *et al.* (1996) for other macromolecules. The successful crystallization of this small protein is a strong indication of the formation of an ordered structure, which had previously been demonstrated by spectroscopic methods at nearly neutral pH (Maubach *et al.*, 1997). The method might be useful for the crystallization of other proteins and peptides with similar solubility properties. Unfortunately, the low resolution of the diffraction data and the high number of monomers in the asymmetric unit have not yet allowed successful molecular replacement calculations using the propeptide from the entire proenzyme procathepsin S, which we have recently determined (unpublished results). In order to determine the cathepsin S propeptide structure, optimization of the crystals is planned.

We would like to thank G. Maubach for providing the plasmid encoding the cathepsin S propeptide. This work was partly supported by the Deutsche Forschungsgemeinschaft (Schi 593/1-1). Rolf Hilgenfeld thanks the Fonds der Chemischen Industrie.

References

Collaborative Computional Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
García-Ruiz, J. M. & Moreno, A. (1994). *Acta Cryst.* D**50**, 484–490.
Giegé, R, Ng, J. D., Lorber, B., Théobald-Dietrich, A., Kern, D. & Witz, J. (1996). *J. Cryst. Growth*, **168**, 50–62.
Karrer, K. M., Pfeiffer, S. L. & DiThomas, M. E. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 3063–3067.
Kreusch, S., Fehn, M., Maubach, G., Nissler, K., Rommerskirch, W., Schilling, K., Weber, E., Wenz, I. & Wiederanders, B. (2000). *Eur. J. Biochem.* **267**, 2965–2972.
Kirschke, H., Schmidt, I. & Wiederanders, B. (1986). *Biochem. J.* **240**, 455–459.
Kirschke, H., Wiederanders, B., Bromme, D. & Rinne, A. (1989). *Biochem. J.* **264**, 467–473.
McPherson, A. (1982). *The Preparation and Analysis of Protein Crystals.* New York: Wiley.
McPherson, A. (1999). *Crystallization of Biological Macromolecules.* New York: Cold Spring Harbor Laboratory Press.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Maubach, G., Schilling, K., Rommerskirch, W., Wenz, I., Schultz, J. E., Weber, E. & Wiederanders, B. (1997). *Eur. J. Biochem.* **250**, 745–750.
Ostwald, W. (1897). *Z. Phys. Chem.* **22**, 289–330.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pietschmann, S., Fehn, M., Kaulmann, G., Wenz, I., Wiederanders, B. & Schilling, K. (2002). *Biol. Chem.* **383**, 1453–1458.
Sauter, N. K., Mau, T., Rader, S. D. & Agard, D. A. (1998). *Nature Struct. Biol.* **5**, 945–950.
Schilling, K., Pietschmann, S., Fehn, M., Wenz, I. & Wiederanders, B. (2001). *Biol. Chem.* **382**, 859–865.
Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination. A Practical Guide*, p. 402. London: Macmillan.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203–205.
Wiederanders, B., Brömme, D., Kirschke, H., von Figura, K., Schmidt, B. & Peters, C. (1992). *J. Biol. Chem.* **267**, 13708–13713.