Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, crystallization and preliminary X-ray analysis of mexicain

M. C. Oliver-Salvador,[a] L. A. González-Ramírez,[b] J. A. Gavira,[b] M. Soriano-García[c] and J. M. García-Ruiz[b]*

[a]Unidad Profesional Interdisciplinaria de Biotecnología, IPN, México DF, México, [b]Laboratorio de Estudios Cristalográficos, CSIC-Universidad de Granada, Granada, Spain, and [c]Instituto de Química, UNAM, México DF, México

Correspondence e-mail: jmgruiz@ugr.es

Received 28 May 2004
Accepted 2 September 2004

Mexicain is a 23.7 kDa papain-like cysteine protease from the tropical plant *Jacaratia mexicana*. Extracted as a mix of proteases from the latex of the fruit, mexicain is isolated after cation-exchange chromatography as the most abundant product. The purified product inhibited with E-64 was crystallized by sitting-drop vapour diffusion in the presence of ethanolamine. Cryoprotected crystals diffracted X-rays from a home source to 1.98 Å and belong to the monoclinic space group $P2_1$, with unit-cell parameters a = 57.36, b = 90.45, c = 80.39 Å, β = 92.64°. The asymmetric unit contains four molecules of mexicain, with a corresponding crystal volume per protein weight (V_M) of 2.24 Å^3 Da^{-1} and a solvent content of 45% by volume. A molecular-replacement model has been determined and refinement is in progress.

1. Introduction

Proteases are classified into five different types according to the group that plays an essential role in their enzymatic activity: serine, threonine, cysteine or an aspartic or metallic group. The cysteine protease family shares a similar enzymatic mechanism with the serine and threonine protease families in which the amino acid acts as the nucleophile in the reaction, while in the other two groups this role is played by an activated water molecule. Cysteine proteases are widely distributed from viruses to mammals. In plants, cysteine proteases are by far the most abundant (Priolo *et al.*, 2000). They are classified into more than 30 families grouped into five clans (CA to CE) according to the amino acids located in the active centre (Barrett & Rawlings, 2001). The best known group of these proteases is the papain family (clan CA, family C1), which includes endoproteases with a broad specificity, such as papain, and a wide range of activities. In plants, the activity of these enzymes is related to germination, the activation of proenzymes, the degradation of defective proteins, nutrition reserves and also defence. In humans, the limited tissue expression of these enzymes implies specific roles in cellular physiology and has been linked to pathological processes such as apoptosis, MHC class II immune response, rheumatoid arthritis and emphysema, amongst others (Chapman *et al.*, 1997; Turk *et al.*, 2000). In addition, proteases have played an important role in medicine as well as in industry for hundreds of years. Because of their high stability at various pH values and temperatures, their high availability, purity and activity, proteases are the most used enzymes, accounting for 60% of the world enzyme market (Haard, 1998).

Mexicain is a cysteine protease from the latex of the tropical plant *Pileus mexicanus* (now also termed *Jacaratia mexicana*) first described by Castañeda-Agulló *et al.* (1942). Mexicain belongs to the papain family and shows a high sequence identity (73.8%) to cysteine protease (CC-III) from *Carica candamarcencis* (mountain papaya) and to chymopapain (69.42%) from *C. papaya*. The four residues that form the active centre, Cys25, Gln19, His157 and Asn178, correspond to those of the papain family. Isolated from the latex of the fruit, mexicain shows remarkable enzymatic activity and stability, higher than those of papain, the enzyme of the cysteine peptidase family that has been most used in industrial applications. The pure white powder extract was named mexicain in 1945 and subjected to crystallization trials, but no results were reported from the earliest crystals obtained by Castañeda-Agulló and coworkers (Castañeda-Agulló *et al.*, 1945). Further purification steps have shown that the original product named mexicain is in fact composed of at least five proteases, named P-I to P-V, of similar molecular weight and isoelectric point. The most abundant fraction, protease IV, was then named mexicain. The purified protease P-IV mexicain covalently bound to the protease inhibitor E-64 has been crystallized, but unfortunately the quality of the crystal

obtained was not sufficient for accurate X-ray analysis (Oliver-Salvador *et al.*, 2000). Here, we report the crystallization and the first crystallographic data of mexicain.

2. Materials and methods

2.1. Purification

Mexicain was isolated from latex of the fruit of the tropical plant *J. mexicana*. The latex was mixed with a solution containing 0.1 *M* NaCl and 0.25 *M* sucrose and homogenized at 2000 rev min^{-1} for 30 min in an ice bath. The homogenate was centrifuged at 10 000*g* for 30 min and the supernatant was separated and centrifuged at 100 000*g* for 60 min. This solution was precipitated with an 18.7%(*w*/*v*) NaCl solution and was allowed to settle for 10 h at 277 K. The precipitate was separated by centrifugation at 10 000*g* for 30 min and dissolved in 50 m*M* phosphate buffer pH 6.3. This procedure was repeated twice to produce the crude extract. The extract, diluted 1:1(*v*:*v*) with 50 m*M* phosphate buffer pH 6.3, was loaded onto a 5 ml strong cation-exchange column (Econo-Pac High S, Bio-Rad) installed in a GradicFrac chromatography system (Pharmacia) pre-equilibrated with the same buffer. Elution was performed at a flow rate of 1 ml min^{-1} with a linear gradient of buffered 0–1.0 *M* NaCl. Five peaks with proteolytic activity were obtained. The mexicain fractions belonging to peak IV were identified, pooled and loaded onto a Protein-Pack SP 8R cation-exchange column (Waters) pre-equilibrated with 50 m*M* phosphate buffer pH 6.3. Elution was performed at a flow rate of 60 ml h^{-1} with a linear gradient of 0–1 *M* NaCl in 50 m*M* phosphate buffer pH 6.3. Dialysis and concentration of the protein during the whole purification process was performed with an Amicon ultrafiltration system with Amicon YM3 membranes (3 kDa molecular weight cutoff). Polyacrylamide gel electrophoresis of the purified protein in the presence of sodium dodecyl sulfate (SDS–PAGE) was performed with a Mini-Protean II System (Bio-Rad) and a Phast System (Pharmacia) (Laemmli, 1970). The gel (12% polyacrylamide) was stained with Coomassie Brilliant Blue R-250. The samples showed a single band with an apparent molecular weight of 24 kDa, which matches the value (23.7 kDa) obtained by mass spectrometry (Oliver-Salvador *et al.*, 2000). Fig. 1 shows an SDS–PAGE of the mexicain.

The final protein concentration was determined by the bicinconic acid method (Smith *et al.*, 1985) using bovine serum albumin as a standard. The protein concentration in chromatographic fractions was estimated by measuring the absorbance of eluates at 280 nm using 1.78 g^{-1} l cm^{-1} as the extinction coefficient (Oliver-Salvador *et al.*, 2000).

2.2. Crystallization

For crystallization trials, the enzyme was inhibited with E-64 [*n*-(3-carboxyoxirane-2-carbonyl)-leucyl-amino(4-guanido)butane] to avoid self-digestion. The purified enzyme was activated with the reductant dithiothreitol (DTT). The specific protease inhibitor E-64 was added to the preparation and stirred at 277 K for 10 h. The sample was then dialysed against 20%(*v*/*v*) methanol–water followed by dialysis against distilled water, concentrated to 15 mg ml^{-1} and then tested for proteolytic activity with casein.

Initial crystal screening was performed using the sparse-matrix system (Jancarik & Kim, 1991) using Hampton Research Crystal Screens I and II and PEG 6000 as well as home-made screening using detergents, buffers and modifiers of the dielectric constant. The experiments were set up with the sitting-drop vapour-diffusion method and kept at 293 K for equilibration. The drops were made by mixing 2 µl 7.9 mg ml^{-1} enzyme solution with an equal volume of the reservoir solutions and were suspended over 1.0 ml reservoir.

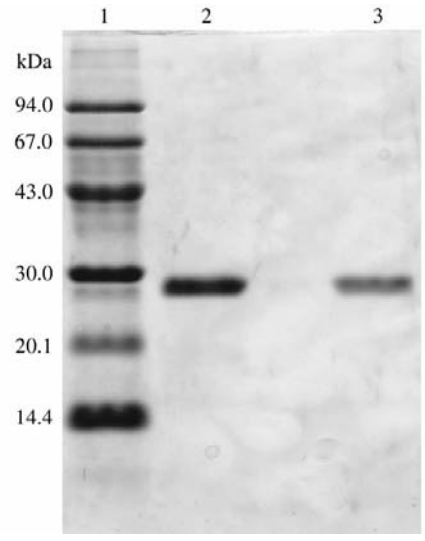

Figure 1
12% SDS–PAGE. Lane 1, standard molecular-weight markers: phosphorylase *b* (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20.1 kDa) and α-lactoalbumin (14.4 kDa). Lanes 2 and 3 show purified mexicain–E-64 (lane 3 is the same as lane 2 but diluted 1:10).

2.3. Data collection

For data collection, selected crystals were transferred to a cryoprotectant solution containing 20%(*v*/*v*) glycerol in the crystallization solution and flash-frozen with a soaking time of less than 60 s. The cryoprotectant 20%(*v*/*v*) glycerol was found to be a necessary ingredient for cryocooling in order to reduce mosaicity. Flash-frozen crystals were placed in a 100 K nitrogen-gas stream produced by an Oxford Cryosystems liquid-nitrogen device for data collection. X-ray diffraction data were recorded on a Kappa CCD2000 135 mm CCD detector using Cu $K\alpha$ radiation from a Bruker–Nonius rotating-anode generator operating at 45 kV and 100 mA and focused with Montel multilayer mirrors. A crystal-to-detector distance of 70 mm and 3 min exposure per frame were used to collect a total of 515 frames at four different values of the κ angle. Data were indexed, integrated and scaled with the *HKL*2000 suite (Otwinowski & Minor, 1997) and further processing was performed with *CNS* (Brünger *et al.*, 1998) and *XTALVIEW* (McRee, 1999).

3. Results and discussion

Initial screening of crystallization conditions was performed with Hampton Research Crystal Screens I and II and PEG 6000 and home-made screening using NaCl, $(NH_4)_2SO_4$ and methanol at various pH values. Only the PEG 6000 screen at pH 4.0 buffered with 100 m*M* sodium citrate produced crystalline material and spherulites. A new screen was designed to fine-tune the variables in order to improve the crystal quality and size. Well faceted crystals suitable for diffraction experiments with maximum dimensions of 0.5 × 0.07 × 0.07 mm (see Fig. 2) appeared within one month under the following conditions. The reservoir contained 20%(*w*/*v*) PEG 6000 in 0.1 *M* citrate buffer pH 4.0 plus 15%(*v*/*v*) ethanolamine. The drop was made by mixing

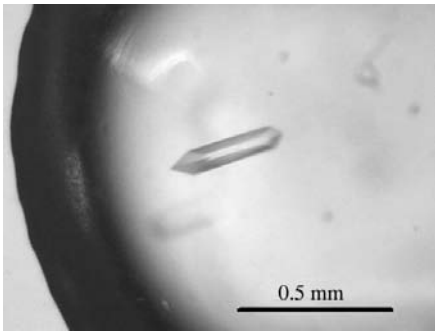

Figure 2
Crystal of mexicain–E-64 complex.

2 µl mexicain solution with 2 µl reservoir solution without ethanolamine.

The minimum distortion index shows that the crystal belongs to the monoclinic system (space group $P2/P2_1$), with unit-cell parameters a = 57.36, b = 90.45, c = 80.39 Å, β = 92.64°. The data were initially scaled in space group $P2$ using *SCALEPACK* (Otwinowski & Minor, 1997), but no significant intensity-to-noise ratio was found for reflections with index (0, k, 0) where k has odd values. The overall $R_{\rm merge}$ of the data set was 9.5 and $R_{\rm merge}$ was 29.6% for the 2.18–2.10 Å resolution shell. There are four molecules in the asymmetric unit which do not exhibit any local point-group symmetry. Data statistics are given in Table 1. Using the molecular weight of 23.7 kDa and assuming the presence of four molecules per asymmetric unit, the Matthews coefficient ($V_{\rm M}$) was calculated to be 2.24 Å^3 Da^{-1} (Matthews, 1968), corresponding to a solvent content of 45.0% (Westbrook, 1985).

The initial model was calculated using the SWISS-MODEL server using *ProMod*II for modelling and *GROMOS*96 for energy minimization (Schwede *et al.*, 2000). Five PDB models with the highest percentage of identity were input, namely 1yal (69.42%; Maes *et al.*, 1996), 1gec (67.25%; O'Hara *et al.*, 1995), 1ppo (64.9%; Pickersgill *et al.*, 1991), 1meg (64.6%; Katerelos *et al.*, 1996) and 1pci (64.6%; Groves *et al.*, 1996), together with the mexicain amino-acid sequence. The fast direct-rotation algorithm from *CNS* was run using the 25–4.0 Å data against the generated model. The top ten peaks were used to search for the translation solution. The final location of the tetramer was determined by visual inspection of the ten top rotation–translation solutions. Considering each monomer as an independent rigid body, the refinement gave an R value of 0.419 and $R_{\rm free}$ = 0.416 using 10% of the data. After manual rebuilding of the model and cycles of refinement and minimization (rigid body, simulated annealing and grouped B value) applying non-crystallographic symmetry, the values of R and $R_{\rm free}$ obtained were 0.28 and 0.29, respectively. Refinement is in progress, as well improvement of crystal quality by use of the counter-diffusion crystallization technique (García-Ruiz, 2004).

Table 1
Summary of X-ray data statistics.

Statistical values for the highest resolution shell (2.18–2.10 Å) are given in parentheses.

Wavelength (Å)	1.54
Space group	$P2_1$
Unit-cell parameters (Å, °)	a = 57.36, b = 90.45, c = 80.39, β = 92.64
Resolution range (Å)	50.00–2.10
No. observations	279184
No. unique reflections	47249
Data completeness (%)	99.3 (97.1)
$R_{\rm merge}$† (%)	9.5 (29.6)
$I/\sigma(I)$	23.9 (1.5)
Redundancy	5.9 (3.1)
Matthews coefficient (Å^3 Da^{-1})	2.24
Solvent content (%)	45.0

† $R_{\rm merge} = \sum_{hkl}\sum_i |\langle I\rangle - I_i|/\sum_{hkl}\sum_i I_i$.

We acknowledge Dr Armando Albert for his support during data collection and analysis at the Instituto Rocasolano (CSIC) in Madrid. This work was supported by Spanish Ministry of Science and Technology Project ESP2003-04759 and by Mexican National Polytechnic Institute (IPN) Projects CGPI 20021341 and 20030510. JAG thanks the help of the Spanish Consejo Superior de Investigaciones Científicas. LAG also thanks CONACyT of Mexico for the fellowship.

References

Barrett, A. J. & Rawlings, N. D. (2001). *Biol. Chem.* **382**, 727–733.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Castañeda-Agulló, M., Gavarrón, F. F. & Balcazar, M. R. (1942). *Science*, **96**, 365–366.
Castañeda-Agulló, M., Hernández, A., Loaeza, F. & Salazar, V. (1945). *J. Biol. Chem.* **159**, 751.
Chapman, H. A., Riese, R. J. & Shi, G.-P. (1997). *Annu. Rev. Physiol.* **59**, 63–88.
García-Ruiz, J. M. (2004). *Methods Enzymol.* **368**, 130–154.
Groves, M. R., Taylor, M. A., Scott, M., Cummings, N. J., Pickersgill, R. W. & Jenkins, J. A. (1996). *Structure*, **4**, 1193–1203.
Haard, N. F. (1998). *Food Biotechnol.* **52**, 64–67.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Katerelos, N. A., Taylor, M. A., Scott, M., Goodenough, P. W. & Pickersgill, R. W. (1996). *FEBS Lett.* **392**, 35–39.
Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
McRee, D. E. (1999). *Practical Protein Crystallography*, 2nd ed. San Diego, CA, USA: Academic Press.
Maes, D., Bouckaert, J., Poortmans, F., Wyns, L. & Looze, Y. (1996). *Biochemistry*, **35**, 16292–16298.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
O'Hara, B. P., Hemmings, A. M., Buttle, D. J. & Pearl, L. H. (1995). *Biochemistry*, **34**, 13190–13195.
Oliver-Salvador, M. C., Moreno, A. & Soriano-García, M. (2000). *Revista de la AMC*, **51**, 12–20.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pickersgill, R. W., Rizkallah, P., Harris, G. W. & Goodenough, P. W. (1991). *Acta Cryst.* B**47**, 766–771.
Priolo, N., Morcelle del Valle, S., Arribére, M. C., López, L. & Caffini, N. (2000). *J. Protein Chem.* **19**, 39–49.
Schwede, T., Diemand, A., Guex, N. & Peitsch, M. C. (2000). *Res. Microbiol.* **151**, 107–112.
Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. & Klenk, D. C. (1985). *Anal. Biochem.* **150**, 76–85.
Turk, B., Turk, D. & Turk, V. (2000). *Biochim. Biophys. Acta*, **1477**, 98–111.
Westbrook, E. M. (1985). *Methods Enzymol.* **114**, 187–196.