Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization, data collection and phasing of infestin 4, a factor XIIa inhibitor

I. T. N. Campos,[a] B. G. Guimarães,[b] F. J. Medrano,[b] A. S. Tanaka[a] and J. A. R. G. Barbosa[b]*

[a]Departamento de Bioquímica, Escola Paulista de Medicina, UNIFESP-EPM, São Paulo, SP, Brazil, and [b]Centro de Biologia Molecular Estrutural (CeBiMe), Laboratório Nacional de Luz Síncrotron (LNLS), CP 6192, Campinas, SP 13084-971, Brazil

Correspondence e-mail: joao@lnls.br, tanaka.bioq@epm.br

Received 21 June 2004
Accepted 1 September 2004

Infestin is a protein from *Triatoma infestans* (kissing bug) composed of seven Kazal-type domains that is further processed to yield several serine protease inhibitors with varying specificities. Infestins 3 and 4 are the last two domains of the infestin gene and are found *in vivo* in the insect's anterior midgut. The last domain, infestin 4, has been cloned, expressed and purified. Here, the crystallization of infestin 4 using the sitting-drop vapour-diffusion method with PEG 8000 as precipitant is described. Crystals belong to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters a = 25.89, b = 45.64, c = 57.41 Å. X-ray diffraction data were collected to a maximum resolution of 1.8 Å using a synchrotron-radiation source. Initial phases were calculated by molecular replacement using an edited rhodniin molecule as the search model. Structure refinement is in progress.

1. Introduction

Hematophagous animals have a vital need to interfere in their hosts' hemostatic system. The first report of hematophagous animal substances able to block or delay vertebrate blood coagulation dates from the 19th century (Moser *et al.*, 1998). Since then, many other substances from these animals have been described, with anti-clotting, anti-platelet aggregation, vasodilatory and fibrinolytic activities (Arocha-Pinango *et al.*, 1999; Crab *et al.*, 2002; Iwanaga *et al.*, 2003; Francischetti *et al.*, 2002; Lee & Vlasuk, 2003). Among them are the serine protease inhibitors, which are able to block the blood-coagulation cascade mainly by inhibiting factor Xa and thrombin.

Serine protease inhibitors from animals have been classified by their primary structure and recently by their tertiary structures into the following families: Kunitz, Kazal, Serpin and Leuko proteinase types (Roberts *et al.*, 1995). Recently, new families such as the Ascaris, *Bombyx* and Pacifastin types have been described (Rhoads *et al.*, 2000; Pham *et al.*, 1996; Simonet *et al.*, 2002).

With very few exceptions, the Kazal family members have several common structural features: a 'canonical' binding loop (Bode & Huber, 1992), a conserved cysteine-distribution pattern, a typical VCG*x*D sequence (Schlott *et al.*, 2002), two to three α-helices connected by turns and a short three-stranded antiparallel β-sheet (Muhlhahn *et al.*, 1994). Some Kazal-type inhibitors from invertebrates, such as leech-derived tryptase inhibitor (LDTI), rhodniin, bdellin B-3 and infestin (Sommerhoff *et al.*, 1994; Friedrich *et al.*, 1993; Fink *et al.*, 1986; Campos *et al.*, 2002), are classified as non-classical Kazal-type inhibitors (Fink *et al.*, 1986) and are characterized by the presence of only one amino-acid residue between the first and second cysteine, compared with the seven amino-acid residues present in the classical Kazal-type inhibitors. Classical and non-classical Kazal-type inhibitors can be composed of one or more domains linked in tandem by loops of variable size.

The infestin gene from the kissing bug *Triatoma infestans*, a Chagas' disease vector, has 1227 base pairs that code for a signal peptide of 17 amino-acid residues and seven Kazal-type domains, constituting 392 amino-acid residues in total, with an average of 56 amino-acid residues per domain. These domains were named 1R, 2R, 3R, 1, 2, 3 and 4 from the N-terminus to the C-terminus of the protein. The protein is found in the insect's anterior midgut and is processed by proteases to release at least three different mature polypeptides *in vivo*: infestin 1R, infestin 1-2 (domains 1 and 2) and infestin 3-4. Infestins 2R and 3R have not yet been found *in vivo*. These inhibitory domains are important to prevent blood-clot formation in the bug's digestive tract. Recently, we showed infestin 3-4 and infestin 4 (residues 354–409) by itself to be the first Kazal family member able to inhibit factor XIIa (Fink *et al.*, 1986; Campos *et al.*, 2004). The three-dimensional structure of factor XIIa is unknown and there are only two inhibitors for this protease listed in the PDB: corn Hageman factor inhibitor (PDB code 1bea; Behnke *et al.*, 1998) and ecotin (PDB code 1ecy; Shin *et al.*, 1996), neither of which are from the Kazal family. The structure of infestin

4 will thus provide information to help elucidate the active site for factor XIIa interaction. Infestin 4 shares 50 and 48% identity to dipetalogastin domain 3 (Mende *et al.*, 1999) and to rhodniin domain 1 (Friedrich *et al.*, 1993), respectively. Both proteins are part of thrombin inhibitors from hematophagous bugs. In this work, we present the crystallization, data collection and preliminary characterization of infestin 4 and its molecular-replacement solution.

2. Methods

Pichia pastoris GS115 harbouring pIC 11.1.1 plasmid containing the infestin 4 gene (PPIC9/inf4; Campos *et al.*, 2004) was grown in 500 ml BMGY media until OD_{600} = 2.0. Cells were harvested, resuspended in 2.5 l BMMY media to an OD_{600} of 1.0 and incubated at 303 K for 5 d following the instructions for the 'Original *Pichia* Expression Kit' from Invitrogen. Methanol was added every 24 h to a final concentration of 0.5%(*v*/*v*) to maintain induction. After fermentation, the culture was centrifuged (4000*g*, 20 min, 277 K), cells were discarded and the supernatant was collected to purify infestin 4. Infestin 4 was purified by affinity chromatography on a trypsin-Sepharose column as described previously (Campos *et al.*, 2002). Briefly, the media containing infestin 4 was applied onto a trypsin-Sepharose column previously equilibrated with 0.1 *M* Tris–HCl buffer pH 8.0. Proteins that bound weakly to the column were washed out with 0.1 *M* Tris–HCl buffer pH 8.0 containing 0.15 *M* NaCl. Finally, the inhibitor was eluted with 0.5 *M* KCl pH 2.0 and immediately neutralized with 1 *M* Tris–HCl buffer pH 8.0. The protein was lyophilized, dialyzed and concentrated using YM-3 Centricon (Millipore) to 10 mg ml^{-1} in 10 m*M* sodium acetate buffer pH 6.0.

Figure 1
Crystal of infestin 4 from *T. infestans* measuring approximately 0.56 mm in the longest dimension.

Crystallization was carried out by vapour diffusion in sitting-drop Cryschem Plates using the Crystal Screen kit (Hampton Research), mixing equal volumes (2 µl) of protein solution and reservoir solution. Crystals from condition 36 [0.1 *M* Tris–HCl pH 8.5, 8%(*w*/*v*) polyethylene glycol 8000] were used for data collection.

Crystallographic data were collected at protein crystallography beamline D03B-CPR at the Laboratório Nacional de Luz Síncrotron (LNLS), Campinas, Brazil. The wavelength was set to 1.427 Å and a MAR CCD detector was used to record the oscillation data with $\Delta\varphi = 1.0^\circ$. Crystals were cryocooled in a stream of nitrogen gas at 110 K in order to minimize radiation damage. The solution in which the crystals were grown provided partial protection against ice formation. The data set was processed using *HKL*2000 (Otwinowski & Minor, 1997) and the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). Molecular replacement was performed with the *MOLREP* program (Vagin & Teplyakov, 1997) using one edited rhodniin molecule obtained from the structure 1tbr (van de Locht *et al.*, 1995) deposited in the PDB. Chain R of this entry, a rhodniin molecule consisting of 103 amino-acid residues forming two domains, was edited based on a sequence alignment to provide an adequate search model. Amino-acid residues 1–5, 34–39 and 49–103 (1tbr numbering) were removed from chain R to yield a search model with one domain lacking the poorly defined N-terminus (1–5), an internal loop (34–39) and the C-terminal domain (49–103). The data used in the molecular replacement were between 30.0 and 3.0 Å resolution.

3. Results and discussion

Cultures of transformed cells with the infestin 4 gene yielded 4 mg of purified protein per litre. Single-step purification by affinity chromatography was enough to produce crystallization-quality protein. Initial crystallization screening of infestin 4 at 10 mg ml^{-1} resulted in crystals after one week from several conditions (Nos. 7, 9, 15, 17, 22, 28 and 36) of the Crystal Screen kit from Hampton Research. All except condition No. 7 contained polyethylene glycol as the precipitant and the crystals grown in condition No. 36 were suitable for diffraction. The best diffraction pattern obtained had a maximum resolution of 1.8 Å (condition No. 36; Fig. 1) and showed the symmetry and systematic absences of the orthorhombic space group $P2_12_12_1$. Weak ice rings were present but did not compromise the data-set quality. Table 1 summarizes the data-collection statistics. The Matthews coefficient (Matthews, 1968) was calculated to be 2.7 Å^{3}Da^{-1} and the solvent content 54.7%, corresponding to one molecule of infestin 4 per asymmetric unit.

Table 1
Data-collection statistics.

Values in parentheses are for the last resolution shell (1.86–1.80 Å).

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 25.89, b = 45.64, c = 57.41
Mosaicity (°)	0.9
Temperature (K)	110
Wavelength (Å)	1.427
Oscillation (°)	1
Resolution limits (Å)	40.00–1.80 (1.86–1.80)
$I/\sigma(I)$ after merging	23.0 (2.7)
Completeness (%)	99.0 (91.5)
Multiplicity	4.6 (3.1)
R_{sym}	5.9 (30.9)
No. reflections	30953
No. unique reflections	6699 (606)

† $R_{sym} = \sum\sum|I_i(h) - \langle I(h)\rangle| / \sum\sum I_i(h)$, where $I_i(h)$ is the observed intensity of the ith measurement of reflection h and $\langle I(h)\rangle$ is the mean intensity of reflection h calculated after loading and scaling.

The search model was constructed from the rhodniin crystallographic structure (PDB code 1tbr) as described in §2. The *MOLREP* algorithm (Vagin & Teplyakov, 1997) was able to find a solution for which the rotation function was 5.5σ and the translation function was 13.9σ with a correlation coefficient of 0.334. Using *CNS* (Brünger *et al.*, 1998), the model was subjected to a protocol consisting of simulated annealing, positional energy minimization and individual temperature-factor refinement, leading to an R factor of 44.3% and an R_{free} of 42.0%. The first electron-density map was calculated using *REFMAC* (Murshudov *et al.*, 1997) and visualized with the graphical modelling package *O* (Jones *et al.*, 1991). The map was good enough to allow the recognition of side chains that differed from the rhodniin search model. At present, the model is being constructed and refined.

This work was supported by FAPESP grants 01/07535-0 (SMolBNet) and 02/13960-8 (regular) and ABTLuS. ITNC was supported by FAPESP fellowship 01/08325-9.

References

Arocha-Pinango, C. L., Marchi, R., Carvajal, Z. & Guerrero, B. (1999). *Blood Coagul. Fibrinolysis*, **10**, 43–68.
Behnke, C. A., Yee, V. C., Trong, I. L., Pedersen, L. C., Stenkamp, R. E., Kim, S. S., Reeck, G. R.

& Teller, D. C. (1998). *Biochemistry*, **37**, 15277–15288.
Bode, W. & Huber, R. (1992). *Eur. J. Biochem.* **204**, 433–451.
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.
Campos, I. T., Amino, R., Sampaio, C. A., Auerswald, E. A., Friedrich, T., Lemaire, H. G., Schenkman, S. & Tanaka, A. S. (2002). *Insect Biochem. Mol. Biol.* **32**, 991–997.
Campos, I. T. N., Tanaka-Azevedo, A. M. & Tanaka, A. S. (2004). Submitted.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Crab, A., Noppe, W., Pelicaen, C., Van Hoorelbeke, K. & Deckmyn, H. (2002). *Thromb. Haemost.* **87**, 899–904.
Fink, E., Rehm, H., Gippner, C., Bode, W., Eulitz, M., Machleidt, W. & Fritz, H. (1986). *Biol. Chem. Hoppe-Seyler*, **367**, 1235–1242.
Francischetti, I. M., Valenzuela, J. G., Andersen, J. F., Mather, T. N. & Ribeiro, J. M. (2002). *Blood*, **99**, 3602–3612.
Friedrich, T., Kroger, B., Bialojan, S., Lemaire, H. G., Hoffken, H. W., Reuschenbach, P., Otte, M. & Dodt, J. (1993). *J. Biol. Chem.* **268**, 16216–16222.
Iwanaga, S., Okada, M., Isawa, H., Morita, A., Yuda, M. & Chinzei, Y. (2003). *Eur. J. Biochem.* **270**, 1926–1934.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard (1991). *Acta Cryst.* A**47**, 110–119.
Lee, A. Y. & Vlasuk, G. P. (2003). *J. Intern. Med.* **254**, 313–321.
Locht, A. van de, Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B., Hoffken, W. & Bode, W. (1995). *EMBO J.* **14**, 5149–5157.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mende, K., Petoukhova, O., Koulitchkova, V., Schaub, G. A., Lange, U., Kaufmann, R. & Nowak, G. (1999). *Eur. J. Biochem.* **266**, 583–590.
Moser, M., Auerswald, E., Mentele, R., Eckerskorn, C., Fritz, H. & Fink, E. (1998). *Eur. J. Biochem.* **253**, 212–220.
Muhlhahn, P., Czisch, M., Morenweiser, R., Habermann, B., Engh, R. A., Sommerhoff, C. P., Auerswald, E. A. & Holak, T. A. (1994). *FEBS Lett.* **355**, 290–296.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pham, T. N., Hayashi, K., Takano, R., Itoh, M., Eguchi, M., Shibata, H., Tanaka, T. & Hara, S. (1996). *J. Biochem. (Tokyo)*, **119**, 428–434.
Rhoads, M. L., Fetterer, R. H., Hill, D. E. & Urban, J. F. Jr (2000). *Exp. Parasitol.* **95**, 36–44.
Roberts, R. M., Mathialagan, N., Duffy, J. Y. & Smith, G. W. (1995). *Crit. Rev. Eukaryot. Gene Expr.* **5**, 385–436.
Schlott, B., Wohnert, J., Icke, C., Hartmann, M., Ramachandran, R., Guhrs, K. H., Glusa, E., Flemming, J., Gorlach, M., Grosse, F. & Ohlenschlager, O. (2002). *J. Mol. Biol.* **318**, 533–546.
Shin, D. H., Song, H. K., Seong, I. S., Lee, C. S., Chung, C. H. & Suh, S. W. (1996). *Protein Sci.* **5**, 2236–2247.
Simonet, G., Claeys, I. & Broeck, J. V. (2002). *Comput. Biochem. Physiol. B Biochem. Mol. Biol.* **132**, 247–255.
Sommerhoff, C. P., Sollner, C., Mentele, R., Piechottka, G. P., Auerswald, E. A. & Fritz, H. (1994). *Biol. Chem. Hoppe-Seyler*, **375**, 685–694.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.