Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Crystallization and preliminary X-ray diffraction analysis of Q4DV70 from *Trypanosoma cruzi*, a hypothetical protein with a putative thioredoxin domain

Camila Ramos dos Santos,[a,b] Melissa Regina Fessel,[a] Leandro de Carvalho Vieira,[a] Marco Aurélio Krieger,[c] Samuel Goldenberg,[c] Beatriz Gomes Guimarães,[d] Nilson Ivo Tonin Zanchin[a] and João Alexandre Ribeiro Gonçalves Barbosa[a,b]*

[a]Center for Structural Molecular Biology (CeBiME), Brazilian Synchrotron Light Laboratory (LNLS), CP 6192, 13084-971 Campinas-SP, Brazil, [b]Instituto de Biologia, Universidade Estadual de Campinas (Unicamp), Campinas-SP, Brazil, [c]Instituto Carlos Chagas, Fiocruz, Curitiba-PR, Brazil, and [d]Synchrotron SOLEIL, Saint-Aubin, France

Correspondence e-mail: joao@lnls.br

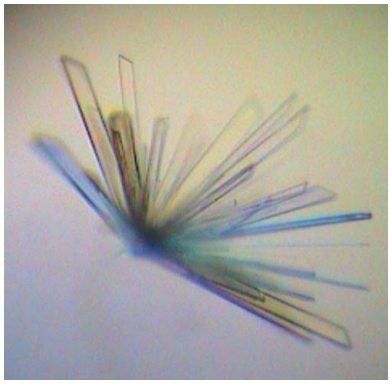
Received 2 March 2009
Accepted 11 May 2009

Q4DV70 is annotated in the *Trypanosoma cruzi* CL Brener genome as a hypothetical protein with a predicted thioredoxin-like fold, although the catalytic cysteine residues that are conserved in typical oxidoreductases are replaced by serine residues. Gene-expression analysis indicates that this protein is differentially expressed during the *T. cruzi* life cycle, suggesting that it plays an important role during *T. cruzi* development. The gene coding for Q4DV70 was cloned and the protein was overexpressed in *Escherichia coli* with an N-terminal His tag. Purification of Q4DV70 was carried out by affinity and size-exclusion chromatography and the His tag was removed by TEV protease digestion. Crystals of Q4DV70 were grown using the sitting-drop vapour-diffusion method. A diffraction data set was collected to 1.50 Å resolution from a single crystal grown in 25% PEG 1500, 200 m*M* sodium thiocyanate pH 6.9, 10 m*M* phenol and 10% ethylene glycol. The crystal belonged to space group $P2_12_12_1$, with unit-cell parameters a = 35.04, b = 50.32, c = 61.18 Å. The Q4DV70 structure was solved by molecular replacement using protein disulfide isomerase from yeast (PDB code 2b5e) as a search model. Initial refinement of the model indicated that the solution was correct. These data are being used for refinement of the model of Q4DV70.

1. Introduction

Chagas disease is caused by the protozoan *Trypanosoma cruzi* (Chagas, 1909) and affects more than 7.7 million people in Latin America. The pathogen's life cycle is complex, involving a mammalian and a triatomine insect host and at least three developmental stages: epimastigotes, trypomastigotes and amastigotes (De Souza, 1984). Metacyclogenesis is the transformation from the replicative stage present in the insect gut (the epimastigotes) to the infective stage (the metacyclic trypomastigotes) (De Souza, 1984). Transition from stages during metacyclogenesis is controlled by the activation of specific genes. In trypanosomatids, gene expression is mostly regulated by post-transcriptional mechanisms (Avila *et al.*, 2003; Minning *et al.*, 2003; Haile & Papadopoulou, 2007; Clayton & Shapira, 2007) and analysis of the level of mRNAs associated with polysomes gives a consistent signature of the genes that are being translated in each stage. Global analysis using DNA microarray indicated that the mRNA encoding Q4DV70 (XP_818273) shows differential association with polysomes during the *T. cruzi* life cycle (Krieger & Goldenberg, unpublished results). mRNAs whose levels of association with polysomes correlate to a single developmental stage are thought to play important functions during metacyclogenesis, in contrast to mRNAs encoding housekeeping proteins, which have similar association with polysomes at all stages.

Although the Q4DV70 protein has been annotated as a conserved hypothetical protein in the *T. cruzi* CL Brener genome, its fold can be predicted to be thioredoxin-like owing to sequence similarity. This fold is composed of a four-stranded β-sheet flanked by three α-helices and usually contains a C*XX*C motif in the active site (Martin, 1995). The thioredoxin fold was named after the *Escherichia coli* thioredoxin structure, which was the first protein structure of this family



to be determined (Holmgren *et al.*, 1975). It is found in a large number of proteins such as thioredoxins (Holmgren *et al.*, 1975), glutharedoxins (Sodano *et al.*, 1991), glutathione *S*-transferases (Reinemer *et al.*, 1991), DsbA (Martin *et al.*, 1993), glutathione peroxidases (Epp *et al.*, 1983) and protein disulfide isomerases (PDIs; Kemmink *et al.*, 1996). The sequence-identity levels between proteins from these groups are very low and do not suggest structural similarity. Furthermore, the groups do not share the same biological functions. Indeed, although Q4DV70 has a thioredoxin-like fold, it does not contain any cysteines in its sequence, rendering it incapable of acting as an oxidoreductase. However, proteins of this family have also been reported to display chaperone activity that is independent of the formation and/or isomerization of disulfides. Thioredoxin from *E. coli* promoted the functional folding of bacterial galactose receptor, a protein without any cysteines (Kern *et al.*, 2003). PDI, DsbA and DsbC assisted the refolding of D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a protein that lacks disulfides (Cai *et al.*, 1994; Zheng *et al.*, 1997; Chen *et al.*, 1999).

Q4DV70 possesses a 30-residue signal peptide in the N-terminal region according to theoretical predictions (*SignalP* 3.0 server; Emanuelsson *et al.*, 2007), indicating that it is transported to the endoplasmic reticulum. There, Q4DV70 may have a chaperone activity independent of oxidoreductase activity towards nascent polypeptide chains, preventing them from aggregating and maintaining them in a favourable conformation for correct folding, similar to other proteins with a thioredoxin-like fold. In this work, we describe the crystallization and solution of the *T. cruzi* Q4DV70 structure, which should help in understanding its molecular function and possible thioredoxin-like chaperone activity.

2. Materials and methods

2.1. Production and purification of recombinant Q4DV70

The Q4DV70 coding sequence, starting from amino acid 30 in order to exclude the signal peptide, was amplified by PCR from *T. cruzi* Dm28c genomic DNA using the following oligonucleotides as primers: Q4DV70F, 5′-CTG ACA TAT GGC GTC AAA TGT GGC AAA TGA C-3′, and Q4DV70R, 5′-TCA GAT CTT CAT GTG TTC TGA AAA ACA AAA CTG-3′. The resulting DNA fragment was cloned into the *Nde*I and *Bam*HI restriction sites of plasmid pET-TEV (Carneiro *et al.*, 2006) to produce pET-TEV/Q4DV70. DNA sequencing of Q4DV70 amplified using DNA from the Dm28c strain as a template revealed a difference from the sequence deposited in the database for the CL Brener strain. In the Dm28c strain Tyr94 is replaced by a histidine. This can be explained by the genetic variability between these two *T. cruzi* strains.

Escherichia coli BL21(DE3)*sly*D$^-$ competent cells were transformed with pET-TEV/Q4DV70. A fresh single colony was grown in 10 ml Luria–Bertani medium (LB) containing 50 µg ml^{-1} kanamycin for 16 h (200 rev min^{-1} and 310 K). This culture was inoculated into 1 l fresh LB with kanamycin, maintained under the same conditions and induced with 1 m*M* isopropyl β-D-1-thiogalactopyranoside when the culture reached an optical density of 0.8 at 600 nm. Following a 4 h induction, the cells were harvested (6000*g*, 15 min, 277 K) and the pellet was suspended in 10 ml buffer *A* (50 m*M* sodium phosphate pH 7.4, 100 m*M* NaCl, 5% glycerol) containing 1 m*M* phenylmethanesulfonylfluoride (PMSF). Cell lysis was performed by lysozyme treatment (80 µg ml^{-1} for 1 h at 277 K) and sonication. The lysates were clarified by centrifugation (20 000*g*, 30 min, 277 K) and submitted to chromatography using a 5 ml HiTrap Chelating HP (GE Healthcare) column coupled to an ÄKTA FPLC (GE Healthcare) system with a flow rate of 1 ml min^{-1}. The column was washed with 15 column volumes (CV) of buffer *A* and eluted with a step gradient of buffer *A* containing 0–500 m*M* imidazole as follows: 0–100 m*M* in 5 CV, 100–300 m*M* in 25 CV and 300–500 m*M* in 2 CV.

Subsequently, the His tag was digested in a reaction mixture containing 0.26 mg ml^{-1} Q4DV70, 0.013 mg ml^{-1} TEV protease, 50 m*M* Tris–HCl pH 8.0, 0.5 m*M* ethylenediaminetetraacetic acid (EDTA) and 1 m*M* dithiothreitol (DTT) for 15 h at room temperature. The reaction was stopped with 0.01% sodium dodecyl sulfate (SDS) and applied onto a Superdex 75 16/60 column (GE Healthcare) equilibrated with 50 m*M* sodium phosphate pH 7.4, 100 m*M* NaCl and 5% glycerol. This chromatography was carried out on an ÄKTA FPLC (GE Healthcare) system with a 1 ml min^{-1} flow rate. A second affinity chromatography step was carried out to adsorb the His-tagged TEV protease and the residual fraction of undigested Q4DV70. This chromatography was carried out on a 5 ml HiTrap Chelating HP (GE Healthcare) column as described above. The flowthrough of this column containing Q4DV70 without the His tag was dialyzed against 20 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 5% glycerol, concentrated by ultrafiltration and used in crystallization assays.

2.2. Crystallization, data collection and processing

Crystallization was carried out by vapour diffusion using a HoneyBee 963 Robot (Genomic Solutions). Sitting drops containing 200 nl of either 3 or 6 mg ml^{-1} Q4DV70 and 200 nl reservoir solution were equilibrated against 80 µl reservoir solution at 291 K. The following crystallization screens were tested: a sparse-matrix screen from the Joint Center for Structural Genomics (JCSG; Page *et al.*, 2003), the PACT Suite for a systematic analysis of the effects of pH, anions and cations (Newman *et al.*, 2005), precipitant synergy (Majeed *et al.*, 2003), Wizard Screens I and II (Hol *et al.*, 2001), SaltRx (Gilliland *et al.*, 1994; Kanaujia *et al.*, 2007) and Crystal Screen and Crystal Screen 2 from Hampton Research (Jancarik & Kim, 1991). Initially, needle-type crystals were obtained using 20% PEG 3350, 200 m*M* sodium thiocyanate pH 6.9 as precipitant. Crystal optimization was performed by changing the concentrations of thiocyanate and PEG and the PEG molecular mass. Sitting drops containing 1 µl Q4DV70 and 1 µl reservoir solution were equilibrated against 80 µl reservoir solution. A second round of optimization was carried out with additive screens (Additive Screens I, II and III; Hampton Research) in the optimized condition (25% PEG 1500, 200 m*M* sodium thiocyanate pH 6.9) using hanging drops containing 2.5 µl Q4DV70, 2.0 µl reservoir solution and 0.5 µl additive solution equilibrated against 200 µl reservoir solution. The best crystals were grown using 10 m*M* phenol as an additive and 10%(*v*/*v*) ethyleneglycol for cryoprotection.

X-ray diffraction data were collected on the W01B-MX2 beamline of the Brazilian Synchrotron Light Laboratory (LNLS), Campinas, Brazil using a MAR 225 CCD detector (MAR Research). Crystals were retrieved from the drops with cryoloops and mounted on the goniometer; a nitrogen-gas stream at 100 K was used for flash-cooling. The first data set contained 360 images and the second contained 129 images; both used an oscillation of 1°. The wavelength of the X-rays was 1.459 Å. Data processing was performed with the *HKL*-2000 package (Otwinowski & Minor, 1997).

2.3. Molecular replacement

The Matthews coefficient (Matthews, 1968) was calculated using 14 894 kDa as the molecular weight. The region Val365–Gly484 of yeast PDI (PDB code 2b5e), which shares 28% identity with

Q4DV70, was selected as a search model for the molecular-replacement phasing method. The model was modified using the 'last common atom' option of the *CHAINSAW* software (Stein, 2008) available in the *CCP*4 package and nonprotein atoms were removed from the coordinate file. Molecular replacement was carried out using the *MOLREP* software (Vagin & Teplyakov, 1997) from *CCP*4. Model refinement was carried out by simulated annealing using *CNS* (Brünger *et al.*, 1998) followed by cycles of manual inspection and model building with *Coot* (Emsley & Cowtan, 2004) and restrained refinement with *REFMAC* (Murshudov *et al.*, 1997).

3. Results and discussion

The recombinant Q4DV70 protein was purified from *E. coli* extracts by IMAC (Fig. 1). The His tag was cleaved with TEV protease, leaving a GHM tripeptide at the N-terminus of the protein. Aggregated proteins were removed by gel filtration. A second IMAC was carried out to adsorb the His-tagged TEV protease and the residual undigested His-tagged Q4DV70 from cleaved Q4DV70; the purified Q4DV70 without a His tag was used for crystallization. Crystals of

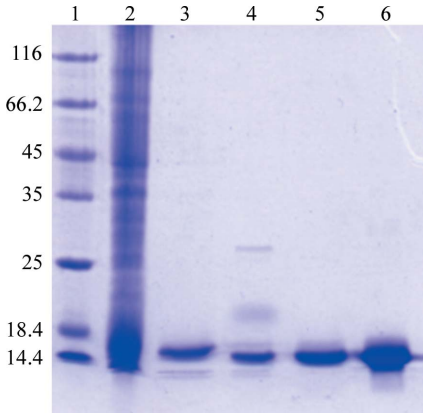


Figure 1
SDS–PAGE analysis of Q4DV70 purification steps. Lane 1, molecular-weight markers (kDa); lane 2, supernatant of induced cells; lane 3, His-tagged Q4DV70 purified by affinity chromatography; lane 4, His-tag cleavage reaction with TEV protease; lane 5, Q4DV70 after gel filtration; lane 6, Q4DV70 in the flowthrough of the affinity chromatography.

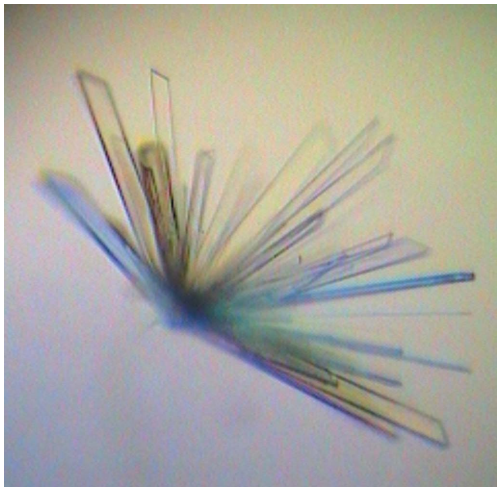


Figure 2
Cluster of crystal plates of Q4DV70; individual plates had approximate dimensions of 5 × 50 × 500 µm.

Table 1
Data-collection and processing statistics for Q4DV70 crystals.

Values in parentheses are for the highest resolution shell.

Crystal	Not cryoprotected	Cryoprotected
Wavelength (Å)	1.459	1.459
Crystal-to-detector distance (mm)	100	60
Temperature (K)	100	100
Images collected	360	129
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	a = 35.04, b = 50.76, c = 60.98	a = 35.04, b = 50.32, c = 61.18
Mosaicity (°)	0.4	0.4
Resolution (Å)	20.00–1.76 (1.82–1.76)	20.00–1.50 (1.55–1.50)
$I/\sigma(I)$	44.1 (8.2)	23.9 (4.7)
Completeness (%)	81.4 (82.7)	99.1 (95.9)
Multiplicity	25.3 (18.3)	4.7 (3.5)
$R_{\rm merge}$†	0.09 (0.27)	0.07 (0.24)
Reflections	233508	84522
Unique reflections	9221 (904)	17813 (1619)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I(hkl)$ is the intensity of reflection hkl, $\sum_{hkl}$ is the sum over all reflections and $\sum_i$ is the sum over i measurements of reflection hkl.

Q4DV70 grew as a cluster of plates in 20% PEG 3350, 200 m*M* sodium thiocyanate pH 6.9 from JCSG (Page *et al.*, 2003) in 1 d at 291 K. Variations of this initial condition and the use of additives did not change the shape of the crystals (Fig. 2). One plate was broken apart from a cluster (grown in 25% PEG 1500, 200 m*M* sodium thiocyanate pH 6.9, 10 m*M* phenol) and placed directly into a 100 K nitrogen-gas stream. As it diffracted very well despite the formation of ice over the crystal, a data set was collected (Table 1). The crystal belonged to space group $P2_12_12_1$, with unit-cell parameters a = 35.04, b = 50.76, c = 60.98 Å. Unfortunately, the ice formation affected the completeness. The calculated Matthews coefficient was 1.82 Å^3 Da^{-1}, which corresponds to one molecule per asymmetric unit and 32.5% solvent content. Molecular replacement carried out using the region Val365–Gly484 of yeast PDI resulted in a translation-function contrast of 4.2. After a few cycles of refinement, 90% of the chain was constructed. A second data set was collected from a crystal grown in 25% PEG 1500, 200 m*M* sodium thiocyanate pH 6.9, 10 m*M* phenol, 10%(*v*/*v*) ethylene glycol. This crystal also belonged to space group $P2_12_12_1$, with a small variation in the unit-cell parameters. This data set is being used for the refinement of the model constructed with the first data set, preserving the same set of reflections used for the calculation of $R_{\rm free}$.

Once the structure of Q4DV70 from *T. cruzi* has been refined, it will be compared with other thioredoxin-like structures, especially those lacking catalytic cysteines, in order to obtain functional information. The hypothesis that this protein might act as a chaperone and the fact that its expression profile changes during metacyclogenesis builds an interesting case for the biology of this human pathogen.

This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), CEPID/CBME-FAPESP, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Associação Brasileira de Tecnologia de Luz Síncrotron (ABTLuS).

References

Avila, A. R., Dallagiovanna, B., Yamada-Ogatta, S. F., Monteiro-Góes, V., Fragoso, S. P., Krieger, M. A. & Goldenberg, S. (2003). *Genet. Mol. Res.* **2**, 159–168.

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.

Cai, H., Wang, C. C. & Tsou, C. L. (1994). *J. Biol. Chem.* **269**, 24550–24552.
Carneiro, F. R. G., Silva, T. L. C., Alves, A. C., Haline-Vaz, T., Gozzo, F. C. & Zanchin, N. I. (2006). *Biochem. Biophys. Res. Commun.* **343**, 260–268.
Chagas, C. (1909). *Mem. Inst. Oswaldo Cruz*, **1**, 159–218.
Chen, J., Song, J. L., Zhang, S., Wang, Y., Cui, D. F. & Wang, C. C. (1999). *J. Biol. Chem.* **274**, 19601–19605.
Clayton, C. & Shapira, M. (2007). *Mol. Biochem. Parasitol.* **156**, 93–101.
De Souza, W. (1984). *Int. Rev. Cytol.* **86**, 197–283.
Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. (2007). *Nature Protoc.* **2**, 953–971.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Epp, O., Ladenstein, R. & Wendel, A. (1983). *Eur. J. Biochem.* **133**, 51–69.
Gilliland, G. L., Tung, M., Blakeslee, D. M. & Ladner, J. E. (1994). *Acta Cryst.* D**50**, 408–413.
Haile, S. & Papadopoulou, B. (2007). *Curr. Opin. Microbiol.* **10**, 569–577.
Hol, W. G. J., Sarfaty, S. H., Stewart, L. J. & Kim, H. (2001). US Patent 6267935.
Holmgren, A., Söderberg, B.-O., Eklund, H. & Brändén, C.-I. (1975). *Proc. Natl Acad. Sci. USA*, **72**, 2305–2309.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kanaujia, S. P., Ranjani, C. V., Jeyakanthan, J., Baba, S., Chen, L., Liu, Z.-J., Wang, B.-C., Nishida, M., Ebihara, A., Shinkai, A., Kuramitsu, S., Shiro, Y., Sekar, K. & Yokoyama, S. (2007). *Acta Cryst.* F**63**, 27–29.
Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. & Creighton, T. E. (1996). *Biochemistry*, **35**, 7684–7691.
Kern, R., Malki, A., Holmgren, A. & Richarme, G. (2003). *Biochem. J.* **371**, 965–972.
Majeed, S., Ofek, G., Belachew, A., Huang, C. C., Zhou, T. & Kwong, P. D. (2003). *Structure*, **11**, 1061–1070.
Martin, J. L. (1995). *Structure*, **3**, 245–250.
Martin, J. L., Bardwell, J. C. A. & Kuriyan, J. (1993). *Nature (London)*, **365**, 464–468.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Minning, T. A., Bua, J., Garcia, G. A., McGraw, R. A. & Tarleton, R. L. (2003). *Mol. Biochem. Parasitol.* **131**, 55–64.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* D**61**, 1426–1431.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C. & Lesley, S. A. (2003). *Acta Cryst.* D**59**, 1028–1037.
Reinemer, P., Dirr, H. W., Ladenstein, R., Schäffer, J., Gallay, O. & Huber, R. (1991). *EMBO J.* **10**, 1997–2005.
Sodano, P., Xia, T. H., Bushweller, J. H., Björnberg, O., Holmgren, A., Billeter, M. & Wüthrich, K. (1991). *J. Mol. Biol.* **221**, 1311–1324.
Stein, N. (2008). *J. Appl. Cryst.* **41**, 641–643.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Zheng, W. D., Quan, H., Song, J. L., Yang, S. L. & Wang, C. C. (1997). *Arch. Biochem. Biophys.* **337**, 326–331.