Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray analysis of the *E. coli* hypothetical protein TdcF

Julia D. Burman,[a] Clare E. M. Stevenson,[a] Katherine A. Hauton,[b] Gary Sawers[b] and David M. Lawson[a]*

[a]Department of Biological Chemistry, John Innes Centre, Norwich NR4 7UH, England, and [b]Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, England

Correspondence e-mail: david.lawson@bbsrc.ac.uk

Received 31 January 2003
Accepted 26 March 2003

Crystals of the hypothetical protein TdcF (subunit MW = 14 007) from *Escherichia coli* were grown by vapour diffusion. The protein crystallizes in space group $P2_12_12$, with unit-cell parameters a = 72.67, b = 86.22, c = 62.62 Å. Native data to a resolution of 2.35 Å were collected from a single crystal at 100 K on a rotating-anode X-ray generator. Preliminary analysis of these data indicated that the asymmetric unit corresponded to a trimer, which was supported by a convincing molecular-replacement solution using the YjgF trimer as the probe structure.

1. Introduction

The anaerobically regulated *tdcABCDEFG* operon of *Escherichia coli* encodes proteins involved in the transport and fermentation of L-serine and L-threonine (Hesslinger *et al.*, 1998; Sawers, 1998). Metabolism of these substrates provides the cell with a source of energy (Hesslinger *et al.*, 1998; Merberg & Datta, 1982). Induction of the *tdc* operon occurs under anaerobic conditions in the presence of the amino acids serine, threonine, valine and isoleucine, as well as fumarate and cyclic AMP (Schweizer & Datta, 1988). The importance of the Tdc enzymes for the metabolism of *E. coli* stems from their ability to produce energy-rich keto acids that are subsequently catabolized to produce ATP *via* substrate-level phosphorylation (Hesslinger *et al.*, 1998; Sawers, 1998). Functional roles have been assigned, albeit tentatively in some cases, to six of the gene products: TdcA is a *trans*-acting positive regulator (Ganduri *et al.*, 1993; Sawers, 2001), TdcB is a threonine dehydratase (Datta *et al.*, 1987), TdcC is an integral membrane protein implicated in the transport of serine and threonine into the cell (Sumantran *et al.*, 1990), TdcD is a propionate kinase (Hesslinger *et al.*, 1998), TdcE is a keto-acid formate-lyase (Hesslinger *et al.*, 1998) and TdcG is a serine dehydratase (Hesslinger *et al.*, 1998). The *tdcF* gene (formerly *yhaR*) encodes a protein of 129 residues in length with a calculated molecular mass of 14 007 Da (Hesslinger *et al.*, 1998). It is the exception in the *tdc* operon in that it has no known function. Nevertheless, it has sequence homologues in both bacteria and eukarya that belong to the YjgF family which, according to the InterPro database (http://www.ebi.ac.uk/interpro/index.html; Apweiler *et al.*, 2001), currently has 107 members. In addition, PSI-BLAST (Altschul *et al.*, 1997) finds four homologues in the *E. coli* genome alone, with identities ranging from 28 to 74% (with a minimum of 64% of the TdcF amino-acid sequence aligned). Whilst several YjgF-like proteins have been linked to certain physiological processes (Ceciliani *et al.*, 1996; Enos-Berlage *et al.*, 1998; Kim *et al.*, 2001; Rappu *et al.*, 1999), the specific biochemical role of none has been ascertained to date. Crystal structures of two homologues have already been determined: those of YjgF itself from *E. coli* (74% identity; Volz, 1999) and of the YabJ protein from *Bacillus subtilis* (47% identity; Sinha *et al.*, 1999). They form closely superposable homotrimers and display structural similarity to the known protein chorismate mutase (Chook *et al.*, 1993). However, the lack of sequence conservation with the active site of chorismate mutase suggests that they do not share similar functions. Nevertheless, this site corresponds to the region of the structure with the highest degree of sequence conservation within the YjgF family, suggesting it has a functional role (Volz, 1999). The co-location of *tdcF* with genes implicated in the metabolism of serine and threonine indicates that it may have a function in this process. Here, we report the crystallization and preliminary X-ray analysis of the *E. coli* TdcF protein, which will form part of a multidisciplinary approach towards determining its biological role.

2. Materials and methods

2.1. Expression and purification

The *E. coli tdcF* gene was amplified by PCR using MC4100 chromosomal DNA as a template. The amplified DNA was cloned into the vector pET-28a (Novagen) to give the plasmid pETHF1, encoding a polypeptide with a thrombin-cleavable N-terminal hexahistidine tag. This added a further 20 residues

to the native protein (with sequence MGSSHHHHHHSSGLVPRGSH), giving a total deduced molecular mass of 16 170 Da. This plasmid was transformed into *E. coli* strain BL21 (DE3) gold (Studier & Moffatt, 1986) and the cells were grown anaerobically in 2 l of Luria–Bertani medium at 310 K. Protein overproduction was induced at an OD_{600} of 0.6–0.8 by the addition of 1 m*M* isopropyl-β-D-thiogalactopyranoside and was continued for 2 h. Harvested cells were resuspended in buffer (20 m*M* Tris–HCl pH 7.5, 250 m*M* NaCl, 5 m*M* imidazole) and were broken open by passage through a French press at 1.03×10^2 MPa. The cell lysate obtained by centrifugation at 13 000*g* for 10 min was loaded onto an Ni^{2+}-charged Hi-trap metal-chelation column and unbound proteins were removed by washing with 5 m*M* imidazole in 20 m*M* Tris–HCl pH 7.5, 250 m*M* NaCl. His-tagged TdcF bound to the column was eluted with a 0–500 m*M* imidazole gradient in 20 m*M* Tris–HCl pH 7.5, 250 m*M* NaCl. The fractions containing TdcF were pooled, dialysed against 10 m*M* Tris–HCl pH 8.0 and concentrated to approximately 10 mg ml^{-1} using a Centriprep-10 concentrator (Amicon). The His tag was not cleaved. The protein was centrifuged through a 0.1 µm Ultrafree filter (Millipore) before use. The solution properties of the sample were subsequently analysed using dynamic light scattering (DLS) with a Dynapro-MSTC instrument (Protein Solutions Inc.).

2.2. Crystallization and X-ray diffraction analysis

Crystallization trials were performed by vapour diffusion in hanging drops using VDX plates (Hampton Research) or sitting drops using CrystalClear strips (Hampton Research) at a constant temperature of 291 K. Drops consisted of 1 µl protein and 1 µl well solution (the well volume was 1 ml for hanging drops and 100 µl for sitting drops). Initial crystallization conditions were sought using five different screens. An in-house screen (consisting of three different concentrations each of ammonium sulfate and PEG 8000 at four different pH values, giving 24 conditions) and X-tal fact1 and X-tal fact2 screens developed by Johan Zeelen (http://www.mpibp-frankfurt.mpg.de/~johan.zeelen/4screen.html) were set up as hanging drops. Hampton Crystal Screens 1 and 2 (Hampton Research) were then set up as sitting drops. The successful conditions were subsequently optimized. Crystals were cryoprotected by soaking for up to 5 min in mother liquor containing ethylene glycol. The crystals were then mounted on cryoloops (Hampton Research) and flash-cooled to 100 K in a stream of gaseous nitrogen produced by an X-Stream cryocooler (Rigaku-MSC). Diffraction data were collected in-house using a MAR 345 image-plate detector (X-ray Research) mounted on a Rigaku RU-H3RHB rotating-anode X-ray generator (operated at 50 kV and 100 mA) fitted with Osmic confocal optics and a copper target (Cu $K\alpha$; λ = 1.542 Å). X-ray data were processed using *DENZO* and merged using *SCALEPACK* (Otwinowski & Minor, 1997). All other data-processing stages and statistical analysis were performed using programs from the *CCP*4 software suite (Collaborative Computational Project, Number 4, 1994).

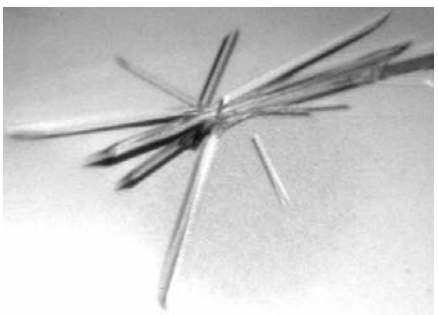

Figure 1
Crystals of *E. coli* TdcF. The largest are approximately 200 × 20 × 20 µm in size.

3. Results and discussion

DLS analysis of purified TdcF gave a monomodal distribution and a relatively low polydispersity (23.8%), indicating that the sample was homogeneous. From these results, the molecular size was estimated to be 40 kDa. This value is approximately halfway between the figures expected for a dimer (32.3 kDa) and a trimer (48.5 kDa) of the His-tagged protein. Of all the crystallization screens tested, only condition No. 7 of Hampton Crystal Screen 2 gave crystals. These were needle-shaped, up to 100 µm in length and grew after about 7 d. The precipitant consisted of an unbuffered solution of 10%(*w*/*v*) PEG 1000 and 10%(*w*/*v*) PEG 8000. Optimization was subsequently performed in hanging-drop format. Lowering the temperature to 277 K prevented crystal growth, although drops equilibrated at this temperature yielded small crystals on moving to 291 K. Larger crystals (200 × 20 × 20 µm) were obtained simply by varying the precipitant concentration, the best growing from 10%(*w*/*v*) PEG 1000 and 8%(*w*/*v*) PEG 8000 (Fig. 1). Cryoprotection was achieved by replacing 20% of the water volume in the mother liquor with ethylene glycol.

Table 1
Summary of TdcF X-ray data.

Values in parentheses are for the outer resolution shell.

Resolution range (Å)	30.0–2.35
Unique reflections	16863
Completeness (%)	99.3 (95.2)
Redundancy	4.4
$R_{\rm merge}$†	0.103 (0.368)
$\langle I\rangle/\langle\sigma(I)\rangle$	12.1 (2.7)

† $R_{\rm merge} = \sum(|I_j - \langle I_j\rangle|)/\sum\langle I_j\rangle$, where I_j is the intensity of an observation of reflection j and $\langle I_j\rangle$ is the average intensity for reflection j.

After flash-cooling in the cold gaseous nitrogen stream on the X-ray camera, data were recorded to a maximum resolution of 2.35 Å. The symmetry was established as primitive orthorhombic, with unit-cell parameters a = 72.67, b = 86.22, c = 62.62 Å. A total of 118 × 1° oscillation images were collected in a single sweep, yielding a data set that was 99.3% complete. Data-collection statistics are summarized in Table 1. After inspection of pseudo-precession plots, the space group was assigned as $P2_12_12$.

Solvent-content estimations for a TdcF dimer in the asymmetric unit gave a $V_{\rm M}$ value of 3.03 Å^3 Da^{-1}, corresponding to a solvent content of 59%, whilst a $V_{\rm M}$ of 2.02 Å^3 Da^{-1} and a solvent content of 39% were obtained for a trimer (Matthews, 1968). However, inspection of a self-rotation function calculated on data in the resolution range 8–5 Å (using the program *POLARRFN*) gave no clear indication of twofold non-crystallographic symmetry, whilst significant peaks were observed in the κ = 120° section consistent with threefold non-crystallographic symmetry. Given the high sequence identity with YjgF (74%), a protein of known structure (Volz, 1999), molecular replacement could be used to solve the structure. The complete YjgF trimer from PDB entry 1qu9, stripped of solvent molecules, was used as the search model in the program *AMoRe*. To verify that we had ascertained the correct space group, all possible enantiomorphs were tested. Only the space group $P2_12_12$ gave a convincing answer: a single trimer was located in the asymmetric unit, giving an initial correlation coefficient of 66.2% and an overall R factor[1] of 40.6% at 2.35 Å resolution. This preliminary model is currently being refined.

[1] R factor = $\sum|(F_{\rm obs} - F_{\rm calc}|)/\sum|F_{\rm obs}| \times 100$, where $F_{\rm obs}$ and $F_{\rm calc}$ are the observed and calculated structure-factor amplitudes, respectively.

Given the ubiquitous nature of the YjgF family members, they are clearly of widespread importance. It is possible that TdcF may have a role in the anaerobic catabolism of serine and threonine, given that its gene is co-transcribed within an operon containing genes implicated in this process (Hesslinger *et al.*, 1998; Sawers, 2001). Through our structural analyses in conjunction with ligand-binding studies, the analysis of protein–protein interaction partners and the rigorous phenotypic examination of a *tdcF* knock-out mutant, we aim to deduce the biochemical function of TdcF and possibly its homologues.

This work is funded by the BBSRC. JDB is supported by a studentship from the John Innes Foundation.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.

Apweiler, R. *et al.* (2001). *Nucleic Acids Res.* **29**, 37–40.

Ceciliani, F., Faotto, L., Negri, A., Colombo, I., Berra, B., Bartorelli, A. & Ronchi, S. (1996). *FEBS Lett.* **393**, 147–150.

Chook, Y. M., Ke, H. & Lipscomb, W. N. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 8600–8603.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Datta, P., Goss, T. J., Omnaas, J. R. & Patil, R. V. (1987). *Proc Natl Acad. Sci. USA*, **84**, 393–397.

Enos-Berlage, J. L., Langendorf, M. J. & Downs, D. M. (1998). *J. Bacteriol.* **180**, 6519–6528.

Ganduri, Y. L., Sadda, S. R., Datta, M. W., Jambukeswaran, R. K. & Datta, P. (1993). *Mol. Gen. Genet.* **240**, 395–402.

Hesslinger, C., Fairhurst, S. A. & Sawers, G. (1998). *Mol. Microbiol.* **27**, 477–492.

Kim, J. M., Yoshikawa, H. & Shirahige, K. (2001). *Genes Cells*, **6**, 507–517.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Merberg, D. & Datta, P. (1982). *J. Bacteriol.* **150**, 52–59.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Rappu, P., Shin, B. S., Zalkin, H. & Mantsala, P. (1999). *J. Bacteriol.* **181**, 3810–3815.

Sawers, G. (1998). *Arch. Microbiol.* **171**, 1–5.

Sawers, G. (2001). *Mol. Microbiol.* **39**, 1285–1298.

Schweizer, H. P. & Datta, P. (1988). *J. Bacteriol.* **170**, 5360–5363.

Sinha, S., Rappu, P., Lange, S. C., Mantsala, P., Zalkin, H. & Smith, J. L. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 13074–13079.

Studier, F. W. & Moffatt, B. A. (1986). *J. Mol. Biol.* **189**, 113–130.

Sumantran, V. N., Schweizer, H. P. & Datta, P. (1990). *J. Bacteriol.* **172**, 4288–4294.

Volz, K. (1999). *Protein Sci.* **8**, 2428–2437.