Recombinant production, purification and crystallization of the *Toxoplasma gondii* coronin WD40 domain

Juha Pekka Kallio^a and Inari Kursula^{a,b*}

^aCentre for Structural Systems Biology, Helmholtz Centre for Infection Research and German Electron Synchrotron (DESY), Notkestrasse 85, Building 25b, 22607 Hamburg, Germany, and

^bFaculty of Biochemistry and Molecular Medicine, University of Oulu, PO Box 3000, 90014 Oulu, Finland

Correspondence e-mail:
inari.kursula@helmholtz-hzi.de

Received 14 February 2014

Accepted 6 March 2014

Toxoplasma gondii is one of the most widely spread parasitic organisms in the world. Together with other apicomplexan parasites, it utilizes a special actin-myosin motor for its cellular movement, called gliding motility. This actin-based process is regulated by a small set of actin-binding proteins, which in Apicomplexa comprises only 10–15 proteins, compared with >150 in higher eukaryotes. Coronin is a highly conserved regulator of the actin cytoskeleton, but its functions, especially in parasites, have remained enigmatic. Coronins consist of an N-terminal actin-binding β -propeller WD40 domain, followed by a conserved region, and a C-terminal coiled-coil domain implicated in oligomerization. Here, the WD40 domain and the conserved region of coronin from *T. gondii* were produced recombinantly and crystallized. A single-wavelength diffraction data set was collected to a resolution of 1.65 Å. The crystal belonged to the orthorhombic space group $C22_1$, with unit-cell parameters $a = 55.13$, $b = 82.51$, $c = 156.98$ Å.

1. Introduction

Toxoplasma gondii is an apicomplexan parasite that is one of the most successful parasitic organisms in the world, able to infect and multiply in a wide variety of warm-blooded animals. Approximately one-third of the human population is infected with *T. gondii*, but it only causes a disease, toxoplasmosis, in the developing foetus and in individuals suffering from a weakened immune defence. *T. gondii* can sexually reproduce only in cats, which are the primary host of the parasite (Innes, 2010). All members of the Apicomplexa phylum are obligate parasites, many of which are of medical or economic importance because they cause disease in either humans or cattle. The most notorious members of the phylum are the *Plasmodium* species, which cause malaria, one of the deadliest diseases affecting mankind.

Apicomplexan parasites share a common method of actin-dependent cellular movement, called gliding motility, which is also thought to be closely linked to host-cell entry (Dobrowolski & Sibley, 1996). Gliding requires extraordinarily short actin filaments, located beneath the parasite plasma membrane, that are linked to transmembrane receptors that recognize surface molecules on the host cell (reviewed in Sattler *et al.*, 2011). No specialized motor organelles or visible changes in the cell shape are involved (Håkansson *et al.*, 1999). Despite the importance of actin in gliding, the parasite cytoplasm is practically free of microfilaments (Schüler & Matuschewski, 2006). Recently, it has been suggested that, as opposed to gliding, host-cell entry may not be directly dependent on actin or other components of the gliding machinery (Andenmatten *et al.*, 2013; Meissner *et al.*, 2013), which is a partly controversial result and calls for a re-evaluation of many of the current mechanistic hypotheses concerning host-cell invasion by apicomplexan parasites.

Actin dynamics and regulation in apicomplexan parasites differ significantly from higher eukaryotes. The set of actin-interacting regulatory proteins is limited to ~10–15 in Apicomplexa compared with at least 150 in higher eukaryotes. Furthermore, some classical regulators, mainly nucleation and branching factors, such as Arp2/3,

© 2014 International Union of Crystallography
All rights reserved

Table 1
Recombinant protein production information.

In the primers, the sequence given in upper-case letters denotes the insert-specific sequence, and the lower-case letters denote the overhang required for the SLIC vector. In the protein sequence, the coronin sequence is indicated in bold, and the 6×His tag and the 3C cleavage site are in italics. Thus, the final protein contains an additional GPG in the N terminus after proteolytic cleavage of the 6×His tag.

Source organism	<i>T. gondii</i>
DNA source	Synthetic (codon-optimized for insect cells)
Forward primer	5'-cag gga ccc ggt GCC GAC GCT GTA GAC G
Reverse primer	5'-cga gga gaa gcc cgg tta AGG TTT GAC ACT ACG TCT AC
Cloning vector	pBS II SK(+)
Expression vector	pFastBacNKI-his-3c-LIC
Expression host	<i>Spodoptera frugiperda</i> (Sf21)
Complete amino-acid sequence of the construct produced	MAHHHHHSAALEVLFQGPADAVDVPLIKN- LYAEAWKQYSDLRRLSTKQTESCGLAANTE- YIAAPWDVGGGGVGLRLADIGRNPAAVAK- IKGHTASIQDTNFSPPFYRDILATACEDTIV- RIWQLPEEVTGTTELKEPIATLTGALKKVL- SAEWNPAVSGILASGCFDGTVAFWNVEKNE- NFASVKFQESLLSAKWSWGDLACTTKDK- ALNIVDPRAAQVVGSAVCHDGSCKACKCTWI- DGLAGRDGHVFTTGFGKMQEREMAIWD- TRKFDKPVYHAEIDRGSSPLYPIFDETTG- MLYVCGKGDSSCRYYYQHGGTLLRSVDA- YRSSVPKNCFCIPKLAVDQMRAEIGRMLKQ- ENGNVLQPSIFIVPRKNQDVQADLYPPAP- DVEPSMTAEWFKGENKAIRRRSVKP

Wiskott–Aldrich syndrome protein (WASp) and WASp-homology 2 (WH2) domains, are missing in apicomplexan parasites (Sattler *et al.*, 2011). The apicomplexan actins are among the least conserved of all actins and have peculiar polymerization properties. These actins form only very short filaments, and *T. gondii* actin has been proposed to polymerize in an isodesmic manner, which is a fundamental difference from all other actins (Skillman *et al.*, 2013). The set of regulatory proteins must have evolved hand in hand with actin, and the apicomplexan actin-binding proteins have divergent sequences, structures and functions compared with canonical actin regulators (Sattler *et al.*, 2011).

Coronin is one of the few actin regulatory proteins present in apicomplexan parasites. Coronin was originally isolated from *Dictyostelium discoideum* (de Hostos *et al.*, 1991) and is a WD-repeat protein that typically contains a seven-bladed β -propeller domain in its N terminus (Appleton *et al.*, 2006), followed by a conserved region and a unique region, including a coiled-coil domain, in the C-terminal end (Uetrecht & Bear, 2006). The β -propeller domain contains the primary binding site for F-actin, and the C-terminal domains mediate putative interactions with other actin-regulatory proteins and actin as well as self-assembly. The coiled-coil domain has also been predicted to have a secondary, lower-affinity binding site for actin (Gandhi *et al.*, 2009). To date, there are only two crystal structures from murine coronin 1 available in the Protein Data Bank (PDB), and both of them lack major parts of the unique region and the coiled-coil domain (Appleton *et al.*, 2006). *T. gondii* coronin shares less than 30% sequence identity with mouse coronin 1A (Fig. 1).

We have expressed and purified a truncated version of *T. gondii* coronin (hereafter TgCor Δ cc), which contains the WD40 and conserved domains, using baculovirus-infected Sf21 insect cells and purification based on affinity and size-exclusion chromatography. The recombinant coronin was folded, as analyzed by circular-dichroism (CD) spectroscopy, was successfully crystallized, and a high-resolution diffraction data set was collected.

2. Materials and methods

2.1. Recombinant protein production and characterization

A synthetic gene (UniProt ID Q5Y1E7) coding for full-length *T. gondii* coronin (Fig. 1), codon-optimized for insect-cell expression, was ordered from Eurofins MWG Operon, Germany. Several constructs, including that encoding amino acids 2–392, were cloned into an insect-cell expression vector using a sequence- and ligation-independent method (SLIC; Li & Elledge, 2012). The pFastBacNKI-his-3c-LIC vector (NKI Protein Facility, the Netherlands) was

Figure 1
Sequence alignment of *T. gondii* coronin and mouse coronin 1A. The WD40 domain is shown in red, the conserved region in green and the coiled-coil domain in orange. Identical residues are displayed as white letters surrounded by blue boxes.

Table 2
Crystallization details of *TgCorΔcc*.

Method	Sitting-drop vapour diffusion
Plate type	Swissci MRC 2 96-well plates
Temperature (K)	295
Protein concentration (mg ml ⁻¹)	8
Buffer composition of protein solution	100 mM NaCl, 100 mM ammonium sulfate, 50 mM sodium phosphate buffer pH 6.5, 1 mM TCEP
Composition of reservoir solution	200 mM magnesium acetate, 25% PMME 2000, 100 mM Tris-HCl pH 7.5
Volume and ratio of drop	0.3 µl protein + 0.3 µl reservoir
Volume of reservoir (µl)	40

amplified in DH5- α *Escherichia coli* cells and purified using a QIAprep Spin Miniprep Kit (Qiagen). The vector was then linearized using the *KpnI* enzyme. The coding region of each construct was amplified by PCR using the primers described in Table 1 for *TgCorΔcc*. *DpnI* was added to the PCR products to digest the template DNA. The PCR-amplified insert and the linearized vector were purified from an agarose gel using the QIAquick gel extraction kit (Qiagen). To create single-stranded overhangs, the digested vector

and inserts were treated separately with T4 polymerase (New England Biolabs) at 296 K for 20 min. The reaction was terminated by adding 25 mM ethylenediaminetetraacetic acid (EDTA), followed by inactivation of the T4 polymerase at 348 K for 20 min. The T4-treated vector was mixed with the insert, and the annealing reaction was performed at 338 K for 10 min, after which the reaction mixture was slowly cooled to 295 K in order to improve the annealing efficiency. The reaction mixture was then used to transform NEB 5- α heat-competent cells (New England Biolabs) in order to amplify the plasmid. DNA sequencing at Eurofins MWG Operon was used to check the correctness of the clones after amplification.

Recombinant *TgCorΔcc* protein with a hexahistidine tag was expressed in baculovirus-infected insect cells using the protocol described by Bieniossek *et al.* (2008). DH10EMBacY (EMBL, Grenoble) cells were used for bacmid generation. Sf21 (Invitrogen) cells were transfected with the *TgCorΔcc* bacmid, using the FuGene 6 transfection reagent (Promega), and virus particles were harvested after 3 d of infection. The recombinant primary virus was amplified to a high-titre viral stock. Approximately 9 µl (1×10^6 cells) of the high-titre virus were used to infect the insect-cell culture at a cell density of $0.6\text{--}0.8 \times 10^6$ cells per millilitre. The cells were harvested after 72 h and the cell pellets were stored at 253 K. For purification, the cells were resuspended in ice-cold lysis buffer consisting of 100 mM NaCl, 100 mM ammonium sulfate, 50 mM sodium phosphate buffer pH 6.5, 5 mM β -mercaptoethanol (BME), complemented with 10 mM imidazole, 8 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), and the cOMplete Mini protease-inhibitor cocktail (Roche). The cells were disrupted by sonication, after which the cell debris was removed by centrifugation at 277 K.

The His-tagged protein in the clarified lysate was bound to Co²⁺-charged iminodiacetic acid (Co-IDA) agarose (Jena Bioscience) in a gravity-flow column, washed with 20 mM imidazole, 500 mM NaCl, 100 mM ammonium sulfate, 50 mM sodium phosphate buffer pH 6.5, 5 mM BME and eluted with 500 mM imidazole, 100 mM NaCl, 100 mM ammonium sulfate, 50 mM sodium phosphate buffer pH 6.5, 5 mM BME. The pooled fractions containing the fusion protein were concentrated to 2.5 ml, applied onto a PD-10 desalting column (GE Healthcare) to remove imidazole, and eluted with 100 mM NaCl, 100 mM ammonium sulfate, 50 mM sodium phosphate buffer pH 6.5, 5 mM BME. Up to this point, the purification steps were performed at room temperature using cold buffers. The N-terminal 6 $\times$ His tag was cleaved using HRV 3C protease (Novagen) for 16 h at 277 K. 5–10 µl of the protease in 45–65% glycerol was added to 3.5 ml of the uncleaved 6 $\times$ His-*TgCorΔcc* solution from the desalting step. 500 µl of 50% Co-IDA slurry was added to the cleaved protein solution in order to bind the cleaved tag, the protease used, and other remaining contaminants that bound to the Co²⁺ matrix, after which the solution was filtered. Final purification was performed at 277 K by size-exclusion chromatography using a Superdex 200 10/300 GL column (GE Healthcare), equilibrated with 100 mM NaCl, 100 mM ammonium sulfate, 50 mM sodium phosphate buffer pH 6.5, 1 mM tris(2-carboxyethyl)phosphine (TCEP). Pure *TgCorΔcc* was used freshly for following experiments or stored on ice for a few days.

A CD spectrum of the purified and cleaved *TgCorΔcc* from 260 to 190 nm was measured at 293 K using an Applied Photophysics Chirascan Plus spectropolarimeter equipped with a thermal control unit (Quantum Northwest, TC125), a direct temperature probe and a 0.5 mm path-length quartz cuvette (Hellma). The spectrum was recorded at a concentration of 0.25 mg ml⁻¹ in 25 mM NaCl, 25 mM ammonium sulfate, 12.5 mM sodium phosphate buffer pH 6.5, 0.25 mM TCEP. The *DichroWeb* server (Lobley *et al.*, 2002) was used for secondary-structure determination using the *CDSSTR* algorithm

Figure 2
Purification and CD analysis of *TgCorΔcc*. (a) Size-exclusion chromatogram and Coomassie-stained denaturing gel of the peak fractions eluting at ~17 ml from a Superdex 200 10/300 GL column. The positions of molecular-weight markers (in kDa) are indicated at the side of the gel. (b) The CD spectrum of *TgCorΔcc* from 260 to 190 nm measured at 293 K.

Table 3

Data-collection and processing statistics.

Values in parentheses are for the outer shell.

Diffraction source	P14 at PETRA III/DESY
Wavelength (Å)	0.976
Temperature (K)	100
Detector	PILATUS 6M
Crystal-to-detector distance (mm)	266.6
Rotation range per image (°)	0.2
Total rotation range (°)	200
Space group	$C22_1$
Unit-cell parameters (Å, °)	$a = 55.13$, $b = 82.51$, $c = 156.98$, $\alpha = \beta = \gamma = 90$
Resolution range (Å)	50–1.65 (1.7–1.65)
Total No. of reflections	168873
No. of unique reflections	43134
Completeness (%)	99.2 (99.1)
Multiplicity	3.9
$\langle I/\sigma(I) \rangle^\dagger$	12 (1.2)
$R_{\text{meas}}^\ddagger$ (%)	7.7 (134)
$CC_{1/2}^\S$ (%)	99.9 (35.3)
Overall B factor from Wilson plot (Å ²)	29.7

† The mean $I/\sigma(I)$ falls below 2.0 at 1.75 Å resolution. ‡ R_{meas} is the redundancy-independent R factor as defined by Diederichs & Karplus (1997) and Weiss & Hilgenfeld (1997). $\sum_{hkl} \{N(hkl)/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. § $CC_{1/2}$ is defined as the correlation coefficient between two random half data sets, as described by Karplus & Diederichs (2012).

and set 4 optimized for 190–240 nm as a reference data set (Compton & Johnson, 1986).

2.2. Crystallization

Crystallization conditions for *TgCorΔcc* were screened using the sitting-drop vapour-diffusion method in Swissci MRC 2 96-well plates. Crystallization experiments were set up by manually mixing 0.3 µl of both reservoir solution and protein solution at a concentration of 4–9 mg ml^{−1}, as determined by UV absorption at 280 nm and the extinction coefficient calculated from the *TgCorΔcc* sequence, and the plates were incubated at 293 K. The following screens were used: PEG/Ion, Crystal Screen Lite (Hampton Research, Aliso Viejo, USA) and ProPlex (Molecular Dimensions, Altamonte Springs, USA). Optimization of promising conditions was performed by screening the effect of various salts with polyethylene glycol (PEG) 3350 or polyethylene glycol monomethyl ether (PMME) 2000 at two different pH values (6.5 and 7.5). Data-collection-quality crystals grew in 200 mM magnesium acetate,

20–25% PMME 2000, 100 mM Tris–HCl pH 7.5 at 293 K. For identifying protein crystals, an X-taLight 100 UV source (Molecular Dimensions, Altamonte Springs, USA) was used to detect the tryptophan fluorescence of protein crystals (Dierks *et al.*, 2010). Crystallization details are given in Table 2.

2.3. Data collection and processing

Crystals of *TgCorΔcc* were flash-cooled in liquid nitrogen after soaking in a cryoprotectant solution with 25%(v/v) glycerol in the reservoir solution. Preliminary X-ray diffraction tests and native data collection to 1.65 Å resolution were performed on the EMBL beamline P14 at PETRA III, Hamburg, Germany. Diffraction images were recorded on a PILATUS 6M detector (DECTRIS Ltd, Switzerland), using a wavelength of 0.976 Å, an oscillation angle of 0.2° per frame and a crystal-to-detector distance of 266.6 mm. The diffraction images were indexed, integrated and scaled using the *XDS* package (Kabsch, 2010) and *XDSi* (Kursula, 2004).

3. Results and discussion

TgCorΔcc (394 residues; molecular weight 43.3 kDa) was successfully overexpressed in *Sf21* insect cells and purified to homogeneity using Co-IDA agarose in gravity-flow columns followed by size-exclusion chromatography (Fig. 2a). *TgCorΔcc* eluted from the Superdex 200 10/300 GL column at an elution volume corresponding to the molecular weight of a monomeric WD40 domain, which was expected in the absence of the predicted coiled-coil domain presumed to be responsible for coronin self-assembly into dimers or trimers. The identity of the purified protein was verified by peptide-mass fingerprinting using mass spectrometry at the Biocenter Oulu Proteomics Core Facility. The typical yield was approximately 3.5 mg of pure *TgCorΔcc* from 1 l of *Sf21* cells at a density of $0.6\text{--}0.8 \times 10^6$ cells per millilitre at the point of infection. The cell count normally doubled from that before growth arrest.

The folding and secondary-structure content of the purified *TgCorΔcc* were analyzed using CD spectroscopy. The CD spectrum indicates the protein to be folded (Fig. 2b). *TgCorΔcc* was estimated to contain 41% β -strands and 22% turns, which is similar to the corresponding region in the homologous murine coronin 1 crystal structure, which consists of 42% β -strands and 15% turns (PDB entries 2aq5 and 2b4e; Appleton *et al.*, 2006).

Figure 3

TgCorΔcc protein crystals nucleate from salt crystals. (a) Visible light image of a *TgCorΔcc* crystal attached to the surface of a salt crystal. The protein crystal size is approximately $100 \times 100 \times 50$ µm. (b) Corresponding image showing the detection of the protein tryptophan fluorescence signal while illuminating the crystal with UV light.

Screening for crystallization conditions was performed using commercially available screens. No crystals were obtained directly from the screens, but promising-looking drops with microcrystalline precipitate were found, especially from the PEG/Ion screen. Screening was continued by optimizing these conditions using various salts with PEG 3350 or PMME 2000 at two different pH values, 6.5 and 7.5. Crystals with the longest dimension varying from 50 to 200 μm were obtained from the optimized condition within 7 d (Fig. 3). A peculiar feature of the crystals was that they all grew using salt crystals as a nucleation point. In the drops, plate-shaped salt crystals appeared immediately after setting up the drops, and approximately after 7 d, additional crystals were detected (Fig. 3a). The crystals were observed under UV light, where a clear tryptophan fluorescence signal was visible (Fig. 3b). The protein crystals could be separated from the salt crystals for X-ray diffraction measurements. A native data set to 1.65 Å resolution was recorded and used for molecular-replacement trials.

The TgCor Δcc crystals belong to space group $C222_1$, with unit-cell parameters $a = 55.1$, $b = 82.5$, $c = 156.98$ Å, $\alpha = \beta = \gamma = 90^\circ$. The Matthews coefficient of $2.06 \text{ Å}^3 \text{ Da}^{-1}$ with a solvent content of 40.3% (Matthews, 1968) suggests that the asymmetric unit contains one TgCor Δcc molecule. Data-collection and processing statistics are given in Table 3. The data quality was evaluated using *phenix.xtriage* (Adams *et al.*, 2010) and no pathological behaviour was detected. Structure determination will be started using molecular replacement using the homologous murine coronin 1 structure (PDB entry 2aq5) as a model.

The upcoming structure of TgCor Δcc will be the first crystal structure of a parasite coronin to be determined and only the second for a WD40 domain from the whole coronin protein family. The crystallized domain from *T. gondii* coronin shares only 30% sequence identity with the corresponding murine coronin 1. Thus, it is expected to reveal valuable information on the diversity of coronin WD40 domains and how, for example, the suggested actin-binding patches located within the WD40 domain of higher eukaryotic coronins differ in apicomplexan coronins.

We thank Dr Ulrich Bergmann for performing the mass spectrometry. We are grateful for the excellent user support at the EMBL

Hamburg beamline P14. We would also like to acknowledge the NKI Protein Facility, supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO; grant No. 175.010.2007.012), for providing the pFastBac-LIC vector. This work was financially supported by the Academy of Finland, the Sigrid Jusélius Foundation, the Emil Aaltonen Foundation and the German Ministry for Education and Research (BMBF).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Andenmatten, N., Egarter, S., Jackson, A. J., Jullien, N., Herman, J. P. & Meissner, M. (2013). *Nature Methods*, **10**, 125–127.
- Appleton, B. A., Wu, P. & Wiesmann, C. (2006). *Structure*, **14**, 87–96.
- Bieniossek, C., Richmond, T. J. & Berger, I. (2008). *Curr. Protoc. Protein Sci.*, Unit 5.20. doi:10.1002/0471140864.ps0520s51.
- Compton, L. A. & Johnson, W. C. (1986). *Anal. Biochem.* **155**, 155–167.
- Diederichs, K. & Karplus, P. A. (1997). *Nature (London)*, **4**, 269–275.
- Dierks, K., Meyer, A., Oberthür, D., Rapp, G., Einspahr, H. & Betzel, C. (2010). *Acta Cryst.* **F66**, 478–484.
- Dobrowolski, J. M. & Sibley, L. D. (1996). *Cell*, **84**, 933–939.
- Gandhi, M., Achard, V., Blanchoin, L. & Goode, B. L. (2009). *Mol. Cell*, **34**, 364–374.
- Håkansson, S., Morisaki, H., Heuser, J. & Sibley, L. D. (1999). *Mol. Biol. Cell*, **10**, 3539–3547.
- Hostos, E. L. de, Bradtke, B., Lottspeich, F., Guggenheim, R. & Gerisch, G. (1991). *EMBO J.* **10**, 4097–4104.
- Innes, E. A. (2010). *Zoonoses Public Health*, **57**, 1–7.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Kursula, P. (2004). *J. Appl. Cryst.* **37**, 347–348.
- Li, M. Z. & Elledge, S. J. (2012). *Methods Mol. Biol.* **852**, 51–59.
- Lobley, A., Whitmore, L. & Wallace, B. A. (2002). *Bioinformatics*, **18**, 211–212.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Meissner, M., Ferguson, D. J. & Frischknecht, F. (2013). *Curr. Opin. Microbiol.* **16**, 438–444.
- Sattler, J. M., Ganter, M., Hliscs, M., Matuschewski, K. & Schüler, H. (2011). *Eur. J. Cell Biol.* **90**, 966–971.
- Schüler, H. & Matuschewski, K. (2006). *Trends Parasitol.* **22**, 146–147.
- Skillman, K. M., Ma, C. I., Fremont, D. H., Diraviyam, K., Cooper, J. A., Sept, D. & Sibley, L. D. (2013). *Nature Commun.* **4**, 2285.
- Uetrecht, A. C. & Bear, J. E. (2006). *Trends Cell Biol.* **16**, 421–426.
- Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203–205.