The LRR-Roc-COR module of the *Chlorobium tepidum* Roco protein: crystallization and X-ray crystallographic analysis

Egon Deyaert,^{a,b} Arjan Kortholt^c and Wim Versées^{a,b*}

Received 8 June 2017

Accepted 17 August 2017

Edited by M. S. Weiss, Helmholtz-Zentrum Berlin für Materialien und Energie, Germany

Keywords: Roco proteins; *Chlorobium tepidum*; LRR-Roc-COR.

^aStructural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium, ^bVIB–VUB Center for Structural Biology, Pleinlaan 2, 1050 Brussels, Belgium, and ^cDepartment of Cell Biochemistry, University of Groningen, 9747 AG Groningen, The Netherlands. *Correspondence e-mail: wim.versees@vub.be

Roco proteins are characterized by the presence of a Roc-COR supradomain harbouring GTPase activity, which is often preceded by an LRR domain. The most notorious member of the Roco protein family is the Parkinson's disease-associated LRRK2. The Roco protein from the bacterium *Chlorobium tepidum* has been used as a model system to investigate the structure and mechanism of this class of enzymes. Here, the crystallization and crystallographic analysis of the LRR-Roc-COR construct of the *C. tepidum* Roco protein is reported. The LRR-Roc-COR crystals belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 95.6$, $b = 129.8$, $c = 179.5$ Å, $\alpha = \beta = \gamma = 90^\circ$, and diffracted to a resolution of 3.3 Å. Based on the calculated Matthews coefficient, Patterson map analysis and an initial molecular-replacement analysis, one protein dimer is present in the asymmetric unit. The crystal structure of this protein will provide valuable insights into the interaction between the Roc-COR and LRR domains within Roco proteins.

1. Introduction

Roco proteins are present in all domains of life. Apart from the characteristic Roc-COR supradomain (Roc, **R**as of **c**omplex proteins; COR, **C**-terminal of **R**oc), the various members of this protein family contain several other domains. Based on the domain topology, the family has been divided into three groups (Bosgraaf & Van Haastert, 2003). Proteins of the first group are built up from the Roc-COR module fused to an N-terminal leucine-rich repeat (LRR) domain. Members of the second group have an additional kinase domain and contain a variety of extra N- and C-terminal domains, such as ankyrin repeats, WD40 repeats, RhoGEF and RhoGAP domains (Bosgraaf & Van Haastert, 2003). The last group is more distinct and consists of proteins of the death-associated protein kinase (DAPk) family. Remarkably, DAPk proteins are the only Roco proteins that lack an LRR domain preceding the Roc-COR module, suggesting an important function for the LRR-Roc-COR module within Roco proteins of groups 1 and 2 (Bosgraaf & Van Haastert, 2003).

The human LRRK2 protein belongs to the second Roco protein group. Mutations in LRRK2 are the most common cause of familial Parkinson's disease and are linked to the more frequently occurring sporadic form of the disease (Lin & Farrer, 2014; Dächsel & Farrer, 2010; Zimprich *et al.*, 2004; Paisán-Ruiz *et al.*, 2004; Ross *et al.*, 2011). Biochemical studies using LRRK2 remain very challenging because of its severe tendency towards aggregation, its low purification yield and the rapid degradation of the protein. Instead, bacterial Roco proteins from group 1 have been successfully used to study the

mechanism of the Roco protein family (Gotthardt *et al.*, 2008; Terheyden *et al.*, 2015).

The Roco protein from the bacterium *Chlorobium tepidum* (CtRoco) contains the core LRR-Roc-COR module and a stretch of approximately 150 amino acids of unknown function at the C-terminus (Bosgraaf & Van Haastert, 2003). The crystal structure of the *C. tepidum* Roc-COR domain (amino acids 412–946) in the nucleotide-free state revealed a dimeric arrangement of this module (Gotthardt *et al.*, 2008). However, only one Roc domain was resolved in this structure, probably owing to flexibility of the other Roc domain. This Roc domain interacts with both COR domains of the dimer (Gotthardt *et al.*, 2008). The COR domain in turn comprises N- and C-terminal subdomains coupled *via* a long flexible linker (Gotthardt *et al.*, 2008). Within the COR domain it is exclusively the C-terminal subdomains that contribute to the dimer interface. The crystal structure of the isolated *C. tepidum* leucine-rich repeat domain has also been solved (Guaitoli *et al.*, 2016). It adopts the hallmark curved α/β -solenoid structure. In this study, we report the crystallization and initial X-ray analysis of a *C. tepidum* LRR-Roc-COR construct (CtLRR-Roc-COR). Since the LRR-Roc-COR module is present in nearly all Roco proteins, determination of the CtLRR-Roc-COR crystal structure may provide insights into the general intra/intermolecular organization and regulation of Roco proteins.

2. Materials and methods

2.1. Macromolecule production

A plasmid coding for a *C. tepidum* LRR-Roc-COR construct (CtLRR-Roc-COR; residues 1–946) was generated by inserting a stop codon into the pProEX plasmid with the *C. tepidum* Roco gene described in Gotthardt *et al.* (2008) using QuikChange mutagenesis (Table 1). This N-terminally His-tagged construct was expressed in *Escherichia coli* BL21 (DE3) cells. The cells were grown in Terrific Broth; at an OD_{600 nm} of 0.7, 0.1 mM IPTG was added and the cells were then grown overnight at 26°C. After centrifugation, the cells were resuspended in 50 mM Tris, 150 mM NaCl, 5 mM MgCl₂, 1 mM β -mercaptoethanol, 10 mM imidazole, 10% glycerol, 100 μ M GDP (pH 7.5). The cells were lysed with a cell disruptor (Constant Systems) and after centrifugation the supernatant was loaded onto an Ni-NTA column. This column was first washed with 50 mM Tris, 150 mM NaCl, 5 mM MgCl₂, 1 mM β -mercaptoethanol, 10% glycerol, 10 mM imidazole (pH 7.5) and subsequently with the same buffer including 300 mM KCl and 1 mM ATP to remove copurified chaperones. The target protein was eluted with a linear gradient of imidazole and was subsequently dialysed against 20 mM HEPES, 2 mM MgCl₂, 5 mM β -mercaptoethanol, 5% glycerol (pH 7.5). This protein solution was loaded onto a Source 30Q anion-exchange column and eluted using a NaCl gradient. Finally, the pure protein fractions were pooled and subjected to size-exclusion chromatography (Superdex 200) with 20 mM HEPES, 150 mM NaCl, 5 mM MgCl₂, 1 mM DTT, 5% glycerol

Table 1

Macromolecule-production information.

Source organism	<i>C. tepidum</i>
Expression vector	pProEX
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced	MSYYHHHHHHYDIPPTTENLYFQGAMGSM DLDVIRQIEQELGMQLEPVDKLKWYSKG YKLDKQDQRTAIGLYDCGSDTLDRIIQP LESLSLSLSLSLSNQTIDISPLASLNS LSMLWLDNRNQITDIAPLASLNSLSMLWL FGNKISDIAPLESLSLSLTELQSSNQIT DIAPLASLSLSLTELSSGNNISDIAPLE SLKSLTELSSNQTIDDIAPLASLSLTL ELSSNQSISDIAPLESLSLSLTELQLSR NQISDIAPLESLSLSLTELQSSNQITDI APLASLSLSLTELQSRNQISDIAPLES NSLSKLWLNQITDIAPLASLNSLTEL ELSSNQTIDDIAPLASLSLSLTLWSSNQ ISDIAPLASLESLSLSLSSNQSISDISP LASLSLTLGFDVRRNPILKRLPETITGFD MEILWDFSSSGFITFFDNPLESPPEI VKQGEAVRQYFQSIEEARSKEALVHL QEIKVHLIGDMAGKTSLLKQLIGETFD PKESQTHGLNVVTKQAPNIKLENDDEL KECLFHFWDFGGQEIHASHQFMTTRSS VYMLLLDSRTDSNKHYWLRIEYKGGKS PVIVVMNKIDENPSYNIQKKINERFPA IENRFHRSCKNGDGVESIAKSLKSAVL HPDSIYGTPLAPSWIKVKEKLVEATTAQ RYLNRTEVEKICNDSGITDPGERKTLLG YLNNGLIVLYFEALDLSIYVLDPHWVT IGVYRIINSSKTKNGHLNLSALGYILNE EQIRCDEYDPAKNNKFTYTLLEQRYLLD IMKQFELCYDEGKGLFIIPSNLPTQIDN EPEITEGEPLRFIMKYDYLPTIIPRLM IAMQHQLDRMQWRYGMVLKSQDHEGAL AKVVAETKDSITIAIQGEPRCKREYLS IIWYEIKKINANFTNLDVKEFIPLPGHP DELVEYKELLGLEKMGREYVSGKLEKV FSVSKMLDSVISKEERNKER

(pH 7.5). The nucleotide load of the purified protein was determined using a C18 reversed-phase column (Jupiter 5 μ m C18 300 Å, Phenomenex) coupled to an HPLC system (Waters). 100 mM KH₂PO₄ pH 6.4, 10 mM tetrabutylammonium bromide, 7.5% acetonitrile was used as the mobile phase. A 100 μ M protein sample was heated at 95°C for 3 min. After centrifugation, 50 μ l of the supernatant with potentially copurified nucleotides was injected onto the column. Nucleotide elution was followed by the absorbance at 254 nm. The nucleotide-detection limit of this setup is 0.5 μ M.

2.2. Crystallization

The JCSG-*plus*, Clear Strategy I, Clear Strategy II, ProPlex and Wizard crystallization screens from Molecular Dimensions were used for initial screening (Page *et al.*, 2003). Two-well sitting-drop vapour-diffusion crystallization plates were set up manually at 293 K using 8 mg ml⁻¹ CtLRR-Roc-COR protein with two different protein:mother liquor ratios for each condition: 0.5 μ l protein solution with 0.5 μ l mother liquor and 0.5 μ l protein solution with 1 μ l mother liquor. The initial hit in 0.8 M sodium formate, 0.1 M sodium acetate pH 5.5, 10% PEG 8000, 10% PEG 1000 could only be reproduced using seeds produced with microseed beads (Molecular

Dimensions). Crystals were optimized by modifying the precipitant concentration and the pH.

2.3. Data collection and processing

The crystals were cryoprotected using the mother-liquor solution supplemented with 15% PEG 200. Data collection was performed on the PROXIMA2 beamline at the SOLEIL synchrotron, Paris, France using an EIGER detector (Dectris). The obtained data set was indexed and processed with *XDS* (Kabsch, 2010).

3. Results and discussion

The LRR-Roc-COR construct of the *C. tepidum* Roco protein (CtLRR-Roc-COR; amino acids 1–946) was purified from

E. coli BL21 (DE3) cells using a three-step purification protocol. The purity of the obtained sample was checked on SDS–PAGE (Fig. 1*a*), and analysis of the supernatant after

Table 2 Crystallization.	
Method	Sitting-drop vapour diffusion with seeding
Plate type	Two-well Intelli-Plate
Temperature (K)	293
Protein concentration (mg ml ^{−1})	8
Buffer composition of protein solution	20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM MgCl ₂ , 1 mM DTT, 5% glycerol
Composition of reservoir solution	0.8 M sodium formate, 0.1 M sodium acetate pH 5.5, 10% PEG 8000, 10% PEG 1000
Volume and ratio of drop	0.5 µl protein solution/1 µl reservoir solution
Volume of reservoir (µl)	80

Figure 1
SDS–PAGE and initial crystal hits for the CtLRR-Roc-COR protein. (*a*) Silver-stained SDS–PAGE of CtLRR-Roc-COR after a three-step purification protocol. The first lane contains the PageRuler Prestained Protein Ladder, with the relevant molecular weights indicated in kDa. The last lane contains the highly pure CtLRR-Roc-COR protein (molecular weight 109.6 kDa). (*b*, *c*, *d*) Different crystal forms of CtLRR-Roc-COR: (*b*) diamond-shaped crystals appearing after two months in 1.6 M sodium citrate tribasic dihydrate pH 6.5, (*c*) needle clusters appearing after 10 d in 0.8 M sodium formate, 0.1 M sodium acetate pH 5.5, 10% PEG 8000, 10% PEG 1000 and (*d*) plate-like crystals appearing after 12 d in 0.3 M sodium acetate, 0.1 M sodium cacodylate pH 6.5, 25% PEG 2000 MME. The scale bar in (*b*) is applicable to all pictures.

boiling showed that the protein was in its apo form after purification (no nucleotides were bound to the Roc domain). Crystallization trials with the purified apo protein were set up at 8 mg ml^{-1} . Crystals with different morphologies were obtained, including diamonds, needle clusters and plates (Figs. 1*b*, 1*c* and 1*d*). Initial in-house diffraction experiments with the diamond-shaped and plate-like crystals resulted in diffraction up to $8 \text{ \AA}$ resolution. Starting from these initial hits, several optimization strategies were implemented to obtain larger crystals. Finally, the crystal with the best diffraction was obtained after 10 d in 0.8 M sodium formate, 0.1 M sodium

acetate pH 5.5, 10% PEG 8000, 10% PEG 1000 using seeds from the plate-like crystals (Fig. 2 and Table 2).

A $3.3 \text{ \AA}$ resolution data set was collected on the PROXIMA2 beamline at the SOLEIL synchrotron by rotating the crystal in increments of 0.1° over 120° (Table 3). The space group was determined to be $P2_12_12_1$, with unit-cell parameters $a = 95.6$, $b = 129.8$, $c = 179.5 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$. The unit-cell volume is consistent with the presence of one dimer in the asymmetric unit, with a solvent content of 51.89% and a Matthews coefficient of $2.56 \text{ \AA}^3 \text{ Da}^{-1}$ (Matthews, 1968; Kantardjieff & Rupp, 2003; Weichenberger & Rupp, 2014).

Figure 2 CtLRR-Roc-COR crystals after optimization and the resulting diffraction. (a) The CtLRR-Roc-COR crystals were grown in 0.8 M sodium formate, 0.1 M sodium acetate pH 5.5, 10% PEG 8000, 10% PEG 1000 using crystal seeds from the plate-shaped crystals. (b) A diffraction pattern from an optimized CtLRR-Roc-COR crystal obtained on the PROXIMA2 beamline.

Figure 3 Self-rotation function and native Patterson map analysis. (a) The self-rotation function was calculated using *MOLREP* from the *CCP4* package with data in the $47.2\text{--}3.5 \text{ \AA}$ resolution range for $\chi = 180^\circ$ (Vagin & Teplyakov, 2010; Winn *et al.*, 2011). The three twofold crystallographic symmetry axes of the $P2_12_12_1$ space group are clearly revealed. (b) Harker section with $w = 0.5$ from the native Patterson map. The native Patterson map was calculated using *FFT* and viewed using *MAPSLICER*, both of which are from the *CCP4* package. The clear peak at position $u, v, w = (0.3, 0.25, 0.5)$ indicates the presence of translational noncrystallographic symmetry (Winn *et al.*, 2011; Read & Schierbeek, 1988).

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	PROXIMA2, SOLEIL
Wavelength (Å)	0.9801
Detector	EIGER
Rotation range per image (°)	0.1
Total rotation range (°)	120
Space group	$P2_12_12_1$
a, b, c (Å)	95.6, 129.8, 179.5
α, β, γ (°)	90, 90, 90
Resolution range (Å)	47.79–3.29 (3.41–3.29)
Total No. of reflections	155879 (13110)
No. of unique reflections	34302 (3210)
Completeness (%)	0.99 (0.96)
Multiplicity	4.5 (4.1)
Mean $I/\sigma(I)$	9.3 (1.5)
R_{meas}	0.158 (1.103)
$CC_{1/2}$	0.995 (0.539)

To obtain further insights into the organization of the asymmetric unit, the self-rotation function was analysed using *MOLREP* (Vagin & Teplyakov, 2010; Fig. 3a). Apart from the three twofold crystallographic rotation axes, no noncrystallographic symmetry (NCS) was observed at first sight. However, a native Patterson map was then calculated (Fig. 3b). The large non-origin peak in the Harker section with $w = 0.5$ indicates translational noncrystallographic symmetry, thus confirming the presence of two molecules in the asymmetric unit.

The phase problem was solved by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the structures of the individual CtLRR (amino acids 1–441), CtRoc (amino acids 442–614) and CtCOR (amino acids 615–946) domains as search models [PDB entries 3dpu (Gotthardt *et al.*, 2008) and 5il7 (Guaitoli *et al.*, 2016)]. As predicted by the Matthews coefficient and the native Patterson map, two LRR-Roc-COR modules were present in the asymmetric unit. The solutions of the *Phaser* runs were LLG = 836, TFZ = 29 for the two LRR domains, LLG = 666, TFZ = 25 for the two COR domains and LLG = 1650, TFZ = 33 for the two Roc domains. Further refinement of the structure will yield the very first experimental structural model of the highly conserved LRR-Roc-COR domain arrangement within a dimeric nucleotide-free

Roco protein, and will be instrumental in understanding the interaction between the LRR and Roc-COR domains in this protein family.

Acknowledgements

We want to thank the staff of the PROXIMA2 beamline at SOLEIL and David Young for their help during data collection. We are grateful to Remy Loris for helpful discussions.

Funding information

Funding for this research was provided by: Fonds Wetenschappelijk Onderzoek; BioStruct-X, European Commission Seventh Framework Programme; Strategic Research Program Financing of the VUB; Hercules Foundation; Michael J. Fox Foundation for Parkinson's Research; NWO-VIDI grant.

References

- Bosgraaf, L. & Van Haastert, P. J. M. (2003). *Biochim. Biophys. Acta*, **1643**, 5–10.
- Dächsel, J. C. & Farrer, M. J. (2010). *Arch. Neurol.* **67**, 542–547.
- Gotthardt, K., Weyand, M., Kortholt, A., Van Haastert, P. J. M. & Wittinghofer, A. (2008). *EMBO J.* **27**, 2239–2249.
- Guaitoli, G. *et al.* (2016). *Proc. Natl Acad. Sci. USA*, **113**, E4357–E4366.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Lin, M. K. & Farrer, M. J. (2014). *Genome Med.* **6**, 48.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C. & Lesley, S. A. (2003). *Acta Cryst.* **D59**, 1028–1037.
- Paisán-Ruiz, C. *et al.* (2004). *Neuron*, **44**, 595–600.
- Read, R. J. & Schierbeek, A. J. (1988). *J. Appl. Cryst.* **21**, 490–495.
- Ross, O. A. *et al.* (2011). *Lancet Neurol.* **10**, 898–908.
- Terheyden, S., Ho, F., Gilsbach, B., Wittinghofer, A. & Kortholt, A. (2015). *Biochem. J.* **465**, 139–147.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Weichenberger, C. X. & Rupp, B. (2014). *Acta Cryst.* **D70**, 1579–1588.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Zimprich, A. *et al.* (2004). *Neuron*, **44**, 601–607.