Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Structure of the N-terminal domain of the effector protein LegC3 from *Legionella pneumophila*

Deqiang Yao,[a]‡ Maia Cherney[a] and Miroslaw Cygler[a,b]*

[a]Department of Biochemistry, University of Saskatchewan, 107 Wiggins Road, Saskatoon, SK S7N 5E5, Canada, and [b]Department of Biochemistry, McGill University, Montreal, Quebec, Canada

‡ Present address: Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, People's Republic of China.

Correspondence e-mail: miroslaw.cygler@usask.ca

Received 24 September 2013
Accepted 31 October 2013

PDB reference: N-terminal domain of LegC3, 4mu6

Legionella pneumophila secretes over 300 effectors during the invasion of human cells. The functions of only a small number of them have been identified. LegC3 is one of the identified effectors, which is believed to act by inhibiting vacuolar fusion. It contains two predicted transmembrane helices that divide the protein into a larger N-terminal domain and a smaller C-terminal domain. The function of LegC3 has been shown to be associated primarily with the N-terminal domain, which contains coiled-coil sequence motifs. The structure of the N-terminal domain has been determined and it is shown that it is highly α-helical and contains a helical bundle followed by a long antiparallel coiled-coil. No similar protein fold has been observed in the PDB. A long loop at the tip of the coiled-coil distal from the membrane is disordered and may be important for interaction with an as yet unidentified protein.

1. Introduction

The intracellular human pathogen *Legionella pneumophila* is the cause of Legionnaires' disease, a severe form of pneumonia. It enters the cell through phagocytosis and modulates the normal conversion pathway of the *Legionella*-containing vacuole (LCV) into the phago-lysosome by preventing its fusion with the lysosome. The interference with normal cell function is orchestrated through effector proteins secreted by the bacterium through the Icm/Dot type IV secretion system (Vogel & Isberg, 1999). *Legionella* has an unusually large repertoire of over 300 effectors (Isberg *et al.*, 2009; Hubber & Roy, 2010), many of which are functionally still uncharacterized. Among *Legionella* effectors, there are approximately 40 that show similarities to eukaryotic proteins (de Felipe *et al.*, 2008), and it has been proposed that they have been acquired through horizontal gene transfer (Gomez-Valero *et al.*, 2011). Several of these proteins have sequence features of coiled-coil proteins and three of them, LegC2, LegC3 and LegC7, have been shown to interfere with organellar trafficking and yeast homotypic vacuolar fusion (de Felipe *et al.*, 2008; Bennett *et al.*, 2013). Of these, LegC3 showed the most profound effect. The predicted presence of coiled-coil supersecondary-structural elements suggested that this protein might interact with SNARE proteins; however, no associations with SNARE proteins has been detected (Bennett *et al.*, 2013).

LegC3 is 560 amino acids in length with two predicted transmembrane helices (residues 373–395 and 402–424). When expressed in mammalian cells, the protein is localized in punctate structures and this localization was dependent on the presence of the transmembrane helices; their deletion led to a diffuse cytoplasmic distribution and a lack of effect on

vacuoles (de Felipe *et al.*, 2008). The effect of LegC3 on vacuolar fusion was investigated *in vivo* in yeast cells and an inhibition of fusion was observed; the transmembrane domain was essential for the proper localization of LegC3 and for its function *in vivo* (Bennett *et al.*, 2013). Secondary-structure predictions indicate that both the N- and C-terminal domains are predominantly α-helical and that a coiled-coil motif is present in the N-terminal domain. Further experiments *in vitro* using vacuoles isolated from yeast showed that the features responsible for fusion inhibition are located mainly within the N-terminal domain, as this domain alone was sufficient to inhibit vacuole fusion, and that this domain remains associated with vacuoles (Bennett *et al.*, 2013). To explain the different behaviour observed *in vivo* and *in vitro*, the authors speculated that the transmembrane helices are necessary *in vivo* for proper trafficking of LegC3 within the host compartment and/or increase the local concentration of the effector at the site of action on the vacuolar membrane surface (Bennett *et al.*, 2013).

Here, we present the structure of the N-terminal domain of LegC3 and show that it is composed of a long discontinuous coiled-coil capped by an antiparallel four-helix bundle that is rigidly connected to the coiled-coil segment. The coiled-coil motif observed here differs from such motifs observed to date.

2. Experimental procedures

2.1. Construction of an expression vector encoding the LegC3 N-terminal domain

The sequence corresponding to amino acids 2–367 of LegC3 (LegC3^{N}) was PCR-amplified from the genome of *L. pneumophila* strain Philadelphia 1 using the following primer set: 5′-TACTTCCAATCCAATgccATTATGTTTTTG-GCCAACTGCAATATC-3′ and 5′-TTATCCACTTCCAAT-gttaCGCTATCTCATTAACTGTTTCTTCTT-3′. The resultant PCR product was ligated into LIC vector pMCSG7 after treatment with T4 polymerase. The constructs were verified by sequencing and confirmed to contain the N-terminal domain with an N-terminal hexahistidine tag.

2.2. Expression and purification of the LegC3 N-terminal domain protein

Escherichia coli BL21 Star (DE3) cells transformed with the above expression plasmid were grown to mid-log phase, cooled to 20°C and induced with 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside. The cells were collected by centrifugation at 5000*g* for 20 min after overnight growth. The cell pellets were resuspended in buffer *A* (50 m*M* Tris pH 8.8, 300 m*M* NaCl, 5% glycerol, 10 m*M* imidazole with 1 m*M* *p*-aminobenzamidine and 1 μ*M* leupeptin as inhibitors) and the cells were broken by sonication. Cell debris was removed by centrifugation for 1 h at 14 000*g* and the supernatant was incubated with Ni–NTA beads for 1 h at room temperature. The beads were washed with ten column volumes of buffer *A*. LegC3^{N} protein was eluted with buffer *B* (20 m*M* Tris pH 8.8, 300 m*M* NaCl) with the addition of 250 m*M* imidazole. The protein was further purified by size-exclusion chromatography on a Superdex 200 10/300 GL column (GE Healthcare) in buffer *B*. The purified protein was concentrated to 10 mg ml^{-1} in the same buffer, flashed-cooled in liquid nitrogen and stored at −80°C. Selenomethionine-containing protein was obtained using a similar protocol by growing B834 cells on LeMaster medium (Hendrickson *et al.*, 1990).

Table 1
X-ray data collection for LegC3^{N} protein treated with endoprotease Glu-C.

	LegC3^{N}	LegC3^{N} + MMA
Space group	$P2_12_12$	$P2_12_12$
Unit-cell parameters (Å)	a = 108.9, b = 150.2, c = 24.2	a = 107.0, b = 149.5, c = 24.2
Wavelength (Å)	0.97887	0.99186
Resolution (Å)	45.50–2.08 (2.18–2.08)	50.00–2.60 (2.64–2.60)
Observed reflections	266588	86709
Unique reflections	23598	12829
Completeness (%)	94.9 (84.9)	99.7 (96.3)
Multiplicity	11.3 (10.3)	6.8 (6.0)
$R_{\rm merge}$†	0.144 (0.932)	0.203 (0.905)
$\langle I/\sigma(I)\rangle$	16.6 (2.3)	9.5 (1.7)
$R_{\rm work}$‡	0.208	
$R_{\rm free}$§	0.275	
Wilson B factor (Å^2)	25.49	
B factor (Å^2)		
Protein	44.4	
Solvent	49.8	
No. of atoms		
Protein	2250	
Solvent	288	
Ramachandran plot¶, residues in (%)		
Favoured regions	99.26	
Allowed regions	0.74	
Disallowed regions	0	
R.m.s. deviations		
Bonds (Å)	0.007	
Angles (°)	0.99	
PDB code	4mu6	

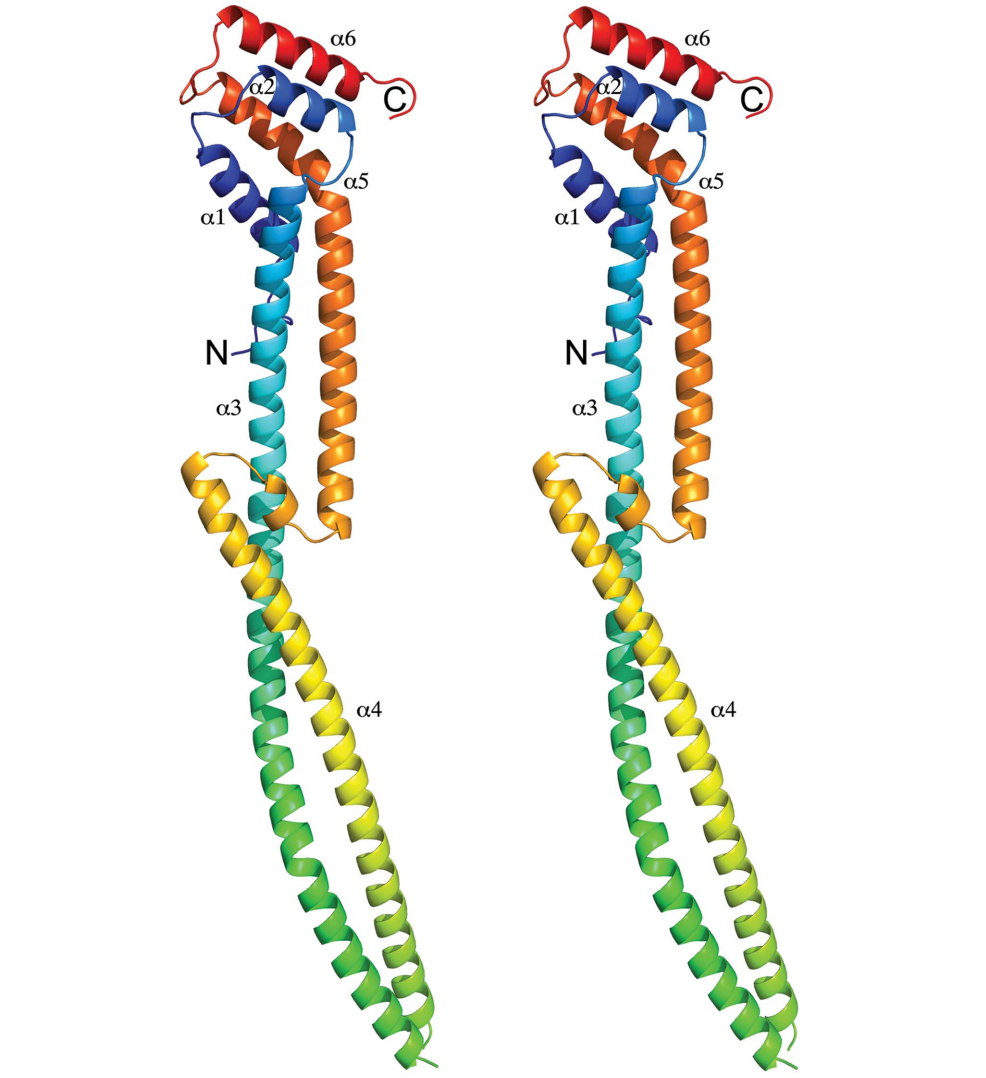
† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm work} = \sum_{hkl}||F_{\rm obs}| - |F_{\rm calc}||/\sum_{hkl}|F_{\rm obs}|$. § $R_{\rm free}$ is the same as $R_{\rm work}$ but calculated for a random set of 8.69% of the unique reflections. ¶ Validated by *MolProbity* (Chen *et al.*, 2010).

2.3. Crystallization

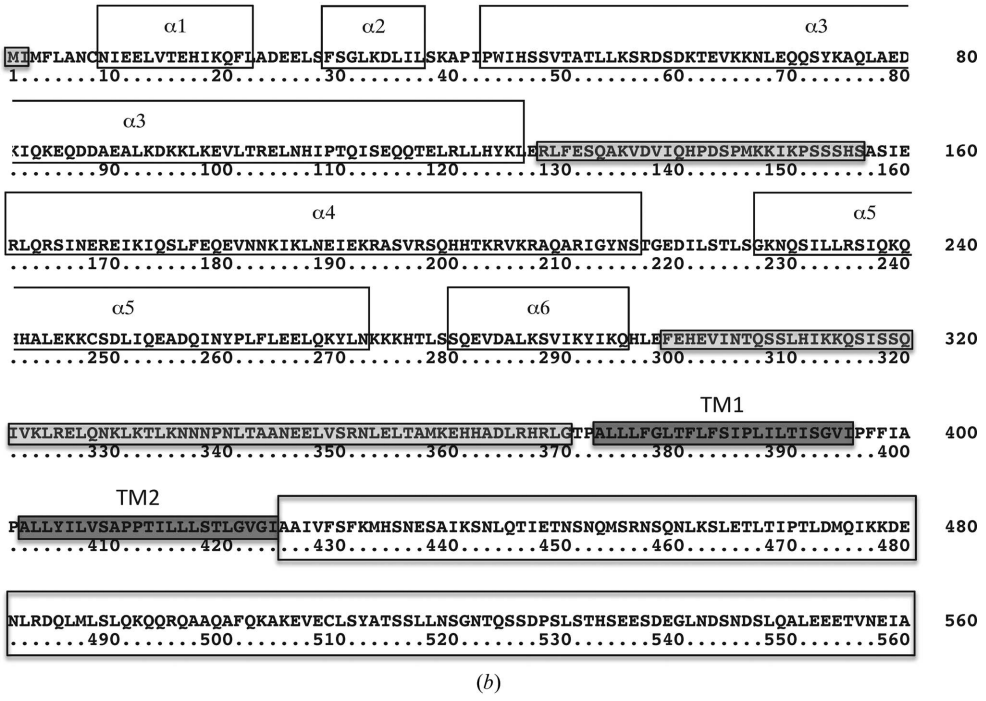
The initial crystallization conditions for LegC3^{N} were identified using the JCSG screen and were optimized using the hanging-drop vapour-diffusion method. The crystals were improved by adding endoprotease Glu-C to the crystallization drop in a 1:100(*w*:*w*) ratio, which removed highly mobile segments through limited proteolysis. The best LegC3^{N} crystals were grown by the equilibration of 1 µl protein solution (at 10 mg ml^{-1}) in 20 m*M* Tris buffer pH 8.8, 300 m*M* NaCl with 1 µl reservoir solution consisting of 20% PEG 3300, 0.1 *M* Tris–HCl pH 8.5, 0.2 *M* tripotassium citrate, 30 m*M* diammonium hydrogen phosphate. To solve the structure, we prepared a heavy-atom derivative by soaking the crystals in 10 m*M* methylmercuric acetate (MMA) for 30 s.

2.4. X-ray data collection, structure solution and refinement

Crystals of native and MMA-soaked LegC3^{N} were flash-cooled in liquid nitrogen and diffraction data were collected



(*a*)



(*b*)

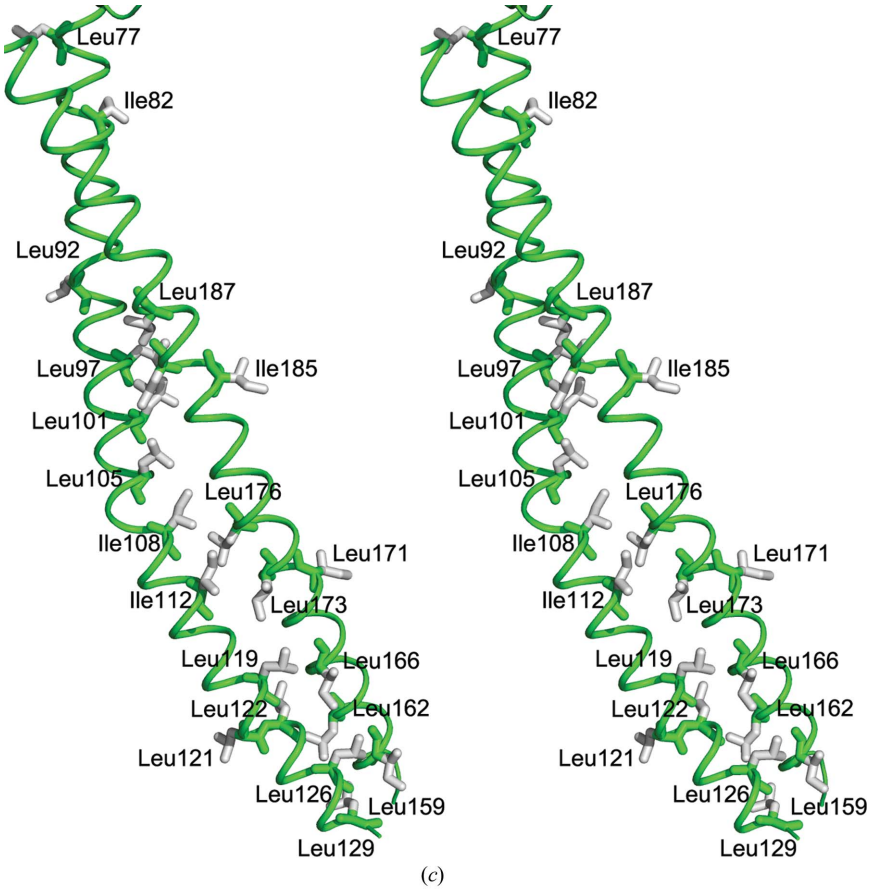
to 2.10 and 2.60 Å resolution, respectively, on the CMCF-1 beamline at the Canadian Light Source (Grochulski *et al.*, 2011). The positions of four Hg atoms were found using *SHELXD* (Sheldrick, 2008). Initial phasing and model building was performed in *PHENIX* (Adams *et al.*, 2010). The molecular-replacement model-completion procedure, as implemented in the *OASIS* software (He *et al.*, 2007), was used with the native data set to improve the phases and to complete model building. Iterative cycles of refinement were carried out using *PHENIX* and *Coot* (Emsley & Cowtan, 2004). The final R_{work} and R_{free} were 20.4 and 26.3%, respectively. The structure was validated with *MolProbity* (Chen *et al.*, 2010). Data-collection and refinement statistics are given in Table 1.

3. Results and discussion

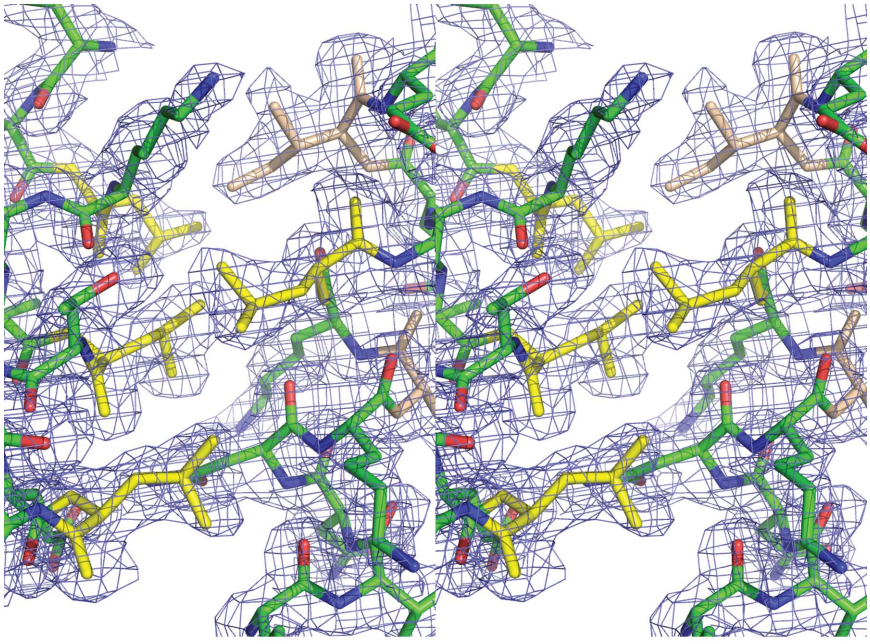
The initial crystals of LegC3^{N} diffracted to a low resolution which was insufficient for structure determination. Much better diffracting crystals (2.1 Å resolution) were obtained when a small amount of endoprotease Glu-C was added to the crystallization drop. Endoprotease Glu-C cleaves

Figure 1
Stereoview of the N-terminal domain of LegC3. (*a*) Cartoon representation of the molecule. The four-helix bundle is at the top of the coiled-coil and extends through almost the entire range of the molecule. The molecule is coloured progressively from blue at the N-terminus to red at the C-terminus. The helical bundle is composed of two helices from the N-terminus and two from the C-terminus, and the coiled-coil constitutes an insertion between them. (*b*) Amino-acid sequence of LegC3. The expressed LegC3^{N} contained residues 2–367. The light-shaded boxes show residues that are not observed in the crystal structure. The α-helices are boxed. The predicted transmembrane helices (TMs) are enclosed in dark-shaded boxes. The C-terminal domain starts at Ala425.

the peptide bond after a glutamic acid or aspartic acid residue. As the last C-terminal residue visible in the electron-density map is Phe299, which is within the glutamate-rich sequence Glu298-Phe-Glu-His-Glu-Val, the protease evidently removed the C-terminal segment His301–His367 or Val303–His367 of the expressed construct. The crystallized protein therefore contains residues Ile2–Glu300 or Ile2–Glu302, with the residues beyond Phe299 being disordered. There is also a missing segment in the electron density that encompasses residues Glu131–Ser154, which are predicted to form a loop. The beginning of the missing segment contains several acidic residues (Glu131, Asp137 and Asp143) and it is likely that endoprotease Glu-C introduced a nick within this flexible loop after one of these residues.



(*c*)



(*d*)

Figure 1 (continued)
(*c*) The leucines and isoleucines (shown in stick representation) contribute to the packing of helices within the coiled-coil. (*d*) Representative $2F_o - F_c$ electron density contoured at the 1σ level around a segment of the coiled-coil encompassing residues 95–106 and 180–191. Leucines are shown in yellow and isoleucines in wheat.

3.1. The N-terminal domain of LegC3 is a coiled-coil protein

The structure of LegC3^N is extremely elongated, with dimensions of 150 × 33 × 20 Å, and contains six α-helices and connecting loops (Figs. 1*a* and 1*b*). It can be divided into two parts: a four-helical bundle at one end of the molecule and a long antiparallel coiled-coil. The coiled-coil segment is based on a 120 Å long helix α3 with 22 turns and two shorter helices, the 15-turn α4 and the eight-turn α5, running in the opposite direction. A 24-residue segment connecting helices α3 and α4 is disordered in the crystal. Helices α3 and α4 associate with a left-handed superhelical twist, while α3 and α5 form a right-handed superhelical twist. Several clusters of Leu and Ile residues are found within the coiled-coil helical interface; in particular, the C-terminal half of α3 has Leu/Ile in nearly every turn at positions directed towards α4 (Fig. 1*c*).

The four-helix antiparallel bundle at one end of the molecule consists of shorter helices with 2–4 turns. Three of them, α2, α6 and α7, form a triangle with each helix contacting the other two. They contact each other predominantly through Leu and Ile side chains. The fourth, N-terminal, helix α1 contacts only helix α6 of this bundle, again through side chains of Leu, Ile and Phe. This helix also makes contacts with the tips of helices α3 and α5 (again through side chains of Leu and Ile) and stabilizes and rigidifies the 'neck' between the helical bundle and the coiled-coil.

A search for proteins with a similar fold conducted with *DALI* (Holm & Sander,

1995) and *PDBeFold* (http://pdbe.org/fold/) found no structural homologues to the entire LegC3^{N}. The closest structural relative of the coiled-coil consisting of helices α3, α4 and α5 is the cytolytic α-helical toxin cytolysin A (ClyA; PDB entry 1qoy) from *Staphylococcus aureus* in its pore-forming conformation (Mueller *et al.*, 2009); α3 and α4 of LegC3 overlap well with αC and αD of ClyA, while α5 is near the position of the C-terminal third of αB (Fig. 2*a*). These helices in ClyA extend away from the membrane and participate in side-by-side contacts that create a cylindrical structure consisting of 12 protomers (Mueller *et al.*, 2009). The closest structural neighbour of the four-helix bundle is the C-terminal domain (amino acids 1156–1251) of the DNA-repair protein Rev1 (PDB entry 4gk0; Xie *et al.*, 2012; Fig. 2*b*). LegC3^{N} has a long insertion (coiled-coil) within the loop between helices α2 and α3 of the Rev1 bundle.

Although the sequence of LegC3^{N} was recognized as containing a coiled-coil, the predicted models are incorrect. For example, the model generated by *ModBase* (Pieper *et al.*, 2006) is based on a molecular motor (PDB entry 1g8x; Kliche *et al.*, 2001) which, while containing α-helices, matches neither the secondary structure nor the arrangement of helices (Fig. 2*c*).

3.2. Functional implications

Previous investigations have shown that although the N-terminal segment of LegC3 comprising residues 1–370 does not contain the transmembrane helices, it partially associates with isolated yeast vacuoles and prevents vacuole fusion (Bennett *et al.*, 2013). The prediction that LegC3 contains coiled-coils suggested the possibility of its interaction with SNARE proteins, leading to abrogation of fusion. However, *in vitro* experiments showed no such association and suggested that the disruption of vacuolar fusion involves a different mechanism (Bennett *et al.*, 2013). SNARE interaction involves four partners, each contributing a long helical segment, which associate into a parallel coiled-coil with a right-handed twist. The LegC3^{N} coiled-coil contains an antiparallel coiled-coil and therefore would not mimic two of the SNARE compo-

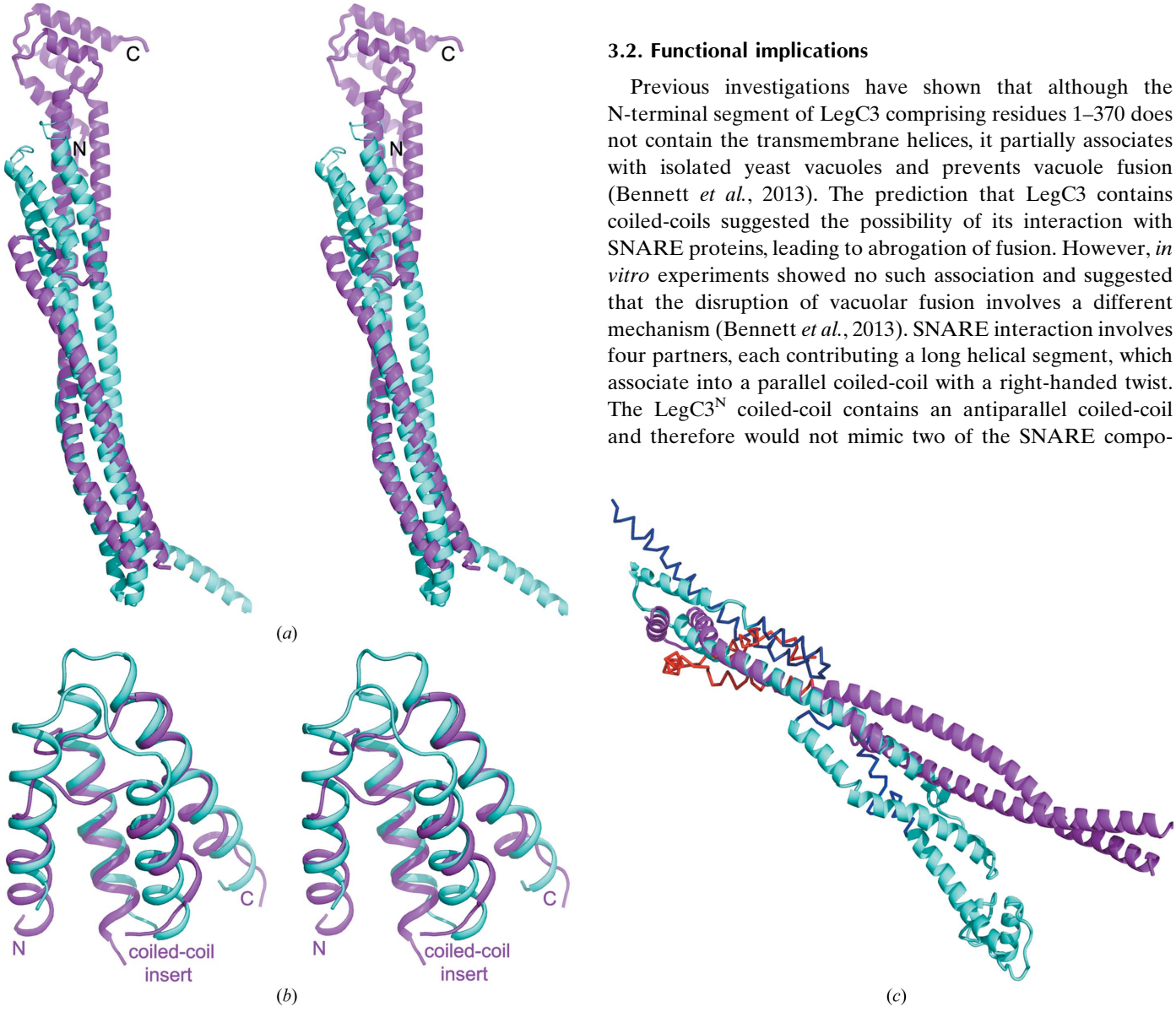


Figure 2
The closest structural neighbours of LegC3^{N}. (*a*) Stereoview of the superposition of the coiled-coil of LegC3^{N} (magenta) and the protomer of the *S. aureus* cytolytic α-helical toxin cytolysin A (ClyA; cyan). The structures overlap only partially and the third helix of ClyA is not present in LegC3. The ClyA protomers are arranged side-by-side into 12-mers forming long cylinders. LegC3^{N} is monomeric in solution and in the crystal. (*b*) Superposition of the four-helix bundle of LegC3^{N} (magenta) with the C-terminal domain of the DNA-repair protein Rev1 (cyan). The coiled-coil domain of LegC3 is inserted between helices α2 and α3 of Rev1. (*c*) Superposition of LegC3^{N} (magenta) and the model generated by *ModBase* (cyan). The predicted model corresponds to residues 61–362, while the experimental structure contains residues 3–127 and 157–298. The modelled residues 299–362 that are not present in the crystal structure are coloured dark blue and the loop 128–156, which is disordered in the experimental structure, is coloured lighter blue; the region 3–60 of the experimental structure that is not present in the model is coloured red.

nents in the complex, in accord with the observed lack of such an interaction. The structure of LegC3^{N} suggests another possible mechanism. The helical bundle top portion of the molecule contains the C-terminal helix of LegC3^{N} and is connected by a ∼70-residue segment to the predicted transmembrane helices. Secondary-structure predictions indicate that this connecting segment contains several α-helices and places the helical bundle in proximity to the membrane. This domain is rigidly attached to the coiled-coil region, and we speculate that the elongated LegC3^{N} molecule extends away from the membrane. It has recently been postulated that the large number of disordered segments that are predicted and observed in bacterial effectors play a role in evasion of the host innate immune system and provide a mechanism of host-protein mimicry by folding in the presence of target protein (Marín *et al.*, 2013). We speculate that the long disordered segment at the opposite end of the coiled-coil may play a role in the inhibitory function of LegC3 by interacting with another protein that remains to be identified. Moreover, the similarity of the helical bundle to the Rev1 C-terminal domain suggests that, like Rev1, this domain is involved in protein–protein interactions. An examination of the effect of various parts of LegC3^{N} on vacuolar fusion is in progress.

This work was supported by grant GSP-48370 from the Canadian Institutes of Health Research (to MC). Research described in this paper was performed using beamline 08ID-1 at the Canadian Light Source, which is supported by the Natural Sciences and Engineering Research Council of Canada, the National Research Council Canada, the Canadian Institutes of Health Research, the Province of Saskatchewan, Western Economic Diversification Canada and the University of Saskatchewan.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Bennett, T. L., Kraft, S. M., Reaves, B. J., Mima, J., O'Brien, K. M. & Starai, V. J. (2013). *PLoS One*, **8**, e56798.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Felipe, K. S. de, Glover, R. T., Charpentier, X., Anderson, O. R., Reyes, M., Pericone, C. D. & Shuman, H. A. (2008). *PLoS Pathog.* **4**, e1000117.
Gomez-Valero, L., Rusniok, C., Jarraud, S., Vacherie, B., Rouy, Z., Barbe, V., Medigue, C., Etienne, J. & Buchrieser, C. (2011). *BMC Genomics*, **12**, 536.
Grochulski, P., Fodje, M. N., Gorin, J., Labiuk, S. L. & Berg, R. (2011). *J. Synchrotron Rad.* **18**, 681–684.
He, Y., Yao, D.-Q., Gu, Y.-X., Lin, Z.-J., Zheng, C.-D. & Fan, H.-F. (2007). *Acta Cryst.* D**63**, 793–799.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Hubber, A. & Roy, C. R. (2010). *Annu. Rev. Cell Dev. Biol.* **26**, 261–283.
Isberg, R. R., O'Connor, T. J. & Heidtman, M. (2009). *Nature Rev. Microbiol.* **7**, 13–24.
Kliche, W., Fujita-Becker, S., Kollmar, M., Manstein, D. J. & Kull, F. J. (2001). *EMBO J.* **20**, 40–46.
Marín, M., Uversky, V. N. & Ott, T. (2013). *Plant Cell*, **25**, 3153–3157.
Mueller, M., Grauschopf, U., Maier, T., Glockshuber, R. & Ban, N. (2009). *Nature (London)*, **459**, 726–730.
Pieper, U., Eswar, N., Davis, F. P., Braberg, H., Madhusudhan, M. S., Rossi, A., Marti-Renom, M., Karchin, R., Webb, B. M., Eramian, D., Shen, M. Y., Kelly, L., Melo, F. & Sali, A. (2006). *Nucleic Acids Res.* **34**, D291–D295.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Vogel, J. P. & Isberg, R. R. (1999). *Curr. Opin. Microbiol.* **2**, 30–34.
Xie, W., Yang, X., Xu, M. & Jiang, T. (2012). *Protein Cell*, **3**, 864–874.