In situ proteolysis, crystallization and preliminary X-ray diffraction analysis of a V_HH that binds *Listeria internalin B*

Ian Huh,^a Robert Gene,^{b,c} Jyothi Kumaran,^{b,c} C. Roger MacKenzie^{b,c} and Cory L. Brooks^{a*}

^aDepartment of Chemistry, California State University Fresno, 2555 E. San Ramon Avenue, Fresno, CA 93740, USA, ^bHuman Health Therapeutics, National Research Council of Canada, 100 Sussex Drive, Ottawa, ON K1A 0R6, Canada, and ^cSchool of Environment Sciences, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada

Correspondence e-mail: cbrooks@csufresno.edu

Received 2 July 2014

Accepted 5 September 2014

The variable region of camelid heavy-chain antibodies produces the smallest known antibody fragment with antigen-binding capability (a V_HH). The V_HH R303 binds internalin B (InlB), a virulence factor expressed by the pathogen *Listeria monocytogenes*. InlB is critical for initiation of *Listeria* infection, as it binds a receptor (c-Met) on epithelial cells, triggering the entry of bacteria into host cells. InlB is surface-exposed and is required for virulence, hence a V_HH targeting InlB has potential applications for pathogen detection or therapeutic intervention. Here, the expression, purification, crystallization and X-ray diffraction of R303 are reported. Crystals of R303 were obtained following *in situ* proteolysis with trypsin. Gel filtration and SDS-PAGE revealed that trypsin removed the C-terminal tag region of R303, facilitating crystal formation. Crystals of R303 diffracted to 1.3 Å resolution and belonged to the monoclinic space group *P*2₁, with unit-cell parameters *a* = 46.4, *b* = 31.2, *c* = 74.8 Å, *β* = 93.8°. The crystals exhibited a Matthews coefficient of 1.95 Å³ Da⁻¹ with two molecules in the asymmetric unit.

1. Introduction

Listeriosis is a food-borne disease caused by the Gram-positive bacterium *Listeria monocytogenes*. Clinical features of listeriosis can include severe gastroenteritis, mother–fetus infection and central nervous system infections resulting in potentially fatal meningitis (Dussurget *et al.*, 2004). Invasion of epithelial cells and cell-to-cell spreading is mediated by the internalin family of virulence factors (Dramsi *et al.*, 1995). Internalin A (InlA) and internalin B (InlB) are primarily responsible for bacterial entry into host cells *via* the binding of cell-surface receptors, triggering receptor-mediated endocytosis and internalization of bacteria (Cossart *et al.*, 2003). The InlB receptor is the hepatocyte growth factor receptor c-Met (Dussurget *et al.*, 2004). InlB is a multi-domain protein consisting of 630 amino acids (67 kDa; Dramsi *et al.*, 1995). The domain structure includes a cap domain, a leucine-rich region (LRR) domain, an inter-repeat (IR) region which stabilizes the LRR by shielding hydrophobic residues from solvent and a Gly-Trp (GW) domain that anchors the protein to the cell wall (Marino *et al.*, 2002). The LRR domain is critical for *Listeria* invasion and is the key domain mediating interaction of InlB with the c-Met receptor (Shen *et al.*, 2000; Niemann *et al.*, 2007). The LRR domain consists of repeating segments of 3₁₀-helix and *β*-strand, resulting in a concave structure (Schubert *et al.*, 2001). Given the importance of the InlB LRR domain in *Listeria* virulence, interference of the bacterial interaction with the c-Met receptor may offer a potential route to inhibit bacterial invasion.

As InlB is surface-exposed and important for virulence, it represents an attractive target for antibody-mediated detection or therapeutic intervention. Heavy-chain single-domain antibodies (V_HH) targeting InlB provide an innovative solution for *Listeria* detection. The V_HH derived from the camelid immune systems are emerging as powerful biotechnological tools for a variety of applications, including pathogen detection and therapeutic functions (Vaneycken *et al.*, 2011). With molecular weights ranging from 12 to 15 kDa, they are tenfold smaller than conventional antibodies (150 kDa) and have a prolate shape, allowing them to insert into cavities on the surface of protein antigens (Muyldermans, 2013). Despite lacking a light chain, V_HH remain soluble and stable owing to the substitution of hydro-

Table 1

Macromolecule-production information.

Source organism	<i>Camelus dromedarius</i>
DNA source	Naive phage display library (Kumaran <i>et al.</i> , 2012)
Forward primer	CAGGTAAAGCTGGAGGAGTCTG
Reverse primer	GTTTCAGATCTTCTCGCTGATCA
Cloning vector	pSJF2H
Expression vector	pET-22b
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced†	QVKLEESGGGSVQAGGSLRLSCAASGHTYSTYCMGWFRQVPG-KEREGVARINVGGSSTWYADSVDRFTISQDNAKNTVYLQ-MNSLKLEDTAIYYCTLHRFCNTWSLGLTNVWGQGTQVTVSSGSEQ <u>KLISEEDLN</u> HHHHH

† The c-Myc tag and His tag are in bold; the underlined lysine is the likely trypsin cleavage site.

phobic amino acids that would have participated in V_H - V_L domain-pairing interactions with smaller hydrophilic amino acids (Muyldermans, 2013). V_H H possess complementarity-determining region (CDR) loops that are longer and more flexible than conventional antibodies, creating a large antigen-interaction surface. These unusually long loops facilitate binding of antigens with concave shapes and can penetrate surface cavities such as enzyme active sites (De Meyer *et al.*, 2014). This range of properties makes V_H H ideally suited for targeting the LRR domain of the *Listeria* virulence factor InlB.

Here, we report the expression, purification and crystallization of R303, a V_H H that binds InlB with high affinity. Difficulties in obtaining diffraction-quality crystals were overcome by using *in situ* proteolysis. The structure may be useful for future co-crystallization or docking studies with InlB LRR to dissect the nature of the interaction between R303 and InlB.

2. Materials and methods

2.1. Macromolecule production

The InlB-specific V_H H antibody fragment R303 was obtained by panning a V_H H phage display library derived from the non-immune repertoires of llamas, alpacas and camels against the *L. monocytogenes* virulence factor internalin B (Kumaran *et al.*, 2012; Gene, 2012). For expression of soluble R303, the gene was subcloned using the Clone EZ PCR cloning kit (GenScript, USA) into the expression vector pET-22b (EMD Millipore, USA) that contains a pelB secretion signal for periplasmic expression (Table 1). For purification of R303, the expression plasmid was transformed into *Escherichia coli* BL21 (DE3) and an overnight culture was used to inoculate 6 l 2×YT medium. The culture was grown until mid-log phase (37°C, 225 rev min⁻¹, OD₆₀₀ = 0.5) and induced with 0.4 mM IPTG. Fermentation was carried out for 1 h at 30°C and 225 rev min⁻¹. For periplasmic extraction of R303, the cells were harvested by centrifugation (5000g, 4°C, 10 min) and resuspended in ice-cold TES buffer (0.2 M Tris pH 8.0, 0.5 M sucrose, 0.5 mM EDTA). The cells were mixed on ice for 30 min and an equal volume of ice-cold water was added to the resuspended cells and mixed for a further 30 min. The periplasmic fraction was recovered by centrifugation (20 000g, 4°C, 30 min) and dialyzed overnight against 50 mM Tris pH 8.0, 0.3 M NaCl. R303 was purified using IMAC (immobilized metal-ion affinity chromatography), the dialyzed periplasmic fraction was batch bound to 1 ml of His-Pur Ni-NTA resin (Thermo Scientific, USA) for 1 h at 4°C. Unbound protein was removed by washing with 35 column volumes of wash buffer (50 mM Tris pH 8.0, 0.3 M NaCl, 10 mM imidazole). R303 was eluted using a step gradient consisting of 0.1,

0.25, 0.5 and 1 M imidazole. Protein purity was assessed by SDS-PAGE.

Figure 1
Purification and characterization of V_H H R303. (a) Highly purified R303 from nickel-affinity chromatography. The three lanes correspond to imidazole concentrations: (i) 0.25 M, (ii) 0.5 M, (iii) 1 M. The left lane contains molecular-weight marker (labelled in kDa). (b) Analytical gel filtration of R303 (red solid line) results in a retention volume of 12.3 ml with a shoulder peak at 11.9 ml and trypsin-treated R303 (blue, dashed line) results in a single monodisperse peak at 11.9 ml. (c) SDS-PAGE of R303 before trypsin treatment (i) and after trypsin treatment (ii).

Table 2

Crystallization conditions.

Method	Sitting-drop vapour diffusion
Plate type	96-well plate
Temperature (°C)	20
Protein composition (mg ml ⁻¹)	12
Buffer composition of protein solution	10 mM HEPES pH 7.5
Composition of reservoir solution	0.1 M MES pH 6.5, 25% PEG 4K
Volume and ratio of drop	1 µl, 1:1 protein:reservoir
Volume of reservoir (µl)	50

2.2. Analysis of *in situ* proteolysis

The monodispersity of R303 before and after limited proteolysis with trypsin [1:100(w:w) trypsin:V_HH, 1 h] was examined using an Enrich SEC 70 10 × 300 gel-filtration column (Bio-Rad, USA) on an NGC Quest chromatography system (Bio-Rad, USA). The column was equilibrated with 20 mM Tris pH 8.0, 0.15 M NaCl. Untreated and trypsin-treated R303 (200 µl at 1 mg ml⁻¹) was injected onto the column at a flow rate of 1 ml min⁻¹. The retention volumes of the peaks were determined using the *NGC ChromLab* software (Bio-Rad, USA). Aliquots from each peak were analyzed with a 14% SDS-PAGE gel. Gels were imaged using the ChemiDoc MP imaging system (Bio-Rad, USA) and molecular-weight analysis of the gels was carried out using the accompanying *ImageLab* software (Bio-Rad, USA).

2.3. Crystallization

R303 was dialyzed into 10 mM HEPES pH 7.5 and concentrated to 12 mg ml⁻¹ for crystallization trials. Initial screening of R303 was conducted using 0.5 µl protein solution with 0.5 µl mother liquor in 96-well Intelli-Plates (Art Robbins Instruments, USA) using the commercial screens The PEGs, PEGs II and JCSG+ Suites (Qiagen, USA). Initial screening failed to yield diffraction-quality crystals. The protein was then rescreened using *in situ* proteolysis by combining R303 with trypsin [100:1(w:w)] immediately prior to crystallization trials. This approach yielded diffraction-quality crystals using The PEGs Suite condition 34 (0.1 M MES pH 6.5, 25% PEG 4K) at 20°C (Table 2).

2.4. Data collection and processing

Crystals of R303 were soaked in cryoprotectant (mother liquor supplemented with 25% glycerol) and flash-cooled in liquid nitrogen. X-ray data were collected on beamline 08ID-1 at the Canadian Light Source equipped with a Rayonix MX300 CCD X-ray detector (Grochulski *et al.*, 2011). The beamline was controlled using the *MaxDC* software (Fodje *et al.*, 2012) and the data were processed using *xia2* (Winter *et al.*, 2013) (Table 3). Molecular-replacement trials were carried out with *Phaser* (McCoy *et al.*, 2007) as implemented in *PHENIX* (Adams *et al.*, 2010).

3. Results and discussion

R303 was purified to homogeneity from the *E. coli* periplasmic fraction using a single nickel-affinity chromatography step (Fig. 1a). However, despite a high level of protein purity, initial crystallization trials using three commercial screens failed to produce crystals. Recombinant R303 contained a C-terminal polyhistidine tag preceded by a c-Myc tag. We postulated that this relatively large and potentially unstructured region of the protein was impeding proper crystal formation. *In situ* proteolysis is a well recognized approach for removing unstructured regions to facilitate protein crystallization.

Table 3

Data-collection statistics for R303.

Values in parentheses are for the highest resolution shell.

Diffraction source	O8ID-1, Canadian Light Source
Wavelength (Å)	0.98
Temperature (°C)	−178
Detector	Rayonix MX300 CCD
Crystal-to-detector distance (mm)	162
Rotation range per image (°)	1
Total rotation range (°)	141
Exposure time per image (s)	0.5
Space group	<i>P</i> 2 ₁
<i>a</i> , <i>b</i> , <i>c</i> (Å)	46.4, 31.2, 74.8
α , β , γ (°)	90, 93.8, 90
Mosaicity (°)	0.53
Resolution range (Å)	40.59–1.30 (1.33–1.30)
Total No. of reflections	157369
No. of unique reflections	52559
Completeness (%)	99.1 (95.4)
Multiplicity	3.0 (2.4)
$\langle I/\sigma(I) \rangle$	10.4 (2.5)
<i>R</i> _{meas} (%)	5.9 (34.9)
Overall <i>B</i> factor from Wilson plot (Å ²)	15.0

The method incorporates a limited amount of protease immediately preceding crystallization trials. The protease can remove flexible regions of the protein, thereby promoting crystal formation (Wernimont & Edwards, 2009; Dong *et al.*, 2007).

R303 was treated with trypsin (100:1 V_HH:protease ratio) immediately prior to crystallization trials. The trypsin-treated protein was rescreened using the same three commercial crystal screens, and crystals formed in The PEGs Suite (Qiagen) within 36 h. These crystals were readily reproducible in larger drop sizes (Fig. 2). To examine the effects of trypsin treatment, purified R303 was assessed by analytical gel filtration before and after protease exposure. Week-old R303 was injected onto the column and displayed an asymmetric peak (retention volume of 11.9 ml) with a distinct shoulder region at 12.3 ml (Fig. 1b). In contrast, trypsin-treated R303 exhibited a symmetrical peak with a shifted retention volume of 12.29 ml (Fig. 1b). SDS-PAGE analysis of untreated R303 from the gel-filtration peak revealed two bands at 16 and 13 kDa (Fig. 1c). This was surprising as freshly purified R303 migrated as a single band on SDS-PAGE (Fig. 1a). The predicted mass of R303 with C-terminal c-Myc and His tags is 15.77 kDa, while the predicted mass of R303 without its tag is 13.3 kDa. Thus, R303 appears to be susceptible to slow proteolysis during storage at 4°C, which is likely to result in a partial

Figure 2

Crystals of R303 following *in situ* proteolysis with trypsin. Crystals were grown in mother liquor consisting of 0.1 M MES pH 6.5, 25% PEG 4K at 20°C.

Figure 3

X-ray diffraction pattern from R303 crystal collected at the Canadian Light Source. The outer black circle corresponds to a resolution of 1.34 Å.

loss of the c-Myc and His tags. The heterogeneity resulting from this mixture of tagged and untagged R303 may explain the initial difficulty in obtaining crystals. Gel filtration of trypsin-treated R303 resulted in a single monodisperse peak corresponding to a single protein band on SDS-PAGE at 13.2 kDa (Figs. 1b and 1c). The C-terminus contains a single lysine residue that is likely to represent the trypsin cleavage site resulting in removal of the c-Myc and His tags (Table 1). The combination of removing the flexible tag and increasing the homogeneity of the sample were crucial in obtaining well diffracting crystals of R303.

The crystals of trypsin-treated R303 belonged to the monoclinic space group $P2_1$ and diffracted to 1.3 Å resolution on beamline 08ID-1 at the Canadian Light Source (Fig. 3, Table 3). A summary of the data-collection statistics is provided in Table 3. Initial molecular-replacement trials were carried out with *Phaser* using the crystal structure of a llama V_HH as a model (Park *et al.*, 2012; PDB entry 3stb; 71% sequence identity). A molecular-replacement solution was found with two copies in the asymmetric unit (RFZ = 7.1, TFZ = 8.1, LLG = 178.02) with a Matthews coefficient (V_M) of $1.95 \text{ Å}^3 \text{ Da}^{-1}$ (solvent content of 37%; Matthews, 1968). Modelling and refinement of the structure are ongoing. The modelled structure by itself may be useful for docking studies with the previously solved structure of *Listeria* InlB LRR (Marino *et al.*, 1999; Niemann *et al.*, 2007). Future work will focus on determining a co-crystal structure of R303 in

complex with InlB LRR. The structure combined with docking studies or a co-structure with InlB will provide valuable insight into the nature of the interaction between R303 and the *Listeria* virulence factor InlB.

This research was supported by a new investigator award from CSUPERB and start-up funds to CLB from the College of Science and Mathematics, California State University Fresno. We thank Teresa and Penny Brooks for helpful discussions. X-ray data collection 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. This work was also supported by a research grant from the Natural Sciences and Engineering Council of Canada (NSERC) and through the network grant SENTINEL (Bioactive paper Network).

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
 Cossart, P., Pizarro-Cerdá, J. & Lecuit, M. (2003). *Trends Cell Biol.* **13**, 23–31.
 De Meyer, T., Muyldermans, S. & Depicker, A. (2014). *Trends Biotechnol.* **32**, 263–270.
 Dong, A. *et al.* (2007). *Nature Methods*, **4**, 1019–1021.
 Dramsi, S., Biswas, I., Maguin, E., Braun, L., Mastroeni, P. & Cossart, P. (1995). *Mol. Microbiol.* **16**, 251–261.
 Dussurget, O., Pizarro-Cerdá, J. & Cossart, P. (2004). *Annu. Rev. Microbiol.* **58**, 587–610.
 Fodje, M., Janzen, K., Berg, R., Black, G., Labiuk, S., Gorin, J. & Grochulski, P. (2012). *J. Synchrotron Rad.* **19**, 274–280.
 Gene, R. (2012). MSc thesis, University of Guelph, Canada.
 Grochulski, P., Fodje, M. N., Gorin, J., Labiuk, S. L. & Berg, R. (2011). *J. Synchrotron Rad.* **18**, 681–684.
 Kumaran, J., MacKenzie, C. R. & Arbabi-Ghahroudi, M. (2012). *Single Domain Antibodies*, edited by D. Saerens & S. Muyldermans, pp. 105–124. New York: Springer.
 Marino, M., Banerjee, M., Jonquière, R., Cossart, P. & Ghosh, P. (2002). *EMBO J.* **21**, 5623–5634.
 Marino, M., Braun, L., Cossart, P. & Ghosh, P. (1999). *Mol. Cell*, **4**, 1063–1072.
 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.
 Muyldermans, S. (2013). *Annu. Rev. Biochem.* **82**, 775–797.
 Niemann, H. H., Jäger, V., Butler, P. J. G., van den Heuvel, J., Schmidt, S., Ferraris, D., Gherardi, E. & Heinz, D. W. (2007). *Cell*, **130**, 235–246.
 Park, Y.-J., Pardon, E., Wu, M., Steyaert, J. & Hol, W. G. J. (2012). *Nucleic Acids Res.* **40**, 1828–1840.
 Schubert, W. D., Göbel, G., Diepholz, M., Darji, A., Kloer, D., Hain, T., Chakraborty, T., Wehland, J., Domann, E. & Heinz, D. W. (2001). *J. Mol. Biol.* **312**, 783–794.
 Shen, Y., Naujokas, M., Park, M. & Ireton, K. (2000). *Cell*, **103**, 501–510.
 Vaneycken, I., D'huyvetter, M., Hernot, S., De Vos, J., Xavier, C., Devoogdt, N., Cavelliers, V. & Lahoutte, T. (2011). *Curr. Opin. Biotechnol.* **22**, 877–881.
 Wernimont, A. & Edwards, A. (2009). *PLoS One*, **4**, e5094.
 Winter, G., Lobley, C. M. C. & Prince, S. M. (2013). *Acta Cryst.* **D69**, 1260–1273.