Purification, crystallization, crystallographic analysis and phasing of the CRISPR-associated protein Csm2 from *Thermotoga maritima*

Gloria Gallo,^a Gilles Augusto,^a Giulliana Rangel,^a André Zelanis,^a Marcelo A. Mori,^b Cláudia Barbosa Campos^a and Martin Würtele^{a*}

Received 22 June 2015
Accepted 6 August 2015

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

Keywords: RNA-guided interference; CRISPR–Cas; Csm2; *Thermotoga maritima*.

^aDepartment of Science and Technology, Federal University of São Paulo, Rua Talim 330, 12231-280 São José dos Campos-SP, Brazil, and ^bDepartment of Biophysics, Federal University of São Paulo, Rua Botucatu 862, 04023-062 São Paulo-SP, Brazil. *Correspondence e-mail: martin.wurtele@unifesp.br

The clusters of regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated proteins (Cas) system consists of an intriguing machinery of proteins that confer bacteria and archaea with immunity against phages and plasmids *via* an RNA-guided interference mechanism. Here, the cloning, recombinant expression in *Escherichia coli* BL21 (DE3), purification, crystallization and preliminary X-ray diffraction analysis of Csm2 from *Thermotoga maritima* are reported. Csm2 is thought to be a component of an important protein complex of the type IIIA CRISPR–Cas system, which is involved in the CRISPR–Cas RNA-guided interference pathway. The structure of Csm2 was solved *via* cadmium single-wavelength anomalous diffraction (Cd-SAD) phasing. Owing to its involvement in the CRISPR–Cas system, the crystal structure of this protein could be of importance in elucidating the mechanism of type IIIA CRISPR–Cas systems in bacteria and archaea.

1. Introduction

Recent reports have characterized the bacterial clusters of regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated proteins (Cas) system as a prokaryotic adaptive immune system based on RNA-guided interference (Mojica *et al.*, 2005; Bolotin *et al.*, 2005; Pourcel *et al.*, 2005; Makarova *et al.*, 2006; Barrangou *et al.*, 2007). The CRISPR–Cas system is effective in protecting cells against nucleotide uptake from bacteriophages and plasmids (Barrangou *et al.*, 2007; Marraffini & Sontheimer, 2008). In summary, foreign DNA is integrated into the CRISPR elements of the prokaryotic host genome (Pourcel *et al.*, 2005; Bolotin *et al.*, 2005; Barrangou *et al.*, 2007) and transcripts of this DNA in the form of processed crRNA (CRISPR RNA) are involved in an RNA-guided interference mechanism to degrade target DNA or RNA (Brouns *et al.*, 2008; Wiedenheft *et al.*, 2009; Hale *et al.*, 2009; Marraffini & Sontheimer, 2010).

Depending on the occurrence of three signature proteins, Cas3, Cas9 and Cas10, CRISPR–Cas systems are classified as type I, type II and type III systems, respectively (Makarova, Aravind *et al.*, 2011; Makarova, Haft *et al.*, 2011). Type III CRISPR–Cas systems are further subdivided into two different subtypes called IIIA and IIIB. Whereas type I and type II CRISPR–Cas systems target DNA (Garneau *et al.*, 2010; Sinkunas *et al.*, 2011; Gasiunas *et al.*, 2012), type III systems are thought to target both DNA and RNA (Marraffini & Sontheimer, 2008; Hale *et al.*, 2009; Staals *et al.*, 2013, 2014;

© 2015 International Union of Crystallography

Table 1

Macromolecule-production information.

Source organism	<i>T. maritima</i>
DNA source	<i>T. maritima</i>
Forward primer (BamHI restriction site)	5'-GGCCGGGATCCGAGTTTCTCAGGGTGTTC
Reverse primer (HindIII restriction site)	5'-GGCCGGAAGCTTATCTTCTGGTGTCTTCTGTTG
Cloning vector	pQtev
Expression vector	pQtev
Expression host	<i>E. coli</i> BL21 (DE3)
Complete amino-acid sequence of the construct produced	GSAVSQGVSLKEDLVRKAAEEIGRELSGKLKT-NQLRKFGHGLTKIWSNYIYKKDYRDNPEKFN-EEILNELHFMKIFLAYQVGRDIEGISELKEIL-EPLIDEIKTPDEFKFKFYDAILAYHKFHSE-SEKSNRRRTARR

Taylor *et al.*, 2015; Samai *et al.*, 2015). The type IIIA system includes Csm Cas proteins (Makarova, Haft *et al.*, 2011), which were identified as being expressed in *Mycobacterium tuberculosis* (Haft *et al.*, 2005). The type IIIB system includes Cmr proteins (Hale *et al.*, 2009). Processing and target recognition by crRNA is carried out by several Cas proteins, which usually form a complex of ribonucleoproteins (RNPs) like the so-called Cascade complex in type I systems (Brouns *et al.*, 2008) and the Cmr CRISPR RNP (crRNP) complex in type IIIB systems (Hale *et al.*, 2012; Staals *et al.*, 2013). Electron-microscopy data and protein crystallographic studies have revealed the composition and function of the Cascade (Jore *et al.*, 2011; Wiedenheft *et al.*, 2011; Jackson *et al.*, 2014) and the Cmr crRNP complexes (Staals *et al.*, 2013; Osawa *et al.*, 2015) in great detail. Furthermore, electron-microscopy and mass-spectrometric data have defined the Csm crRNP complex from *Sulfolobus solfataricus*, which is formed by up to seven different proteins (Rouillon *et al.*, 2013). Similar experiments have defined a homologous complex in *Thermus thermophilus* that consists of five different proteins: Csm1–Csm5 (Staals *et al.*, 2014). Cmr, Csm and Cascade complexes show overall structural similarity, possibly owing to the fact that they are composed of structurally related subunits (Rouillon *et al.*, 2013). To obtain further insight into the Csm crRNP complex, we have cloned, expressed, purified, crystallized and solved the structure of the Csm2 protein from *Thermotoga maritima* MSB8.

2. Materials and methods

2.1. Cloning and expression

Full-length Csm2 (GenBank entry AKE29563.1) was amplified *via* PCR with primer overhangs from total genomic extracts of *T. maritima* MSB8 (Huber *et al.*, 1986) obtained from DSMZ (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany). DNA oligonucleotides (Exxtend, Paulínia, Brazil) are described in Table 1. PCR fragments were cloned into the pQtev His-tag expression vector (Protein Structure Factory, Berlin, Germany) using BamHI and HindIII cloning sites. The construct was verified by sequencing (Proteobras, Paulínia, Brazil). Csm2 was expressed in LB medium for 6 h at 37°C in *Escherichia coli*

Table 2

Crystallization.

Method	Hanging drop
Plate type	24-well VDX plate
Temperature (K)	293.15
Protein concentration (mg ml ⁻¹)	15
Buffer composition of protein solution	25 mM Tris–HCl pH 7.5, 100 mM NaCl, 3 mM DTT, 10% glycerol
Composition of reservoir solution	100 mM sodium acetate pH 4.6, 100 mM CdCl ₂ , 21% PEG 400
Volume and ratio of drop	4 µl, 1:1 ratio
Volume of reservoir (ml)	1

strain BL21 (DE3) after induction with 1 mM β -D-1-thiogalactopyranoside (IPTG).

2.2. Protein purification

The cells were harvested by centrifugation and resuspended in lysis buffer consisting of 50 mM Tris–HCl pH 7.5, 400 mM NaCl, 10 mM imidazole, 1 mM PMSF, 5 mM β -mercaptoethanol. The cells were lysed using an M-110L high-pressure homogenizer (Microfluidics, Westwood, USA). After separation of the soluble fraction by centrifugation, the protein was purified by affinity chromatography using a 5 ml HisTrap Sepharose column (GE Healthcare) on an ÄKTAprime Plus liquid-chromatography system (GE Healthcare). After elution with imidazole and dialysis in 50 mM Tris–HCl pH 7.5, 400 mM NaCl, 3 mM DTT, 10% glycerol, the His tag was cleaved using TEV (*Tobacco etch virus*) protease (Blommel & Fox, 2007). After a subsequent affinity-chromatography step to remove the protease and the His tag, Csm2 was concentrated using Amicon Ultra-15 centrifugal filters (Millipore) and submitted to gel-filtration chromatography on a HiLoad 26/600 Superdex 75 prep-grade column with a nominal volume of 320 ml (GE Healthcare) in buffer consisting of 25 mM Tris–HCl pH 7.5, 100 mM NaCl, 3 mM DTT, 10% glycerol. Finally, the protein was concentrated to 15 mg ml⁻¹. The molecular mass of native Csm2 (650 ng) was confirmed by electrospray-ionization (ESI) mass spectrometry using a hybrid quadrupole (Q)-IM-ToF MS instrument (Synapt G2 HDMS mass spectrometer, Waters). Deconvolution of the MS protein spectrum yielded an isotope-averaged molecular mass of 16 732.78 Da, which is similar to the value obtained by SDS–PAGE using SeeBlue Plus2 Pre-stained protein standard as a marker (Invitrogen, Life Technologies).

2.3. Crystallization

Csm2 was crystallized by the hanging-drop method in 24-well plates after screening trials using Crystal Screen 2 (Hampton Research, Aliso Viejo, USA) at 20°C. Initial screening conditions were optimized, leading to optimal crystallization conditions consisting of 100 mM cadmium chloride, 21% PEG 400, 100 mM sodium acetate buffer pH 4.6, as described in Table 2.

2.4. Data collection and processing

Diffraction data were collected on the MX-1 beamline at the Brazilian Synchrotron Light Laboratory (LNLS, Campinas,

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	MX-1, LNLS
Wavelength (Å)	1.458
Temperature (K)	100
Detector	MAR CCD 165
Crystal-to-detector distance (mm)	120
Rotation range per image (°)	1
Data range (images)	1–190
Space group	$P3_121$
a, b, c (Å)	77, 77, 160
α, β, γ (°)	90, 90, 120
Mosaicity (°)	0.2
Resolution range (Å)	50–2.9 (3.08–2.90)
Total No. of reflections	143907 (22874)
No. of unique reflections	23467 (3813)
Completeness (%)	99.9 (99.6)
Multiplicity	6.13 (5.99)
$\langle I/\sigma(I) \rangle$	13.77 (2.62)
R_{meas} (%)	11.0 (81.3)
Overall B factor from Wilson plot (Å ²)	81.2

Brazil; Polikarpov *et al.*, 1998) after flash-cooling in liquid nitrogen using crystallization buffer supplemented with 12% glycerol. A highly redundant data set was collected at a wavelength of 1.458 Å. Crystallographic data were processed with *XDS* (Kabsch, 2010) and *Adxv* (Arvai, 2015). The structure of Csm2 was solved by single-wavelength anomalous diffraction (SAD) using *Phaser* (McCoy *et al.*, 2007) within the *AutoSol* module of *PHENIX* (Adams *et al.*, 2010), using the fact that the crystallization conditions contained cadmium ions. Electron-density maps were inspected using *Coot* (Emsley *et al.*, 2010) and figures were rendered using *PyMOL* (DeLano, 2002).

Figure 1

Purification of Csm2. (a) Gel-filtration chromatography profile of Csm2 using a HiLoad 26/600 Superdex 75 prep-grade column (GE Healthcare). The last peak corresponds to the crystallized Csm2. (b) Molecular-mass determination of native Csm2 by mass spectrometry. Deconvolution of the MS protein spectrum yielded an isotope-averaged molecular mass of 16 732.78 Da. (c) SDS-PAGE of Csm2 after gel filtration using SeeBlue Plus2 Pre-stained protein standard (Invitrogen, Life Technologies) as a marker (labelled in kDa).

3. Results and discussion

In recent years, several outstanding crystal structures of RNA-processing enzymes and protein complexes have been obtained using isoforms from thermophilic bacteria (Wang *et al.*, 2008; Mulepati & Bailey, 2011; Lintner *et al.*, 2011; Cocozaki *et al.*, 2012; Osawa *et al.*, 2013). To obtain further data for the Csm CRISPR–Cas type IIIA crRNP complex, we have cloned, expressed, purified and crystallized a major protein from this complex, Csm2.

T. maritima MS8 Csm2 was expressed in soluble form in *E. coli* strain BL21 (DE3) in relatively high quantities of up to 30 mg per litre of culture. However, the protein could initially not be concentrated owing to precipitation. This problem was

(a)

(b)

Figure 2

Csm2 crystals and X-ray diffraction pattern. (a) Representative native crystals of Csm2; scale bar = 0.5 mm. (b) Diffraction image of a Csm2 crystal.

Figure 3
Electron-density maps. Stereoview of the *AutoSol* density-modified electron-density map of the Csm2 crystal structure contoured at a 1.5σ (a) and 3σ (b) level showing a cadmium ion (yellow) linked to an α -helix with a C^α trace (green).

circumvented by using glycerol (10%) and high concentrations of NaCl (up to 400 mM) during all purification steps. To improve the chance of crystallization, after purifying the protein the concentration of NaCl in the final gel-filtration step was reduced to 100 mM (Fig. 1). The protein could then be concentrated to 15 mg ml⁻¹. Small monocrystals obtained in initial screens could be improved in size through refinement of the crystallization conditions, as shown in Fig. 2.

A highly redundant data set was collected from one of these crystals on the MX-1 beamline at the Brazilian Synchrotron Light Laboratory (LNLS), Campinas, Brazil. Crystallographic data for this crystal are given in Table 3. The crystals tested diffracted to a limited resolution of 2.9 Å in space group $P3_121$ and contained an estimated three subunits in the asymmetric unit (Matthews coefficient of 2.7 Å³ Da⁻¹ and 54% solvent content). As there are no known close structural homologues of Csm2 in the PDB, the structure of Csm2 could not be solved by molecular replacement. However, as our crystals grew at high concentrations (100 mM) of CdCl₂, we hypothesized that Cd²⁺ ions possibly formed strong linkages with the numerous charged amino acids in Csm2. As several structures have been solved using Cd-SAD and/or S-SAD (Yogavel *et al.*, 2010; Robbins *et al.*, 1991; Medlock *et al.*, 2009), Cd-SAD was used together with the *PHENIX* software package (Adams *et al.*, 2010) to try to solve the structure. Indeed, the measured data set indicated the presence of an anomalous signal, quantified by an overall anomalous correlation of 31% and an overall

mean anomalous difference in units of standard deviation (SigAno in *XDS*) of 1.085. Consequently, *phenix.hyss* was able to detect six cadmium ions. Automatic refinement and phasing with *Phaser* and subsequent phase density modifications then led to the first experimental electron-density maps. This solution was characterized by an overall figure of merit of 0.319, a correlation of local r.m.s. density of 0.82 and a meaningful map skew of 0.10.

Most importantly, the experimental electron-density maps after density modification promptly showed meaningful electron density, and several α -helical structures could be detected in these maps (Fig. 3). However, the resolution of the data set has thus far hindered the assignment of amino-acid side chains. Improvement of the diffraction resolution will thus be necessary in order to build and refine an unambiguous model.

As Csm2 is a homologue of *S. solfataricus* SSo1424 (18% sequence identity) and *T. thermophilus* Csm2 (22% sequence identity), which play important roles in the assembly of Csm crRNP complexes (Rouillon *et al.*, 2013; Staals *et al.*, 2014), the results described here have an impact in helping to understand the structure and function of Csm crRNP complexes of type IIIA CRISPR–Cas systems.

Acknowledgements

This work was supported by research grants 11/50963-4 from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, São Paulo Research Foundation), 480411/2011-5 from

CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, National Council for Scientific and Technological Development, Brazil) and 448833/2014-0 from CNPq. The authors thank the LNLS beamline staff for help with the measurements and Dr A. Tashima from the proteomics laboratory of UNIFESP for help with the MS measurements.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Arvai, A. (2015). *Adxv – A Program to Display X-ray Diffraction Images*. <http://www.scripps.edu/tainer/arvai/adxv.html>.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. (2007). *Science*, **315**, 1709–1712.
- Blommel, P. G. & Fox, B. G. (2007). *Protein Expr. Purif.* **55**, 53–68.
- Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. (2005). *Microbiology*, **151**, 2551–2561.
- Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V. & van der Oost, J. (2008). *Science*, **321**, 960–964.
- Cocozaki, A. I., Ramia, N. F., Shao, Y., Hale, C. R., Terns, R. M., Terns, M. P. & Li, H. (2012). *Structure*, **20**, 545–553.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org/>.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H. & Moineau, S. (2010). *Nature (London)*, **468**, 67–71.
- Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. (2012). *Proc. Natl Acad. Sci. USA*, **109**, E2579–E2586.
- Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. (2005). *PLoS Comput. Biol.* **1**, e60.
- Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A. M., Glover, C. V., Graveley, B. R., Terns, R. M. & Terns, M. P. (2012). *Mol. Cell*, **45**, 292–302.
- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M. & Terns, M. P. (2009). *Cell*, **139**, 945–956.
- Huber, R. L., Langworthy, T. A., König, H., Thomm, M., Woese, C. R., Sleytr, U. B. & Stetter, K. O. (1986). *Arch. Microbiol.* **144**, 324–333.
- Jackson, R. N., Golden, S. M., van Erp, P. B., Carter, J., Westra, E. R., Brouns, S. J., van der Oost, J., Terwilliger, T. C., Read, R. J. & Wiedenheft, B. (2014). *Science*, **345**, 1473–1479.
- Jore, M. M. *et al.* (2011). *Nature Struct. Mol. Biol.* **18**, 529–536.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copié, V., Young, M. J., White, M. F. & Lawrence, C. M. (2011). *J. Biol. Chem.* **286**, 21643–21656.
- Makarova, K. S., Aravind, L., Wolf, Y. I. & Koonin, E. V. (2011). *Biol. Direct*, **6**, 38.
- Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. (2006). *Biol. Direct*, **1**, 7.
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., van der Oost, J. & Koonin, E. V. (2011). *Nature Rev. Microbiol.* **9**, 467–477.
- Marraffini, L. A. & Sontheimer, E. J. (2008). *Science*, **322**, 1843–1845.
- Marraffini, L. A. & Sontheimer, E. J. (2010). *Nature Rev. Genet.* **11**, 181–190.
- 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.
- Medlock, A. E., Carter, M., Dailey, T. A., Dailey, H. A. & Lanzilotta, W. N. (2009). *J. Mol. Biol.* **393**, 308–319.
- Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. (2005). *J. Mol. Evol.* **60**, 174–182.
- Mulepati, S. & Bailey, S. (2011). *J. Biol. Chem.* **286**, 31896–31903.
- Osawa, T., Inanaga, H. & Numata, T. (2013). *J. Mol. Biol.* **425**, 3811–3823.
- Osawa, T., Inanaga, H., Sato, C. & Numata, T. (2015). *Mol. Cell*, **58**, 418–430.
- Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. C. & Craievich, A. (1998). *J. Synchrotron Rad.* **5**, 72–76.
- Pourcel, C., Salvignol, G. & Vergnaud, G. (2005). *Microbiology*, **151**, 653–663.
- Robbins, A. H., McRee, D. E., Williamson, M., Collett, S. A., Xuong, N.-H., Furey, W. F., Wang, B.-C. & Stout, C. D. (1991). *J. Mol. Biol.* **221**, 1269–1293.
- Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilstein-Edmands, V., Cannone, G., Graham, S., Robinson, C. V., Spagnolo, L. & White, M. F. (2013). *Mol. Cell*, **52**, 124–134.
- Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A. & Marraffini, L. A. (2015). *Cell*, **161**, 1164–1174.
- Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. & Siksnys, V. (2011). *EMBO J.* **30**, 1335–1342.
- Staals, R. H. *et al.* (2013). *Mol. Cell*, **52**, 135–145.
- Staals, R. H. *et al.* (2014). *Mol. Cell*, **56**, 518–530.
- Taylor, D. W., Zhu, Y., Staals, R. H., Kornfeld, J. E., Shinkai, A., van der Oost, J., Nogales, E. & Doudna, J. A. (2015). *Science*, **348**, 581–585.
- Wang, Y., Juranek, S., Li, H., Sheng, G., Tuschl, T. & Patel, D. J. (2008). *Nature (London)*, **456**, 921–926.
- Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S. J., van der Oost, J., Doudna, J. A. & Nogales, E. (2011). *Nature (London)*, **477**, 486–489.
- Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W. & Doudna, J. A. (2009). *Structure*, **17**, 904–912.
- Yogavel, M., Nithya, N., Suzuki, A., Sugiyama, Y., Yamane, T., Velmurugan, D. & Sharma, A. (2010). *Acta Cryst.* **D66**, 1323–1333.