

Stefano Stella,^{a‡} Rafael
Molina,^{a‡} Claudia Bertonatti,^b
Alexandre Juillerrat^b and
Guillermo Montoya^{a*}

^aMacromolecular Crystallography Group,
Structural Biology and Biocomputing
Programme, Spanish National Cancer Research
Centre (CNIO), Calle de Melchor Fernández
Almagro 3, 28029 Madrid, Spain, and
^bCellectis, 8 Rue de la Croix Jarry, 75013 Paris,
France

‡ These authors contributed equally.

Correspondence e-mail: gmontoya@cnio.es

Received 15 October 2013

Accepted 5 December 2013

Expression, purification, crystallization and preliminary X-ray diffraction analysis of the novel modular DNA-binding protein BurrH in its apo form and in complex with its target DNA

Different genome-editing strategies have fuelled the development of new DNA-targeting molecular tools allowing precise gene modifications. Here, the expression, purification, crystallization and preliminary X-ray diffraction of BurrH, a novel DNA-binding protein from *Burkholderia rhizoxinica*, are reported. Crystallization experiments of BurrH in its apo form and in complex with its target DNA yielded crystals suitable for X-ray diffraction analysis. The crystals of the apo form belonged to the primitive hexagonal space group $P3_1$ or its enantiomorph $P3_2$, with unit-cell parameters $a = b = 73.28$, $c = 268.02$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. The BurrH–DNA complex crystallized in the monoclinic space group $P2_1$, with unit-cell parameters $a = 70.15$, $b = 95.83$, $c = 76.41$ Å, $\alpha = \gamma = 90$, $\beta = 109.51^\circ$. The self-rotation function and the Matthews coefficient suggested the presence of two protein molecules per asymmetric unit in the apo crystals and one protein–DNA complex in the monoclinic crystals. The crystals diffracted to resolution limits of 2.21 and 2.65 Å, respectively, using synchrotron radiation.

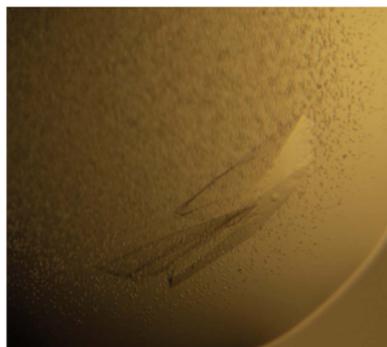
1. Introduction

New exciting possibilities in biotechnology and medicine are emerging thanks to advances in genome sequencing along with the ability to modify cellular functions through the addition, removal or exchange of DNA sequences in the genome (Perez-Pinera *et al.*, 2012). Tools such as the modular transcription activator-like effector (TALE; Boch *et al.*, 2009; Bogdanove & Voytas, 2011; Moscou & Bogdanove, 2009) have demonstrated their versatility for generating specific DNA-binding proteins. The ease with which it is possible to design new specificities and their modular assembly promotes the broader use of these tools, especially in the field of synthetic biology. Although a large number of fully sequenced bacterial genomes are available, TALE proteins have only been identified in two plant pathogens: *Xanthomonas* spp. and *Ralstonia solanacearum*. The constant release of new genomic data prompted us to search for new modular DNA-binding proteins. Thus, we looked for DNA-binding proteins containing differences in their modular sequence. We found a protein fulfilling this condition in *Burkholderia rhizoxinica*, a symbiotic bacterium that lives in the cytosol of *Rhizopus microsporus*, which has previously been identified (Schornack *et al.*, 2013; hereafter termed BurrH; UniProtKB E5AV36_BURRH; gene ID 9979518; gene symbol RBRH_01844). BurrH shows a modular organization; however, the sequences of the modules display few conserved residues. The protein contains 19 repeats able to recognize a 19 bp DNA target. Here, we report the expression, purification, crystallization and preliminary X-ray diffraction analysis of this novel modular domain in its apo form and in complex with the DNA target.

2. Materials and methods

2.1. Protein expression and purification

Escherichia coli BL21 (DE3) cells were transformed with pET-24d(+) vector containing the gene coding for BurrH. The polypeptide chain expressed is the UniProt E5AV36 sequence with six histidines at the C-terminus (81 kDa including the tag). Protein expression tests in Luria–Bertani medium showed high levels of BurrH expression at 310 K. After induction with 0.5 mM isopropyl β -D-1-thiogalacto-



pyranoside (IPTG), the cultures were grown for 2 h and the cells were harvested by centrifugation (9000g for 20 min at 277 K). For the preparation of selenomethionine-substituted protein, the cells were grown in SelenoMethionine Medium Complete (Molecular Dimensions) including 40 $\mu\text{g ml}^{-1}$ selenomethionine at 310 K. When the culture reached an OD of 0.8 at 600 nm, protein expression was induced with 1 mM IPTG for 3 h. Finally, the culture was harvested by centrifugation as described previously. The pellet was flash-frozen and stored at 193 K.

Before purification, the pellet was resuspended in 25 mM HEPES-HCl pH 8.0, 150 mM NaCl, 0.05% Triton X-100, 0.2 mM TCEP in the presence of 2 \times cComplete EDTA-free tablets (Roche). The cells were disrupted by sonication and the lysate was clarified by centrifugation at 40 000g for 1 h. The protein was predominantly found in the soluble fraction. The purification procedure consisted of three consecutive chromatography steps. After centrifugation, the supernatant was applied onto a Ni-NTA agarose column (Qiagen) equilibrated in buffer *A* (25 mM HEPES-HCl pH 8.0, 150 mM NaCl, 0.2 mM TCEP, 1 \times cComplete EDTA-free tablet, Roche). The Ni-NTA column was washed with buffer *A* plus 5 mM imidazole. The elution was performed applying a step gradient of 10, 25, 50 and 100% buffer *B* (buffer *A* plus 1 M imidazole). Enriched protein fractions corresponding to 25 and 50% buffer *B* were pooled together and applied onto a 5 ml HiTrap Heparin HP column (GE Healthcare) equilibrated with buffer *A*. The protein was eluted with a linear gradient of 0–100% buffer *H* (buffer *A* plus 1 M NaCl) in ten column volumes. Protein-rich fractions were collected and concentrated (using 30 kDa cutoff Centriprep Amicon Ultra devices) and subsequently loaded onto a HiLoad 16/60 200 Superdex column (GE Healthcare) equilibrated in buffer *A*. The protein peaks were concentrated (using 30 kDa cutoff Centriprep Amicon Ultra devices), flash-frozen in liquid nitrogen and stored at 193 K. The protein concentration was determined using the theoretical molar extinction coefficient at 280 nm calculated from the amino-acid composition. An overloaded SDS-PAGE stained with SimplyBlue (Invitrogen) displayed a very pure protein preparation (Fig. 1*a*).

2.2. Size-exclusion chromatography–multiangle laser light scattering

Size-exclusion chromatography–multiangle laser light scattering (SEC–MALLS) experiments were performed at 293 K using a Superdex 200 10/300 GL column (GE Healthcare) attached to a

DAWN HELEOS light-scattering detector and an Optilab rEX differential refractive-index detector (Wyatt Technology). The column was equilibrated in buffer *A* with 0.03% (*w/v*) NaN_3 and the SEC–MALLS system was calibrated with a sample of BSA at 1 mg ml^{-1} in the same buffer. A sample of 100 μl of BurrH at 1 mg ml^{-1} in the same buffer was injected at a flow rate of 0.5 ml min^{-1} . Data acquisition and analysis were carried out using the *ASTRA* software (Wyatt). The SEC–MALLS measurements indicated that the protein is a monomer with an experimental molecular mass of 78.6 kDa (Fig. 1*b*).

2.3. Apo-BurrH sample preparation

Both the native and the selenomethionine-substituted (hereafter referred to as SeMet-BurrH) proteins were initially concentrated to 12 mg ml^{-1} and dialyzed against 20 mM MES pH 6.0, 200 mM NaCl, 5 mM MgCl_2 at 277 K. A light white precipitate was formed and eliminated by filtration using a 0.2 μm cutoff filter (Millipore) before crystallization. The final concentration of the protein preparation ranged between 8 and 10 mg ml^{-1} after filtration.

2.4. BurrH–DNA complex formation

The BurrH DNA target is a blunt duplex oligonucleotide of 23 bp (5'-TTAAGAGAAGCAAATACGTTAA-3'). The target sequence was determined using the RVD code previously described for TALE (Boch *et al.*, 2009). The duplex DNA was purchased from IDT (Integrated DNA Technologies, Madrid) and resuspended in 25 mM HEPES-HCl pH 8.0, 150 mM NaCl. The complex was obtained by mixing BurrH protein and the duplex in the presence of 5 mM MgCl_2 in a 1:1.3 protein:DNA molar ratio. The mixture was incubated for 1 h at 293 K. The final concentration of the BurrH–DNA complex solution was 8 mg ml^{-1} . A band-shift assay was used to test the binding of BurrH to the oligonucleotide (Fig. 1*c*). The protein in the same buffer was incubated with the DNA duplex as previously described and loaded onto a native polyacrylamide gel.

2.5. Crystallization

Crystallization screening was performed immediately after protein and protein–DNA complex preparation using the sitting-drop method with a Cartesian MicroSys robot (Genomic Solutions) in 96-well MRC plates. The nanodrops consisted of 0.2 μl protein solution plus 0.2 μl reservoir solution and were equilibrated against a

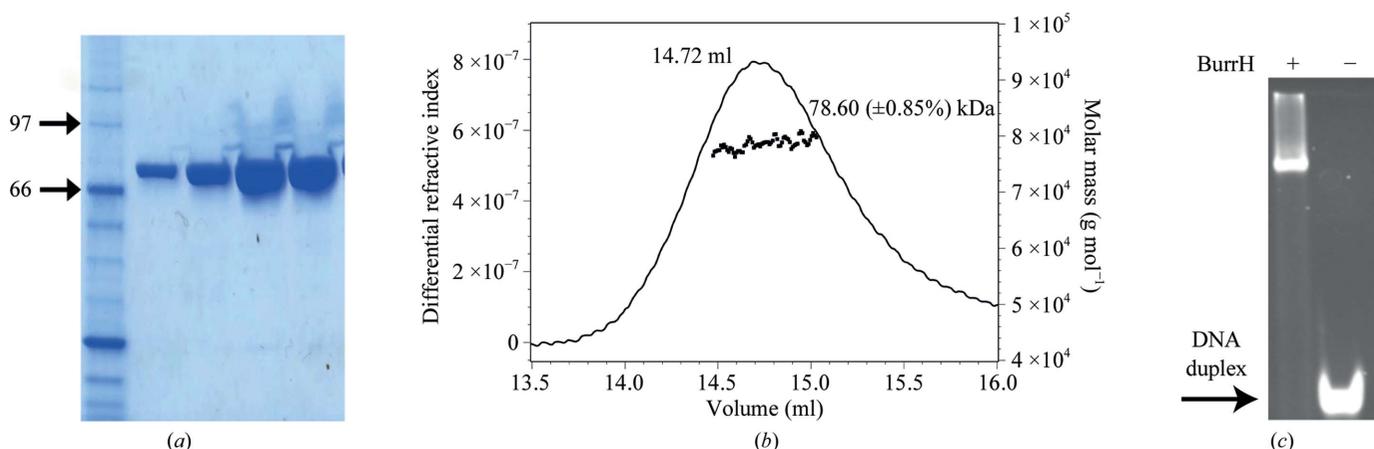


Figure 1 BurrH expression and purification and protein–DNA complex formation. (*a*) SDS gel showing the purified BurrH protein. The left lane contains molecular-mass marker (labelled in kDa). (*b*) SEC–MALLS chromatogram. (*c*) Band-shift assay. The DNA target was incubated with and without BurrH protein for 1 h at 298 K. The mixture was separated on a native polyacrylamide gel. The arrow indicates the band corresponding to free DNA target.

reservoir volume of 60 μl . The initial screens used for crystallization of apo-BurrH and the BurrH–DNA complex were Crystal Screen and Crystal Screen 2, Crystal Screen Cryo and Crystal Screen Lite (Hampton Research), JBScreen Classic 1, 2, 3, 4, 5, 6, 7 and 8 (Jena Bioscience), PACT and the Protein Complex Suite (Qiagen). Crystals of apo-BurrH were obtained in condition No. 76 of PACT [0.2 *M* potassium thiocyanate, 0.1 *M* bis-tris propane pH 7.5, 20% (*w/v*) PEG 3350]. The crystallization condition was optimized and the best diffracting crystals were grown in a reservoir solution consisting of 0.2 *M* potassium thiocyanate, 0.1 *M* bis-tris propane pH 7.5, 20–26% (*w/v*) PEG 3350 with 0.02 *M* cobalt(II) chloride hexahydrate or 0.2 *M* sodium iodide as an additive (conditions 8 and 20 of Hampton Research Additive Screen). Fresh crystals were harvested and cryo-protected by adding 28% ethylene glycol to the reservoir solution before data collection. Diffraction experiments were performed at the Swiss Light Source (SLS, Villigen) and ALBA (Barcelona).

The initial crystals were later reproduced using the hanging-drop method. Drops of 2 μl were obtained by mixing 1 μl reservoir and sample solutions and produced rod-shaped crystals with maximum dimensions of $0.2 \times 0.2 \times 0.025$ mm (Fig. 2*a*). The best diffracting crystals were obtained using the hanging-drop method and grew using 0.2 *M* potassium thiocyanate, 0.1 *M* bis-tris propane pH 7.5,

22% (*w/v*) PEG 3350 in the presence of 0.2 *M* sodium iodide. These crystals appeared after 1 d and grew for 5 d. These specimens diffracted to 2.5 Å resolution (data not shown). Apo SeMet-BurrH was crystallized using the same condition as used to crystallize the apo native protein. These crystals (Fig. 2*b*) diffracted to 2.21 Å resolution (see Table 1).

For the crystallization of BurrH–DNA, only the native protein was used to reconstitute the complex. Initial crystals were detected in condition No. 32 of The Protein Complex Suite (Qiagen) [0.2 *M* NaCl, 0.1 *M* Tris pH 8.0, 20% (*w/v*) PEG 4000]. These crystals were optimized and the best diffracting specimens were obtained using 0.27 *M* NaCl, 0.1 *M* Tris–HCl pH 8.2, 26% (*w/v*) PEG 4000 in the reservoir solution (Fig. 2*c*). The crystals were cryoprotected by adding 20% ethylene glycol to the reservoir solution. These crystals appeared after 5–7 d, grew for 12–15 d and diffracted to 2.65 Å resolution (see Table 1). To confirm the presence of both the protein and the DNA in the crystals, some specimens were washed five to seven times in the reservoir solution and dissolved in SDS loading dye. This sample was analysed by SDS–PAGE, using the crystallization oligonucleotide and the protein as controls. After electrophoresis the gel was stained with SimplyBlue Safe Stain (Invitrogen) to detect the protein (Fig. 2*d*) and also with SYBR–Gold (Invitrogen)

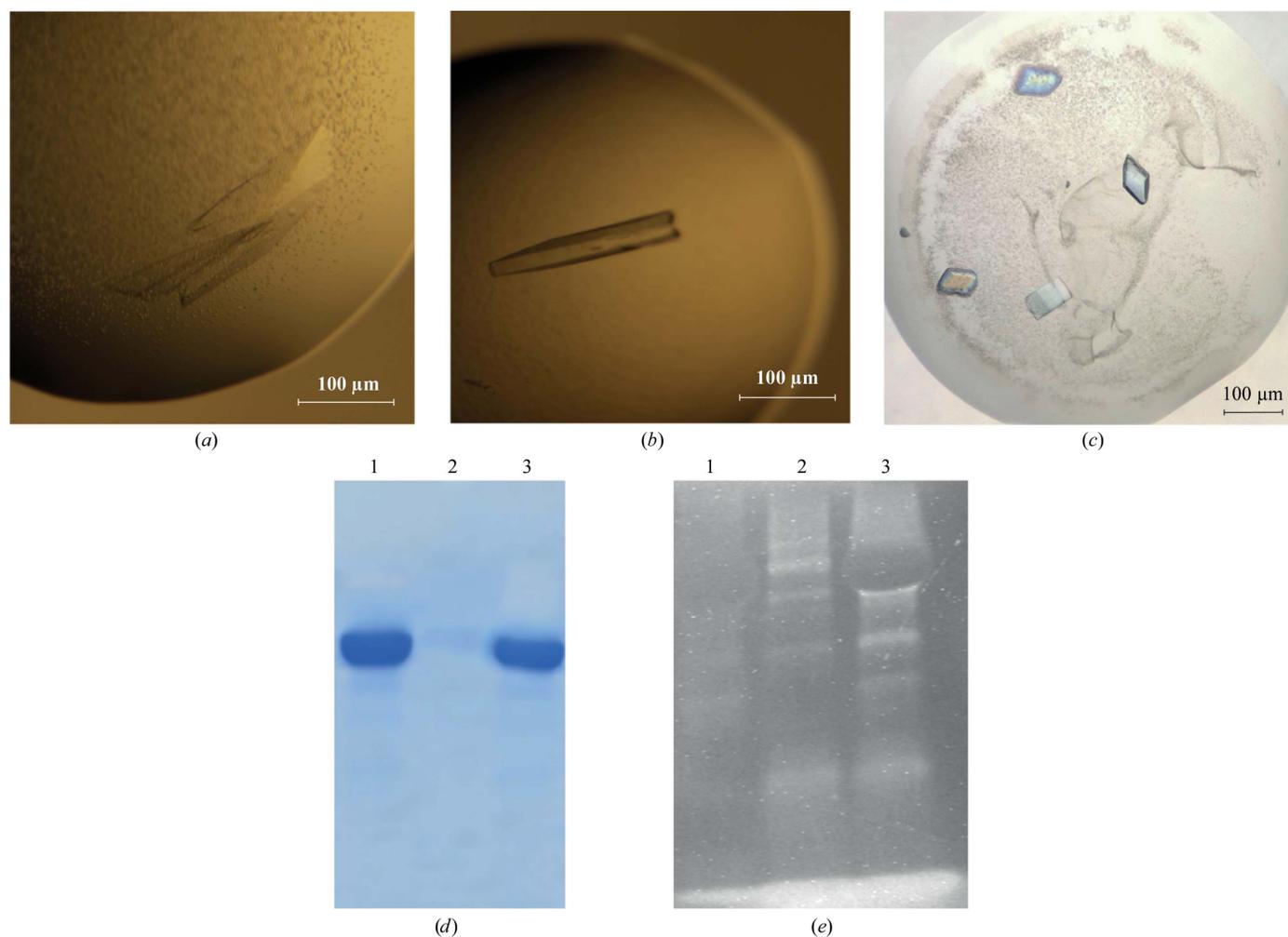


Figure 2 BurrH crystallization. (*a*) Crystal of BurrH in the apo form. (*b*) Selenomethionine-substituted BurrH crystals in the apo form. (*c*) Crystals of native BurrH in complex with its target DNA. (*d*) SDS–PAGE gel confirming the crystallization of the BurrH–DNA complex stained with SimplyBlue Safe Stain to visualize the protein. (*e*) SDS–PAGE gel confirming the crystallization of the BurrH–DNA complex stained with SYBR–Gold to visualize the DNA. In both gels, lanes 1, 2 and 3 contain BurrH protein, the DNA target and the dissolved crystals, respectively.

to detect DNA (Fig. 2e). The additional SYBR-Gold-stained bands are owing to dissociation of the duplex DNA target in the presence of SDS.

2.6. Data collection

Cryoprotected crystals were flash-cooled in liquid nitrogen. The diffraction data were collected using a Pilatus 6M detector on the PX

Table 1

Data-collection statistics for the apo form of BurrH and its complex with DNA.

Values in parentheses are for the outermost resolution shell.

	Apo SeMet-BurrH	BurrH–DNA complex
Space group	$P3_1/P3_2$	$P2_1$
Unit-cell parameters		
a (Å)	73.28	70.15
b (Å)	73.28	95.83
c (Å)	268.02	76.41
α (°)	90	90
β (°)	90	109.51
γ (°)	120	90
Data collection		
Temperature (K)	100	100
Wavelength (Å)	0.97974	1.00
Beamline	XS06A, SLS	XS06A, SLS
Resolution (Å)	46.08–2.21 (2.33–2.21)	47.92–2.65 (2.79–2.65)
Total reflections	277915 (40222)	92439 (13676)
Unique reflections	79217 (11699)	27134 (4019)
Multiplicity	3.5 (3.4)	3.4 (3.4)
Completeness (%)	98.7 (99.8)	97.7 (99.9)
Mean $I/\sigma(I)$	7.0 (2.4)	12.1 (1.7)
R_{merge}^\dagger	0.112 (0.42)	0.065 (0.61)

$^\dagger R_{\text{merge}}$ is defined according to XDS (Kabsch, 2010). $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$.

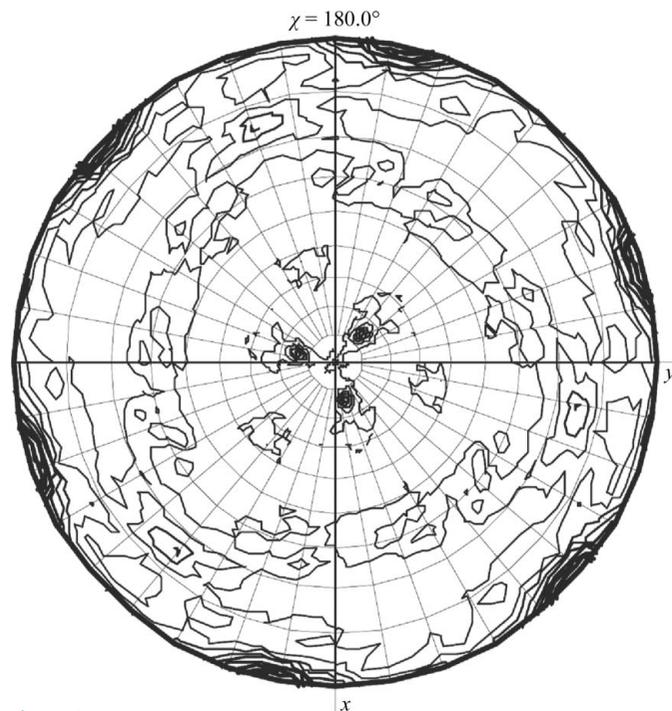
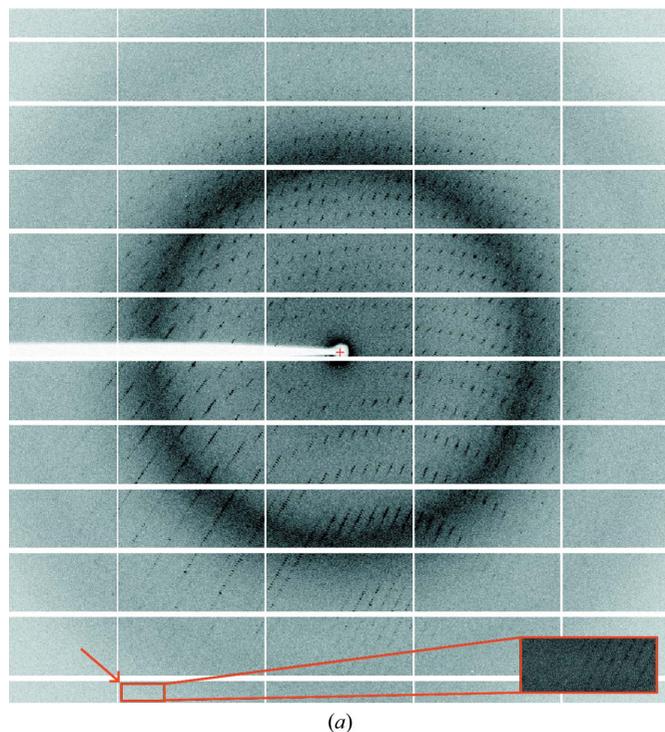


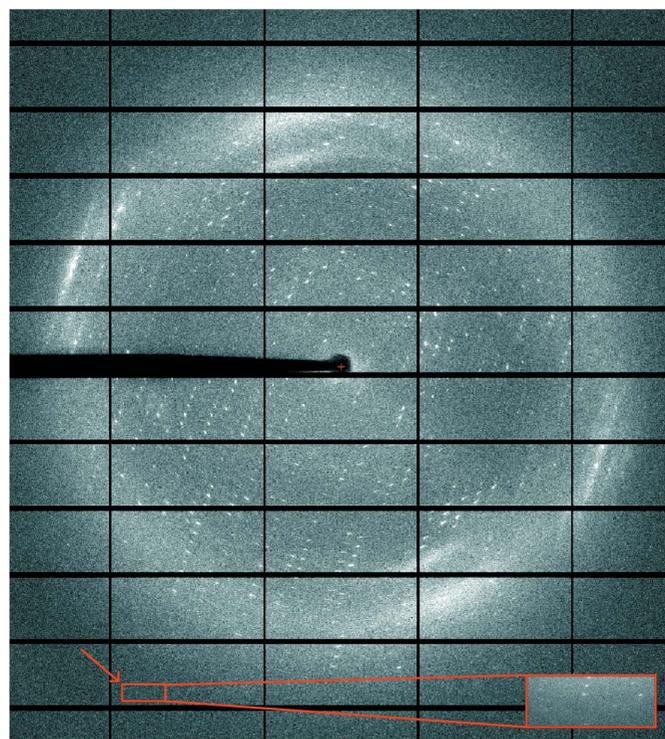
Figure 3

Self-rotation function of the apo-BurrH crystals. *SFCHECK* (Vaguine *et al.*, 1999) and *POINTLESS* (Evans, 2006) were used to check the data, indicating the absence of a translational Patterson peak and no twinning. We computed the self-rotation function of the apo-BurrH crystals using *MOLREP* (Vagin & Teplyakov, 2010). Inspection of the stereographic projection detected strong peaks in the $\chi = 180^\circ$ section and an absence of peaks in the $\chi = 120^\circ$, $\chi = 90^\circ$ and $\chi = 60^\circ$ sections. The peaks indicated the presence of twofold symmetry axes, which together with the Matthews coefficient data and the absence of substantial peaks in other sections suggested the presence of a dimer in the asymmetric unit.

XS06 beamline (SLS, Villigen) and at XALOC (ALBA, Barcelona). The best data set for the apo crystals was collected with $\Delta\varphi = 0.5^\circ$ at the Se absorption peak (0.97974 Å wavelength). In the case of the protein–DNA complex, a good data set was collected with $\Delta\varphi = 0.5^\circ$ at 1.0 Å wavelength. Data processing was accomplished using *XDS* (Kabsch, 2010) and *SCALA* from the *CCP4* package (Evans, 2006;



(a)



(b)

Figure 4

X-ray diffraction patterns of both crystal types. (a) Selenomethionine-substituted apo-BurrH. (b) Native BurrH in complex with DNA. The insets indicated by arrows show reflections at 2.21 and 2.65 Å resolution, respectively.

Winn *et al.*, 2011). The data-collection statistics for each crystal type are summarized in Table 1.

3. Results and discussion

The crystals of SeMet-BurrH in its apo form belonged to space group $P3_1$ or $P3_2$, with unit-cell parameters $a = b = 73.28$, $c = 268.02$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. These enantiomorphic space groups cannot be distinguished at this stage. The Matthews coefficient ($V_M = 2.31$ Å³ Da⁻¹; Matthews, 1968) and self-rotation function (Rossmann & Blow, 1962) suggested the presence of two protein molecules per asymmetric unit and a solvent content of 52% (Fig. 3). The data were 98.7% complete, with a multiplicity of 3.5 and an overall mean $I/\sigma(I)$ of 7.0 (Fig. 4a). Detailed statistics for the data set are summarized in Table 1.

The crystals of BurrH in complex with its DNA target belonged to space group $P2_1$, with unit-cell parameters $a = 70.15$, $b = 95.83$, $c = 76.41$ Å, $\beta = 109.51^\circ$. The Matthews coefficient ($V_M = 2.42$ Å³ Da⁻¹) and self-rotation function (data not shown) suggested the presence of one protein–DNA complex per asymmetric unit and a solvent content of 55%. No extra peaks apart from those arising from the crystallographic symmetry could be observed. The data were 97.7% complete, with a multiplicity of 3.4 and an overall mean $I/\sigma(I)$ of 12.1 (Fig. 4b). Detailed statistics for the data set are summarized in Table 1.

These are the first crystals of this novel DNA-binding protein in its apo form and in complex with its DNA target. Furthermore, preliminary SAD experiments yielded good-quality data for solving the phase problem. We believe that these studies will help to elucidate

the molecular mechanism of DNA recognition by this novel DNA-binding domain. BurrH may constitute a new platform for the design of DNA-binding proteins with new specificities, which may improve the current tools available for biotechnological and therapeutic purposes.

We thank the Swiss Light Source and ALBA beamline staff for their support. This work was supported by EU Marie Curie ‘SMARTBREAKER’ (2010-276953 to SS), Ministerio de Educación (SB2010-0105 to SS), Ministerio de Economía y Competitividad (JCI-2011-09308 to RM, BFU2011-23815/BMC to GM), the Fundación Ramón Areces and the Comunidad Autónoma de Madrid (CAM-S2010/BMD-2305 to GM).

References

- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. & Bonas, U. (2009). *Science*, **326**, 1509–1512.
- Bogdanove, A. J. & Voytas, D. F. (2011). *Science*, **333**, 1843–1846.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Moscou, M. J. & Bogdanove, A. J. (2009). *Science*, **326**, 1501.
- Perez-Pinera, P., Ousterout, D. G. & Gersbach, C. A. (2012). *Curr. Opin. Chem. Biol.* **16**, 268–277.
- Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
- Schornack, S., Moscou, M. J., Ward, E. R. & Horvath, D. M. (2013). *Annu. Rev. Phytopathol.* **51**, 383–406.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* **D66**, 22–25.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* **D55**, 191–205.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.