CrossMark

electronic reprint

STRUCTURAL BIOLOGY
COMMUNICATIONS

ISSN: 2053-230X
journals.iucr.org/f

pHluorin-assisted expression, purification, crystallization and X-ray diffraction data analysis of the C-terminal domain of the HsdR subunit of the *Escherichia coli* type I restriction-modification system EcoR124I

Pavel Grinkevich, Iuliia Iermak, Nicholas A. Luedtke, Jeroen R. Mesters, Rüdiger Ettrich and Jost Ludwig

Acta Cryst. (2016). F**72**, 672–676

IUCr Journals
CRYSTALLOGRAPHY JOURNALS ONLINE

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see http://journals.iucr.org/services/authorrights.html

STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Received 6 April 2016
Accepted 16 July 2016

Edited by M. L. Pusey, University of Alabama, USA

Keywords: restriction-modification system; EcoR124I; HsdR; pHluorin; GFP; *Escherichia coli*.

pHluorin-assisted expression, purification, crystallization and X-ray diffraction data analysis of the C-terminal domain of the HsdR subunit of the *Escherichia coli* type I restriction-modification system EcoR124I

Pavel Grinkevich,[a,b]* Iuliia Iermak,[a,b] Nicholas A. Luedtke,[a,c] Jeroen R. Mesters,[d] Rüdiger Ettrich[a,b]* and Jost Ludwig[a,b]

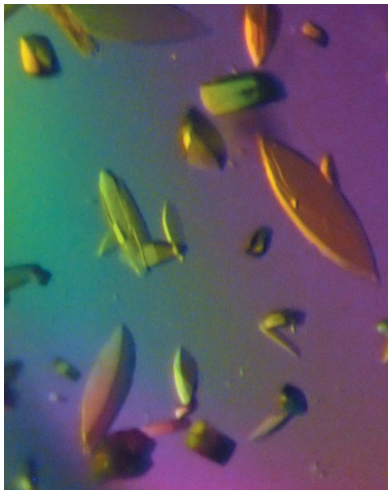
[a]Center for Nanobiology and Structural Biology, Institute of Microbiology ASCR, v.v.i., Zamek 136, 37333 Nove Hrady, Czech Republic, [b]Faculty of Science, University of South Bohemia in Ceske Budejovice, Branisovska 31, 37005 Ceske Budejovice, Czech Republic, [c]Ripon College, 300 Seward Street, Ripon, WI 54971, USA, and [d]Institute of Biochemistry, University of Lübeck, Ratzeburger Allee 160, Lübeck, Germany. *Correspondence e-mail: pavel.grinkevich@gmail.com, ettrich@nh.cas.cz

The HsdR subunit of the type I restriction-modification system EcoR124I is responsible for the translocation as well as the restriction activity of the whole complex consisting of the HsdR, HsdM and HsdS subunits, and while crystal structures are available for the wild type and several mutants, the C-terminal domain comprising approximately 150 residues was not resolved in any of these structures. Here, three fusion constructs with the GFP variant pHluorin developed to overexpress, purify and crystallize the C-terminal domain of HsdR are reported. The shortest of the three encompassed HsdR residues 887–1038 and yielded crystals that belonged to the orthorhombic space group $C222_1$, with unit-cell parameters a = 83.42, b = 176.58, c = 126.03 Å, $\alpha = \beta = \gamma = 90.00^\circ$ and two molecules in the asymmetric unit (V_M = 2.55 Å^3 Da^{-1}, solvent content 50.47%). X-ray diffraction data were collected to a resolution of 2.45 Å.

1. Introduction

EcoR124I belongs to the type I restriction-modification (RM) systems, a diverse group of DNA-cleaving enzymes that are present in many species of bacteria and archaea. The type I RM systems are further subdivided into five families: A–E (for a recent review, see Loenen *et al.*, 2014). EcoR124I and the almost identical EcoR124II are prototypical of the plasmid-encoded type IC family, their natural host being *Escherichia coli* (Youell & Firman, 2008). The fully assembled enzyme complex is able to recognize particular sites on DNA and read out the methylation states of adenine residues within these recognition sequences. Based on methylation information, the enzyme differentiates whether the DNA belongs to the host bacterium, in which case methylation of the complementary strand may occur, or to the invading agent (normally a bacteriophage), which triggers the whole complex to translocate thousands of base pairs and cleave the DNA at random sites (Horiuchi & Zinder, 1972). Many molecular and structural details of this complex process remain largely unknown.

All of the aforementioned functions are performed by three kinds of subunits, HsdS, HsdM and HsdR, that dynamically assemble into different complexes carrying out various tasks. One HsdS and two HsdMs form methyltransferase (MTase), which is capable of sequence recognition and methylation, while the addition of two HsdRs enables the complex to



© 2016 International Union of Crystallography

electronic reprint

Table 1
Macromolecule-production information.

Construct	pHluorin-HsdR705	pHluorin-HsdR867	pHluorin-HsdR887
Source organism	*E. coli*	*E. coli*	*E. coli*
DNA source	pTrcR124	pTrcR124	pTrcR124
Vector	pUC-Kan-pH	pUC-Kan-pH	pUC-Kan-pH
Expression host	*E. coli*	*E. coli*	*E. coli*
Full length (amino acids)	580	418	398
pHluorin tag length (amino acids)	246	246	246
Partial HsdR length (amino acids)	334	172	152
Molecular weight (Da)	66797.71	47778.64	45442.12
Forward primer† (5′→3′)	**GAATTC**TTCCGGGATCTGGAACG	AAAGAATTCGAGAAATCAACCACTGACTG	AAAGAATTCGAAATAAACCTGGATTATATC
Reverse primer‡ (5′→3′)	**GGATCC**CTATATTTTTCCGCCTACGCC	GGTTGATTTCTCGAATTCTTTGTATAGTTC	CAGGTTTATTTCGAATTCTTTGTATAGTTC
Complete amino-acid sequence of the construct produced§	MHHHHHH*SKGEELFTGVVPILVELDGDVNGH-KFSVSGEGEGDATYGKLTLKFICTTGKLP-VPWPTLVTTFSYGVQCFSRYPDHMKRHDF-FKSAMPEGYVQERTIFFKDDGNYKTRAEV-KFEGDTLVNRIELKGIDFKDDGNILGHKL-EYNYNEHLVYIMADKQKNGTKAIFQVHHN-IEDGGVQLADHYQQNTPIGDGPVLLPDNH-YLHTQSALSKDPNEKRDHMVLLEFVTAAG-ITHGMDELYK*EFFRDLERSTIDAITLFGD-KNTKNVVLEKSYTEYMEGFTDAATGEAKR-GFMTVVSELEQRFPDPTSIESEKEKKDFV-KLFGEYLRAENILQNYDEFATLKALQQID-LSDPVAVEKFKAEHYVDDEKFAELQTIRL-PADRKIQDYRSAYNDIRDWQRREKEAEKK-EKSTTDWDDVVFEVDLLKSQEINLDYILG-LIFEHNRQNKGKGEMIEEVKRLIRSSLGN-RAKEGLVVDFIQQTNLDDLPDKASIIDAF-FTFAQREQQREAEALIKEENLNEDAAKRY-IRTSLKREYATENGTELNETLPKLSPLNP-QYKTKKQAVFQKIVSFIEKFKGVGGKI	MHHHHHH*SKGEELFTGVVPILVELDGDVNGH-KFSVSGEGEGDATYGKLTLKFICTTGKLP-VPWPTLVTTFSYGVQCFSRYPDHMKRHDF-FKSAMPEGYVQERTIFFKDDGNYKTRAEV-KFEGDTLVNRIELKGIDFKDDGNILGHKL-EYNYNEHLVYIMADKQKNGTKAIFQVHHN-IEDGGVQLADHYQQNTPIGDGPVLLPDNH-YLHTQSALSKDPNEKRDHMVLLEFVTAAG-ITHGMDELYK*EFEKSTTDWDDVVFEVDLL-KSQEINLDYILGLIFEHNRQNKGKGEMIE-EVKRLIRSSLGNRAKEGLVVDFIQQTNLD-DLPDKASIIDAFFTFAQREQQREAEALIK-EENLNEDAAKRYIRTSLKREYATENGTEL-NETLPKLSPLNPQYKTKKQAVFQKIVSFI-EKFKGVGGKI	MHHHHHH*SKGEELFTGVVPILVELDGDVNGH-KFSVSGEGEGDATYGKLTLKFICTTGKLP-VPWPTLVTTFSYGVQCFSRYPDHMKRHDF-FKSAMPEGYVQERTIFFKDDGNYKTRAEV-KFEGDTLVNRIELKGIDFKDDGNILGHKL-EYNYNEHLVYIMADKQKNGTKAIFQVHHN-IEDGGVQLADHYQQNTPIGDGPVLLPDNH-YLHTQSALSKDPNEKRDHMVLLEFVTAAG-ITHGMDELYK*EFEINLDYILGLIFEHNRQ-NKGKGEMIEEVKRLIRSSLGNRAKEGLVV-DFIQQTNLDDLPDKASIIDAFFTFAQREQ-QREAEALIKEENLNEDAAKRYIRTSLKRE-YATENGTELNETLPKLSPLNPQYKTKKQA-VFQKIVSFIEKFKGVGGKI

† The EcoRI restriction site is in bold. ‡ The BamHI restriction site is in bold. § The first underlined sequence is the 6×His tag, the GFP sequence is shown in italics and the second underlined sequence is the HsdR sequence.

translocate DNA in an ATP-dependent manner and cleave DNA (Seidel *et al.*, 2008). Knowledge of how HsdR interacts with MTase is crucial in furthering our understanding of how the whole complex works.

The first crystal structure of HsdR (Lapkouski *et al.*, 2009; PDB entry 2w00) revealed a planar arrangement of four domains with pronounced grooves between them. Around 150 residues at the C-terminus (residues 893–1038) were not resolved in this structure. Another four structures of mutant HsdRs published in the PDB since then (PDB entries 4be7, 3h1t, 4beb and 4bec; Csefalvay *et al.*, 2015; Uyen *et al.*, 2009) also lacked structural information for this part of the protein. A role for the C-terminus in complex assembly has been postulated (Dryden *et al.*, 2001; Obarska-Kosinska *et al.*, 2008). Experiments with proteolytic fragments of HsdR and MTase from the related type I RM enzyme EcoKI revealed that the C-terminal region of HsdR is required for binding to MTase (Davies *et al.*, 1999), indicating the role of the C-terminus in complex assembly. A crystal structure of the C-terminus will allow the prediction of residues that are potentially involved in complex assembly and would be valuable for further, more sophisticated studies of the EcoR124I complex and of the type I family as a whole.

Since initial attempts to express the C-terminus of HsdR alone failed, fusion with a readily crystallizable protein was chosen as a well recognized and highly successful approach to aid protein overexpression, purification and crystallization (Smyth *et al.*, 2003; Kobe *et al.*, 2015). Constructs based on fusing C-terminal parts of HsdR of varying lengths with the green fluorescent protein (GFP) variant pHluorin (Miesenböck *et al.*, 1998) allowed easy expression and purification. In this paper, vector design, expression, purification, crystallization and X-ray diffraction data are reported for one of these pHluorin-fusion proteins.

2. Materials and methods

2.1. Cloning

HsdR is a rather large protein, consisting of 1038 amino acids, with a molecular weight of about 120.1 kDa. The C-terminal part of the *hsdR* gene (corresponding to amino acids 705–1038; HsdR705) was amplified from the pTrcR124 plasmid routinely used for HsdR overexpression (Janscak *et al.*, 1996) with primers containing EcoRI and BamHI restriction sites (Table 1) and was cloned into a variant of pUC-18 (Yanisch-Perron *et al.*, 1985) that carries the gene encoding the pH-sensitive GFP variant ratiometric pHluorin (Miesenböck *et al.*, 1998) and 6×His as an N-terminal fusion tag, and in which the ampicillin-resistance gene was replaced by a kanamycin-resistance gene (pUC-Kan-pH; J. Ludwig, unpublished work). The resulting plasmid contained (from 5′ to 3′) a

electronic reprint

6×His tag, *gfp* and partial *hsdR* sequences. The sequence of this plasmid is available from the authors. Two shorter constructs were obtained, containing stretches of *hsdR* that code for amino acids 867–1038 (HsdR867) and 887–1038 (HsdR887), respectively, by means of one-step PCR-based deletions (Qi & Scholthof, 2008). All constructs were verified by DNA sequencing.

Few guidelines exist on how the length of the linker between domains of a fusion protein impacts crystallization (Kobe *et al.*, 2015). As a first possibility, a very short linker of two amino acids (Glu-Phe) between the pHluorin and HsdR sequences was chosen based on a proximate EcoRI restriction site, with the idea of reducing conformational heterogeneity and maximizing interactions between the pHluorin and HsdR domains.

2.2. Expression and purification

E. coli BL21-Gold(DE3) cells transformed with the plasmid containing the fusion construct were grown overnight (16–18 h) at 37°C in 2 l LB medium with kanamycin (30 mg l^{-1}). After harvesting the cells by centrifugation for 20 min at 4000*g* at 4°C, they were resuspended in NPI-10 buffer (50 m*M* NaH_2PO_4, 300 m*M* NaCl, 10 m*M* imidazole pH 8.0), incubated on ice for 30 min with 1 mg l^{-1} lysozyme and disrupted with a French press. The lysate was centrifuged for 1 h at 23 000*g* and 4°C. The supernatant was filtered through a sterile syringe filter (Fisherbrand, Fisher Scientific, USA) and loaded onto Ni–NTA agarose matrix (Qiagen, Germany) pre-equilibrated with five volumes of NPI-10 buffer using an ÄKTApurifier (GE Healthcare, USA). After washing with 5–10 volumes of NPI-20 buffer containing 20 m*M* imidazole, the sample was eluted from the column by increasing the concentration of imidazole to 100 m*M*. After exchanging the buffer to 20 m*M* Tris, 50 m*M* NaCl, 0.1 m*M* EDTA, 1 m*M* DTT pH 8, the sample was loaded onto a Q Sepharose column (GE Healthcare, USA) and eluted with an NaCl gradient (from 50 m*M* to 1 *M*) to increase the overall purity. The purified protein was concentrated using a 30 kDa centrifugal filter unit (Millipore, Germany) and the buffer was exchanged to 20 m*M* phosphate buffer pH 8.

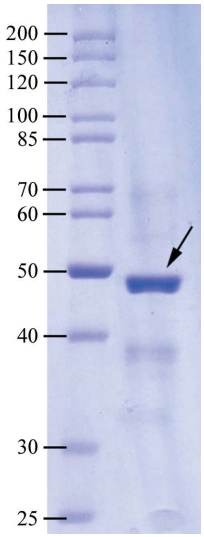


Figure 1
SDS–PAGE analysis of purified pHluorin-HsdR887. The numbers next to the marker lane indicate molecular masses in kDa. The theoretical molecular weight of pHluorin-HsdR887 (indicated with an arrow) is 45.4 kDa.

2.3. Crystallization

Initial screening was performed with a Gryphon crystallization robot (Art Robbins Instruments, USA) using the commercial screens Morpheus (Molecular Dimensions, UK) and PEG/Ion (Hampton Research, USA). Protein at a concentration of 12 mg ml^{-1} was mixed with precipitant in ratios of 1:1 and 2:1 (final drop volume 0.4 or 0.6 µl, reservoir volume 70 µl) on MRC 2-well crystallization plates (Hampton Research, USA) using the sitting-drop vapour-diffusion technique. Further optimization of growth conditions was performed by varying the protein and precipitant concentration and by using additives. The same sitting-drop vapour-diffusion method was employed in 24-well CombiClover crystallization plates (Jena Bioscience, Germany) with a drop volume of 5 µl and a reservoir volume of 700 µl. Additives from the Additive Screen (Hampton Research, USA) were added according to the manufacturer's recommendations.

2.4. Data collection and processing

X-ray diffraction data were collected on beamline P13 operated by EMBL at the PETRA III X-ray radiation source at the DESY campus in Hamburg, Germany equipped with a PILATUS 6M-F detector. 1–3 µl 50%(*w*/*v*) PEG 3350 was added directly to the drop with crystals for cryoprotection; after 20 min the crystals were mounted in LithoLoops (Molecular Dimensions, England) and flash-cooled in liquid nitrogen. Diffraction data images were integrated in *XDS* (Kabsch, 2010) and the space group was determined with *POINTLESS* from the *CCP*4 software package (Winn *et al.*, 2011). The Matthews coefficient (Matthews, 1968) was determined using *MATTHEWS_COEF* from *CCP*4.

3. Results and discussion

Three GFP-fusion constructs were successfully expressed and purified using a two-step purification protocol and were concentrated to 20–30 mg ml^{-1} (Fig. 1). Initial crystallization screening revealed that the shortest construct containing HsdR amino acids 887–1038 (pHluorin-HsdR887) is readily crystallizable in a wide range of conditions (crystals formed in over 40% of the conditions from both the Morpheus and the PEG/Ion screens) and therefore further work focused on this construct. Based on the shape and size of the crystals, two conditions from the PEG/Ion screen consisting of 20%(*w*/*v*) PEG 3350 and 0.2 *M* KH_2PO_4 or 0.2 *M* KI were optimized by varying the PEG and protein concentrations [16, 20 and 24%(*w*/*v*) PEG and 12 and 18 mg ml^{-1} protein] as well as the protein:precipitant ratio (1:1 and 2:1). The condition consisting of 20%(*w*/*v*) PEG 3350, 0.2 *M* KH_2PO_4 with

electronic reprint

12 mg ml^{-1} protein and a protein:precipitant ratio of 1:1 was chosen to further optimize the crystals using Additive Screen. Crystals with a similar morphology grew in over 60% of conditions with different additives, and based on their size and shape, conditions H9, H10, G10 and G11 [containing 0.025%(*v*/*v*) dichloromethane, 0.7%(*v*/*v*) 1-butanol, 3%(*v*/*v*) ethanol and 3%(*v*/*v*) 2-propanol in the reservoir solution, respectively] were selected for measurements. The best-diffracting crystal was obtained from condition H9 (Fig. 2) after adding 3 µl 50%(*w*/*v*) PEG 3350 for cryoprotection. Other crystals from the aforementioned conditions diffracted to a much lower resolution (>7 Å) or did not diffract at all.

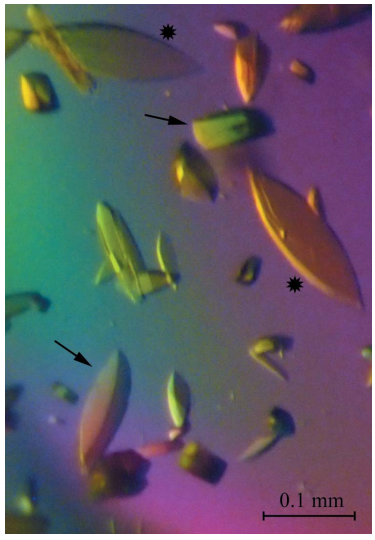


Figure 2
Crystals of pHluorin-HsdR887. Three-dimensional crystals (marked with arrows) were preferred for measurements over flat crystals (marked with asterisks). 20%(*w*/*v*) PEG 3350, 0.2 *M* KH_2PO_4 reservoir solution was mixed with 12 mg ml^{-1} pHluorin-HsdR887 in a 1:1 ratio (the total drop volume was 5 µl); dichloromethane was added to the reservoir solution to a final concentration of 0.025%(*v*/*v*).

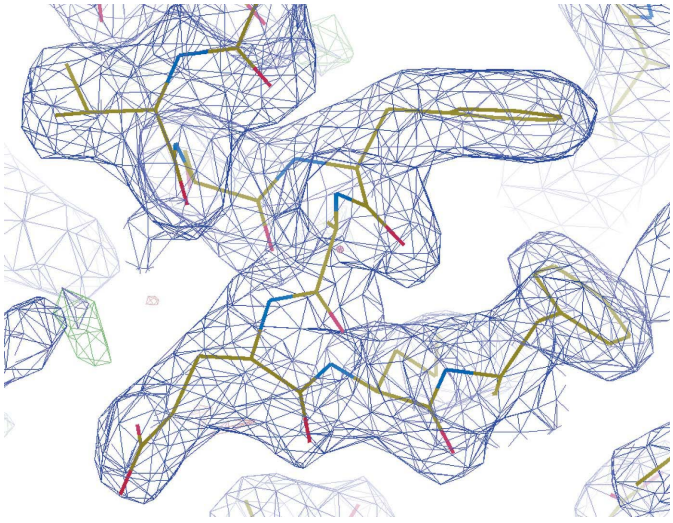


Figure 3
An example of the $2F_o - F_c$ electron-density map, contoured at 1σ, around a fragment of the C-terminal domain.

Table 2
Data-collection and processing statistics for pHluorin-HsdR887.

Values in parentheses are for the outer shell.

Wavelength (Å)	0.9763
Temperature (K)	100
Detector	Pilatus 6M-F
Crystal-to-detector distance (mm)	381.41
Rotation range per image (°)	0.1
Total rotation range (°)	360
Exposure time per image (s)	0.04
Space group	$C222_1$
a, *b*, *c* (Å)	83.42, 176.58, 126.03
α, β, γ (°)	90.00, 90.00, 90.00
Mosaicity (°)	0.158
Resolution range (Å)	20.0–2.45 (2.59–2.45)
Total No. of reflections	451284 (66141)
No. of unique reflections	34663 (5161)
Completeness (%)	98.7 (93.6)
Multiplicity	13.02 (12.82)
$CC_{1/2}$ (%)	99.8 (92.5)
$\langle I/\sigma(I)\rangle$	14.35 (3.41)
R_{meas}† (%)	17.2 (84.9)
Overall *B* factor from Wilson plot (Å^2)	41.27

† R_{meas} is the redundancy-independent merging *R* factor (Diederichs & Karplus, 1997; Weiss & Hilgenfeld, 1997; Weiss *et al.*, 1998). $R_{meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $\langle I(hkl)\rangle$ is the mean of the $N(hkl)$ individual measurements $I_i(hkl)$ of the intensity of reflection *hkl*.

The data set was processed in *XDS* (Table 2). Space group $C222_1$ and unit-cell parameters (Table 2) were determined using *POINTLESS* from the *CCP*4 software package (Winn *et al.*, 2011). The Matthews coefficient (V_M = 2.55 Å^3 Da^{-1}) revealed the presence of two molecules in the asymmetric unit with a solvent content of 50.47%. Molecular-replacement and refinement efforts are currently under way using a GFP structure (PDB entry 1w7s; van Thor *et al.*, 2005) as a template. An example of a $2F_o - F_c$ electron-density map at the current stage of refinement around a short fragment of the C-terminal domain is shown in Fig. 3.

Acknowledgements

The authors gratefully acknowledge support from the Czech Science Foundation (P207/12/2323; http://www.gacr.cz/); NL was supported by the REU Site Molecular Biophysics award (REU1358737) to Jannette Carey, Princeton University, USA. Access to instruments and other facilities was supported by the Czech Research Infrastructure for Systems Biology C4SYS (project No. LM2015055). Data collection was carried out at the PETRA III light source at DESY, a member of the Helmholtz Association (HGF). The authors would like to thank Dr Johanna Kallio for assistance in using beamline P13.

References

Csefalvay, E., Lapkouski, M., Guzanova, A., Csefalvay, L., Baikova, T., Shevelev, I., Bialevich, V., Shamayeva, K., Janscak, P., Kuta Smatanova, I., Panjikar, S., Carey, J., Weiserova, M. & Ettrich, R. (2015). *PLoS One*, **10**, e0128700.
Davies, G. P., Martin, I., Sturrock, S. S., Cronshaw, A., Murray, N. E. & Dryden, D. T. F. (1999). *J. Mol. Biol.* **290**, 565–579.
Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.
Dryden, D. T. F., Murray, N. E. & Rao, D. N. (2001). *Nucleic Acids Res.* **29**, 3728–3741.

electronic reprint

Horiuchi, K. & Zinder, N. D. (1972). *Proc. Natl Acad. Sci. USA*, **69**, 3220–3224.
Janscak, P., Abadjieva, A. & Firman, K. (1996). *J. Mol. Biol.* **257**, 977–991.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 133–144.
Kobe, B., Ve, T. & Williams, S. J. (2015). *Acta Cryst.* F**71**, 861–869.
Lapkouski, M., Panjikar, S., Janscak, P., Kuta Smatanova, I., Carey, J., Ettrich, R. & Csefalvay, E. (2009). *Nature Struct. Mol. Biol.* **16**, 94–95.
Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A. & Wilson, G. G. (2014). *Nucleic Acids Res.* **42**, 20–44.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Miesenböck, G., De Angelis, D. A. & Rothman, J. E. (1998). *Nature (London)*, **394**, 192–195.
Obarska-Kosinska, A., Taylor, J. E., Callow, P., Orlowski, J., Bujnicki, J. M. & Kneale, G. G. (2008). *J. Mol. Biol.* **376**, 438–452.
Qi, D. & Scholthof, K. B. (2008). *J. Virol. Methods*, **149**, 85–90.
Seidel, R., Bloom, J. G., Dekker, C. & Szczelkun, M. D. (2008). *EMBO J.* **27**, 1388–1398.
Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. & Kobe, B. (2003). *Protein Sci.* **12**, 1313–1322.
Thor, J. J. van, Georgiev, G. Y., Towrie, M. & Sage, J. T. (2005). *J. Biol. Chem.* **280**, 33652–33659.
Uyen, N. T., Park, S.-Y., Choi, J.-W., Lee, H.-J., Nishi, K. & Kim, J.-S. (2009). *Nucleic Acids Res.* **37**, 6960–6969.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203–205.
Weiss, M. S., Metzner, H. J. & Hilgenfeld, R. (1998). *FEBS Lett.* **423**, 291–296.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). *Gene*, **33**, 103–119.
Youell, J. & Firman, K. (2008). *Microbiol. Mol. Biol. Rev.* **72**, 365–377.

electronic reprint