Jia Quyen Truong, Santosh Panjikar, Linda Shearwin-Whyatt, John B. Bruning and Keith E. Shearwin
Combining random microseed matrix screening and the magic triangle for the efficient structure solution of a potential lysin from bacteriophage P68
Acta crystallographica Section D. Structural biology, 2019; 75(7):670-681

Originally published at: http://doi.org/10.1107/s2059798319009008

STRUCTURAL BIOLOGY
ISSN 2059-7983

Combining random microseed matrix screening and the magic triangle for the efficient structure solution of a potential lysin from bacteriophage P68

Jia Quyen Truong,[a] Santosh Panjikar,[b] Linda Shearwin-Whyatt,[a] John B. Bruning[c] and Keith E. Shearwin[a]*

[a]School of Biological Sciences, The University of Adelaide, North Terrace, Adelaide, South Australia 5005, Australia, [b]MX, Australian Synchrotron, 800 Blackburn Road Clayton, Melbourne, VIC 3168, Australia, and [c]Institute of Photonics and Advanced Sensing, School of Biological Sciences, The University of Adelaide, North Terrace, Adelaide, South Australia 5005, Australia. *Correspondence e-mail: keith.shearwin@adelaide.edu.au

Received 1 March 2019
Accepted 24 June 2019

Edited by B. Kobe, University of Queensland, Australia

Keywords: crystallography; seeding; random microseed matrix screening; I3C; magic triangle; phasing; lysin.

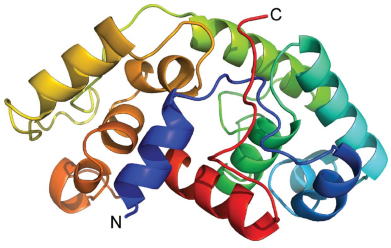
PDB references: lysin protein from *Staphylococcus* phage P68, 6o43; hen egg-white lysozyme, 6pbb

Supporting information: this article has supporting information at journals.iucr.org/d

Two commonly encountered bottlenecks in the structure determination of a protein by X-ray crystallography are screening for conditions that give high-quality crystals and, in the case of novel structures, finding derivatization conditions for experimental phasing. In this study, the phasing molecule 5-amino-2,4,6-triiodoisophthalic acid (I3C) was added to a random microseed matrix screen to generate high-quality crystals derivatized with I3C in a single optimization experiment. I3C, often referred to as the magic triangle, contains an aromatic ring scaffold with three bound I atoms. This approach was applied to efficiently phase the structures of hen egg-white lysozyme and the N-terminal domain of the Orf11 protein from *Staphylococcus* phage P68 (Orf11 NTD) using SAD phasing. The structure of Orf11 NTD suggests that it may play a role as a virion-associated lysin or endolysin.

1. Introduction

X-ray crystallography is commonly used to determine the three-dimensional structures of proteins, with over 148 000 structures deposited in the Protein Data Bank (PDB) as of February 2019 (Burley *et al.*, 2019). However, there are many bottlenecks in successful structure determination. For a protein for which no X-ray crystal structure exists, the conditions that will give rise to crystals are not known *a priori*. Many crystallization conditions are screened to find one that provides diffracting crystals. However, the large number of factors that affect crystallization, including protein concentration, precipitant, buffer, salt concentrations and temperature, make an exhaustive screening expensive and laborious (Jancarik & Kim, 1991). Sparse-matrix screens are often used to efficiently find crystallization conditions (Jancarik & Kim, 1991). Sparse matrix is a data-mining approach that chooses crystallization conditions based on known and published crystallization conditions. It can be generalized or tailored to the macromolecule, for example membrane proteins or DNA-binding proteins. Many sparse-matrix screens are readily available commercially. Ireton & Stoddard (2004) pioneered the microseed matrix screening (MMS) technique of transferring crystal seeds from existing crystals to new crystallization conditions to create new conditions in which crystals can grow. D'Arcy *et al.* (2007) further developed the technique for use with random crystallization screens, such as the sparse-matrix screens already described. This approach was named random microseed matrix screening (rMMS) and was found



to be an effective method for increasing the number of hit crystallization conditions. In many cases, improved diffraction-quality crystals could be obtained.

A protein structure can only be obtained using X-ray crystallography if the phase problem can be solved. This presents another bottleneck. If a suitable search model is available through a homologous structure, the phase problem can be solved by molecular replacement (MR; Rossmann, 1990; McCoy *et al.*, 2007; Bibby *et al.*, 2012). However, many protein targets being studied do not have a suitable template and require experimental phasing.

Several experimental phasing techniques have been developed. The traditional phasing method is isomorphous replacement, which involves soaking crystals in solutions containing heavy atoms such as lead, mercury or uranium. Another phasing technique is anomalous dispersion at a single wavelength (SAD; Wang, 1985) or multiple wavelengths (MAD; Hendrickson, 1991). The anomalous dispersion technique requires atoms to be incorporated into the crystal lattice that scatter anomalously at the wavelengths available at the X-ray source. These can be incorporated through soaks or co-crystallization with heavy-atom solutions, chemical modification (for example selenium modification of nucleic acids) or labeling during protein expression (selenomethonine and selenocysteine). The theories of isomorphous replacement and anomalous dispersion are described in Taylor (2010). Suitable compounds and conditions for heavy-atom derivatization are found empirically by screening. This process can be very laborious, although some rational approaches have been devised to guide the screening process (Lu & Sun, 2014).

Beck *et al.* (2008, 2010) synthesized the 'magic triangle' (I3C) and 'MAD triangle' (B3C) compounds specifically for phasing, showing that they could phase several model proteins (Fig. 1). Currently, there are 19 structures in the Protein Data Bank that have been solved using I3C. The triangular arrangement of anomalous scatterers can easily be identified in the substructure. The two carboxylate groups and one amino group can interact with the protein through hydrogen bonding to either the protein backbone or to side chains to facilitate more specific binding.

In this study, we incorporated I3C into an rMMS screen to increase the chance of obtaining diffraction-quality crystals while simultaneously incorporating heavy atoms into the crystal for experimental phase determination. This method aims to overcome the two identified bottlenecks in structure determination of novel targets in one step. We demonstrated this technique using hen egg-white lysozyme (HEWL) and the N-terminal domain of Orf11 from *Staphylococcus* bacteriophage P68 (Orf11 NTD). Orf11 NTD is a good example of a target that does not have a suitable template in the PDB to use for MR and that did not produce sufficient hit conditions in an initial screen. Our devised method very quickly gave diffraction-quality crystals that were already derivatized with a heavy-atom compound and allowed the structure to be solved.

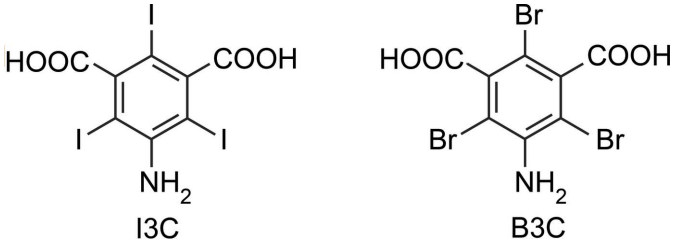


Figure 1
Chemical structures of I3C (5-amino-2,4,6-triiodoisophthalic acid) and B3C (5-amino-2,4,6-tribromoisophthalic acid).

2. Methods

2.1. Crystallization of HEWL

HEWL was commercially acquired from Sigma–Aldrich (catalog No. L6876) as a lyophilized powder. The powder was dissolved in TBS (50 m*M* Tris–HCl pH 7.6, 150 m*M* NaCl) to a final concentration of 30 mg ml^{-1} as determined using UV absorbance at 280 nm.

Lysozyme crystals were grown via hanging-drop vapor diffusion. 1 µl protein solution was mixed with 1 µl reservoir solution [0.2 *M* ammonium tartrate dibasic pH 7.0, 20%(*w*/*v*) polyethylene glycol 3350] and equilibrated against 500 µl reservoir solution. The crystals obtained were crushed to generate an rMMS seed stock as described by D'Arcy *et al.* (2007).

Sitting-drop vapor-diffusion crystallization screening of HEWL was performed in 96-well Intelli-Plates (Art Robbins) using the commercial Index HT screen (Hampton Research). Four crystallization screens were carried out corresponding to lysozyme without seeds or I3C, lysozyme with seeds, lysozyme with I3C, and lysozyme with seeds and I3C.

For screens containing I3C, 1 *M* I3C stock was directly added to the protein to give a final concentration of 20 m*M* I3C. Sitting-drop vapor-diffusion trays were set up with the I3C-containing protein stock. 1 µl protein solution was added to 1 µl reservoir solution and 0.1 µl seed stock and equilibrated over 75 µl reservoir solution.

A seeded crystal from condition C6 of Index HT grown in the presence of I3C showed suitable diffraction. These crystals appeared after three days and reached their maximum size within two weeks. The size of the crystals was estimated to be 150 µm. Crystals were mounted on cryoloops (Hampton Research), passed through Paratone (Hampton Research) for cryoprotection and flash-cooled in liquid nitrogen for data collection (Teng, 1990).

2.2. Data collection, structure solution and refinement of HEWL

A 1.87 Å resolution SAD data set was collected at a wavelength of 1.459 Å using an ADSC Quantum 210r CCD detector on the macromolecular beamline MX1 at the Australian Synchrotron (McPhillips *et al.*, 2002). This wavelength allows iodine to have a large anomalous signal (f'' = 6.3 e and f' = −0.13 e). 360 diffraction images with 1° oscillation width were collected at a crystal-to-detector distance of 120 mm. The diffraction data from two crystals were processed

Table 1
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Protein	HEWL	Orf11 NTD
Diffraction source	MX1, Australian Synchrotron	MX1, Australian Synchrotron
Wavelength (Å)	1.459	1.459
Temperature (K)	100	100
Detector	ADSC Quantum 210r CCD detector	ADSC Quantum 210r CCD detector
Rotation range per image (°)	1	1
Total rotation range (°)	360	360
Space group	$P4_32{1}2$	$P3_121$
a, b, c (Å)	77.16, 77.16, 38.20	61.397, 61.397, 101.606
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution range (Å)	38.20–1.87 (1.92–1.87)	36.73–2.08 (2.13–2.08)
Total No. of reflections	428577 (7041)	267451 (14604)
No. of unique reflections	9810 (475)	13689 (853)
Completeness (%)	98.4 (76.4)	98.3 (80.0)
Multiplicity	43.7 (14.8)	19.5 (17.1)
$\langle I/\sigma(I)\rangle$	23.0 (1.8)	22.2 (3.5)
Wilson B factor (Å^2)	17.501	42.71
$R_{\rm p.i.m.}$	0.019 (0.198)	0.013 (0.134)
$R_{\rm meas}$	0.127 (0.850)	0.060 (0.578)
$R_{\rm anom}$	0.0777 (0.171)	0.0528 (0.160)
$CC_{\rm anom}$	0.452 (−0.120)	0.291 (−0.038)
$CC_{1/2}$	0.999 (0.938)	1 (0.983)

Table 2
Structure-solution statistics.

Values in parentheses are for the highest resolution shell.

Protein	HEWL	Orf11 NTD
Molecular mass (kDa)	14.3	22.8
Reflections used in refinement	9547 (921)	13296 (1219)
Reflections used for $R_{\rm free}$	913 (97)	893 (94)
$R_{\rm work}$	0.1967 (0.3442)	0.1792 (0.2498)
$R_{\rm free}$	0.2404 (0.3874)	0.2245 (0.3030)
$CC_{\rm work}$	0.948 (0.887)	0.966 (0.910)
$CC_{\rm free}$	0.923 (0.778)	0.938 (0.866)
No. of non-H atoms		
Total	1142	1675
Macromolecules	1004	1535
I3C	64	16
Solvent	73	124
No. of protein residues	129	199
R.m.s.d., bonds (Å)	0.003	0.003
R.m.s.d., angles (°)	0.54	0.50
Ramachandran favored (%)	99.21	96.95
Ramachandran allowed (%)	0.79	3.05
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.00	2.00
Clashscore	5.39	3.32
Average B factor (Å^2)		
Overall	36.01	54.11
Macromolecules	33.76	53.53
Ligands	67.83	90.66
Solvent	38.66	56.48
No. of TLS groups	5	5

using *XDS* (Kabsch, 2010) and combined and scaled using *AIMLESS* (Evans & Murshudov, 2013) (Table 1).

The structure was solved using the SAD protocol of *Auto-Rickshaw*, the EMBL Hamburg automated crystal structure-determination platform (Panjikar *et al.*, 2005). The diffraction data processed using *AIMLESS* (Evans & Murshudov, 2013) were used as input. $F_{\rm A}$ values were calculated using *SHELXC* (Sheldrick, 2015). Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 1.8 Å. 17 potential heavy-atom sites out of the maximum number of 20 heavy atoms requested were found using *SHELXD* (Schneider & Sheldrick, 2002). The correct hand for the substructure was determined using *ABS* (Hao, 2004) and *SHELXE* (Sheldrick, 2002). Initial phases were calculated after density modification using *SHELXE* (Sheldrick, 2002). 89.48% of the model was built using *ARP*/*wARP* (Perrakis *et al.*, 1999; Morris *et al.*, 2004).

The resulting structure was used as a starting model in the MRSAD module (Panjikar *et al.*, 2009) of *Auto-Rickshaw* in space group $P4_32{1}2$ for further phase improvement, model completion and refinement. The molecular-replacement step of the pipeline was skipped. Refinement of the structure was carried out using *CNS* (Brünger *et al.*, 1998), *REFMAC*5 (Murshudov *et al.*, 2011) and *phenix.refine* (Afonine *et al.*, 2012) within *Auto-Rickshaw*. A search for and refinement of heavy atoms were conducted using *Phaser* (McCoy *et al.*, 2007) and *MLPHARE* (Winn *et al.*, 2011), which identified 13 sites. Nine sites were subsequently determined to correspond to the three I atoms from each of three I3C ligands, and four sites corresponded to the S atoms of two methionine and two cysteine residues in the structure. Density modification was performed in *Pirate* (Cowtan, 2000). Model building was conducted using *SHELXE* (Sheldrick, 2002), *RESOLVE* (Terwilliger, 1999, 2000) and *Buccaneer* (Cowtan, 2006), which collectively built 128 out of 129 residues that were correctly docked in the electron density. *REFMAC*5 and *phenix.refine* were used to further refine the structure, which resulted in an R and $R_{\rm free}$ of 31.48% and 34.53%, respectively. The structure was then iteratively rebuilt and refined using *Coot* (Emsley & Cowtan, 2004) and *phenix.refine* to an R and $R_{\rm free}$ of 19.67% and 24.04%, respectively. Structure-solution statistics are summarized in Table 2.

2.3. Expression of Orf11 NTD

Staphylococcus phage P68 Orf11 N-terminal domain (residues 2–200) was expressed in *Escherichia coli* BL21(λDE3) cells carrying the pLysS and pET-15b Orf11 NTD plasmids (strain JT438). JT438 cells were grown overnight in 5 ml LB medium supplemented with 100 µg ml^{-1} ampicillin. The cells were diluted 1:200 into fresh prewarmed LB medium supplemented with 100 µg ml^{-1} ampicillin. The culture was grown at 37°C with shaking until an OD_{600} of 0.6 was reached and were then induced overnight at 16°C using isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 200 µ*M*. The cells were harvested by centrifugation at 5000*g* for 20 min. The cell pellets were stored at −80°C.

2.4. Purification of Orf11 NTD

The cell pellets were thawed and resuspended in START buffer [20 m*M* Tris pH 7.6, 500 m*M* NaCl, 20 m*M* imidazole, 5 m*M* β-mercaptoethanol (BME) pH 7.2]. Cell resuspensions were lysed by sonication on ice (5 × 30 s, 50% duty cycle,

Sonifier Cell Disruptor B-30). The lysate was clarified by centrifugation at 40 000*g* for 1 h and subsequent filtration through 0.45 and 0.2 µm syringe filters (Sartorius Minisart) using a disposable syringe. The supernatant was subjected to Ni^{2+}–IMAC chromatography on a 5 ml HisTrap FF column (GE Healthcare) using an NGC FPLC system (Bio-Rad). The column was washed with ten column volumes of START buffer. The protein was eluted from the column with START buffer with an imidazole gradient from 70 to 220 m*M* over eight column volumes, followed by an imidazole gradient from 220 to 500 m*M* over two column volumes. Fractions showing UV absorbance A_{280} above baseline were tested for purity using SDS–PAGE. Fractions containing the purified protein were pooled and dialysed against 25 m*M* Tris pH 7.5, 150 m*M* NaCl, 10% glycerol, 5 m*M* BME. The protein was concentrated to 50 mg ml^{-1} using Amicon Ultra-15 centrifugation units (Millipore) and was stored at −80°C.

2.5. Crystallization of Orf11 NTD using the rMMS protocol in the presence of I3C

Crystallization screening for the Orf11 NTD protein was conducted at 16°C by the sitting-drop vapor-diffusion method in 96-well Intelli-Plates (Art Robbins) using the commercial PEG/Ion HT screen (Hampton Research). The stored protein was diluted to 30 mg ml^{-1} in 25 m*M* Tris pH 7.5, 150 m*M* NaCl, 10% glycerol, 5 m*M* BME for crystallization. 1 µl protein solution was added to 1 µl reservoir solution and equilibrated over 75 µl reservoir solution. Crystals from condition G12 of PEG/Ion HT were used to generate an rMMS seed stock, as described by D'Arcy *et al.* (2007). A 1 *M* stock of lithium I3C solution was prepared as described in Beck *et al.* (2008). The protein stock was diluted to 30 mg ml^{-1} using heavy-atom buffer (25 m*M* Tris pH 7.5, 150 m*M* NaCl, 10% glycerol, 5 m*M* BME, 50 m*M* I3C). The protein stock contained I3C at a final concentration of 20 m*M*. Sitting-drop vapor-diffusion trays were set up with the I3C-containing protein stock. 1 µl protein solution was added to 1 µl reservoir solution and 0.1 µl seed stock and was equilibrated over 75 µl reservoir solution. For crystallization screens without I3C, I3C was omitted from the buffer used to dilute the protein. The sizes of the crystals used for data acquisition were estimated to be between 50 and 75 µm. Crystals appeared after one day and reached their maximum size within a week. If too many crystals were observed in the drop, the seed stock was diluted with PEG/Ion HT condition G12 reservoir solution in subsequent optimization steps.

Crystals were mounted on cryoloops (Hampton Research), passed through Paratone (Hampton Research) for cryoprotection and flash-cooled in liquid nitrogen for data collection (Teng, 1990).

2.6. Data collection, structure solution and refinement of Orf11 NTD

A seeded crystal from condition H3 of PEG/Ion HT grown in the presence of I3C showed suitable diffraction. A 2.0 Å resolution SAD data set was collected at a wavelength of 1.459 Å using an ADSC Quantum 210r CCD detector on the macromolecular beamline MX1 at the Australian Synchrotron (McPhillips *et al.*, 2002). At this wavelength, iodine has an f'' and f' of 6.3 and −0.13 e, respectively. 360 diffraction images with 1° oscillation width were collected at a crystal-to-detector distance of 120 mm. The diffraction data were processed using *XDS* (Kabsch, 2010) and scaled with *AIMLESS* (Evans & Murshudov, 2013) (Table 1).

The structure was solved using the SAD protocol of *Auto-Rickshaw* (Panjikar *et al.*, 2005). The input diffraction data were prepared and converted for use in *Auto-Rickshaw* using programs from the *CCP*4 suite (Winn *et al.*, 2011). F_A values were calculated using *SHELXC* (Sheldrick, 2015). Based on an initial analysis of the data, the maximum resolution for substructure determination and initial phase calculation was set to 3.4 Å. Five heavy atoms out of the maximum number of nine heavy atoms requested were found using *SHELXD* (Schneider & Sheldrick, 2002). The correct hand for the substructure was determined using *ABS* (Hao, 2004) and *SHELXE* (Sheldrick, 2002). Based on the analysis of the hand of the substructure, the space group of the data was changed from $P3_221$ to $P3_121$. Initial phases were calculated after density modification using *SHELXE* (Sheldrick, 2002). 84.54% of the model was built using *ARP*/*wARP* (Perrakis *et al.*, 1999; Morris *et al.*, 2004).

The resulting structure was used as a starting model in the MRSAD module (Panjikar *et al.*, 2009) of *Auto-Rickshaw* in space group $P3_121$ for further phase improvement, model completion and refinement. Refinement of the structure was carried out using *CNS* (Brünger *et al.*, 1998), *REFMAC*5 (Murshudov *et al.*, 2011) and *phenix.refine* (Afonine *et al.*, 2012) within *Auto-Rickshaw*. A search for and refinement of heavy atoms were conducted using *Phaser* (McCoy *et al.*, 2007) and *MLPHARE* (Winn *et al.*, 2011), which identified four sites. Three sites were subsequently determined to correspond to the three I atoms from a single I3C, and one site corresponded to the S atom of a methionine. Density modification was performed in *Pirate* (Cowtan, 2000). Model building was conducted using *SHELXE* (Sheldrick, 2002), *RESOLVE* (Terwilliger, 1999, 2000) and *Buccaneer* (Cowtan, 2006), which collectively built 194 out of 207 residues that were correctly docked in the electron density. *REFMAC*5 and *phenix.refine* were used to further refine the structure, which resulted in an R and R_{free} of 26.75% and 32.27%, respectively. The structure was then iteratively rebuilt and refined using *Coot* (Emsley & Cowtan, 2004) and *phenix.refine* to an R and R_{free} of 17.34% and 22.64%, respectively. Structure-solution statistics are summarized in Table 2.

2.7. PDB codes

The coordinates and structure factors for Orf11 NTD and HEWL have been deposited in the Protein Data Bank as entries 6o43 and 6pbb, respectively.

3. Results

3.1. I3C can be added to a random microseed matrix screen and still give rise to more crystals than unseeded protein

The effect of adding I3C to an rMMS optimization screen was tested using HEWL as a model crystallization protein. rMMS optimization has previously been successfully applied to lysozyme to increase the number of conditions supporting crystal growth in the JCSG, PACT and Morpheus screens (Till *et al.*, 2013). Consistent with previous studies, we observe that rMMS increases the number of conditions in which lysozyme can crystallize in the Index HT screen (Hampton Research). The rMMS screen generated 35 more conditions that supported crystal growth [Fig. 2(*a*)]. Adding I3C without seeding to the screen did not notably increase the number of crystallization conditions found. Eight conditions no longer

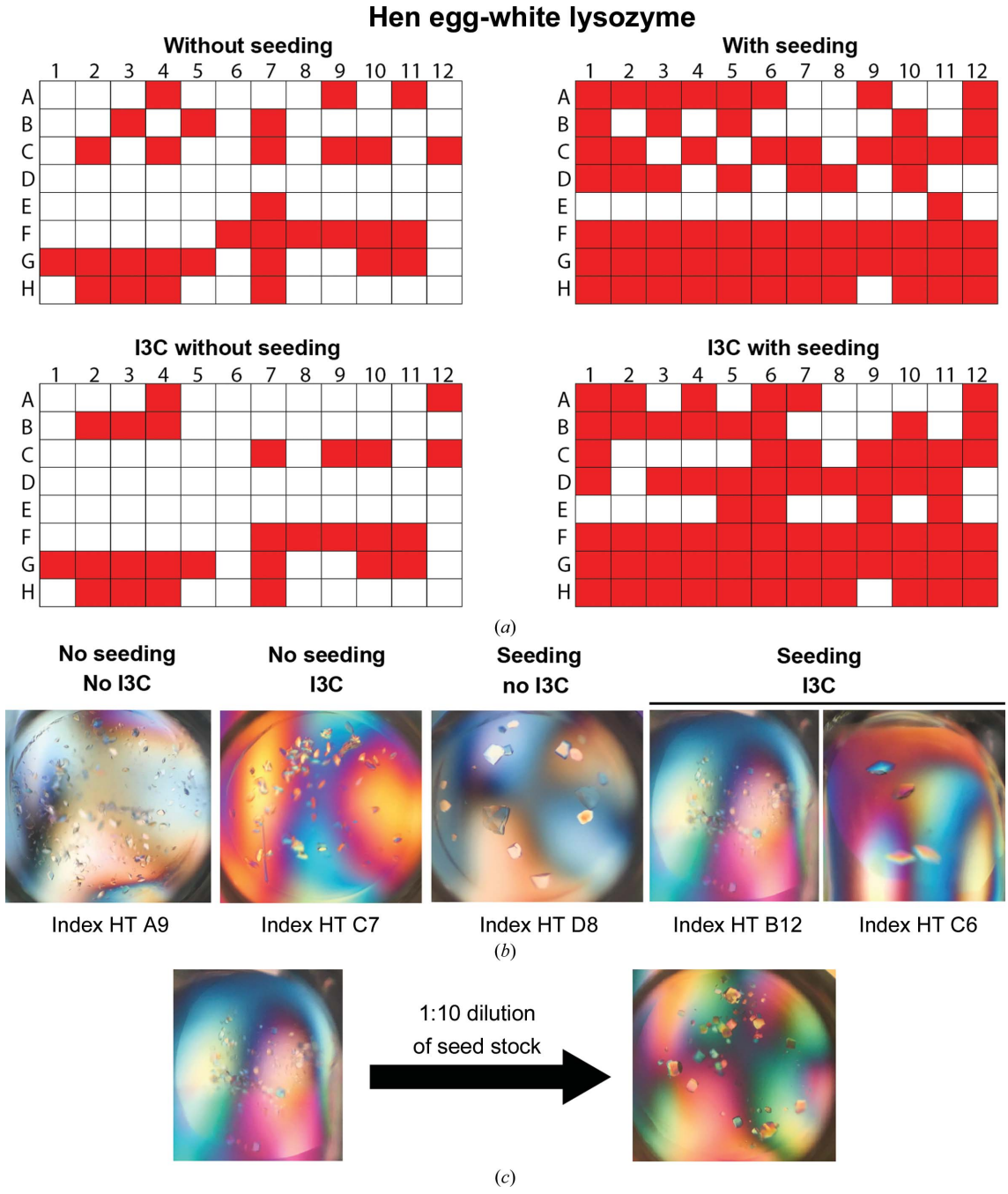


Figure 2
rMMS can generate new conditions that support the crystal growth of HEWL in the presence of I3C. (*a*) Crystallization conditions supporting the crystal growth of HEWL in the Index HT screen are shaded in red. Seed stock was made from lysozyme crystals grown in 0.2 *M* ammonium tartrate dibasic pH 7.0, 20%(*w*/*v*) polyethylene glycol 3350. (*b*) Representative photographs showing crystals that formed with and without seeding both in the presence and absence of I3C. (*c*) Dilution of seed stock was used to reduce the number of seeds to influence the crystal size. Seed stock was diluted 1:10 to give fewer and larger crystals in condition B12 of Index HT.

supported crystal growth and three new conditions were found in the presence of I3C. Adding I3C to the lysozyme rMMS optimization screen generated substantially more conditions than screens without seeding with and without I3C [Fig. 2(*a*)]. Having used HEWL as a test case, we tested this technique with a protein with significantly lower crystallizability, Orf11 NTD.

Only one hit condition was obtained from the initial screen with Orf11 NTD using the PEG/Ion HT screen (Hampton Research). This condition generated a shower of microcrystals that were too small to mount for X-ray diffraction. Adding I3C to the screen resulted in the same hit condition but gave no new conditions. The crystalline material from the unseeded screen was used to make a seed stock for an rMMS screen using PEG/Ion HT with and without I3C added. Five and six new crystallization hits were identified in these optimization screens, respectively [Fig. 3(*a*)].

The new hit condition H3 from PEG/Ion found by adding undiluted seed stock produced too many crystals and was optimized for crystal number by the dilution of seed stock as detailed in Fig. 2. The growth of fewer crystals resulted in larger crystals [Fig. 3(*c*)]. These two screens using HEWL and

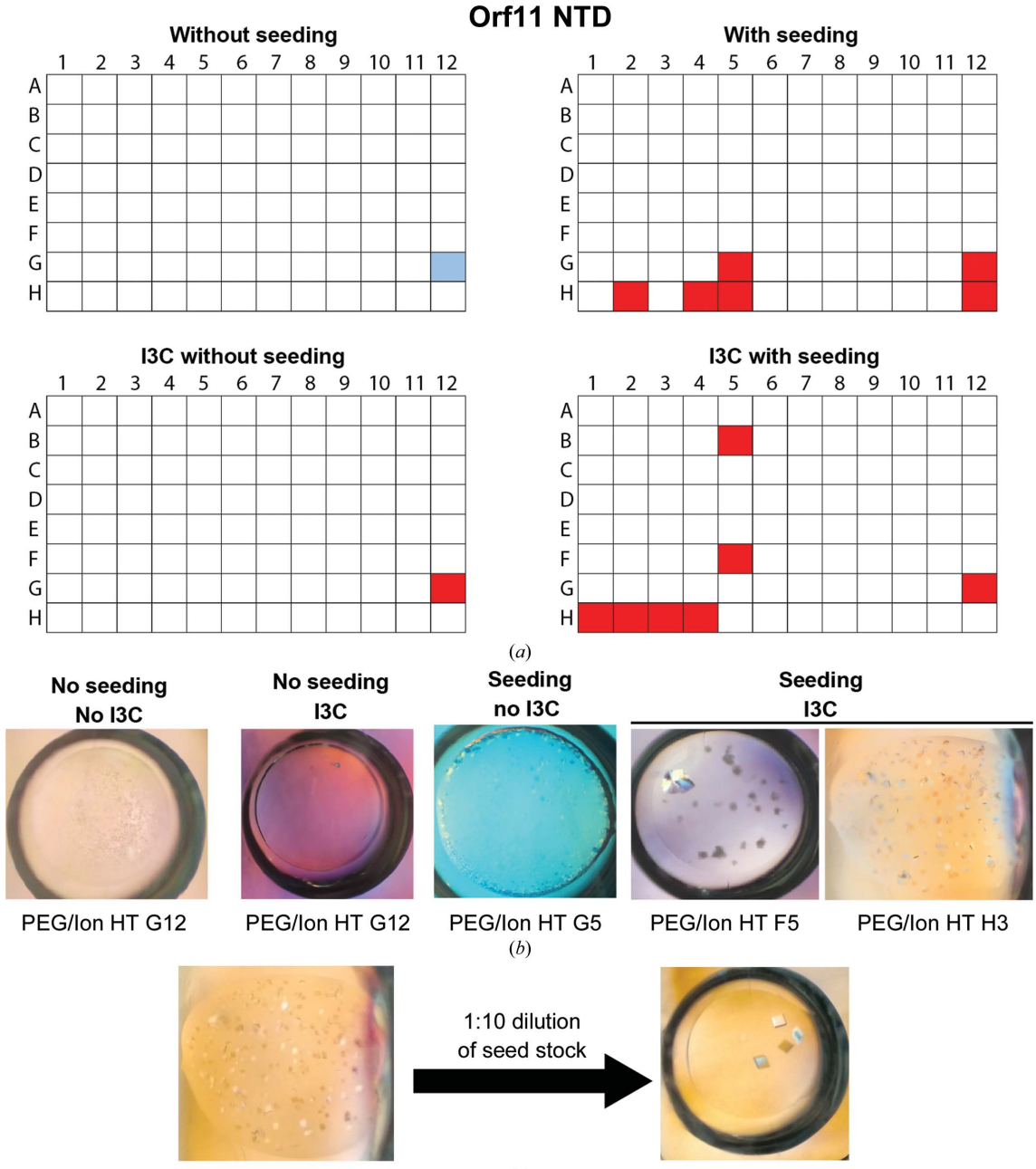


Figure 3
rMMS was used to find new hit conditions that support the crystal growth of Orf11 NTD in the presence of I3C. (*a*) Crystallization conditions supporting crystal growth of Orf11 NTD in the PEG/Ion HT screen are colored. The crystallization condition used for seed stock is shown in blue. (*b*) Representative photographs showing Orf11 NTD crystals that formed without and with seeding. (*c*) Dilution of seed stock can reduce excess nucleation to give larger crystals. Seed stock was diluted 1:10 to give fewer but larger crystals in condition H3 of PEG/Ion HT.

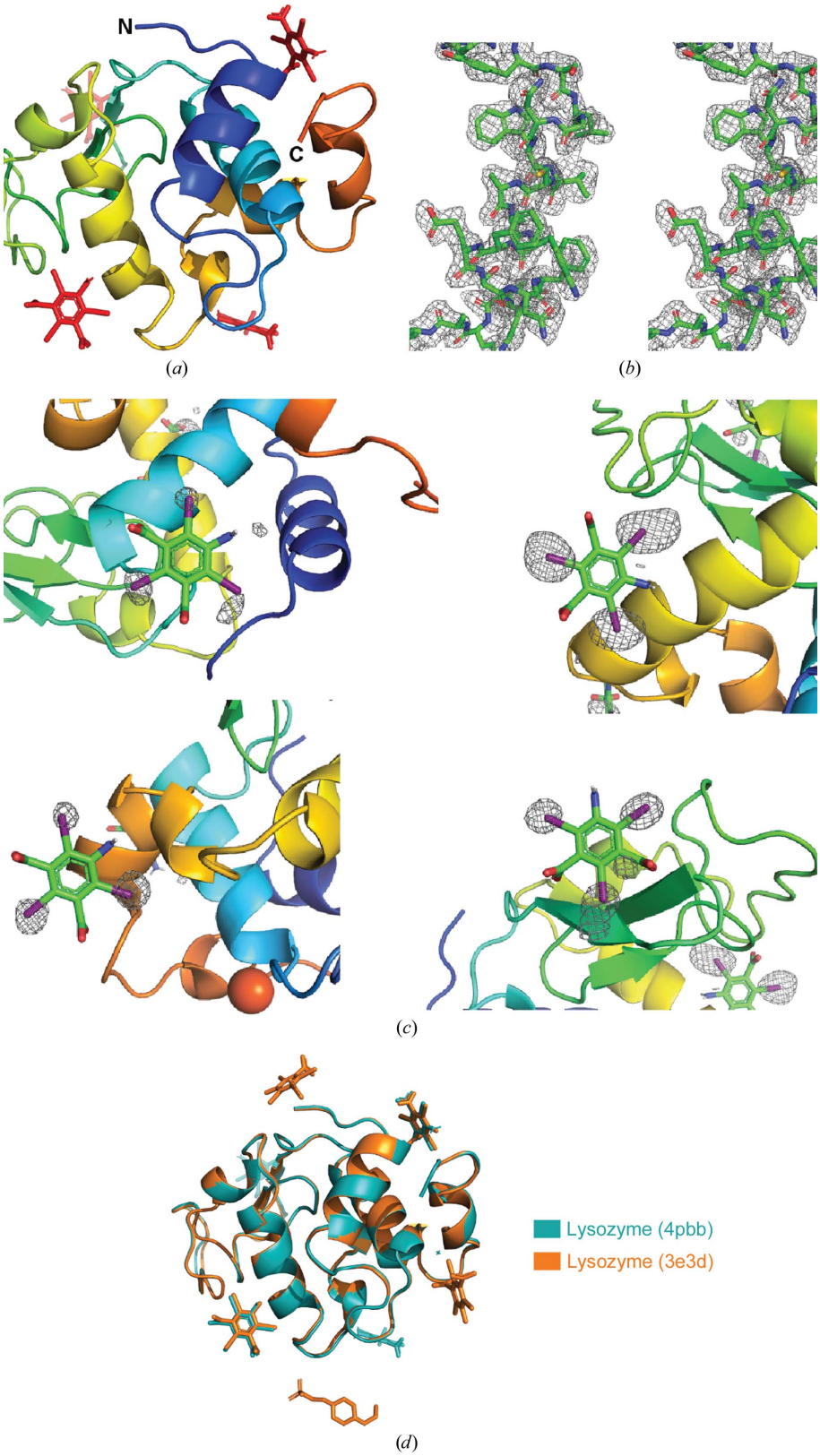


Figure 4
Structure solution of HEWL. (*a*) The crystal structure of HEWL. (*b*) Stereoview of a composite OMIT $2mF_o - DF_c$ electron-density map of HEWL. The contour level was set to 1σ. (*c*) Substructure density of each I3C molecule. The anomalous difference map is contoured at 5σ. The map was generated using *phenix.maps*. (*d*) The crystal structure of HEWL (PDB entry 6pbb) was superimposed onto a previously solved structure of lysozyme soaked with I3C (PDB entry 3e3d) and shows that two I3C molecules are bound in the same position.

Orf11 NTD provide a proof of concept that I3C-derivatized crystals can be grown and optimized using seeding from a nonderivatized crystal.

To confirm that I3C was incorporated into the crystals, crystals from the I3C rMMS screen were harvested and the crystal structures of the two proteins were solved using the anomalous signal without any external template information.

3.2. Structure of HEWL

The structure of HEWL was solved by SAD phasing using diffraction data from a crystal grown using condition C6 of Index HT [Fig. 4(*a*)]. The structure showed four I3C molecules bound to a single HEWL monomer in the asymmetric unit, with occupancies of 57, 54, 32 and 26% [Fig. 4(*c*)]. A structural superposition of the structure with a previously solved structure of HEWL from a crystal co-crystallized with I3C (PDB entry 3e3d; Beck *et al.*, 2008) showed that two of the four I3C molecules bind to the same positions in the protein [Fig. 4(*d*)].

3.3. Structure of Orf11 NTD

The structure of Orf11 NTD was obtained by SAD phasing using diffraction data from a crystal grown using condition H3 of PEG/Ion HT [Fig. 3(*c*)] in an rMMS screen in the presence of I3C. This structure is shown in Fig. 5(*a*), with representative electron density shown in Fig. 5(*b*). One I3C molecule was bound to the protein with an occupancy of 53% and provided sufficient signal to phase the structure. An intrinsic S atom from Met138 provided a fourth anomalous scattering atom. All scattering atoms displayed clear density in the anomalous difference map [Figs. 5(*c*) and 5(*d*)].

4. Discussion

4.1. The I3C–rMMS method shows promise

Molecular-replacement (MR) phasing is the most popular method for solving protein structures; however, in many cases the protein target lacks a suitable homology or *ab initio* model for

successful MR phasing. In such cases, experimental phasing is the method of choice for solving the phase problem. For proteins that do not intrinsically possess scattering atoms, external anomalous scatterers can be incorporated into the crystal by soaking or co-crystallizing the protein with heavy atoms. If heavy-atom ions or small molecules are incorporated into a crystal via co-crystallization, it usually results in higher occupancy. Another benefit of co-crystallization over soaking is that it reduces manual crystal-manipulation steps that may damage the crystal.

However, soaking is often preferred over co-crystallization to derivatize crystals. Co-crystallization requires the set up of additional screens. In addition, attempting to co-crystallize a heavy-atom ligand in the same condition that yields underivatized crystals can fail to yield crystals. This observation is unsurprising as the ligand can change the equilibrium of the crystallization condition or the crystal contacts (Garman & Murray, 2003; McPherson & Cudney, 2014). In this study, we present an efficient method for screening for crystallization conditions that co-crystallize a phasing compound into a protein crystal.

rMMS often substantially increases the number of hit conditions and produces improved diffraction-quality crystals in new crystallization conditions (D'Arcy *et al.*, 2007; Obmolova *et al.*, 2010). I3C has previously been demonstrated to bind to proteins within a protein crystal, often via multiple functional groups. The amino group and carboxyl groups of I3C can form hydrogen bonds directly to the protein backbone and side chains or via water bridges to the protein. The benzene ring and iodine allow π–π interactions and halogen bonds, respectively. The ability to form multiple interactions results in I3C binding with high specificity and occupancy in a protein crystal (Beck *et al.*, 2008).

It is possible that I3C can also stabilize a lattice and/or generate new crystal contacts, resulting in a protein crystal with improved mechanical and diffraction properties. In the HEWL crystal, three of the four I3C molecules interact with three different protein monomers [Fig. 6(*a*)]. In the Orf11 NTD crystal, I3C was found at the interface between three Orf11 NTD molecules and makes interactions with all three protein molecules [Fig. 6(*b*)]. Such contacts between the protein molecules bridged by I3C may explain why this crystallization condition was only found when rMMS was combined with I3C and not with rMMS alone. We are aware of one published case in which protein crystals would only form in the presence of I3C and would fail without it (Leverrier *et al.*, 2011). Thus, it seems likely that I3C can increase the number of hits found by rMMS by creating new crystal contacts. In this study, a small increase in the number of hits was found by adding I3C to rMMS over a standard rMMS screen. However, this is not a clear positive effect as some hit conditions were lost upon adding I3C.

By adding I3C to the protein solution and using rMMS, we are efficiently searching for crystallization conditions in which I3C can bind the protein. rMMS increases the likelihood of obtaining crystals, and any such protein

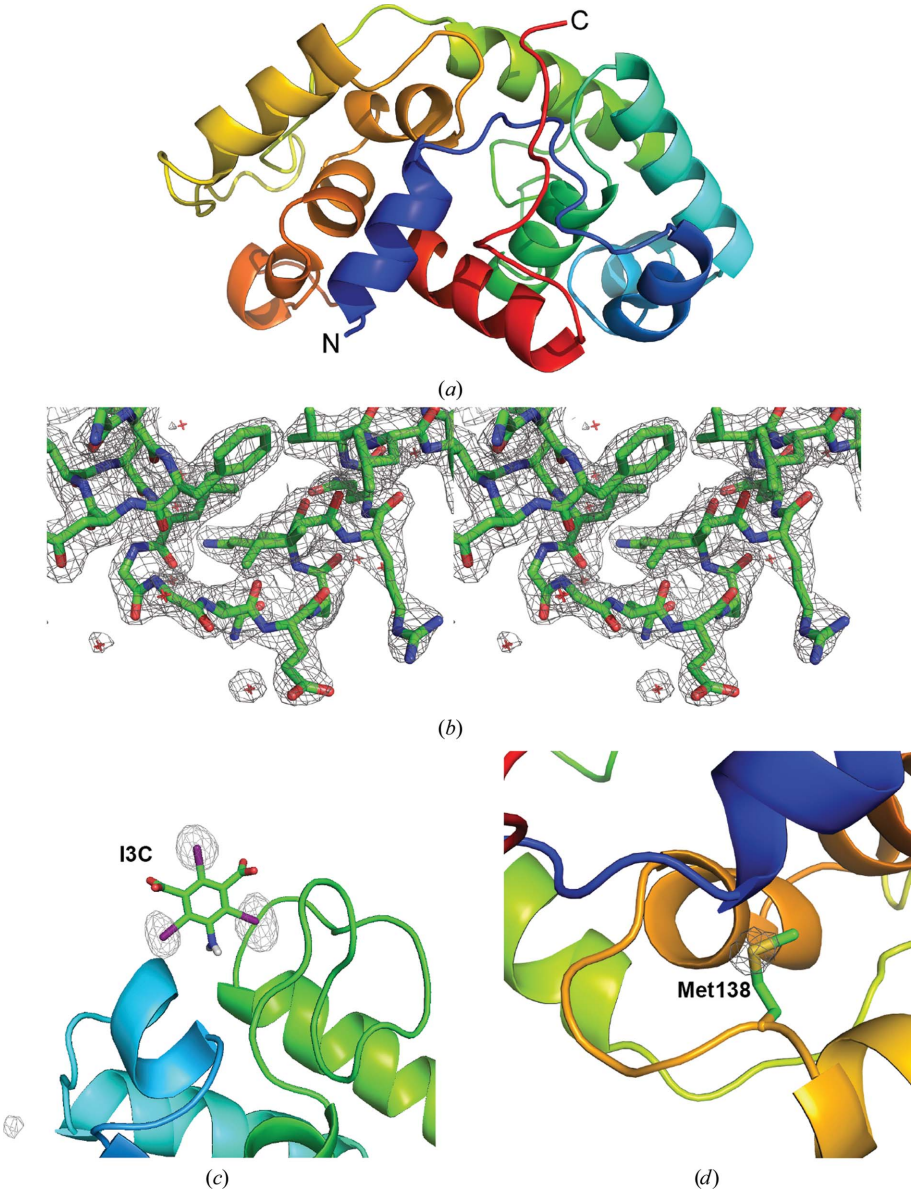


Figure 5
Structure solution of Orf11 NTD. (*a*) The crystal structure of Orf11 NTD. (*b*) Stereoview of a composite OMIT $2mF_o - DF_c$ electron-density map of Orf11 NTD. The contour level was set to 1σ. Substructure density of I3C and the intrinsic S atom are shown in (*c*) and (*d*), respectively. The anomalous difference map is contoured at 5σ. The map was generated using *phenix.maps*.

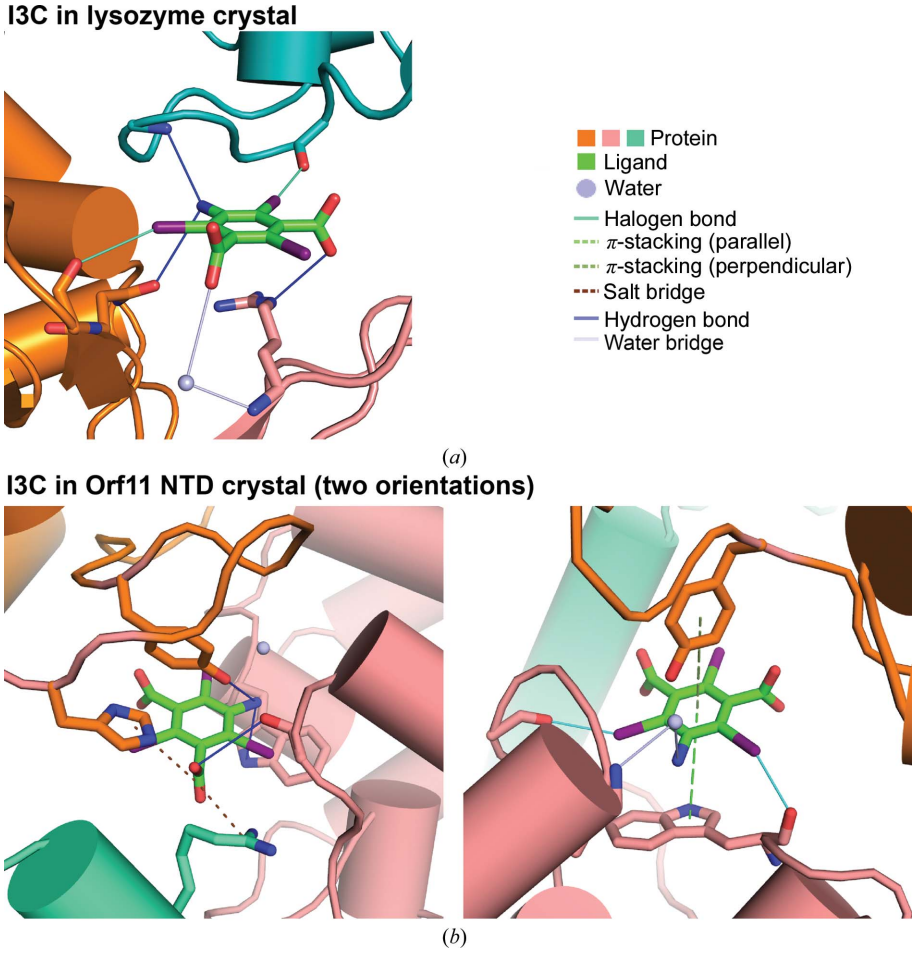


Figure 6
I3C can mediate contacts between protein molecules in a crystal. One I3C molecule from (*a*) the HEWL crystal and (*b*) the Orf11 NTD crystal are shown. The same I3C molecule from the Orf11 NTD crystal is displayed in two different orientations for clarity. I3C–protein interactions include hydrogen bonding, π–π stacking interactions, salt bridges, water bridges and halogen bonding. In both cases, each I3C molecule forms interactions with three different protein molecules (each shown in a different color), which could assist in lattice packing. Protein–ligand interaction analysis was conducted using the *PLIP* web server (Salentin *et al.*, 2015).

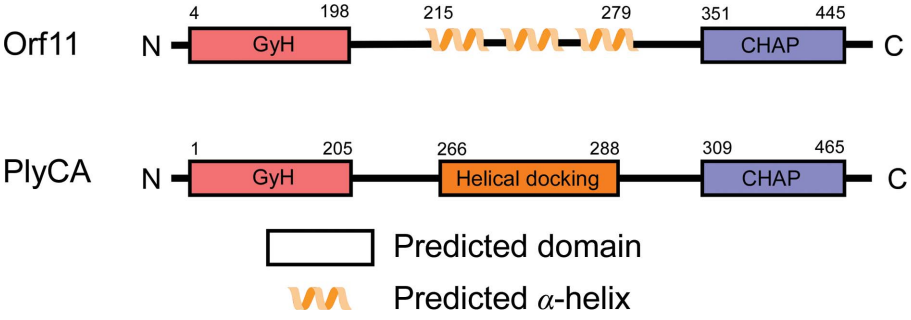


Figure 7
Orf11 is predicted to have a GyH domain and a CHAP domain, connected via a linker containing three α-helices. This predicted domain architecture of Orf11 is similar to that of the endolysin protein PlyCA from *Streptococcus* phage C1. The PlyCA domain architecture was defined using the crystal structure of the protein (PDB entry 4f88; McGowan *et al.*, 2012). The domain architecture of Orf11 was assigned using multiple bioinformatics tools. The GyH domain was identified using the *FFAS-3D* homology-detection server (Xu *et al.*, 2014). The CHAP domain was determined using the Pfam protein-domain database (Finn *et al.*, 2016). The α-helical content within the linker region was defined using the *JPred*4 secondary-structure prediction server (Drozdetskiy *et al.*, 2015). The residue numbers for the domain boundaries are annotated above the diagrams.

crystals obtained are likely to have I3C incorporated, ready for SAD phasing. This approach has successfully been applied to both the HEWL and Orf11 NTD proteins. In both cases, an increase in the number of hit conditions was identified in the rMMS with I3C screens compared with the unseeded screens with and without I3C. For both HEWL and Orf11 NTD, one of these new conditions was confirmed to produce I3C-derivatized crystals that allowed solution of the protein structure.

There are several benefits to the use of I3C with rMMS screening compared with many of the compounds that are commonly used for derivatization. Firstly, I3C is inexpensive and readily available. Secondly, many of the commonly used heavy-metal salts such as uranyl acetate and platinum potassium chloride are incompatible with our rMMS approach. Many of these metal ions will form insoluble compounds when mixed with sulfate and phosphate buffers. Citrate and acetate in certain crystallization conditions can also chelate divalent metal ions to reduce their effective concentration (Pike *et al.*, 2016). Sulfate, phosphate, citrate and acetate are all commonly used in crystallization conditions and would preclude many heavy metals from our rMMS approach. Finally, some heavy atoms, but not I3C, react with HEPES and Tris and/or bind to DTT and β-mercaptoethanol, which are used in many protein preparations (Pike *et al.*, 2016).

I3C intrinsically provides two benefits when solving structures. Each I3C molecule provides three heavy scattering atoms, providing a significant anomalous signal for phasing. The heavy atoms are arranged in an equilateral triangle with sides of 6 Å. This arrangement allows the presence of I3C to be confirmed in the substructure-determination stage. If a triangle with these dimensions is found, it indicates a correct substructure. The specificity of I3C binding also makes it preferable to halide or alkali metal ions, which bind to many positions on the protein. This lack of specificity can result in many poorly occupied sites, making substructure determination difficult.

Another benefit of using iodine is that it provides a large anomalous signal ($f'' = 6.9$ e) at the Cu $K\alpha$ wavelength used by many home-source X-ray generators (and can also be used on home sources for isomorphous replacement). This property would allow the collection of anomalous data for phasing without the need to wait for synchrotron beamtime.

This screening approach can be applied to identify new co-crystallization conditions using other heavy-atom molecules. One possibility is the MAD triangle (Beck *et al.*, 2010), which has a similar molecular structure with iodine substituted by bromine. This phasing molecule has the benefit of allowing MAD phasing as the Br K edge is accessible at many

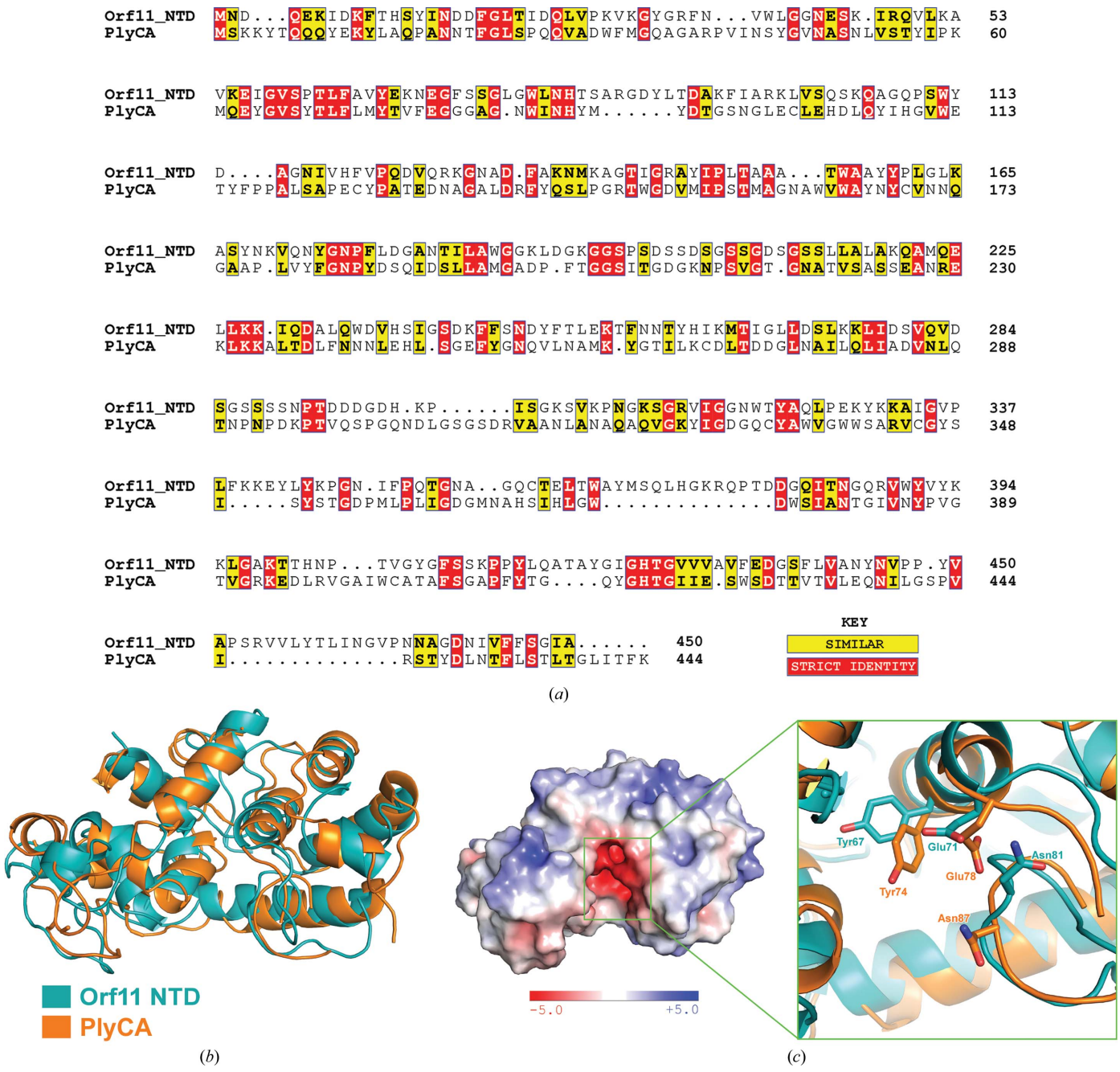


Figure 8
Orf11 shows limited sequence similarity and high structural similarity to the PlyCA protein from *Streptococcus* phage C1. (*a*) Sequences of the PlyCA GyH domain (residues 1–213) and the Orf11 NTD domain (residues 1–201) were aligned using *T-Coffee* (Armougom *et al.*, 2006), with an identity of 22% and a sequence similarity of 40%. Alignments were displayed using *ESPript* (Gouet *et al.*, 1999) and similarity was calculated using the Risler matrix (Risler *et al.*, 1988). (*b*) A superposition of the PlyCA GyH domain and Orf11 NTD structures shows that the proteins adopt similar tertiary folds. The two proteins were superimposed with *PyMOL* (DeLano, 2002) using C^{α} positions, with an r.m.s.d. of 3.5 Å over 127 atoms. (*c*) The conserved catalytic residues of the glucosaminidase domain of GyH are Glu78, Tyr74 and Asn87. The equivalent residues Glu71, Tyr67 and Asn81 in Orf11 NTD appear in similar positions. The residues appear in an electronegative cleft in the protein. The electrostatic potentials of Orf11 NTD were calculated using the *APBS* plugin in *PyMOL*.

synchrotron beamlines. It would also be of interest to test the tantalum bromide cluster (Knäblein *et al.*, 1997), which provides an extremely large amount of phasing power, especially for large proteins or protein–protein/DNA complexes.

4.2. Orf11 is likely to be a lysin

Bacteriophage P68 Orf11 was identified as a putative endolysin using a bioinformatics approach. Endolysins are enzymes that degrade the peptidoglycan structure in bacterial cell walls and, along with holins that form pores in the inner membrane, form the escape system for bacteriophages to exit their host via lysis (Young, 2014). A homology search using the *FFAS* (Xu *et al.*, 2014) and Pfam (Finn *et al.*, 2016) web servers identified two putative domains in the Orf11 protein: an N-terminal glycosyl hydrolase (GyH) domain and a C-terminal cysteine, histidine-dependent aminohydrolase/peptidase (CHAP) domain (Fig. 7). Both domains commonly appear in endolysin proteins (Oliveira *et al.*, 2013). Another possibility is that Orf11 corresponds to a virion-associated lysin protein (VAL). Some phages, particularly those that infect Gram-positive bacteria, use glycosyl hydrolases associated with the virion to degrade the peptidoglycan layer (Fernandes & São-José, 2018). This process clears a path to enable the phage to find the cell membrane and allows the phage machinery to inject DNA into the cell.

The structure of Orf11 NTD is very similar to that of the GyH domain of PlyC. PlyC is an endolysin protein from *Streptococcus* phage C1 and is the most potent endolysin protein discovered to date (McGowan *et al.*, 2012; Riley *et al.*, 2015). PlyC consists of one PlyCA polypeptide mounted on a PlyCB octamer. A structural superposition of the GyH domain of PlyCA (PDB entry 4f88; McGowan *et al.*, 2012) with Orf11 NTD reveals similar overall folds, despite a low sequence identity between the domains of 22% [Figs. 8(*a*) and 8(*b*)]. However, attempts to solve the data set via molecular replacement using the PlyC GyH domain as a model failed to yield a solution. Small deviations from the true structure in the template may have prevented the algorithm from finding an initial solution. The glycosyl hydrolase domain of PlyCA has three key catalytic residues, Glu78, Tyr74 and Asn87, thought to form the catalytic center. In the structure of Orf11 NTD these same residues appeared at similar spatial positions within an electronegative cleft in the protein [Fig. 8(*c*)]. From this, it is possible that the Orf11 NTD is a glycosyl hydrolase.

A secondary-structure prediction of Orf11 shows that a region containing three α-helices lies between the two domains (Fig. 7). This topology is similar to that of PlyCA, which has a helical docking domain consisting of three α-helices that anchors it to the cell-wall-binding protein PlyCB. All of this, taken together, suggests that Orf11 is a lysin protein that perhaps requires an additional protein to direct it to its bacterial target.

Acknowledgements

This research was undertaken in part using the MX1 beamline at the Australian Synchrotron, which is part of ANSTO. Author contributions are as follows. JQT performed the crystallographic experiments and structural modeling and prepared the manuscript. JQT and LS performed the bioinformatics analysis. SP provided support with crystallographic data processing. KES and JBB contributed to manuscript preparation.

Funding information

The following funding is acknowledged: Australian Research Council (grant Nos. DP150103009 and DP160101450 to Keith Shearwin); University of Adelaide (Adelaide Postgraduate Scholarship to Jia Quyen Truong).

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V. & Notredame, C. (2006). *Nucleic Acids Res.* **34**, W604–W608.
Beck, T., Gruene, T. & Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 374–380.
Beck, T., Krasauskas, A., Gruene, T. & Sheldrick, G. M. (2008). *Acta Cryst.* D**64**, 1179–1182.
Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D. & Rigden, D. J. (2012). *Acta Cryst.* D**68**, 1622–1631.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Costanzo, L., Christie, C., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Liang, Y., Lowe, R., Peisach, E., Periskova, I., Randle, C., Rose, A., Sekharan, M., Shao, C., Tao, Y., Valasatava, Y., Voigt, M., Westbrook, J., Young, J., Zardecki, C., Zhuravleva, M., Kurisu, G., Nakamura, H., Kengaku, Y., Cho, H., Sato, J., Kim, J. Y., Ikegawa, Y., Nakagawa, A., Yamashita, R., Kudou, T., Bekker, G., Suzuki, H., Iwata, T., Yokochi, M., Kobayashi, N., Fujiwara, T., Velankar, S., Kleywegt, G. J., Anyango, S., Armstrong, D. R., Berrisford, J. M., Conroy, M. J., Dana, J. M., Deshpande, M., Gane, P., Gáborová, R., Gupta, D., Gutmanas, A., Koča, J., Mak, L., Mir, S., Mukhopadhyay, A., Nadzirin, N., Nair, S., Patwardhan, A., Paysan-Lafosse, T., Pravda, L., Salih, O., Sehnal, D., Varadi, M., Vařeková, R., Markley, J. L., Hoch, J. C., Romero, P. R., Baskaran, K., Maziuk, D., Ulrich, E. L., Wedell, J. R., Yao, H., Livny, M. & Ioannidis, Y. E. (2019). *Nucleic Acids Res.* **47**, D520–D528.
Cowtan, K. (2000). *Acta Cryst.* D**56**, 1612–1621.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
D'Arcy, A., Villard, F. & Marsh, M. (2007). *Acta Cryst.* D**63**, 550–554.
DeLano, W. (2002). *CCP4 Newsl. Protein Crystallogr.* **40**, 82–92.
Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. (2015). *Nucleic Acids Res.* **43**, W389–W394.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Fernandes, S. & São-José, C. (2018). *Viruses*, **10**, 396.
Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. & Bateman, A. (2016). *Nucleic Acids Res.* **44**, D279–D285.
Garman, E. & Murray, J. W. (2003). *Acta Cryst.* D**59**, 1903–1913.
Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.

Hao, Q. (2004). *J. Appl. Cryst.* **37**, 498–499.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Ireton, G. C. & Stoddard, B. L. (2004). *Acta Cryst.* D**60**, 601–605.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Knäblein, J., Neuefeind, T., Schneider, F., Bergner, A., Messerschmidt, A., Löwe, J., Steipe, B. & Huber, R. (1997). *J. Mol. Biol.* **270**, 1–7.
Leverrier, P., Declercq, J.-P., Denoncin, K., Vertommen, D., Hiniker, A., Cho, S.-H. & Collet, J.-F. (2011). *J. Biol. Chem.* **286**, 16734–16742.
Lu, J. & Sun, P. D. (2014). *FEBS J.* **281**, 4021–4028.
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.
McGowan, S., Buckle, A. M., Mitchell, M. S., Hoopes, J. T., Gallagher, D. T., Heselpoth, R. D., Shen, Y., Reboul, C. F., Law, R. H. P., Fischetti, V. A., Whisstock, J. C. & Nelson, D. C. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 12752–12757.
McPherson, A. & Cudney, B. (2014). *Acta Cryst.* F**70**, 1445–1467.
McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *J. Synchrotron Rad.* **9**, 401–406.
Morris, R. J., Zwart, P. H., Cohen, S., Fernandez, F. J., Kakaris, M., Kirillova, O., Vonrhein, C., Perrakis, A. & Lamzin, V. S. (2004). *J. Synchrotron Rad.* **11**, 56–59.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Obmolova, G., Malia, T. J., Teplyakov, A., Sweet, R. & Gilliland, G. L. (2010). *Acta Cryst.* D**66**, 927–933.
Oliveira, H., Melo, L. D. R., Santos, S. B., Nóbrega, F. L., Ferreira, E. C., Cerca, N., Azeredo, J. & Kluskens, L. D. (2013). *J. Virol.* **87**, 4558–4570.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* D**61**, 449–457.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2009). *Acta Cryst.* D**65**, 1089–1097.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Pike, A. C. W., Garman, E. F., Krojer, T., von Delft, F. & Carpenter, E. P. (2016). *Acta Cryst.* D**72**, 303–318.
Riley, B. T., Broendum, S. S., Reboul, C. F., Cowieson, N. P., Costa, M. G. S., Kass, I., Jackson, C., Perahia, D., Buckle, A. M. & McGowan, S. (2015). *PLoS One*, **10**, e0140219.
Risler, J. L., Delorme, M. O., Delacroix, H. & Henaut, A. (1988). *J. Mol. Biol.* **204**, 1019–1029.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. (2015). *Nucleic Acids Res.* **43**, W443–W447.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Sheldrick, G. M. (2015). *Acta Cryst.* C**71**, 3–8.
Taylor, G. L. (2010). *Acta Cryst.* D**66**, 325–338.
Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Terwilliger, T. C. (1999). *Acta Cryst.* D**55**, 1863–1871.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Till, M., Robson, A., Byrne, M. J., Nair, A. V., Kolek, S. A., Shaw Stewart, P. D. & Race, P. R. (2013). *J. Vis. Exp.*, 50548.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.
Xu, D., Jaroszewski, L., Li, Z. & Godzik, A. (2014). *Bioinformatics*, **30**, 660–667.
Young, R. (2014). *J. Microbiol.* **52**, 243–258.