STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Received 27 February 2018
Accepted 6 May 2018

Edited by P. Dunten, Stanford Synchrotron Radiation Lightsource, USA

Keywords: FAD-dependent halogenase; hemihedral twinning; perfect twin; rotational pseudosymmetry; merohedral twinning.

Perfect merohedral twinning combined with noncrystallographic symmetry potentially causes the failure of molecular replacement with low-homology search models for the flavin-dependent halogenase HalX from *Xanthomonas campestris*

Maren Buss,[a] Christina Geerds,[a] Thomas Patschkowski,[b,c] Karsten Niehaus[b,c] and Hartmut H. Niemann[a]*

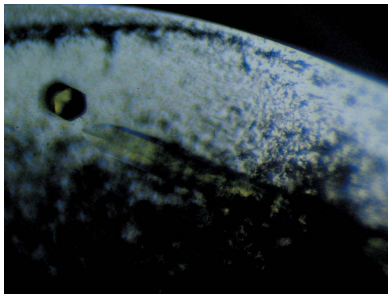
[a]Department of Chemistry, Bielefeld University, Universitätsstrasse 25, 33615 Bielefeld, Germany, [b]Faculty of Biology, Proteome and Metabolome Research, Bielefeld University, Universitätsstrasse 25, 33615 Bielefeld, Germany, and [c]Centre of Biotechnology (CeBiTec), Bielefeld University, Universitätsstrasse 27, 33615 Bielefeld, Germany. *Correspondence e-mail: hartmut.niemann@uni-bielefeld.de

Flavin-dependent halogenases can be used as biocatalysts because they regioselectively halogenate their substrates under mild reaction conditions. New halogenases with novel substrate specificities will add to the toolbox of enzymes available to organic chemists. HalX, the product of the *xcc-b100*_4193 gene, is a putative flavin-dependent halogenase from *Xanthomonas campestris*. The enzyme was recombinantly expressed and crystallized in order to aid in identifying its hitherto unknown substrate. Native data collected to a resolution of 2.5 Å showed indications of merohedral twinning in a hexagonal lattice. Attempts to solve the phase problem by molecular replacement failed. Here, a detailed analysis of the suspected twinning is presented. It is most likely that the crystals are trigonal (point group 3) and exhibit perfect hemihedral twinning so that they appear to be hexagonal (point group 6). As there are several molecules in the asymmetric unit, noncrystallographic symmetry may complicate twinning analysis and structure determination.

1. Introduction

Enzymatic halogenation by flavin-dependent halogenases (FDHs) has several advantages over chemical halogenation (Weichold *et al.*, 2016). It proceeds under mild conditions, requiring only $FADH_2$, O_2 and harmless halide salts such as NaCl or NaBr as substrates, and it occurs with high regioselectivity that cannot easily be achieved otherwise. This makes FDHs attractive tools for organic chemists in the transition from classical processes to green chemistry (Latham *et al.*, 2018). However, there are also drawbacks that limit the use of FDHs in biocatalysis, such as low enzyme activity and poor long-term stability. Nevertheless, regioselective chlorination of tryptophan has already been successful on the gram scale using cross-linked enzyme aggregates (Frese & Sewald, 2015). The crystal structures of the tryptophan halogenases PrnA (Dong *et al.*, 2005), RebH (Yeh *et al.*, 2007; Bitto *et al.*, 2008) and PyrH (Zhu *et al.*, 2009) have greatly advanced our understanding of the reaction mechanism and the regioselectivity.

To broaden the scope of FDHs as biocatalysts, enzymes accepting substrates other than tryptophan are required (Latham *et al.*, 2018). Some FDHs, such as CndH (Buedenbender *et al.*, 2009), PltA (Pang *et al.*, 2015), Bmp2 and Mpy16 (El Gamal *et al.*, 2016), are part of multi-enzyme systems. They



belong to the so-called variant B FDHs that accept their substrate only if bound to a carrier protein, which limits their general applicability. More recently, several halogenases accepting free substrates other than tryptophan have been identified (Zeng & Zhan, 2010; Menon *et al.*, 2017; Fraley *et al.*, 2017). Some of these have a broader substrate spectrum than other FDHs. Novel FDHs can be identified from genomes or metagenomes by sequence analysis as they contain conserved signature sequences (Zehner *et al.*, 2005; Smith *et al.*, 2017).

We annotated the *xcc-b100*_4193 gene from *Xanthomonas campestris* pathovar *campestris* strain B100 (Vorhölter *et al.*, 2008; Alkhateeb *et al.*, 2017) as a putative FDH and called the resulting protein HalX owing to its origin from *Xanthomonas* and its unknown substrate. In HalX the conserved sequences G*x*G*xx*G and W*x*W*x*IP are ^{12}GGGLAG^{17} and ^{237}WFWLIP^{242}, respectively. Sequence alignment also showed that a lysine residue (Lys79) is present in HalX at the position where the catalytically essential lysine residue is invariantly located in other FDHs. HalX has 21 and 22% sequence identity to the prototypic tryptophan 7-halogenases PrnA and RebH, respectively. In the Protein Data Bank (PDB), the FDH CmlS is the protein with the highest sequence identity to HalX (Podzelinska *et al.*, 2010). We assume that the putative FDH HalX is an active enzyme and accepts as yet unidentified substrates that most likely differ from tryptophan. HalX may thus be a welcome addition to the biocatalysis toolbox. To aid in identifying the substrate of HalX, we set out to crystallize the protein for structure determination by X-ray crystallography. Here, we describe the expression, purification and crystallization of HalX. Unexpectedly, structure solution by molecular replacement failed. We attribute this to suspected perfect merohedral twinning in combination with non-crystallographic symmetry. The indications of twinning are discussed in detail.

Table 1
Macromolecule-production information.

Soucre organism	*X. campestris* pv. *campestris* strain B100
DNA source	1626 bp *halX* gene in pET-28a
Forward primer†	5′-AGGAGCC **ATG** GGC *AGT GCA CAA CGC GCA GAT GTC*-3′
Reverse primer†	5′-AGGAGAATTC **TCA** *CGC CGC GTC GGC ATA CC*-3′
Cloning vector	pET-28a
Expression vector	pETM-11
Expression host	*E. coli* BL21-CodonPlus(DE3)-RIL
Complete amino-acid sequence of the construct produced‡	*MKHHHHHHPMSDYDIPTTENLYFQ*|GAMG**S AQRADVVITGGGLAGLSLALQLRQRDPA LAITVLERRAHPVREAAFKVGESTVEIG AHYFADVLGLREHLEAEQIRKFGFRFFF SDKRDDIDRCTELGVSKILPTPSWQIDR GRFENFLGERARAQGIVFLDSCSVKGVD LSDDDASDHRVRYERGGEAGTLDARWVV DASGRAGLLKRKLGLTQDNAHDANAVWW RVEGLIDPNAWSQDSSWLQRCTPPDRWR STNHMCGPGYWFWLIPLSSGAHSLGIVC DAAMHPLETMNTHDKAMAWLRTHQPQVA ATLDQTDYRLQDFLFLRNFSYGCKQVFS TQRWALTGEAGVFLDPFYSPGSDFIAIS NTYICELIGRDRAGRSLSPYVELYEQLY FSFYENTLSLYQDQYALFGDAQVMPVKV IWDYTYYWSLLAPLFCSGRIADLGLLAR MKADFFYARDMNLAMQRVLHDWGQHNTA LGTVTGDGRLLDQYLIGWFNELNGALHD TLDDEAFAARIHANIARMAVLAREILEQ ARQRHPALPDHGLAALTANAVGDPILTA AWYADAA**

† In the primer sequences, the restriction sites for NcoI (forward primer) and EcoRI (reverse primer) are underlined. The in-frame ATG of the NcoI site in the forward primer and the stop codon in the reverse primer are shown in bold. Codon boundaries are indicated by spaces. ‡ In the amino-acid sequence, the TEV recognition sequence is underlined. The cleavage site is indicated by a vertical bar and the cleaved tag is shown in italics. The HalX sequence is shown in bold. Residues N-terminal to the HalX sequence that remain after TEV cleavage are owing to the cloning procedure. The amino-acid sequence of HalX used in this publication is based on predictions published in 2008 (Vorhölter *et al.*, 2008). The exact same amino-acid sequence is predicted in many other *X. campestris* pv. *campestris* strains. A refined annotation of the complete genome of *X. campestris* pv. *campestris* strain B100 based on RNA-sequencing data now predicts a HalX protein that is eight amino acids shorter at the N-terminus (Alkhateeb *et al.*, 2017). This refined sequence, which is currently accessible in the public database, would require the codon GTG coding for Val9 in our sequence to act as a start codon.

2. Materials and methods

2.1. Macromolecule production

The sequence coding for HalX was subcloned from a pET-28a expression vector with a thrombin-cleavable N-terminal His_6 tag into the expression vector pETM-11 (Gunter Stier, EMBL Heidelberg) using the primers listed in Table 1. The resulting fusion protein has an N-terminal His_6 tag followed by the recognition sequence for *Tobacco etch virus* (TEV) protease (Table 1). HalX was expressed in *Escherichia coli* BL21-CodonPlus(DE3)-RIL cells. The bacteria were grown to an OD_{600} of 0.6 in LB medium with 30 mg l^{-1} kanamycin and 34 mg l^{-1} chloramphenicol at 310 K. The cultures were shifted to 293 K, induced with 100 µ*M* isopropyl *β*-D-1-thiogalactopyranoside and incubated with shaking (90 rev min^{-1}; IFS-1-V Kühner) overnight. The cells were harvested by centrifugation. After resuspension in 15 ml lysis buffer [phosphate-buffered saline (PBS) supplemented with cOmplete EDTA-free protease inhibitors (Roche) and 400 µg DNaseI (Macherey-Nagel)] per litre of bacterial culture, the cells were lysed in an SPCH ultrahigh-pressure cell homogenizer (Stansted Fluid Power) for three to five cycles at a pressure of approximately 120 MPa. The cell lysate was clarified by centrifugation (60 min, 30 000*g*). All chromatographic steps were performed at 277 K. The supernatant was applied onto 15 ml PureCube Ni–NTA agarose (Cube Biotech) equilibrated with PBS. The matrix was washed three times with 20 ml of each of the following buffers: (i) PBS pH 8.0 plus an additional 300 m*M* NaCl, (ii) 20 m*M* Tris pH 8.0, 20 m*M* NaCl, (iii) 20 m*M* Tris pH 8.0, 20 m*M* NaCl, 10 m*M* imidazole. Immediately after elution with 20 m*M* Tris pH 8.0, 20 m*M* NaCl, 200 m*M* imidazole, EDTA was added to the protein-containing fractions to a final concentration of 5 m*M*. HalX was further purified by anion-exchange chromatography on a self-packed Source Q column (GE Healthcare) equilibrated with 20 m*M* Tris pH 8.0. HalX was eluted with a gradient of 0–40% high-salt buffer (20 m*M* Tris pH 8.0, 1 *M* NaCl) over 20 column volumes. The N-terminal His_6 tag was removed by TEV protease cleavage (HalX:TEV mass ratio of 50:1) at 277 K overnight. Uncleaved HalX and His-tagged TEV protease were removed by binding to 15 ml Ni–NTA agarose at 277 K for 60 min. The flowthrough was collected, concentrated with

Table 2
Crystallization.

	Crystal 1	Crystal 2
Method	Vapour diffusion	Vapour diffusion
Plate type	Swissci MRC 2-well 96-well sitting drop	Cryschem M 24-well sitting drop
Temperature (K)	293	293
Protein concentration (mg ml^{-1})	10	20
Buffer composition of protein solution	12.5 m*M* HEPES pH 8.0, 125 m*M* NaCl, 1 m*M* EDTA, 1 m*M* iodoacetamide	12.5 m*M* HEPES pH 8.0, 125 m*M* NaCl, 1 m*M* EDTA, 1 m*M* iodoacetamide
Composition of reservoir solution	652 m*M* NaH_2PO_4, 448 m*M* K_2HPO_4, 400 m*M* $(NH_4)_2SO_4$ pH 6.4	204 m*M* NaH_2PO_4, 516 m*M* K_2HPO_4, 400 m*M* $(NH_4)_2SO_4$ pH 7.1
Volume and drop ratio	1 µl + 1 µl	2 µl + 2 µl
Volume of reservoir (µl)	70	500
Cryoprotection	Reservoir solution plus 25% glycerol	Reservoir solution plus 40% xylitol

an ultrafiltration device (Vivaspin, molecular-weight cutoff 10 kDa; Sartorius) and applied onto a HiLoad 16/600 Superdex 200 prep-grade column (GE Healthcare) equilibrated with 12.5 m*M* HEPES pH 8.0, 125 m*M* NaCl. The protein was concentrated to 24 mg ml^{-1} by ultrafiltration as above and stored in small aliquots at 253 K. The protein concentration was determined using absorption measurements with a NanoDrop 1000 spectrophotometer (PEQLAB) at a wavelength of 280 nm and the theoretical molar extinction coefficient of HalX as calculated by the *ProtParam* tool on the *ExPASy* server (Gasteiger *et al.*, 2005). The yield of purified protein from 1 l of bacterial culture was 15 mg.

2.2. Crystallization

For crystallization, the frozen protein stock was diluted in 12.5 m*M* HEPES pH 8.0, 125 m*M* NaCl and supplemented with 1 m*M* EDTA and 1 m*M* iodoacetamide to suppress the nonspecific proteolysis that was observed upon prolonged storage at 293 K in the absence of protease inhibitors. Crystal screening was performed in MRC 2-well plates (Swissci) using commercial screens and a Crystal Gryphon robot (Art Robbins) with drop sizes of 100 nl protein solution and 100 nl reservoir solution or 100 nl protein solution and 200 nl reservoir solution. Crystal optimization with larger drops took place in MRC 2-well plates or Cryschem M 24-well sitting-drop plates. The final crystallization conditions are listed in Table 2.

2.3. Data collection and processing

For cryoprotection, we initially used a solution containing sodium potassium phosphate and ammonium sulfate at the same concentrations as in the reservoir solution plus an additional 25% glycerol. Later, we switched to reservoir solution plus 40% xylitol, which somewhat reduced the problem of the rapid formation of phosphate crystals during the mounting procedure (Table 2). Crystals were flash-cooled in liquid nitrogen. Fine-sliced data were collected at 100 K on various beamlines at the ESRF using pixel-array detectors (Dectris). Data were indexed and integrated with *XDS* and scaled with *XSCALE* (Kabsch, 2010). Twinning was analysed with *phenix.xtriage* (Adams *et al.*, 2010). Molecular replacement was attempted with *Phaser* (McCoy *et al.*, 2007) run through the *CCP*4*i* interface (Winn *et al.*, 2011) and the *MorDa* automatic molecular-replacement pipeline that relies on the program *MOLREP* (Vagin & Teplyakov, 2010).

3. Results and discussion

3.1. Protein production and crystallization

HalX was recombinantly expressed as a His_6-tagged protein and was purified by Ni–NTA affinity chromatography, anion-exchange chromatography and gel filtration, from which it eluted as a single peak with an apparent molecular weight consistent with that of a monomer. The protein appeared as a single band on Coomassie-stained gels, but showed degradation upon prolonged storage at 293 K. This nonspecific proteolysis could be suppressed by the addition of 1 m*M* EDTA and 1 m*M* iodoacetamide. Initial crystals grew as small needles at 293 K with a protein concentration of 10 mg ml^{-1} in conditions A12 and B1 of the commercial screen The MbClass Suite (Qiagen) and condition B1 of The MbClass II Suite (Qiagen). All conditions contained 1 *M* sodium potassium phosphate and 0.1 *M* ammonium phosphate or ammonium sulfate at pH values of 7.5–8.5. The crystallization conditions were further refined to those described in Table 2. Crystal growth varied greatly and somewhat erratically between neighbouring conditions in finely sampled two-dimensional grid screens of pH (steps of 0.1) *versus* precipitant concentration (steps of 10 m*M* sodium potassium phosphate). For

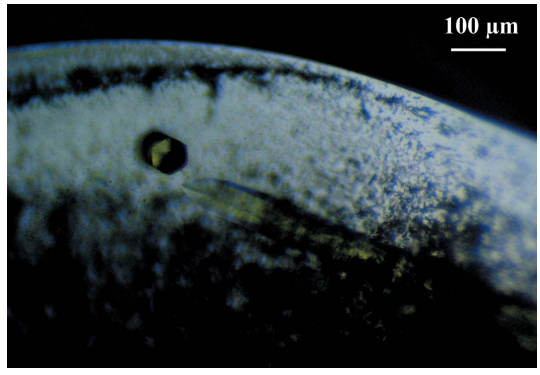


Figure 1
Crystals of HalX similar to crystal 1 described in this work. The crystals appeared as hexagonal needles. The left crystal shows the cross-section looking down the long axis of a needle.

Table 3
Data collection and processing.

	Crystal 1			Crystal 2		
Space group	$P3_1$ or $P3_2$	$P6_2$ or $P6_4$	$P6_222$ or $P6_422$	$P3_1$ or $P3_2$	$P6_2$ or $P6_4$	$P6_222$ or $P6_422$
Diffraction source	ID30B, ESRF			ID30A-3, ESRF		
Wavelength (Å)	0.9725			0.9677		
Temperature (K)	100			100		
Detector	PILATUS 6M-F			EIGER X 4M		
Crystal-to-detector distance (mm)	365.66			144.57		
Rotation range per image (°)	0.05			0.1		
Total rotation range (°)	360			180		
Exposure time per image (s)	0.037			0.01		
Mosaicity (°)	0.200			0.155		
a, b, c (Å)	167.669	167.669	112.707	165.003	165.003	110.811
α, β, γ (°)	90.000	90.000	120.000	90.000	90.000	120.000
Resolution range (Å)	50–2.46 (2.51–2.46)			50–2.13 (2.18–2.13)		
Total No. of reflections	1275608	1271428	957777	997921	992248	970648
No. of unique reflections	127975	65265	33988	188754	96039	50239
Completeness (%)	99.3 (92.9)	99.6 (95.8)	98.6 (93.0)	99.4 (94.2)	99.7 (96.3)	99.8 (97.3)
Multiplicity	9.97	19.48	28.18	5.29	10.33	19.32
$\langle I/\sigma(I)\rangle$	12.37 (1.61)	16.96 (2.18)	9.05 (3.47)	10.83 (1.68)	7.93 (2.13)	5.70 (2.12)
R_{meas} (%)	21.4 (240.3)	21.8 (236.4)	34.8 (92.7)	9.9 (120.4)	19.1 (120.1)	44.9 (121.8)
$CC_{1/2}$	99.7 (65.6)	99.8 (77.4)	99.2 (87.7)	99.8 (47.1)	99.4 (62.7)	98.8 (56.9)
Overall B factor from Wilson plot (Å^2)	56.3	55.2	48.0	48.0	48.2	47.3

example, in one 24-well grid screen sampling pH from 6.9 to 7.4 against sodium potassium phosphate concentration from 690 to 720 m*M* there were drops with no crystals, drops with many small needles and drops with one or a few thick hexagonal needles reaching up to 900 × 60 µm in size (Fig. 1). Occasionally these thick needles appeared to be single crystals, but often their ends suggested that they were actually made up of many aligned thin hexagonal needles. The large crystals were yellowish, most likely owing to the presence of the cofactor FAD.

3.2. Space-group determination and twinning analysis

The data for crystal 1, which grew under conditions similar to the initial screening hits, scaled equally well in point groups 3 and 6 but considerably worse in point group 622 (Table 3). Systematic absences suggested space groups $P3_1$, $P3_2$, $P6_2$ or $P6_4$. Matthews probability analysis suggested six and three molecules per asymmetric unit in points groups 3 and 6, respectively, with a Matthews coefficient of 2.43 Å^3 Da^{-1} and a solvent content of 49%. There was no indication of pseudo-translation, with the largest off-origin peak in the Patterson function being around 3% of the height of the origin peak. The twin indicators in *phenix.xtriage* hinted towards twinning, although not perfect twinning (Table 4). One possibility for the true point group could be 6 with twin operator (h, $-h-k$, $-l$), emulating point group 622. However, the twin-law-dependent twin fraction calculated for this operator by *phenix.xtriage* was very low (Table 4; Britton α 0.04, *H* α 0.06, ML α 0.02). Alternatively, the true point group might be 3 with twinning around a twofold axis along the crystallographic threefold axis [twin law ($-h$, $-k$, l)]. The scaling statistics suggest that twinning around this operator would be (almost) perfect, with a twin fraction close to 0.5, as the R_{meas} for point group 3 and point group 6 is virtually identical (Table 3). However, the twin-law-independent twin indicators calculated by *phenix.xtriage* deviate from the values expected for a perfect twin (Table 4). The twin-law-dependent calculation of the twin fraction for operator ($-h$, $-k$, l) yielded values of between 0.44 and 0.48, suggesting almost perfect twinning (Table 4).

As we were unable to solve the structure by molecular replacement (see below), we collected data from a further 13 crystals, hoping to find a less twinned specimen. This set comprised crystals grown under conditions similar to those for crystal 1 and under conditions similar to those for crystal 2 (Table 2). Since the twin fraction sometimes varies within a crystal (Keller, 2017), we collected several complete data sets at different positions along the long axis of the large needles of a couple of crystals. All crystals but one behaved very similarly to crystal 1. Generally, scaling in point group 6 increased the $\langle I\rangle/\langle\sigma(I)\rangle$ over that from scaling in point group 3, without any significant increase in R_{meas}. Only one crystal (crystal 2 in Tables 2 and 3), which was grown at a higher protein concentration and lower precipitant concentration than crystal 1 and that was block-shaped rather than needle-shaped, showed a different behaviour, with a decreased $\langle I\rangle/\langle\sigma(I)\rangle$ and a doubling of R_{meas} in $P6_x$ compared with $P3_x$. This strongly suggested that twinning occurred around a twofold axis along the crystallographic threefold axis [operator ($-h$, $-k$, l)], which would be consistent with the crystal morphology described above. Although the twin fraction calculated for this particular twin law was lower for crystal 2 (between 0.25 and 0.32; Table 3) than for crystal 1, the twin-law-independent twin indicators were closer to the values expected for a perfect twin for crystal 2 than for crystal 1 (Table 3). The correlation between the data sets of crystals 1 and 2 in point group 3 (Table 3) reported by *XSCALE* was 0.997. Interestingly, a second data set collected from a different position of crystal 2 showed scaling behaviour similar to that of all of the other 13 crystals, with a higher $\langle I\rangle/\langle\sigma(I)\rangle$ and no significant increase of R_{meas} in $P6_x$ compared with scaling in $P3_x$. Thus, the degree of

Table 4
Twin indicators as reported by *phenix.xtriage*.

	Crystal 1		Crystal 2	
Space group	$P3_{1/2}$	$P6_{2/4}$	$P3_{1/2}$	$P6_{2/4}$
$\langle I^2\rangle/\langle I\rangle^2$ (untwinned, 2.0; perfect twin, 1.5)	1.934	1.918	1.847	1.769
$\langle F^2\rangle/\langle F\rangle^2$ (untwinned, 0.785; perfect twin, 0.885)	0.816	0.818	0.836	0.849
$\langle\lvert E^2-1\rvert\rangle$ (untwinned, 0.736; perfect twin, 0.541)	0.714	0.683	0.660	0.628
$\langle\lvert L\rvert\rangle$ (untwinned, 0.500; perfect twin, 0.375)	0.453	0.447	0.424	0.406
$\langle L^2\rangle$ (untwinned, 0.333; perfect twin, 0.200)	0.278	0.272	0.247	0.229
Multivariate Z-score L-test	2.885	3.197	5.228	7.282
Twin operator ($-h$, $-k$, l)	x	—	x	—
R_{obs}†	0.048		0.181	
Britton α	0.444		0.308	
H α	0.459		0.309	
ML α	0.478		0.242	
Twin operator (h, $-h-k$, $-l$)	x	x	x	x
R_{obs}†	0.437	0.437	0.409	0.397
Britton α	0.038	0.040	0.070	0.084
H α	0.049	0.055	0.067	0.079
ML α	0.022	0.022	0.022	0.022
Twin operator (–k, –h, –l)	x	—	x	—
R_{obs}†	0.438		0.412	
Britton α	0.038		0.071	
H α	0.048		0.065	
ML α	0.022		0.022	

† $R_{obs} = \sum_{hkl}|I(hkl) - I(hkl')|/\sum_{hkl}[I(hkl) + I(hkl')]$, where hkl and hkl' are related by the potential twin operator (Lebedev *et al.*, 2006).

twinning appears to vary throughout crystal 2. The correlation between the two data sets of crystal 2 in point group 3 reported by *XSCALE* was 1.000.

3.3. Attempts to solve the structure

The structures of several other FDHs have been solved by molecular replacement despite low sequence identity with the respective search models (Pang *et al.*, 2015; El Gamal *et al.*, 2016). In the case of MibH this approach allowed the location of four chains in the asymmetric unit (Ortega *et al.*, 2017). For molecular replacement in *Phaser*, we created an ensemble of superposed search models based on the four best hits found by *HHpred* (Alva *et al.*, 2016), namely MalA (PDB entry 5wgx; Fraley *et al.*, 2017), CndH (PDB entry 3e1t; Buedenbender *et al.*, 2009), Bmp2 (PDB entry 5bva; El Gamal *et al.*, 2016) and PltA (PDB entry 5dbj; Pang *et al.*, 2015). Flexible loops and regions that show a high divergence in secondary-structure elements were deleted manually in *Coot* (Emsley *et al.*, 2010). A very similar approach had previously allowed us to solve the phase problem for crystals of an α/β-hydrolase with less than 20% sequence identity to any known structure and three molecules in the asymmetric unit (Thoms *et al.*, 2011). For HalX, we tried molecular replacement in various possible space groups. We put particular effort into solving the structure in $P3_1$ and $P3_2$. The *Phaser* run took several days, indicating major problems in solving the structure. We also ran the automated molecular-replacement pipeline *MorDa* in space groups $P3$, $P3_1$ and $P3_2$. We did not obtain a promising solution that refined to reasonable R factors or allowed further model building using either *Phaser* or *MorDa*.

4. Conclusion

Molecular replacement is more difficult for twinned crystals, and it is especially tricky with low-homology search models (Zwart *et al.*, 2008). Trying to find a crystal even less twinned than crystal 2 is not a viable option, as 13 of the 14 crystals that were tested showed (almost) perfect twinning at all positions from which we collected data. Only one of two data sets collected from crystal 2 showed a lower twin fraction, estimated at between 0.25 and 0.32 (Table 4). Even with this lower twin fraction, molecular replacement failed. Thus, structure solution will have to await the availability of better search models or the generation of another crystal form that does not suffer from twinning. As an alternative, we are pursuing the production of selenomethionine (SeMet)-labelled protein for use in anomalous diffraction experiments. Phasing by multiple-wavelength anomalous diffraction (MAD) or single-wavelength anomalous diffraction (SAD) is more difficult with twinned crystals (McCoy & Read, 2010). Nevertheless, MAD phasing worked for hemihedrally twinned crystals with twin fractions of about 0.36–0.4 without detwinning the data (Yang *et al.*, 2000; Rudolph *et al.*, 2003). SAD phasing was even possible from model data that were artificially perfectly twinned (Dauter, 2003).

For the two crystals that are described here, the twin-law-independent twin indicators do not seem to behave as expected. Scaling statistics suggest that crystal 1 is (almost) perfectly twinned, while crystal 2 must have a lower twin fraction. Still, the twin-law-independent twin indicators of crystal 2 are closer to the values expected for a perfect twin than those of crystal 1. This unexpected behaviour could be owing to the presence of rotational noncrystallographic symmetry (NCS) elements parallel to the twin axis (Lebedev *et al.*, 2006). As the asymmetric unit in point group 3 is most likely to contain six molecules, the presence of a non-crystallographic twofold, threefold or even sixfold axis parallel to the crystallographic threefold axis and the twin axis seem possible or even likely. Rotational NCS is often present in crystals that exhibit merohedral twinning (Lebedev *et al.*, 2006). Rotational NCS has been shown to result in an underestimation of the twin fraction by twin-law-independent twin indicators because it relates reflection intensities that are assumed to be unrelated (Lebedev *et al.*, 2006). The presence or absence of rotational NCS parallel to the crystallographic threefold axis and the suspected twofold twin axis cannot be established from the self-rotation function, as the NCS axis would be masked by the crystallographic threefold axis and the twin operator. Finally, if the true point group is 3, not only hemihedral but also tetartohedral twinning with four twin domains instead of only two will be possible (Roversi *et al.*, 2012). The determination of the true space group, the kind of twinning present and the actual twin fractions will have to await structure solution and refinement.

Acknowledgements

The diffraction experiments were performed on beamlines ID30B (crystal 1) and ID30A-3 (crystal 2) at the European

Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to Cristiano de Sousa Mota and Antoine Royant at the ESRF for providing assistance in using beamlines ID30B and ID30A-3, respectively. We thank Roman Fedorov (Medizinische Hochschule Hannover) for collecting the data for crystal 1.

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.

Alkhateeb, R. S., Rückert, C., Rupp, O., Pucker, B., Hublik, G., Wibberg, D., Niehaus, K., Pühler, A. & Vorhölter, F. J. (2017). *J. Biotechnol.* **253**, 55–61.

Alva, V., Nam, S.-Z., Söding, J. & Lupas, A. N. (2016). *Nucleic Acids Res.* **44**, W410–W415.

Bitto, E., Huang, Y., Bingman, C. A., Singh, S., Thorson, J. S. & Phillips, G. N. (2008). *Proteins*, **70**, 289–293.

Buedenbender, S., Rachid, S., Müller, R. & Schulz, G. E. (2009). *J. Mol. Biol.* **385**, 520–530.

Dauter, Z. (2003). *Acta Cryst.* D**59**, 2004–2016.

Dong, C., Flecks, S., Unversucht, S., Haupt, C., Van Pée, K.-H. & Naismith, J. H. (2005). *Science*, **309**, 2216–2219.

El Gamal, A., Agarwal, V., Diethelm, S., Rahman, I., Schorn, M. A., Sneed, J. M., Louie, G. V., Whalen, K. E., Mincer, T. J., Noel, J. P., Paul, V. J. & Moore, B. S. (2016). *Proc. Natl Acad. Sci. USA*, **113**, 3797–3802.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Fraley, A. E., Garcia-Borràs, M., Tripathi, A., Khare, D., Mercado-Marin, E. V., Tran, H., Dan, Q., Webb, G. P., Watts, K. R., Crews, P., Sarpong, R., Williams, R. M., Smith, J. L., Houk, K. N. & Sherman, D. H. (2017). *J. Am. Chem. Soc.* **139**, 12060–12068.

Frese, M. & Sewald, N. (2015). *Angew. Chem. Int. Ed.* **54**, 298–301.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Keller, J. P. (2017). *Acta Cryst.* D**73**, 22–31.

Latham, J., Brandenburger, E., Shepherd, S. A., Menon, B. R. K. & Micklefield, J. (2018). *Chem. Rev.* **118**, 232–269.

Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). *Acta Cryst.* D**62**, 83–95.

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.

McCoy, A. J. & Read, R. J. (2010). *Acta Cryst.* D**66**, 458–469.

Menon, B. R. K., Brandenburger, E., Sharif, H. H., Klemstein, U., Shepherd, S. A., Greaney, M. F. & Micklefield, J. (2017). *Angew. Chem. Int. Ed.* **56**, 11841–11845.

Ortega, M. A., Cogan, D. P., Mukherjee, S., Garg, N., Li, B., Thibodeaux, G. N., Maffioli, S. I., Donadio, S., Sosio, M., Escano, J., Smith, L., Nair, S. K. & Van Der Donk, W. A. (2017). *ACS Chem. Biol.* **12**, 548–557.

Pang, A. H., Garneau-Tsodikova, S. & Tsodikov, O. V. (2015). *J. Struct. Biol.* **192**, 349–357.

Podzelinska, K., Latimer, R., Bhattacharya, A., Vining, L. C., Zechel, D. L. & Jia, Z. (2010). *J. Mol. Biol.* **397**, 316–331.

Roversi, P., Blanc, E., Johnson, S. & Lea, S. M. (2012). *Acta Cryst.* D**68**, 418–424.

Rudolph, M. G., Kelker, M. S., Schneider, T. R., Yeates, T. O., Oseroff, V., Heidary, D. K., Jennings, P. A. & Wilson, I. A. (2003). *Acta Cryst.* D**59**, 290–298.

Smith, D. R. M., Uria, A. R., Helfrich, E. J. N., Milbredt, D., Van Pée, K.-H., Piel, J. & Goss, R. J. M. (2017). *ACS Chem. Biol.* **12**, 1281–1287.

Thoms, S., Hofhuis, J., Thöing, C., Gärtner, J. & Niemann, H. H. (2011). *J. Struct. Biol.* **175**, 362–371.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Vorhölter, F. J. *et al.* (2008). *J. Biotechnol.* **134**, 33–45.

Weichold, V., Milbredt, D. & van Pée, K.-H. (2016). *Angew. Chem. Int. Ed.* **55**, 6374–6389.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Yang, F., Dauter, Z. & Wlodawer, A. (2000). *Acta Cryst.* D**56**, 959–964.

Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. (2007). *Biochemistry*, **46**, 1284–1292.

Zehner, S., Kotzsch, A., Bister, B., Süssmuth, R. D., Méndez, C., Salas, J. A. & van Pée, K. H. (2005). *Chem. Biol.* **12**, 445–452.

Zeng, J. & Zhan, J. (2010). *Chembiochem*, **11**, 2119–2123.

Zhu, X., De Laurentis, W., Leang, K., Herrmann, J., Ihlefeld, K., van Pée, K.-H. & Naismith, J. H. (2009). *J. Mol. Biol.* **391**, 74–85.

Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 99–107.