

## **Structural studies of a surface-entropy reduction mutant of O-GlcNAcase.**

MALES, Alexandra <<http://orcid.org/0000-0002-7250-8300>> and DAVIES, Gideon J <<http://orcid.org/0000-0002-7343-776X>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<https://shura.shu.ac.uk/32403/>

---

This document is the Published Version [VoR]

### **Citation:**

MALES, Alexandra and DAVIES, Gideon J (2019). Structural studies of a surface-entropy reduction mutant of O-GlcNAcase. *Acta crystallographica. Section D, Structural biology*, 75 (Pt 1), 70-78. [Article]

---

### **Copyright and re-use policy**

See <http://shura.shu.ac.uk/information.html>



## Structural studies of a surface-entropy reduction mutant of O-GlcNAcase

Alexandra Males and Gideon J. Davies\*

Department of Chemistry, University of York, York YO10 5DD, England. \*Correspondence e-mail: gideon.davies@york.ac.uk

Received 24 September 2018

Accepted 21 November 2018

Edited by M. Czjzek, Station Biologique de Roscoff, France

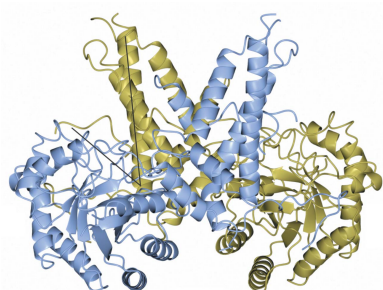
**Keywords:** O-GlcNAc; O-GlcNAcase; neurodegeneration; surface-entropy reduction; crystallization.**PDB reference:** surface-entropy reduction mutant of O-GlcNAcase, 6hki**Supporting information:** this article has supporting information at journals.iucr.org/d

The enzyme O-GlcNAcase catalyses the removal of the O-GlcNAc co/post-translational modification in multicellular eukaryotes. The enzyme has become of acute interest given the intimate role of O-GlcNAcylation in tau modification and stability; small-molecular inhibitors of human O-GlcNAcase are under clinical assessment for the treatment of tauopathies. Given the importance of structure-based and mechanism-based inhibitor design for O-GlcNAcase, it was sought to test whether different crystal forms of the human enzyme could be achieved by surface mutagenesis. Guided by surface-entropy reduction, a Glu602Ala/Glu605Ala variant [on the Gly11–Gln396/Lys535–Tyr715 construct; Roth *et al.* (2017), *Nature Chem. Biol.* **13**, 610–612] was obtained which led to a new crystal form of the human enzyme. An increase in crystal contacts stabilized disordered regions of the protein, enabling 88% of the structure to be modelled; only 83% was possible for the wild-type construct. Although the binding of the C-terminus was consistent with the wild type, Lys713 in monomer *A* was bound in the –1 subsite of the symmetry-related monomer *A* and the active sites of the *B* monomers were vacant. The new crystal form presents an opportunity for enhanced soaking experiments that are essential to understanding the binding mechanism and substrate specificity of O-GlcNAcase.

## 1. Introduction

The regulation of O-GlcNAc cycling on thousands of nuclear and cytoplasmic proteins is coordinated by two enzymes. O-GlcNAc transferase (OGT) catalyses the addition of GlcNAc, derived from UDP-GlcNAc, to serine and threonine residues, and O-GlcNAcase (OGA; CAZY database family GH84) cleaves O-GlcNAc (Torres & Hart, 1984; Holt & Hart, 1986; Kreppel *et al.*, 1997; Dong & Hart, 1994; Lubas *et al.*, 1997; Hart *et al.*, 2007). Two isoforms, OGA-L and OGA-S, are localized to the nucleus/cytoplasm (Comtesse *et al.*, 2001) and to the surface of lipid droplets, respectively. The reciprocal relationship between O-phosphorylation and O-glycosylation on the particular protein tau has been keenly studied in the context of neurodegeneration (Arnold *et al.*, 1996; Yuzwa *et al.*, 2008, 2014; Shen *et al.*, 2012; Griffith & Schmitz, 1995; Gao *et al.*, 2001; Liu *et al.*, 2004). In patients with Alzheimer's disease, tau undergoes hyperphosphorylation, causing it to dissociate from microtubules and aggregate into paired helical filaments (PHF) and neurofibrillary tangles (NFTs) (Grundke-Iqbal *et al.*, 1986; Marotta *et al.*, 2015). O-GlcNAc cycling has also been implicated in tumorigenesis owing to its significant role in orchestrating a vast number of cellular processes, for example transcriptional and cytoskeletal regulation, cell signalling and division, and metabolism (Slawson & Hart, 2011).

Structural characterization of the human O-GlcNAcase orthologue (*HsOGA-L/HsOGA*) showed dimer formation



OPEN ACCESS

both in solution and in the crystal where it has been shown to be promoted by helix-exchange (Roth *et al.*, 2017) in contrast to the non-helix-exchanged bacterial dimers (Dennis *et al.*, 2006; Schimpl *et al.*, 2010; Rao *et al.*, 2006). Composed of two ordered domains, the N-terminal glycoside hydrolase domain forms a ( $\beta/\alpha$ )<sub>8</sub> barrel (Li *et al.*, 2012), while the C-terminal stalk domain (Toleman *et al.*, 2004; He *et al.*, 2014) is composed of six  $\alpha$ -helices.  $\alpha$ 17, consisting of Glu676–Pro694, undergoes a dimer ‘swap’ (Roth *et al.*, 2017), playing a structural role that contributes to dimerization (Li, Li, Lu *et al.*, 2017; Elsen *et al.*, 2017; Roth *et al.*, 2017). Located in a V-shaped cleft formed between the N-terminal domain of monomer *A* and the C-terminal domain of monomer *B* (Fig. 1*a*) are the catalytic residues, Asp174 and Asp175, that initially act as a general acid and a general base through a two-step retaining mechanism (Macauley *et al.*, 2005). The terminal domain, which is not present in OGA-S, has a ‘pseudo’-histone acetyltransferase activity but was not included in the crystallized construct in the three *Hs*OGA structures as it has a high degree of disorder.

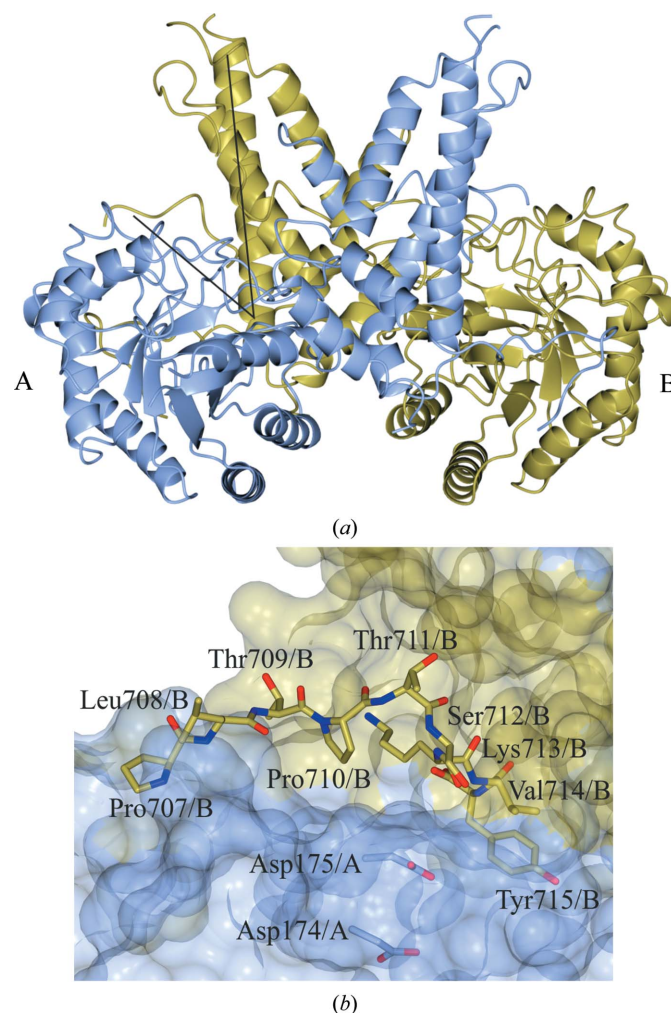
The recognition mode for glycopeptide substrates has been explored by obtaining structures of the catalytically inactive Asp175Asn mutant of *Hs*OGA in complex with a variety of glycosylated peptides (Li, Li, Lu *et al.*, 2017; Li, Li, Hu *et al.*, 2017). The peptides bound in each structure can be segregated into two binding modes with forward or reverse orientations of the peptide (amino to carboxyl or carboxyl to amino, respectively) within the binding site. Initially, crystallographic peptide studies were conducted using an orthologue from the bacterium *Clostridium perfringens* (Schimpl *et al.*, 2012; Mariappa *et al.*, 2015) in complex with TAB1, lamin B1 and p53 glycosylated peptides. These peptides bound in the forward direction. Later, Li, Li, Hu *et al.* (2017) compared the same glycosylated peptides with *Hs*OGA and they were found to act in the same way. However,  $\alpha$ -crystallin B and ELK1 bound in the reverse direction, strengthening the interest in determining how OGA selects its target (Li, Li, Hu *et al.*, 2017).

In the ‘apo’ structure, *Hs*OGA<sub>11–396+535–715</sub> (thus named owing to the co-expression of two plasmids containing residues 11–396 and 535–715), which was published by Roth *et al.* (2017) and deposited as PDB entry 5m7r, the C-terminus of monomer *A* from residues Pro707 to Tyr715 can be traced in the reverse direction into the active-site groove of the symmetry-related molecule of monomer *B* (Fig. 1*b*). When overlaid with the *Hs*OGA–p53 complex (PDB entry 5un8; Li, Li, Lu *et al.*, 2017), the position of the Tyr715 hydroxyl group lies directly on top of the O6 of O-GlcNAc.

To conduct biochemical/biophysical studies and rational drug design, complete and coherent structures are required. However, the published structures of *Hs*OGA have stretches of residues that are incomplete (Elsen *et al.*, 2017; Li, Li, Lu *et al.*, 2017; Li, Li, Hu *et al.*, 2017). For example, the structure described by Roth *et al.* (2017) has regions within the N-terminus (Glu11–Arg58 and Lys341–Thr370) and the C-terminus (Asp696–Pro706) that are not modelled in the structure. Expulsion of the C-terminus of *Hs*OGA<sub>11–396+535–715</sub>

from the active site is required before a competing compound can bind, therefore making it challenging to conduct crystal-soaking experiments; this is a problem for weak-binding compounds/inhibitors.

The work in this paper utilized surface-entropy reduction (SER) to enhance the structural characterization of *Hs*OGA and to contribute towards the hypothesis for the substrate-recognition mode of OGA, in which either the O-GlcNAc moiety or the peptide sequence is important for recognition. Using the *SERp* online server (Goldschmidt *et al.*, 2007), potential pairs of mutations were identified for *Hs*OGA. The rational design of mutating clusters of residues on a protein is a favourable strategy, with the aim of making the protein more susceptible to crystallization (Derewenda, 2004*a,b*; Derewenda & Vekilov, 2006; Cooper *et al.*, 2007). Surface-entropy reduction (SER) is a concept in which flexible, solvent-exposed residues, primarily lysine and glutamate, are mutated to alanine to reduce the entropic loss during the packing of the



**Figure 1**  
*Hs*OGA forms a dimer in solution and in the crystal structure. Monomer *A* is shown in blue and monomer *B* in gold. (*a*) One of the two active sites is indicated by intersecting black lines. (*b*) The C-terminal peptide, Pro707–Tyr715 in chain *B*, bound in the active site of monomer *A* with the catalytic residues, Asp174/*A* and Asp175/*A*, displayed.

protein into a crystalline lattice (Longenecker *et al.*, 2001; Mateja *et al.*, 2002; Vekilov, 2003; Vekilov *et al.*, 2002).

Following the co-expression strategy of Roth *et al.* (2017) using pACYC-Duet-1 Gly11–Gln396 and pET-YSBLIC3C Lys535–Tyr715 plasmids, the successful pair of mutations was Glu602Ala/Glu605Ala (*HsOGA*<sub>E602AE605A</sub>). Unlike *HsOGA*<sub>11–396+535–715</sub>, the C-terminus of monomer A can be traced into the active site of the symmetry-related monomer A, with Lys713 binding in the position of O-GlcNAc. Furthermore, previously disordered loops had become ordered and could be built into the final model. Additionally, the activities and secondary-structure profiles of full-length *HsOGA* (*HsOGA*<sub>FL</sub>), *HsOGA*<sub>11–396+535–715</sub> and *HsOGA*<sub>E602AE605A</sub> were examined to verify the loss of the ‘pseudo-HAT’ domain and that the SER mutation did not alter the activity. The results showed that the SER mutant exhibited similar  $K_m$  values to the full-length enzyme, since the mutation is distant from the active site, highlighting the potential of SER variants for studying the structural and ligand-binding characteristics of *HsOGA*.

## 2. Materials and methods

### 2.1. Macromolecular production

The cloning of the constructs has been described previously (Roth *et al.*, 2017). Homologous DNA for the Lys535–Tyr715 construct, in the vector pET-YSBL-LIC-3C, was mutated using a Q5 Site-Directed Mutagenesis Kit with Q5 Hot Start High-Fidelity DNA Polymerase. The forward primer, A GAT AGC gct AAA ATC gct GAA TGG, was designed to mutate the primary sequence A GAT AGC GAA AAA ATC GAA GAA; the reverse primer was TTA CCC TTG CAG TTA ACC GAA. NEB 5-alpha competent *Escherichia coli* cells were transformed with the Lys535–Tyr715 E602AE605A construct. The DNA was extracted and sequenced to verify the mutation.

The Gly11–Gln396 and Lys535–Tyr715 E602AE605A constructs in the vectors pACYCDUET-1 and pET-YSBL-LIC-3C, respectively, were co-transformed into *E. coli* BL21 (DE3) cells. Luria–Bertani broth was inoculated with a cell suspension and was incubated at 37°C until the OD<sub>600</sub> reached 1.0. The cells were induced with 1 mM IPTG and incubated at 16°C for 20 h.

The purification of *HsOGA*<sub>E602AE605A</sub> followed the same purification protocol as that of *HsOGA*<sub>11–396+535–715</sub> described previously (Roth *et al.*, 2017; Supplementary Figs. S1a and S1b).

### 2.2. Crystallization

*HsOGA*<sub>E602AE605A</sub> was initially crystallized by sitting-drop vapour diffusion at 15 mg ml<sup>−1</sup> under condition E11 of the PACT *premier* screen from Molecular Dimensions (Newman *et al.*, 2005): 0.2 M sodium citrate tribasic, 20% PEG 3350. Further optimization of the conditions to 0.2 M sodium citrate tribasic pH 7.5, 17% polyethylene glycol 3350 in a 48-well MRC MAXI optimization plate improved the crystal shape.

**Table 1**

Data collection and processing.

Values in parentheses are for the outer shell.

|  |                            |
|--|----------------------------|
| Diffraction source   | Beamline I04-1, DLS        |
| Wavelength (Å)   | 0.9282                     |
| Temperature (K)  | 100                        |
| Detector   | PILATUS 6M-F               |
| Rotation range per image (°)                               | 0.10                       |
| Total rotation range (°)                                   | 360                        |
| Exposure time per image (s)                                | 0.040                      |
| Space group  | <i>P</i> 3 <sub>1</sub> 21 |
| <i>a</i> , <i>b</i> , <i>c</i> (Å)                         | 222.2, 222.2, 72.4         |
| $\alpha$ , $\beta$ , $\gamma$ (°)                          | 90.0, 90.0, 120.0          |
| Resolution range (Å)                                       | 192.4–3.3 (3.5–3.3)        |
| Total No. of reflections                                   | 523842 (75864)             |
| No. of unique reflections                                  | 31103 (4440)               |
| Completeness (%)   | 99.8 (98.6)                |
| Multiplicity   | 16.8 (17.1)                |
| $\langle I/\sigma(I) \rangle$                              | 7.3 (1.8)†                 |
| $R_{p.i.m.}$   | 0.11 (0.62)                |
| CC <sub>1/2</sub>  | 0.99 (0.67)                |
| Overall <i>B</i> factor from Wilson plot (Å <sup>2</sup> ) | 58.53                      |

† The mean  $I/\sigma(I)$  in the outer shell fell below <2.0 at 3.4 Å resolution.

The total volume of the drop was 1 µl and the protein:reservoir solution ratio was 1:1; the total volume in the reservoir was 100 µl.

### 2.3. Data collection and processing

Diffraction images were collected on beamline I04-1 at Diamond Light Source (DLS). After data collection, the diffraction images were integrated using the *-3dii* option in *xia2* (Winter, 2010) and reintegrated using *AIMLESS* (Evans, 2006; Evans & Murshudov, 2013) from the *CCP4* software suite (Winn *et al.*, 2011). Data-collection and processing statistics are given in Table 1.

### 2.4. Structure solution and refinement

Molecular replacement against the coordinates of PDB entry 5m7r was conducted using *MOLREP* (Vagin & Teplyakov, 2010). Refinement of the model was conducted using multiple rounds of *REFMAC* (Murshudov *et al.*, 1997, 2011; Pannu *et al.*, 1998; Winn *et al.*, 2003; Vagin *et al.*, 2004; Nicholls *et al.*, 2012) and manual model building in *Coot* (Emsley *et al.*, 2010). Waters were added using Find Waters in *Coot* and validated. The data were processed to a resolution of 3.3 Å (Table 2).

### 2.5. Michaelis–Menten kinetics

Michaelis–Menten kinetics were assayed using *HsOGA*<sub>FL</sub> and *HsOGA*<sub>11–396+535–715</sub> as positive controls against the mutant *HsOGA*<sub>E602AE605A</sub>. In a 200 µl reaction volume, 50 nM protein solution and a serial dilution of the ligand 4-nitrophenyl *N*-acetyl-β-D-glucosaminide (pNP-GlcNAc) from 1500 to 11.7 µM [dissolved in 2.5% DMSO (final concentration)] was added to PBS buffer at pH 7.4.

The reaction was monitored at 405 nM continuously using a Molecular Devices SpectraMax M5 spectrophotometer. The



experiment, which was conducted at 25°C, was duplicated and each ligand concentration was repeated in triplicate.

*GraphPad Prism* v.5 was used to process the data, with nonlinear regression of Michaelian saturation curves. The initial velocities were calculated from the linear range of the reaction-progress curve. A standard curve of 4-nitrophenol was used to extract a molar extinction coefficient.

## 2.6. Circular-dichroism spectroscopy

The protein samples were dialysed overnight into 25 mM sodium phosphate pH 8.0 and diluted to 0.1 mg ml<sup>-1</sup>. The spectra were recorded at 21°C in a QS 248 0.2 mm cuvette with 0.5 s per point and 78 s per read. The wavelength ranged from 195 to 320 nm. The background for each protein was measured immediately before the experiment in the same cuvette and values were taken as averages from triplicate reads.

## 3. Results and discussion

### 3.1. Comparison of the mutant crystal structure with that of the wild type

To incorporate protein molecules into a crystal, a thermodynamic cost is endured to bury hydrophobic residues into a constrained conformation and from the immobilization of the prevalent flexible hydrophilic side chains on the surface (Avbelj & Fele, 1998). Reducing the entropic shield can lead to an increase in the variety of conditions, morphologies and crystallographic space groups (Parthasarathy *et al.*, 2008; Kim *et al.*, 2005). Therefore, crystallization conditions were re-screened using the *HsOGA*<sub>E602AE605A</sub> variant; crystals were obtained in 17% polyethylene glycol 3350, 0.2 M sodium citrate pH 8.0. This is comparable to the wild-type crystallization condition of 0.1–0.2 M triammonium citrate pH 6.5–7.5, 16–24% PEG 3350.

In a different crystal form, flexible loops can become ordered by making backbone crystal contacts or adopting preferential conformations (Derewenda, 2004a). The space group was *P*<sub>3</sub><sub>2</sub><sub>1</sub>, which is a lower symmetry group compared with *P*<sub>4</sub><sub>3</sub><sub>2</sub><sub>1</sub> for the *HsOGA*<sub>11–396+535–715</sub> structure. The data statistics are presented in Tables 1 and 2.

Theoretically, SER should lead to an improvement in the resolution and hence the overall quality of the structure (Parthasarathy *et al.*, 2008). However, the structure was determined to a resolution of 3.3 Å, which was a lower resolution compared with the *HsOGA*<sub>11–396+535–715</sub> and catalytically inactive mutant structures, possibly owing to the thin-rod crystal morphology compared with the large trigonal bipyramid wild-type crystals.

A solvent channel can be observed through the crystal structure (Fig. 2a). The dimers form a trigonal spring such that the C-terminus of monomer *A* binds into the active-site monomer *A* of a symmetry-related dimer, *etc.* (Figs. 2a and 2b). Owing to the different crystal contacts made on the surface of the protein, regions of high disorder could be built into the structure. 88% of the structure was complete, in comparison with 83% of the *HsOGA*<sub>11–396+535–715</sub> structure.

Table 2

Structure solution and refinement.

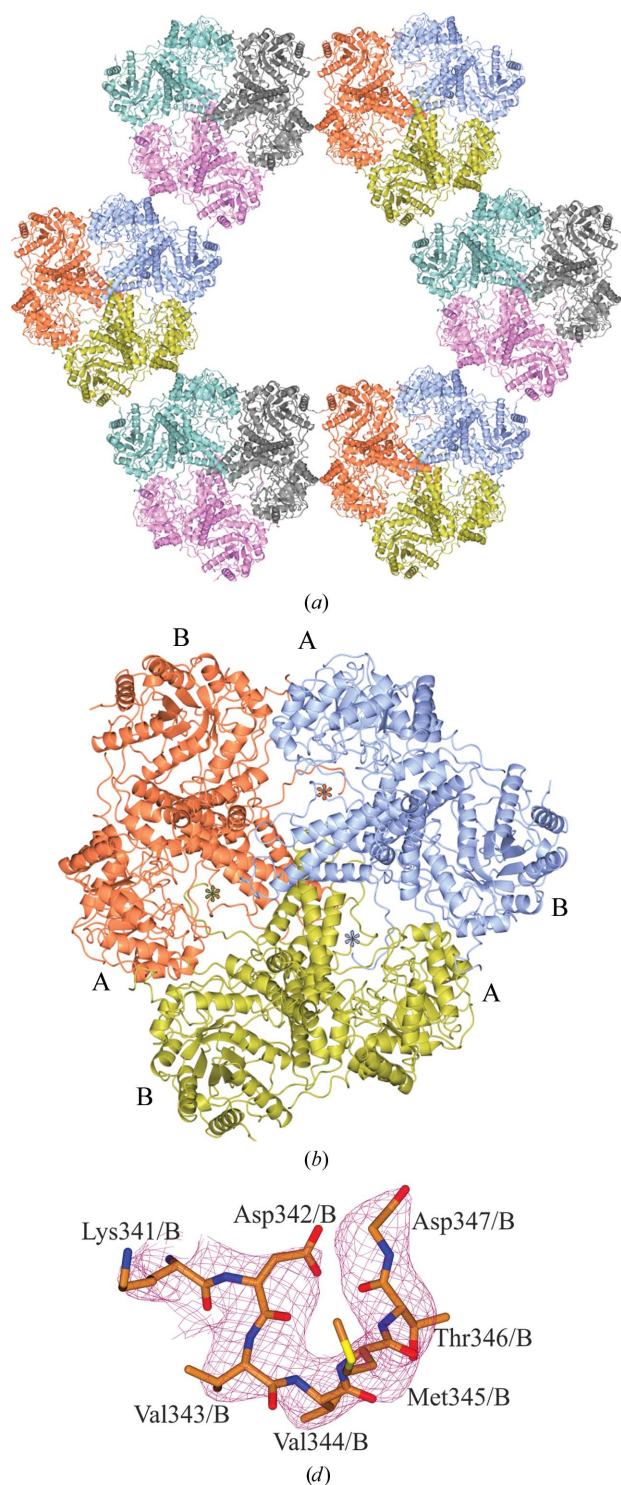
Values in parentheses are for the outer shell.

|  |                     |
|--|---------------------|
| Resolution range (Å)                       | 192.4–3.3 (3.5–3.3) |
| No. of reflections                         | 31064               |
| Final <i>R</i> <sub>cryst</sub>            | 0.17                |
| Final <i>R</i> <sub>free</sub>             | 0.23                |
| Cruickshank DPI                            | 0.40                |
| No. of non-H atoms                         |                     |
| Protein                                    | 7877                |
| Water                                      | 1                   |
| Total                                      | 7878                |
| R.m.s. deviations                          |                     |
| Bonds (Å)                                  | 0.010               |
| Angles (°)                                 | 1.58                |
| Average <i>B</i> factors (Å <sup>2</sup> ) |                     |
| Protein                                    | 75                  |
| Water                                      | 16                  |
| Ramachandran plot                          |                     |
| Most favoured (%)                          | 92                  |
| Allowed (%)                                | 5                   |

The regions of highest disorder in both monomers were between Lys341 and Thr370, in addition to loops on the protein surface. After multiple rounds of refinement, the confidence for the inclusion of residues Lys341–Asp347 increased, enabling further visualization of the disordered region (Figs. 2c and 2d). The residues that were still disordered in monomers *A* and *B* were Gly11–Gly56, Ser348–Glu369, Lys535 and Pro707–Tyr715 for monomer *B* only. In the protein structures described by Elsen *et al.* (2017) and Li, Li, Lu *et al.* (2017) the residues Lys341–Asp371 in monomer *A* (PDB entry 5uhk) and Asn335–Val372 in monomer *A* (PDB entry 5tke), respectively, were also not observed.

When the N-terminal domains of monomer *A* from *HsOGA*<sub>E602AE605A</sub> and *HsOGA*<sub>11–396+535–715</sub> were overlapped, the r.m.s.d. of the N-terminal domains was 1.5 Å and the r.m.s.d. for the C-terminal domains was 7.1 Å, indicating a high degree of flexibility in the latter domain (Fig. 2e).

As mentioned, the C-terminus of monomer *A* was found to bind into the active site of a symmetry-related monomer *A*, aiding the formation of the new crystal packing (Figs. 2b and 3a). Initially, the residues of α17 interact with monomer *B* in a domain swap; the residues from Pro694 to Phe703 then bend back towards the residues of the respective monomer, with Gln704–Tyr715 binding into the active site (Figs. 3a and 3b). Electrostatic interactions between the C-terminus of monomer *A* and other residues of monomer *A*, *B* and a symmetry-related molecule *B* stabilize this formation. Pro707 of *HsOGA*<sub>E602AE605A</sub> has drastically moved position and faces in the opposite direction. There are three consecutive prolines that facilitate the change in direction. Hence, the C-terminus binds to the active site of the symmetry-related monomer *A* rather than monomer *B* (as in the wild type). The C-terminus of monomer *B* could not be built in from Pro707, indicating that it does not bind into an active site because of the crystal packing. In *HsOGA*<sub>11–396+535–715</sub>, residues Asp696–Pro707 in monomer *A* are missing; therefore, the direction of the peptide is ambiguous. In the natural human sequence, the pseudo-histone acetyltransferase domain is connected to the



**Figure 2**

(a, b) Crystal symmetry of *HsOGA*<sub>E602AE605A</sub>. (a) The connections made between the dimers show the trigonal solvent channel, with 50% solvent as calculated from the Matthews coefficient of  $2.51 \text{ \AA}^3 \text{ Da}^{-1}$  (Kantardjieff & Rupp, 2003; Matthews, 1968). (b) Side view of the repeating trigonal dimers showing the linking C-terminus of monomer A binding into monomer A of the next dimer complex. The monomers are labelled A and B, with asterisks indicating the C-termini. (c, d) Disordered loop regions were stabilized in the new crystal form. Monomer A is shown in sea green and monomer B in brown. (c) The regions in purple were built into the *HsOGA*<sub>E602AE605A</sub> structure using PDB entry 5m7r as the template model and were missing from the wild-type structure. (d) Residues Lys341–Gly347 and the maximum-likelihood/ $\sigma_A$ -weighted  $2F_{\text{obs}} - F_{\text{calc}}$  map shown in red contoured at  $0.12 \text{ e \AA}^{-3}$ . (e) Overlap of the N-terminal monomers A from chain A for both *HsOGA*<sub>11–396+535–715</sub> (monomer A, blue; monomer B, gold) and *HsOGA*<sub>E602AE605A</sub> (monomer A, sea green; monomer B, brown) as calculated by CCP4mg Superpose models. The residue range selected for superposition was Arg59/A–His395/A.

stalk domain; therefore, binding of the C-terminus in the active site or alternatively C-termini disorder are possible artefacts of the removal of the HAT domain and of crystal packing.

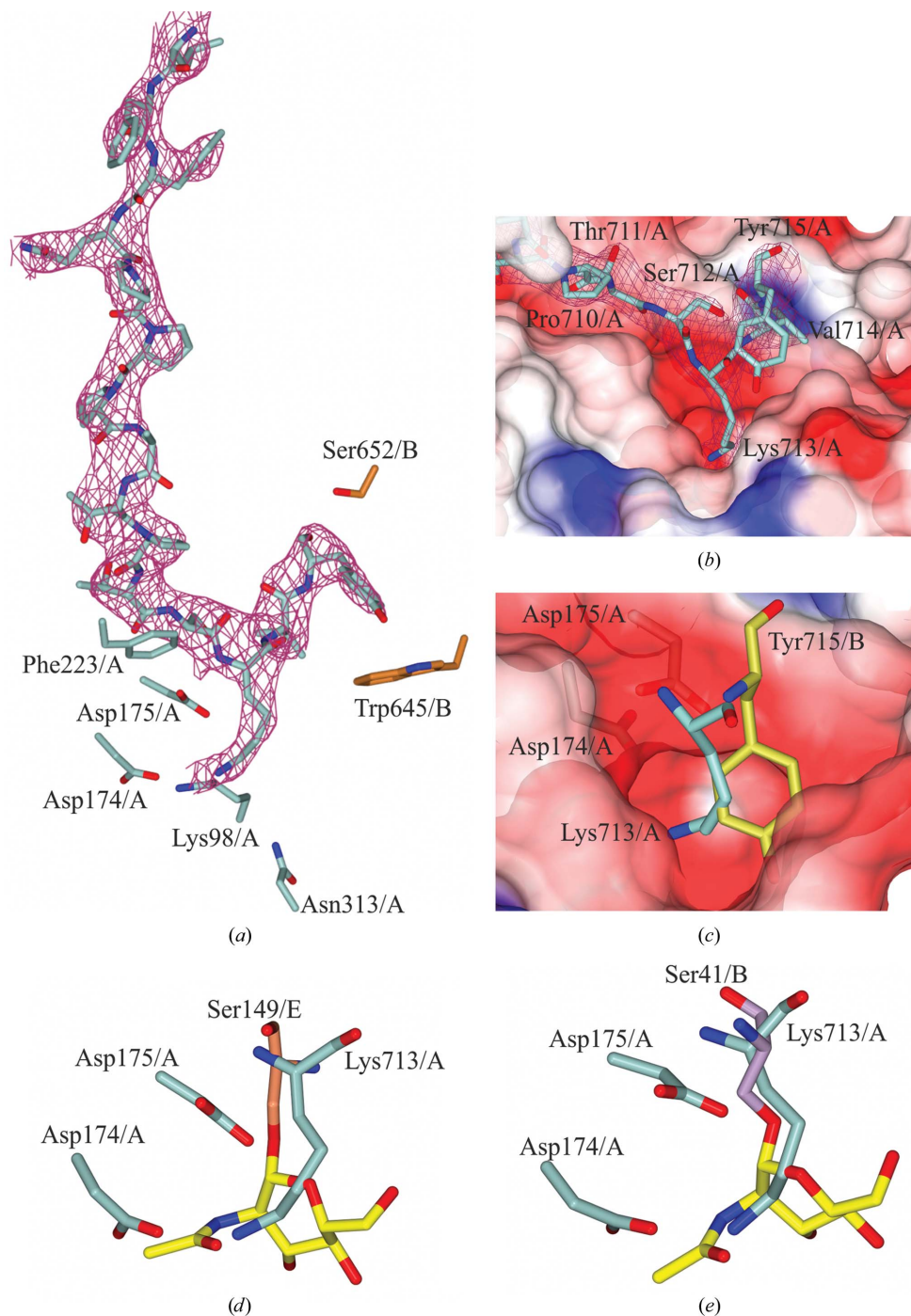
The C-terminal residues Lys713/A and Tyr715/A hydrogen-bond to and make electrostatic interactions with residues of the active site (Fig. 3*b*). Interestingly, the NZ atom of Lys713 is in the same position as the N2 atom of O-GlcNAc and is  $2.97 \text{ \AA}$  away from the  $\text{O}^\delta$  atom of Asp175/A on the symmetry-related monomer (Fig. 3*d*). This pushes Tyr715 into the +2 subsite, where it interacts with the  $\text{O}^\gamma$  atom of Ser652/B of the





symmetry-related monomer (Fig. 3*a*). In comparison, the Tyr715/A hydroxyl group of the wild type lies above O6 of O-GlcNAc and hydrogen-bonds to the O<sup>δ1</sup> atom of Asp285/B (Fig. 3*c*).

In comparison to the crystal structure of *HsOGA* in complex with  $\alpha$ -crystallin B and ELK1 (PDB entries 5vvv and 5vvt, respectively; Li, Li, Lu *et al.*, 2017), the C-terminal residues are in the same reverse direction as the glycosylated



**Figure 3**

The C-terminus of monomer *A* binds into the active site of the symmetry-related monomer *A*. (*a*) Neighbouring residues in the active site of symmetry-related monomer *A* of *HsOGA*<sub>E602AE605A</sub>, with the C-terminal residues bound and the maximum-likelihood/ $\sigma_A$ -weighted  $2F_{\text{obs}} - F_{\text{calc}}$  map shown in red contoured at  $0.17 \text{ e } \text{\AA}^{-3}$ . (*b*) Surface representation of the active site with the C-terminus bound in a negatively charged pocket. (*c*) Overlay of Lys713/A from *HsOGA*<sub>E602AE605A</sub> and Tyr715/B from *HsOGA*<sub>11–396+535–715</sub> (in gold) in the binding pocket. (*d*) Overlay of the C-terminus of chain *A* of *HsOGA*<sub>E602AE605A</sub> and that of *HsOGA* in complex with glycosylated p53 peptide (PDB entry 5un8), showing Ser149/*E* in coral. O-GlcNAc is shown in yellow. (*e*) Overlay of Lys713/A from *HsOGA*<sub>E602AE605A</sub> with *HsOGA* in complex with glycosylated  $\alpha$ -crystallin B (PDB entry 5vvv), showing Ser41/*B* in purple.

Table 3

Kinetics analysis comparing different *HsOGA* constructs.

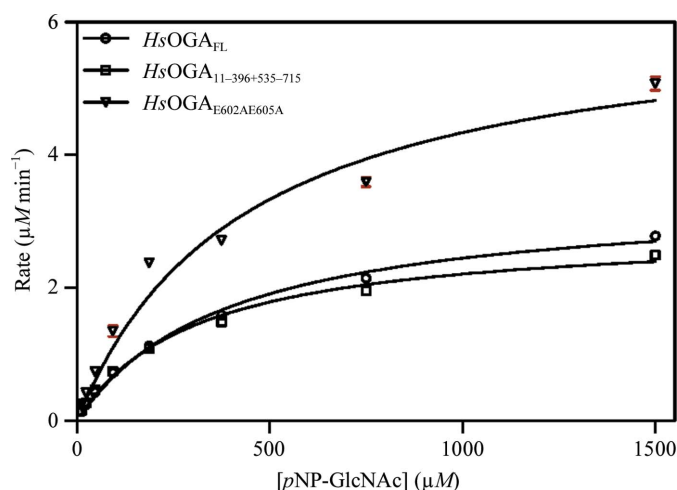
| Construct   | <i>HsOGA</i> <sub>FL</sub> | <i>HsOGA</i> <sub>11-396+535-715</sub> | <i>HsOGA</i> <sub>E602AE605A</sub> |
|---|----------------------------|--|------------------------------------|
| $V_{\max}$ ( $\mu\text{M min}^{-1}$ )                       | $2.96 \pm 0.07$            | $2.50 \pm 0.06$                        | $4.62 \pm 0.15$                    |
| $K_m$ ( $\mu\text{M}$ )                                     | $298 \pm 15$               | $227 \pm 13$                           | $230 \pm 20$                       |
| $k_{\text{cat}}$ ( $\text{min}^{-1}$ )                      | $59.1 \pm 1.4$             | $49.9 \pm 1.2$                         | $92.5 \pm 3.1$                     |
| $k_{\text{cat}}/K_m$ ( $\text{min}^{-1} \mu\text{M}^{-1}$ ) | $0.198 \pm 0.012$          | $0.219 \pm 0.013$                      | $0.401 \pm 0.038$                  |

peptides (Fig. 3e). This is in contrast to the structure containing a glycosylated p53 peptide shown in Fig. 3(d).

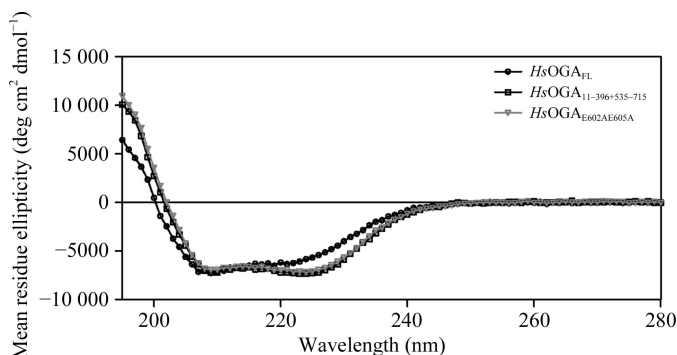
The density for all available *HsOGA* peptide-complex structures supports the notion that OGA can bind peptides in both the forward and reverse directions. Comparison between the different peptide structures shows the versatility of the active site for different peptides.

### 3.2. Comparison of the constructs and mutants

To ensure that the structural stabilization had not occurred owing to a change in the secondary structure of *HsOGA*<sub>E602AE605A</sub> and that the mutation had not affected the activity of the enzyme, the full-length protein (*HsOGA*<sub>FL</sub>),



**Figure 4**  
Michaelis-Menten curves for the kinetic assay of the *HsOGA* mutant. pNP-GlcNAc was used as the substrate at concentrations up to seven times higher than the  $K_m$ .



**Figure 5**  
Circular-dichroism spectra for the different constructs. The experimental data in millidegree units were converted to mean residue ellipticity values with units of  $\text{deg cm}^2 \text{dmol}^{-1}$  using the equation given in Ishtikhari *et al.* (2014).

*HsOGA*<sub>11-396+535-715</sub> and *HsOGA*<sub>E602AE605A</sub> underwent kinetic and secondary-structure analyses.

Comparison of the results for *HsOGA*<sub>FL</sub> and *HsOGA*<sub>11-396+535-715</sub> shows that the split construct has similar activity to the full-length variant, suggesting that co-expression of the two domains does not affect the activity (Fig. 4). Therefore, the ligand-binding data should be an accurate representation of binding to the full-length protein. When comparing *HsOGA*<sub>11-396+535-715</sub> and *HsOGA*<sub>E602AE605A</sub>, the values are very similar. The  $V_{\max}$  is higher, indicating an increase in the reaction rate (Table 3). The increase in  $V_{\max}$  may be owing to a discrepancy in the enzyme concentration ( $[E]$ ), as  $V_0$  is directly proportional to  $[E]$ . The similarity of  $V_{\max}$  and  $K_m$  for the mutant and the wild type suggest that the mutation did not alter the substrate-binding affinity or the enzyme activity of the protein for its substrate.

The CD spectra show that all of the constructs are fully folded and the spectra for *HsOGA*<sub>11-396+535-715</sub> and *HsOGA*<sub>E602AE605A</sub> are very similar, suggesting that there is no change in the composition of the mutant (Fig. 5). The spectral profile of *HsOGA*<sub>FL</sub> differs from those of the split constructs in that it has a less pronounced minimum in the 222 nm region and an overall blue-shifted spectrum in the <210 nm region. This suggests that there could be a lower proportion of  $\alpha$ -helical structure and/or higher structural disorder. This links to the inclusion of the pseudo-HAT domain that is connected to the C-terminal stalk domain. Since the structure of the human homologue of this domain is unknown owing to the inability of *HsOGA*<sub>FL</sub> to crystallize, the structure can only be inferred from the structural homologues from *Oceanicola granulosus* and *Streptomyces sveticus* (He *et al.*, 2014; Rao *et al.*, 2013). An estimate of the secondary-structure content of the proteins suggests a decrease of 6.4% in  $\alpha$ -helical components and an increase of 3.7% in  $\beta$ -sheet components in *HsOGA*<sub>FL</sub> compared with *HsOGA*<sub>11-396+535-715</sub> (Supplementary Table S1). Homology modelling using crystal structures of the HAT domain from the bacterial homologue *O. granulosus* suggests a similar overall structure minus the acetyltransferase activity (Rao *et al.*, 2013). However, the data could be skewed by the inclusion of the linker region to this domain and the potential difference in the homology model structure. The secondary-structure contents of *HsOGA*<sub>11-396+535-715</sub> and *HsOGA*<sub>E602AE605A</sub> are consistent, further confirming that the mutation did not affect the overall structure.

## 4. Conclusions

In this study, surface-entropy reduction has been utilized to produce further structural information on O-GlcNAcase by the incorporation of residues Ala57-Arg58, Lys341-Asp347, Thr370, Glu536, Cys596-Gly598, Gly674-Asp675 and Asp696-Pro707, an increase in the number of observed residues of 5%. Although the binding of the C-terminus to the active site may be an artefact of crystallization, it reveals further details regarding the substrate specificity of OGA, as peptides have been shown to bind in a bidirectional yet conserved conformation. The results described in this study



present an opportunity for further investigation of the binding orientation of peptides within an SER-modified OGA enzyme. Given the progression of hOGA inhibitors into clinical trials, different surface mutants of the enzyme may afford new routes to drug development.

## Acknowledgements

CD data collection was conducted by Dustin King and assistance in conducting the kinetics experiments and analysing the data was provided by Matthew Alteen, both of whom are employed by Professor David Vocado at Simon Fraser University. We thank Diamond Light Source for access to beamline I04-1 (proposal No. mx-13587) that contributed to the results presented here, and Johan Turkenburg and Sam Hart for coordinating data collection.

## Funding information

This research was funded by the Biotechnology and Biological Sciences Research Council (grant BB/M011151/1) for the support of AM. GJD is the Royal Society Ken Murray Research Professor.

## References

- Arnold, C. S., Johnson, G. V. W., Cole, R. N., Dong, D. L.-Y., Lee, M. & Hart, G. W. (1996). *J. Biol. Chem.* **271**, 28741–28744.
- Avbelj, F. & Fele, L. (1998). *J. Mol. Biol.* **279**, 665–684.
- Comtesse, N., Maldener, E. & Meese, E. (2001). *Biochem. Biophys. Res. Commun.* **283**, 634–640.
- Cooper, D. R., Boczek, T., Grelewski, K., Pinkowska, M., Sikorska, M., Zawadzki, M. & Derewenda, Z. (2007). *Acta Cryst. D* **63**, 636–645.
- Dennis, R. J., Taylor, E. J., Macauley, M. S., Stubbs, K. A., Turkenburg, J. P., Hart, S. J., Black, G. N., Vocado, D. J. & Davies, G. J. (2006). *Nature Struct. Mol. Biol.* **13**, 365–371.
- Derewenda, Z. S. (2004a). *Structure*, **12**, 529–535.
- Derewenda, Z. S. (2004b). *Methods*, **34**, 354–363.
- Derewenda, Z. S. & Vekilov, P. G. (2006). *Acta Cryst. D* **62**, 116–124.
- Dong, D. L.-Y. & Hart, G. W. (1994). *J. Biol. Chem.* **269**, 19321–19330.
- Elsen, N. L., Patel, S. B., Ford, R. E., Hall, D. L., Hess, F., Kandula, H., Kornienko, M., Reid, J., Selnick, H., Shipman, J. M., Sharma, S., Lumb, K. J., Soisson, S. M. & Klein, D. J. (2017). *Nature Chem. Biol.* **13**, 613–615.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D* **66**, 486–501.
- Evans, P. (2006). *Acta Cryst. D* **62**, 72–82.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst. D* **69**, 1204–1214.
- Gao, Y., Wells, L., Comer, F. I., Parker, G. J. & Hart, G. W. (2001). *J. Biol. Chem.* **276**, 9838–9845.
- Goldschmidt, L., Cooper, D. R., Derewenda, Z. S. & Eisenberg, D. (2007). *Protein Sci.* **16**, 1569–1576.
- Griffith, L. S. & Schmitz, B. (1995). *Biochem. Biophys. Res. Commun.* **213**, 424–431.
- Grundke-Iqbal, I., Iqbal, K., Tung, Y.-C., Quinlan, M., Wisniewski, H. M. & Binder, L. I. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 4913–4917.
- Hart, G. W., Housley, M. P. & Slawson, C. (2007). *Nature (London)*, **446**, 1017–1022.
- He, Y., Roth, C., Turkenburg, J. P. & Davies, G. J. (2014). *Acta Cryst. D* **70**, 186–195.
- Holt, G. D. & Hart, G. W. (1986). *J. Biol. Chem.* **261**, 8049–8057.
- Ishtikhari, M., Khan, S., Badr, G., Osama Mohamed, A. & Hasan Khan, R. (2014). *Mol. Biosyst.* **10**, 2954–2964.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Kim, A.-R., Dobransky, T., Rylett, R. J. & Shilton, B. H. (2005). *Acta Cryst. D* **61**, 1306–1310.
- Kreppel, L. K., Blomberg, M. A. & Hart, G. W. (1997). *J. Biol. Chem.* **272**, 9308–9315.
- Li, B., Li, H., Hu, C.-W. & Jiang, J. (2017). *Nature Commun.* **8**, 666.
- Li, B., Li, H., Lu, L. & Jiang, J. (2017). *Nature Struct. Mol. Biol.* **24**, 362–369.
- Li, J., Li, Z., Li, T., Lin, L., Zhang, Y., Guo, L., Xu, Y., Zhao, W. & Wang, P. (2012). *Biochemistry (Mosc.)*, **77**, 194–200.
- Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W. & Gong, C.-X. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 10804–10809.
- Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. (2001). *Acta Cryst. D* **57**, 679–688.
- Lubas, W. A., Frank, D. W., Krause, M. & Hanover, J. A. (1997). *J. Biol. Chem.* **272**, 9316–9324.
- Macauley, M. S., Whitworth, G. E., Debowski, A. W., Chin, D. & Vocado, D. J. (2005). *J. Biol. Chem.* **280**, 25313–25322.
- Mariappa, D., Selvan, N., Borodkin, V. S., Alonso, J., Ferenbach, A. T., Shepherd, C., Navratilova, I. H. & van Aalten, D. M. F. (2015). *Biochem. J.* **470**, 255–262.
- Marotta, N. P., Lin, Y. H., Lewis, Y. E., Ambrosio, M. R., Zaro, B. W., Roth, M. T., Arnold, D. B., Langen, R. & Pratt, M. R. (2015). *Nature Chem.* **7**, 913–920.
- Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K., Dauter, Z., Otlewski, J. & Derewenda, Z. S. (2002). *Acta Cryst. D* **58**, 1983–1991.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- 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.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst. D* **53**, 240–255.
- Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst. D* **61**, 1426–1431.
- Nicholls, R. A., Long, F. & Murshudov, G. N. (2012). *Acta Cryst. D* **68**, 404–417.
- Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst. D* **54**, 1285–1294.
- Parthasarathy, G., Cummings, R., Becker, J. W. & Soisson, S. M. (2008). *Acta Cryst. D* **64**, 141–148.
- Rao, F. V., Dorfmueller, H. C., Villa, F., Allwood, M., Eggleston, I. M. & van Aalten, D. M. F. (2006). *EMBO J.* **25**, 1569–1578.
- Rao, F. V., Schüttelkopf, A. W., Dorfmueller, H. C., Ferenbach, A. T., Navratilova, I. & van Aalten, D. M. F. (2013). *Open Biol.* **3**, 130021.
- Roth, C., Chan, S., Offen, W. A., Hemsworth, G. R., Willems, L. I., King, D. T., Varghese, V., Britton, R., Vocado, D. J. & Davies, G. J. (2017). *Nature Chem. Biol.* **13**, 610–612.
- Schimpl, M., Borodkin, V. S., Gray, L. J. & van Aalten, D. M. F. (2012). *Chem. Biol.* **19**, 173–178.
- Schimpl, M., Schüttelkopf, A. W., Borodkin, V. S. & van Aalten, D. M. F. (2010). *Biochem. J.* **432**, 1–7.
- Shen, D. L., Gloster, T. M., Yuzwa, S. A. & Vocado, D. J. (2012). *J. Biol. Chem.* **287**, 15395–15408.
- Slawson, C. & Hart, G. W. (2011). *Nature Rev. Cancer*, **11**, 678–684.
- Toleman, C., Paterson, A. J., Whisenhunt, T. R. & Kudlow, J. E. (2004). *J. Biol. Chem.* **279**, 53665–53673.
- Torres, C. R. & Hart, G. W. (1984). *J. Biol. Chem.* **259**, 3308–3317.
- Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Cryst. D* **60**, 2184–2195.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst. D* **66**, 22–25.
- Vekilov, P. G. (2003). *Methods Enzymol.* **368**, 84–105.
- Vekilov, P. G., Feeling-Taylor, A., Yau, S.-T. & Petsev, D. (2002). *Acta Cryst. D* **58**, 1611–1616.
- 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.
- Winn, M. D., Murshudov, G. N. & Papiz, M. Z. (2003). *Methods Enzymol.* **374**, 300–321.
- Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.
- Yuzwa, S. A., Macauley, M. S., Heinonen, J. E., Shan, X., Dennis, R. J., He, Y., Whitworth, G. E., Stubbs, K. A., McEachern, E. J., Davies, G. J. & Vocadlo, D. J. (2008). *Nature Chem. Biol.* **4**, 483–490.
- Yuzwa, S. A., Shan, X., Jones, B. A., Zhao, G., Woodward, M. L., Li, X., Zhu, Y., McEachern, E. J., Silverman, M. A., Watson, N. V., Gong, C.-X. & Vocadlo, D. J. (2014). *Mol. Neurodegener.* **9**, 42.