Crystal structure of XCC3289 from *Xanthomonas campestris*: homology with the N-terminal substrate-binding domain of Lon peptidase

Rahul Singh, Sonali Deshmukh, Ashwani Kumar, Venuka Durani Goyal and Ravindra D. Makde*

Received 12 August 2020

Accepted 28 August 2020

Edited by J. Newman, Bio21 Collaborative Crystallisation Centre, Australia

Keywords: LonA; cereblon; protein binding; XCC3289; *Xanthomonas campestris*.

PDB reference: XCC3289, 7cay

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

Beamline Development and Application Section, Bhabha Atomic Research Centre, Mumbai 400 085, India.

*Correspondence e-mail: ravimakde@rrcat.gov.in

LonA peptidase is a major component of the protein quality-control mechanism in both prokaryotes and the organelles of eukaryotes. Proteins homologous to the N-terminal domain of LonA peptidase, but lacking its other domains, are conserved in several phyla of prokaryotes, including the Xanthomonadales order. However, the function of these homologous proteins (LonNTD-like proteins) is not known. Here, the crystal structure of the LonNTD-like protein from *Xanthomonas campestris* (XCC3289; UniProt Q8P5P7) is reported at 2.8 Å resolution. The structure was solved by molecular replacement and contains one polypeptide in the asymmetric unit. The structure was refined to an R_{free} of 29%. The structure of XCC3289 consists of two domains joined by a long loop. The N-terminal domain (residues 1–112) consists of an α -helix surrounded by β -sheets, whereas the C-terminal domain (residues 123–193) is an α -helical bundle. The fold and spatial orientation of the two domains closely resembles those of the N-terminal domains of the LonA peptidases from *Escherichia coli* and *Mycobacterium avium*. The structure is also similar to that of cereblon, a substrate-recognizing component of the E3 ubiquitin ligase complex. The N-terminal domains of both LonA and cereblon are known to be involved in specific protein–protein interactions. This structural analysis suggests that XCC3289 and other LonNTD-like proteins might also be capable of such protein–protein interactions.

1. Introduction

Lon peptidases are ATP-dependent serine peptidases that are classified into the S16 family in the MEROPS database (Rawlings *et al.*, 2014). Lon peptidases are crucial for cell survival as they mediate the degradation of mutated, abnormally folded and some short-lived regulatory proteins into small peptides (of 5–10 amino acids in length; Lee & Suzuki, 2008). Type A Lon (LonA) peptidases are composed of three domains: an N-terminal substrate-binding domain, a middle ATPase domain belonging to the AAA+ superfamily and a C-terminal proteolytic domain with a serine–lysine catalytic dyad (Botos *et al.*, 2004). LonA generally remains in an inactive state until a specific substrate binds to an allosteric site found in the N-terminal domain of the enzyme (Wohlever *et al.*, 2014). This binding step briefly activates the enzyme, leading to rapid degradation of the bound protein substrate. Overall, LonA proteins are large multi-domain proteins which are conserved in bacteria, archaea and eukaryotes, including humans (Iyer *et al.*, 2004). Interestingly, a smaller polypeptide (~200 amino acids) with homology to the N-terminal substrate-binding domain of LonA is conserved in many prokaryotic phyla (Supplementary Fig. S1). These smaller

polypeptides (LonNTD-like proteins) are found in all known species of Xanthomonadales and related orders of proteobacteria. In addition to the LonNTD-like protein, organisms from these phyla also encode a full-length LonA in their genome. Despite significant conservation of the LonNTD-like proteins, no function has yet been assigned to these proteins. Nevertheless, the structure of a putative LonNTD-like protein from *Bordetella parapertussis* (BPP1347) has been deposited in the Protein Data Bank by the Northeast Structural Genomics Consortium (PDB entry 1zbo), albeit unaccompanied by a structural or functional analysis.

In this study, we report the crystal structure of a recombinant LonNTD-like protein from *Xanthomonas campestris* (XCC3289; UniProt Q8P5P7) at 2.8 Å resolution. Despite poor sequence identity (less than 18%), the structure of XCC3289 matches very well with the structures of the N-terminal substrate-binding domains of the LonA peptidases from *Escherichia coli* (Li *et al.*, 2010) and *Mycobacterium avium* (Chen *et al.*, 2019). This structural module is also conserved in another protein named cereblon (Kim *et al.*, 2016), which is the substrate-binding module of the E3 ubiquitin ligase complex. Our structural analyses suggest that XCC3289 and other LonNTD-like proteins might be capable of specific interactions with a protein or a group of proteins.

2. Materials and methods

2.1. Cloning, expression and protein purification

The coding DNA corresponding to the LonNTD-like protein (XCC3289; UniProt Q8P5P7) was amplified by the polymerase chain reaction from genomic DNA of *X. campestris* pv. *campestris*. The amplified DNA was cloned into pST50Tr (Tan *et al.*, 2005), a T7-promoter-based expression plasmid, to form an in-frame translational fusion at the N-terminus with a streptavidin-His₆-TEV tag (3.8 kDa). The tagged XCC3289 clone was expressed in *E. coli* BL21(DE3) pLysS cells. The cells were grown at 37°C until the OD_{600 nm} reached 0.6. The cell culture was shifted to 28°C and was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 14 h. The cells were harvested by centrifugation at 8300g. The cell pellet was suspended in P300 buffer (50 mM sodium phosphate pH 8.0, 300 mM NaCl) and frozen using liquid nitrogen. Subsequently, the cell pellet was thawed at room temperature and lysed by ultrasonication. The supernatant of the cell lysate was loaded onto an immobilized metal-affinity chromatography (IMAC) column. The column was washed with P300 buffer containing 20 mM imidazole. The tagged protein was eluted with a gradient of 20–500 mM imidazole in P300 buffer. The streptavidin-His₆-TEV tag of the protein was removed by digestion with TEV peptidase and IMAC chromatography. The protein was concentrated to about 20 mg ml⁻¹ using a 10 kDa cutoff ultrafiltration membrane, supplemented with 20% (v/v) glycerol, flash-frozen in liquid nitrogen and stored at -80°C until further use. To set up the crystallization trials, the protein was further purified by size-exclusion chromatography on a Superdex 200 10/300 GL

Table 1

Data-collection, refinement and model statistics.

Values in parentheses are for the highest resolution shell.

Diffraction source	PX-BL21, Indus-2, India
Wavelength (Å)	0.979470
Space group	I422
<i>a</i> , <i>b</i> , <i>c</i> (Å)	124.92, 124.92, 55.59
α , β , γ (°)	90, 90, 90
Resolution (Å)	44.17–2.80 (2.95–2.80)
No. of protein chains in asymmetric unit	1
No. of observed reflections	46030 (6711)
No. of unique reflections	5595 (795)
Multiplicity	8.2 (8.4)
Completeness (%)	99.0 (99.6)
$\langle I \rangle / \langle \sigma(I) \rangle$	19.9 (2.3)
$R_{\text{merge}}^{\dagger}$	0.076 (0.905)
R_{meas}	0.081 (0.963)
$R_{\text{work}}/R_{\text{free}}^{\ddagger}$	0.262/0.288
CC _{1/2}	0.999 (0.859)
Wilson <i>B</i> factor (Å ²)	73.9
No. of atoms in asymmetric unit	1262
No. of water atoms	0
No. of ligand atoms	0
Bond-length r.m.s.d. (Å)	0.002
Bond-angle r.m.s.d. (°)	0.569
Average <i>B</i> factor (Å ²)	71
Ramachandran allowed (%)	4.91
Ramachandran favoured (%)	95.09
Ramachandran outliers (%)	0.0
PDB code	7cay

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ represent the diffraction-intensity values of the individual measurement and the corresponding mean value, respectively. $\ddagger R_{\text{free}}$ was calculated using 5% of the data, which were excluded from refinement.

column using T200 buffer (20 mM Tris-HCl pH 8.0, 200 mM NaCl). The typical yield of the purified protein was found to be about 10 mg per litre of cell culture. The protein was greater than 95% pure as judged by SDS-PAGE. The purified protein was subsequently used to set up crystallization trials at a concentration of 15 mg ml⁻¹.

2.2. Crystallization

Crystallization was performed by the microbatch-under-oil method at 4°C (Chayen *et al.*, 1992) in 96-well U-bottom plates. 1 µl protein solution was mixed with 1 µl crystallization solution and placed under 50 µl Al's oil [silicone fluid:mineral oil, 1:1(v:v)]. The buffer composition of the protein solution was 20 mM Tris-HCl pH 8.0, 200 mM NaCl. For crystallization trials, Crystal Screen, Crystal Screen 2 and Index from Hampton Research were used. Crystals with good morphology were observed in Crystal Screen condition B7: 100 mM Tris-HCl pH 8.5, 200 mM ammonium acetate, 30% (v/v) 2-propanol. Diffraction data were obtained from a crystal which was grown in the same crystallization condition with 35% (v/v) 2-propanol. The crystals were typically observed after 3–7 days of incubation at 4°C. The crystals were removed from the microbatch drop using a cryoloop and were immersed in Paratone N (Hampton Research). The crystals were coated with Paratone N by removing the aqueous phase from their surfaces and were flash-cooled in liquid nitrogen.

2.3. Data collection, processing, structure determination and refinement

Single-crystal diffraction studies were carried out on the protein crystallography beamline (PX-BL21) at the Indus-2 synchrotron, India (Kumar *et al.*, 2016). Diffraction images were collected from cryocooled crystals (100 K) with an oscillation of 1° at a wavelength of 0.9794 Å using a MAR225 CCD (Rayonix) detector. The data were processed and scaled using *XDS* (Kabsch, 2010) and *AIMLESS* (Winn *et al.*, 2011). The crystal diffracted poorly at high resolution, as can be seen from the data-processing statistics (R_{meas} of 0.963 in the highest resolution shell; Table 1). Thus, we chose to keep the resolution at 2.8 Å for data with an $I/\sigma(I)$ of 2.3 in the highest resolution shell. The initial phases for the structure were obtained by molecular replacement (MR) in *Phaser* (McCoy *et al.*, 2007). We first used the monomer of BPP1347 (PDB entry 1zbo), a structural homolog of XCC3289, for MR solution. The result of MR had a TFZ value of 4.5 and an LLG score of 13.25. We subsequently used the fold-recognition server *Phyre* (Kelley *et al.*, 2015) to generate a model of XCC3289. The *Phyre* model was used as a template for MR and gave an MR solution with a TFZ value of 7.5 and an LLG score of 36.39. Early models were automatically built using *Phenix AutoBuild* (Terwilliger *et al.*, 2008). Further refinement of the crude model was carried out by iterative cycles of manual model building in *Coot* (Emsley *et al.*, 2010) and automated refinement in *phenix.refine* (Afonine *et al.*, 2012). Details of the data processing and refinement are summarized in Table 1.

2.4. Molecular-dynamics simulation

Two loop regions (54–62 and 136–150) of XCC3289 could not be built during structure refinement owing to poor electron density. Homology modelling was used to build the missing regions of XCC3289 using the structure of the LonNTD-like protein from *B. parapertussis* (PDB entry 1zbo) as a template in *MODELLER* 9.23 (Webb & Sali, 2016). The software predicted a number of structures, which were sorted according to their discrete optimized protein energy (DOPE) scores. The model with the highest DOPE score was selected. A molecular-dynamics simulation was carried out on the monomeric model of XCC3289 built by *MODELLER*. The simulation was performed using *GROMACS*-2019.4 (Abraham *et al.*, 2015) and the AMBER99SB-ILDN force field (Lindorff-Larsen *et al.*, 2010). The model was placed in a rhombic dodecahedral unit cell with a minimum solute-box distance of 1 nm. The protein was solvated using the TIP3P water model and the overall charge was neutralized by adding sodium ions at random positions within the simulation box. The solvation was followed by steepest-descent energy minimization using the Verlet cutoff scheme, temperature equilibration for 100 ps using the modified Berendsen thermostat at 300 K and pressure equilibration for another 100 ps using the Parrinello–Rahman algorithm. A final MD production run of 100 ns using a step precision of 2 fs was carried out for all of the calculations.

2.5. Bioinformatics analysis

LonNTD-like protein homologs were identified in various prokaryotic phyla by *BLASTp* (>23% sequence identity and >85% query coverage). Conservation analyses of various sequence motifs among the homologs of XCC3289 were carried out based on multiple sequence alignments generated by *Clustal Omega*. Evolutionary conservation of amino-acid residues was carried out using the *ConSurf* server (Ashkenazy *et al.*, 2016) with 300 aligned sequences from the UniRef90 database (sequence-identity cutoff: maximum 70% and minimum 25%) and the results were visualized on the loop-built model of XCC3289. The *PDBePISA* analysis tool (Krissinel & Henrick, 2007) was used to predict the quaternary structures of XCC3289 and BPP1347. Comparison of the XCC3289 structure with deposited entries in PDB was performed by the *DALI* server (Holm, 2020). The *DSSP* server (Kabsch & Sander, 1983) was used to annotate the secondary structure and *PyMOL* was used for structural analysis and illustrations.

3. Results and discussion

LonNTD-like proteins are comprised of polypeptides containing ~200 residues and are homologous to the N-terminal domain of LonA (Supplementary Fig. S1). The biological function of LonNTD-like proteins is not known. A *BLASTp* search against the nonredundant NCBI protein database shows that these proteins are widely conserved in prokaryotes but are entirely missing in eukaryotes (Supplementary Fig. S2). In archaea, LonNTD-like proteins are only conserved in the phylum Euryarchaeota. In eubacteria, they are present in most of the known phyla of Proteobacteria except for the Chlamydiae and Epsilonproteobacteria. These proteins are widely conserved in known organisms of the Xanthomonadales order, particularly in the *Xanthomonas* genus, where they are almost entirely conserved in all known species. Organisms that encode LonNTD-like proteins in their genome also encode full-length LonA peptidases, suggesting that LonNTD-like proteins play a functional role independent of that of LonA.

3.1. Overall structure of XCC3289

The crystal structure of XCC3289 showed the presence of a single polypeptide in the asymmetric unit. The structure was solved by molecular replacement and was refined to a resolution of 2.8 Å, with a final R_{free} of 29% (Supplementary Fig. S3). Five residues at the N-terminus (¹GSPAE⁴), seven residues in region 1 (⁵⁵SDVGAPA⁶¹) and 13 residues in region 2 (¹³⁷EQVGGEFASAGPG¹⁴⁹) could not be modelled into the final crystal structure owing to poor electron density. The crystal structure of XCC3289 clearly shows the presence of two distinct domains joined by a long loop (residues 102–125) (Fig. 1). The N-terminal domain comprises two β -sheets and an α -helix. The first two β -strands, β_1 (residues 6–15) and β_2 (residues 20–21), are followed by the only α -helix present in this domain, α_1 (residues 30–42). The α_1 helix is positioned

inside a U-shaped cavity formed by twisted β -sheets ($\beta 3$; residues 47–51, $\beta 4$; residues 63–74, $\beta 5$; residues 83–100, $\beta 6$; residues 106–112). The C-terminal domain is all-helical and consists of four helices: $\alpha 2$ (residues 126–134), $\alpha 3$ (residues 154–164), $\alpha 4$ (residues 169–177) and $\alpha 5$ (residues 181–190). The two domains of XCC3289 interact with each other and form a large groove at their interface. These interactions exclude about 1450 Å² of solvent-accessible surface area (Krissinel & Henrick, 2007).

A 100 ns molecular-dynamics simulation was carried out on the loop-built model of XCC3289. The root-mean-square deviation (r.m.s.d.) was calculated relative to the minimized

equilibrated structure using all backbone atoms of the XCC3289 protein, and shows that the overall structure remains stable over time (Fig. 2*a*) and exhibits no significant inter-domain motion. The root-mean-square fluctuation (r.m.s.f.) was calculated from the MD simulation data using all backbone atoms of the protein. Region *a* (residues 56–59), region *b* (residues 78–81) and region *c* (residues 139–148) were identified as three highly mobile regions with molecular motion of greater than 2 Å (Fig. 2*b*). Region *a* is part of a large loop connecting strands $\beta 3$ and $\beta 4$. Region *c* is part of a helix ($\alpha 2$) and a loop preceding $\alpha 2$. Both of these regions (residues 55–61 and residues 137–149) were not modelled into the final

Figure 1

The XCC3289 protein crystal structure in ribbon representation. The regions that are missing from the crystal structure are represented by dotted lines.

Figure 2

Molecular-dynamics simulation on the loop-built model of XCC3289. (a) R.m.s.d. plot showing the stability of the protein structure. (b) R.m.s.f. plot showing the flexible regions. R.m.s.f. values greater than 2 Å are highlighted.

Figure 3

ConSurf analysis showing the evolutionary conservation of amino-acid residues with respect to the structure of loop-built XCC3289 in cartoon (*a*) and surface (*b*) representations. A colour-coded conservation scale is shown at the bottom.

structure owing to poor electron density, further supporting the high flexibility of these loops. However, we could model region *b* into the final crystal structure.

We performed a *ConSurf* analysis (Ashkenazy *et al.*, 2016) of the XCC3289 structure to estimate the evolutionary conservation of amino-acid positions based on the phylogenetic relations between homologous sequences. These studies show that the surface-exposed residues in a region that corresponds to a long shallow groove running across both the N- and C-terminal domains are conserved (Fig. 3*a*). Additionally, the residues at the interface of the secondary structures that make the core of the protein are conserved (Fig. 3). The solvent-exposed residues Phe28, Glu29, Asp56 and Pro166 are strictly conserved. Most of the other surface-exposed residues are less conserved.

3.2. Structural conservation

The structure of XCC3289 exhibits a significant similarity to the structure of a hypothetical protein from *B. parapertussis* (BPP1347; PDB entry 1zbo; Fig. 4*a*), which was solved at resolution of 2.6 Å with two identical polypeptide chains in the asymmetric unit. The sequence of XCC3289 shares about 34% sequence identity with that of BPP1347. The relative structural topology and domain orientations are similar in both of the proteins (Fig. 4*a*). The structure of BPP1347 also consists of two domains connected by a long loop. The N-terminal domain of BPP1347 is comprised mainly of β -sheets, with only

one α -helix, and the C-terminal domain is composed of only α -helices. The two monomers of BPP1347 in a single asymmetric unit do not show any significant structural variation and superpose with an r.m.s.d. of 0.88 Å on 195 C α pairs. Superposition of XCC3289 on a single polypeptide of BPP1347 (chain *A* of PDB entry 1zbo) by the *DALI* server (Holm, 2020) resulted in an r.m.s.d. of 1.9 Å (Z-score 20.8) for 167 C α pairs. The N-terminal domains of the proteins share greater structural similarity, superimposing with an r.m.s.d. of 1.2 Å (99 C α pairs; Z-score 16.9), in comparison to the C-terminal domains, which have an r.m.s.d. of 1.7 Å (97 C α pairs; Z-score 14.7). Although a *PISA* analysis suggested that dimers might form, the gel-filtration results suggest that the XCC3289 protein is monomeric in solution (Supplementary Figs. S4, S5*a* and S5*b*). *PISA* analysis of the crystal structure of BPP1347 suggests the quaternary structure of the protein to be tetrameric; more precisely, a dimer of dimers (Supplementary Fig. S5*c*).

Structural conservation studies of XCC3289 using the *DALI* server revealed some noteworthy results. XCC3289 exhibits significant structural similarity to the N-terminal domains of LonA from *E. coli* (Fig. 4*b*; PDB entry 3ljc; Li *et al.*, 2010) and *M. avium* (Fig. 4*c*; PDB entry 6ihg; Chen *et al.*, 2019). Superposition of XCC3289 with the N-terminal domain of LonA from *E. coli* shows an r.m.s.d. of 2.3 Å (Z-score 19.2) for 165 C α pairs. Similarly, superposition with the N-terminal domain of LonA from *M. avium* shows an r.m.s.d. of 3.2 Å (Z-score 16.1) for 171 C α pairs. In previous studies, it has been

Figure 4

Structural superposition (ribbon representation) of XCC3289 with the LonNTD-like protein from *B. paraptensis* (PDB entry 1zbo) (a), the N-terminal domain of LonA from *E. coli* (PDB entry 3ljc) (b) and the N-terminal domain of LonA from *M. avium* (PDB entry 6ihg) (c). The residues responsible for Sul20 degron binding are shown in red.

suggested that the N-terminal region of LonA from *E. coli* is involved in oligomerization and substrate recognition/binding (Brown *et al.*, 2019; Wohlever *et al.*, 2013; Roudiak & Shrader, 1998). The N-terminal domain of LonA has also been implicated in allosteric regulation of ATPase activity and substrate hydrolysis (Wohlever *et al.*, 2014). LonA peptidases recognize and degrade some of their protein substrates by binding through degrons: stretches of hydrophobic residues that are uncovered during unfolding or misfolding (Gur & Sauer, 2008, 2009). Sul20 is a well studied degron in *E. coli* which mediates the degradation of the regulatory protein Sula (Higashitani *et*

al., 1997; Ishii & Amano, 2001). It has been shown that the Sul20 degron is specifically recognized by key residues located in the N-terminal subdomain of the N-terminal domain of LonA (Fig. 4b; Wohlever *et al.*, 2014). This subdomain shows significant structural similarity to the N-terminal domain of XCC3289 and the structures superimpose with an r.m.s.d. of 1.7 Å (Z-score 14.4) for 101 C α pairs. The C-terminal α -helical subdomain of the LonA N-terminal domain is also similar to that of XCC3289. This α -helical subdomain has previously been shown to play a crucial role in the formation of a functionally active structure of LonA (Kudzhayev *et al.*, 2017, 2018).

A LonNTD-like structural module is also conserved in the protein cereblon (PDB entry 6boy, chain B; Nowak *et al.*, 2018). Cereblon (UniProt Q96SW2) is one of the polypeptide components of the DCX (DDB1–CUL4–X-box) E3 ubiquitin ligase complex, where it performs the function of substrate recognition and binding (Matyskiela *et al.*, 2016; Nowak *et al.*, 2018). Cereblon consists of three domains: an N-terminal domain, a helical bundle domain and a C-terminal domain (Fig. 5). The substrate-binding site of cereblon is primarily constituted of its N-terminal domain, which is structurally similar to the N-terminal domain of XCC3289 (Fig. 5). They superimpose with each other with an r.m.s.d. of 2.8 Å (Z-score 12.7) for 112 C α pairs. Overall, the N-terminal and helical bundle domains of cereblon show structural similarity to the two-domain structure of XCC3289. The proteins share a sequence similarity of only 21% and their structures superimpose with an r.m.s.d. of 2.6 Å (Z-score 17.2) for 174 C α pairs. Overall, despite poor sequence identity (<21%), the structure of XCC3289 matches significantly with the substrate-binding domains of the LonA peptidases and cereblon. In both cases, the conserved structural module corresponding to XCC3289 carries out the function of substrate recognition and binding. In light of these observations, we speculate that conserved LonNTD-like proteins are capable of specific interactions with

Figure 5

Structural comparison of XCC3289 with cereblon (CRBN; PDB entry 6boy, chain B) by superimposition. NTD, N-terminal domain; HCB, helical bundle domain; CTD, C-terminal domain.

other proteins and that such interactions might be associated with their function, although further studies are needed to confirm this.

4. Conclusion

In this study, we cloned, expressed and purified a LonNTD-like protein from *X. campestris* (XCC3289). We crystallized the protein and determined its crystal structure at 2.8 Å resolution. The structure consists of two domains joined by a long loop. The N-terminal domain consists mostly of β -strands, whereas the C-terminal domain is almost entirely α -helical. XCC3289 shares a similar structural architecture with a LonNTD-like protein from *B. paraptentus*. Despite poor sequence identity, the structure of XCC3289 significantly matches the substrate-binding domains of LonA peptidases. The structural topology of XCC3289 is also conserved in cereblon from the E3 ubiquitin ligase complex. As the N-terminal domains of both LonA and cereblon are known to be involved in specific protein–protein interactions, we speculate that LonNTD-like proteins, including XCC3289, might also be involved in protein binding.

Acknowledgements

We thank the staff at Indus-2, RRCAT, India for providing the synchrotron beam and we acknowledge the use of the PX-BL21 beamline. We thank Dr Pooja Yadav and Bhaskar Sharma for initial work on this project. The authors declare that they have no conflicts of interest with the contents of this article.

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. & Lindahl, E. (2015). *SoftwareX*, **1–2**, 19–25.
- 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.* **D68**, 352–367.
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. & Ben-Tal, N. (2016). *Nucleic Acids Res.* **44**, W344–W350.
- Botos, I., Melnikov, E. E., Cherry, S., Tropea, J. E., Khalatova, A. G., Rasulova, F., Dauter, Z., Maurizi, M. R., Rotanova, T. V., Wlodawer, A. & Gustchina, A. (2004). *J. Biol. Chem.* **279**, 8140–8148.
- Brown, B. L., Vieux, E. F., Kalastavadi, T., Kim, S., Chen, J. Z. & Baker, T. A. (2019). *Protein Sci.* **28**, 1239–1251.
- Chayen, N. E., Shaw Stewart, P. D. & Blow, D. M. (1992). *J. Cryst. Growth*, **122**, 176–180.
- Chen, X., Zhang, S., Bi, F., Guo, C., Feng, L., Wang, H., Yao, H. & Lin, H. (2019). *Protein Sci.* **28**, 1720–1726.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Gur, E. & Sauer, R. T. (2008). *Genes Dev.* **22**, 2267–2277.
- Gur, E. & Sauer, R. T. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 18503–18508.
- Higashitani, A., Ishii, Y., Kato, Y. & Horiuchi, K. (1997). *Mol. Gen. Genet.* **254**, 351–357.
- Holm, L. (2020). *Protein Sci.* **29**, 128–140.
- Ishii, Y. & Amano, F. (2001). *Biochem. J.* **358**, 473–480.
- Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. (2004). *J. Struct. Biol.* **146**, 11–31.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. (2015). *Nat. Protoc.* **10**, 845–858.
- Kim, H. K., Ko, T. H., Nyamaa, B., Lee, S. R., Kim, N., Ko, K. S., Rhee, B. D., Park, C.-S., Nilius, B. & Han, J. (2016). *Pflugers Arch.* **468**, 1299–1309.
- Krisinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Kudzhayev, A. M., Andrianova, A. G., Dubovtseva, E. S., Serova, O. V. & Rotanova, T. V. (2017). *Acta Naturae*, **9**, 75–81.
- Kudzhayev, A. M., Dubovtseva, E. S., Serova, O. V., Andrianova, A. G. & Rotanova, T. V. (2018). *Russ. J. Bioorg. Chem.* **44**, 518–527.
- Kumar, A., Ghosh, B., Poswal, H. K., Pandey, K. K., Jagannath, Hosur, M. V., Dwivedi, A., Makde, R. D. & Sharma, S. M. (2016). *J. Synchrotron Rad.* **23**, 629–634.
- Lee, I. & Suzuki, C. K. (2008). *Biochim. Biophys. Acta*, **1784**, 727–735.
- Li, M., Gustchina, A., Rasulova, F. S., Melnikov, E. E., Maurizi, M. R., Rotanova, T. V., Dauter, Z. & Wlodawer, A. (2010). *Acta Cryst.* **D66**, 865–873.
- Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O. & Shaw, D. E. (2010). *Proteins*, **78**, 1950–1958.
- Matyskiela, M. E., Lu, G., Ito, T., Pagarigan, B., Lu, C., Miller, K., Fang, W., Wang, N., Nguyen, D., Houston, J., Carmel, G., Tran, T., Riley, M., Nosaka, L. A., Lander, G. C., Gaidarova, S., Xu, S., Ruchelman, A. L., Handa, H., Carmichael, J., Daniel, T. O., Cathers, B. E., Lopez-Girona, A. & Chamberlain, P. P. (2016). *Nature*, **535**, 252–257.
- 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.
- Nowak, R. P., DeAngelo, S. L., Buckley, D., He, Z., Donovan, K. A., An, J., Safaei, N., Jedrychowski, M. P., Ponthier, C. M., Ishoe, M., Zhang, T., Mancias, J. D., Gray, N. S., Bradner, J. E. & Fischer, E. S. (2018). *Nat. Chem. Biol.* **14**, 706–714.
- Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. (2014). *Nucleic Acids Res.* **42**, D503–D509.
- Roudiak, S. G. & Shrader, T. E. (1998). *Biochemistry*, **37**, 11255–11263.
- Tan, S., Kern, R. C. & Selleck, W. (2005). *Protein Expr. Purif.* **40**, 385–395.
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* **D64**, 61–69.
- Webb, B. & Sali, A. (2016). *Curr. Protoc. Bioinform.* **54**, 1–5.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krisinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Pottert, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* **D67**, 235–242.
- Wohlever, M. L., Baker, T. A. & Sauer, R. T. (2013). *J. Bacteriol.* **195**, 5622–5628.
- Wohlever, M. L., Baker, T. A. & Sauer, R. T. (2014). *Mol. Microbiol.* **91**, 66–78.