An unexpected co-crystal structure of the calpain PEF(S) domain with Hfq reveals a potential chaperone function of Hfq

Joel Cresser-Brown,[a] Pierre Rizkallah,[a] Yi Jin,[a] Christian Roth,[b] David J. Miller[a] and Rudolf K. Allemann[a]*

[a]School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, UK, and [b]Carbohydrates: Structure and Function, Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Arnimallee 22, 14195 Berlin, Germany. *Correspondence e-mail: allemannrk@cardiff.ac.uk
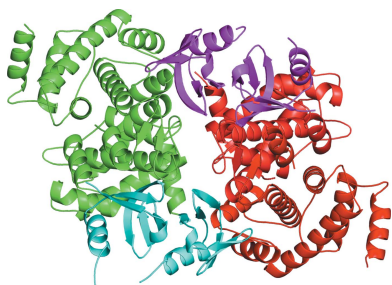
Calpain is a Ca^{2+}-activated, heterodimeric cysteine protease consisting of a large catalytic subunit and a small regulatory subunit. Dysregulation of this enzyme is involved in a range of pathological conditions such as cancer, Alzheimer's disease and rheumatoid arthritis, and thus calpain I is a drug target with potential therapeutic applications. Difficulty in the production of this enzyme has hindered structural and functional investigations in the past, although heterodimeric calpain I can be generated by *Escherichia coli* expression in low yield. Here, an unexpected structure discovered during crystallization trials of heterodimeric calpain I (CAPN1C115S + CAPNS1ΔGR) is reported. A novel co-crystal structure of the PEF(S) domain from the dissociated regulatory small subunit of calpain I and the RNA-binding chaperone Hfq, which was likely to be overproduced as a stress response to the recombinant expression conditions, was obtained, providing unexpected insight in the chaperone function of Hfq.

1. Introduction

Proteases are ubiquitous in cellular processes that respond to a range of stimuli and signalling molecules. Calpains are a family of calcium-activated cysteine proteases consisting of 15 known isoforms in the human body. They respond to intracellular increases in Ca^{2+} by the specific and regulated cleavage of a variety of targets involved in signalling pathways such as apoptosis (Wang *et al.*, 2016), chemotaxis (Hallett & Dewitt, 2007) and cell motility (Santella *et al.*, 1998; Miller *et al.*, 2013). The two most studied isoforms, calpain I and calpain II, are heterodimeric proteins consisting of a large subunit (80 kDa) and a regulatory subunit (CAPNS1; 30 kDa). The large subunit contains the active site (the CysPc domain), a calcium-binding penta-EF-hand domain [PEF(L)], a calpain-type β-sandwich domain (CBSW) and an N-terminal anchor α-helix (Ono *et al.*, 2016). The regulatory subunit CAPNS1 (30 kDa) comprises two domains: a glycine-rich domain (GR) and another calcium-binding penta-EF-hand domain [PEF(S)] (Adams *et al.*, 2015). Calpain I and calpain II are activated *in vitro* by calcium concentrations of 50 and 350 μM, respectively. The large subunits of calpain I and calpain II share 62% sequence identity, while the small subunits of the two proteins are identical (Goll *et al.*, 2003; Hitomi *et al.*, 1998).

Calpain I and calpain II have been targeted to treat various conditions including rheumatoid arthritis, ischemic cell death and cancer (Miller *et al.*, 2013; Bartus *et al.*, 1995; Yamashima, 2004; Luo *et al.*, 2015). The precise involvement of the individual enzyme isoforms in these processes remains unclear.

Table 1
Macromolecule-production information.

The expression host and vectors used for PEF(S)–Hfq protein production.

Source organism	*Homo sapiens*
DNA source	CAPN1C115S (His$_6$ tag)/CAPNS1ΔGR
Expression vector	pET24(+)/pACpET24
Expression host	*E. coli* C41 (DE3) pLysS

Table 2
Crystallization.

Conditions used for PEF(S)–Hfq crystallization trials.

Method	Vapour diffusion
Plate type	Sitting drop
Temperature (K)	293
Protein concentration (mg ml^{-1})	10
Buffer composition of protein solution	20 mM HEPES, 0.5 mM EGTA, 0.5 mM TCEP pH 7.6
Composition of reservoir solution	0.1 M MMT, 25% PEG 1500 pH 9.0
Volume and ratio of drop	200 nl, 1:1
Volume of reservoir (μl)	50

Table 3
Data collection and processing.

Data-collection and processing statistics for the PEF(S)–Hfq data set are presented. Values in parentheses are for the outer shell.

Diffraction source	I04, Diamond Light Source
Wavelength (Å)	0.9795
Temperature (K)	100
Detector	PILATUS3 6M, Dectris
Space group	$P2_13$
a, b, c (Å)	147.607, 147.607, 147.607
α, β, γ (°)	90.0, 90.0, 90.0
Resolution range (Å)	2.32–66.02
Total No. of reflections	1029720 (75756)
No. of unique reflections	46565 (3416)
Completeness (%)	100.0 (100.0)
$\langle I/\sigma(I)\rangle$	19.2 (3.0)
$R_{\text{r.i.m.}}$ (%)	11.6 (114.1)
Overall B factor from Wilson plot (Å^2)	42.6

Structural insights from crystallographic data on the heterodimeric calpain I complex could accelerate the development of therapeutics. This has been hindered by a lack of available material owing to low-yielding recombinant expression as well as challenging protein properties such as subunit dissociation, aggregation and autolysis (Hata *et al.*, 2012; Pal *et al.*, 2001). To tackle such difficult-to-express proteins, *Escherichia coli* strain C41 was developed, which is particularly well suited to the expression of toxic and membrane proteins.

Here, we present our results on the expression, purification and structure determination of human calpain I, which led to an unusual structure of the PEF(S) domain from the regulatory subunit of calpain bound to the RNA chaperone Hfq.

2. Materials and methods

2.1. Protein purification

E. coli C41 (DE3) pLysS cells containing the CAPN1C115S (80 kDa) and CAPNS1ΔGR (20 kDa; CAPNS1 with a truncated GR domain; Hata *et al.*, 2013) genes were grown at 37°C in kanamycin- and ampicillin-selective TB(Enhanced) medium until the OD$_{600}$ reached 0.9. Subsequently, protein production was induced with 1 mM IPTG. The protein was expressed overnight at 20°C and the cells were harvested by centrifugation in a Sorvall RC6 Plus centrifuge (Thermo Fisher Scientific, Massachusetts, USA) using an SLA-3000 rotor at 6080g for 20 min at 4°C. The cells were resuspended in 20 mM HEPES, 0.5 mM EGTA, 0.5 mM TCEP pH 7.6 (buffer A) and lysed by sonication for 5 min (pulsed; 5 s on, 10 s off). The lysate was clarified by centrifugation at 4°C for 40 min at 30310g in a Sorvall RC6 Plus centrifuge. The supernatant was passed through a 0.2 μm syringe filter and applied onto an Ni–NTA column. The unbound protein was washed out with 15 column volumes (CV) of buffer A and the target protein was eluted with 10 CV buffer A supplemented with 250 mM imidazole. The eluted fraction containing the target protein was then applied onto a Mono Q HR 10/10 column (GE Healthcare) pre-equilibrated with buffer A. The protein was eluted with a linear gradient of 0–0.5 M NaCl in buffer A over 7 CV. The purity of the peak fractions was assessed by SDS–PAGE (10%). The solution containing the protein complex was concentrated to 10 mg ml^{-1} using a 30 kDa molecular-weight cutoff Vivaspin concentrator. Macromolecule-production information is summarized in Table 1.

2.2. Crystallization

Crystals of PEF(S)–Hfq were grown at 20°C in sitting-drop vapour-diffusion plates containing equal volumes of precipitant [0.1 M MMT (malic acid, MES and Tris in a 1:2:2 molar ratio), 25%(v/v) PEG 1500 pH 9.0] and the CAPN1C115S–CAPNS1ΔGR protein solution. Crystals were observed after 2–3 weeks. The crystals were cryoprotected with 10% ethylene glycol before being flash-cooled in liquid nitrogen and stored until data collection. Crystallization information is summarized in Table 2.
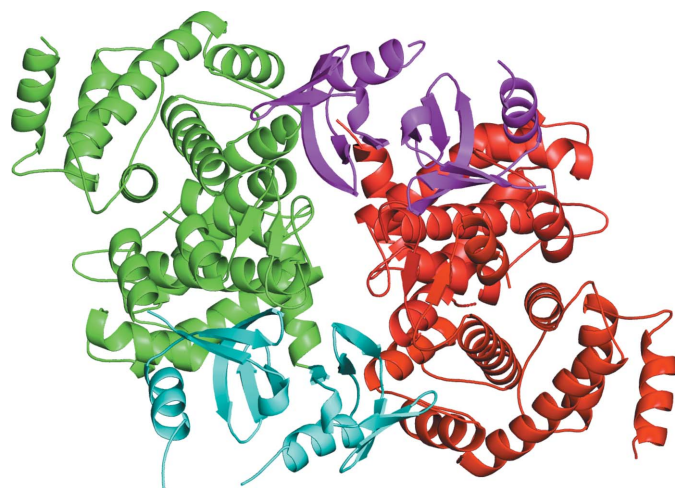
2.3. Data collection and refinement

The resulting crystals were cubic, belonging to space group $P2_13$ (unit-cell parameters $a = b = c = 147.61$ Å), and diffracted to a higher resolution limit ranging from 2.9 to 2.3 Å. The structure was determined by molecular replacement using *Phaser* (McCoy *et al.*, 2007) and *MOLREP* (Vagin & Teplyakov, 2010) as implemented within the *CCP4i2* framework (Potterton *et al.*, 2018). PDB entry 1df0 (Hosfield *et al.*, 1999) was used as a starting model. Only domain VI, PEF(S), could be placed by MR. All attempts to place the second subunit led to stalled or increased R values during molecular replacement and subsequent refinement. The partial model was refined using alternating rounds of reciprocal refinement with *REFMAC*5 (Murshudov *et al.*, 1997, 2011) and real-space refinement in *Coot* (Emsley *et al.*, 2010). A polyalanine model from short stretches of ideal β-strand segments was built in the remaining unexplained electron density, subsequently refined and rebuilt to achieve a near-complete model. A combination of partial sequence information based on interpreted side-chain density, as well as fold comparison using the *DALI* server (Holm & Sander, 1995), pointed to Hfq as the most

likely second domain in the obtained crystal structure. This was confirmed by the successful model building and refinement of the partial model with *Buccaneer* (Cowtan, 2006) and *REFMAC*5 using the correct sequence as input. Further rounds of graphical adjustment of the model using *Coot*, and refinement with *REFMAC*5, led to the final model. Other data analysis was completed with programs in the *CCP*4 package. The electron-density map revealed that the asymmetric unit contained four copies of the PEF(S) domain and four copies of the Hfq chaperone protein, which were noncovalently bound, and the structure was refined to R_{work} and R_{free} values of 0.201 and 0.247, respectively. Data-collection and processing statistics are summarized in Table 3 and refinement statistics are summarized in Table 4.

3. Results and discussion

Crystallization trials using the PACT-*premier* and JCSG-*plus* screens (Molecular Dimensions) produced several hits. The data set presented here corresponds to a $P2_13$ unit cell that is large enough to accommodate the heterodimeric calpain I structure when compared with the previously published calpain II structure (Strobl *et al.*, 2000). After initial molecular replacement and refinement using various domains from the previously reported μ-like calpain (PDB entry 1qxp (Pal *et al.*, 2003) and calpain II structures (PDB entry 1df0), only the PEF(S) domain could be fitted, resulting in models with poor R_{work} and R_{free} values and suggesting an incomplete solution. Manual model building, guided only by the difference electron density, led to a partial model for the remaining unexplained density, which could be identified as the *E. coli* chaperone Hfq. The structure determined from the highest resolution data set is shown in Fig. 1. The impurity (Hfq) was likely to have been overlooked in SDS gels as a small, faint band which was poorly resolved in the 10% acrylamide gels that were used (Fig. 2). It has been observed that the calpain I heterodimer can readily undergo subunit dissociation under mild conditions (Pal *et al.*,
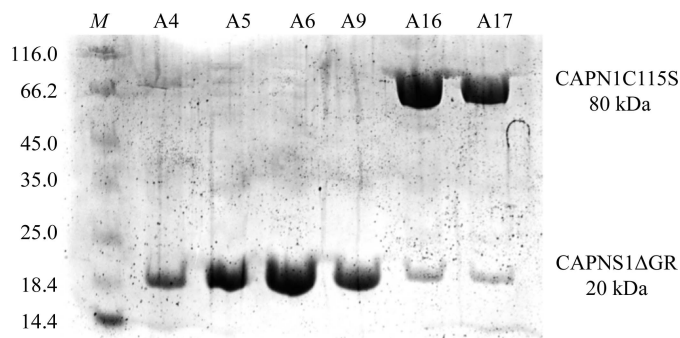
Table 4
Structure solution and refinement.

Refinement statistics for the PEF(S)–Hfq model deposited in the PDB as entry 6qlb.

Resolution range (Å)	2.32
No. of reflections, working set	43741
No. of reflections, test set	2337
Final R_{cryst}	0.201
Final R_{free}	0.247
No. of non-H atoms	7869
R.m.s. deviations	
Bond lengths (Å)	0.013
Angles (°)	1.855
Average B factor (Å^2)	51.2
Ramachandran plot	
Most favoured (%)	97.2
Allowed (%)	2.9

2001), and the stability of a PEF(S) homodimer in solution and the tendency for the large subunit to precipitate are likely to have contributed to the formation of the PEF(S)–Hfq structure. Indeed, significant precipitation was observed in many wells during the crystallization trials, including under the conditions from which crystals were harvested.
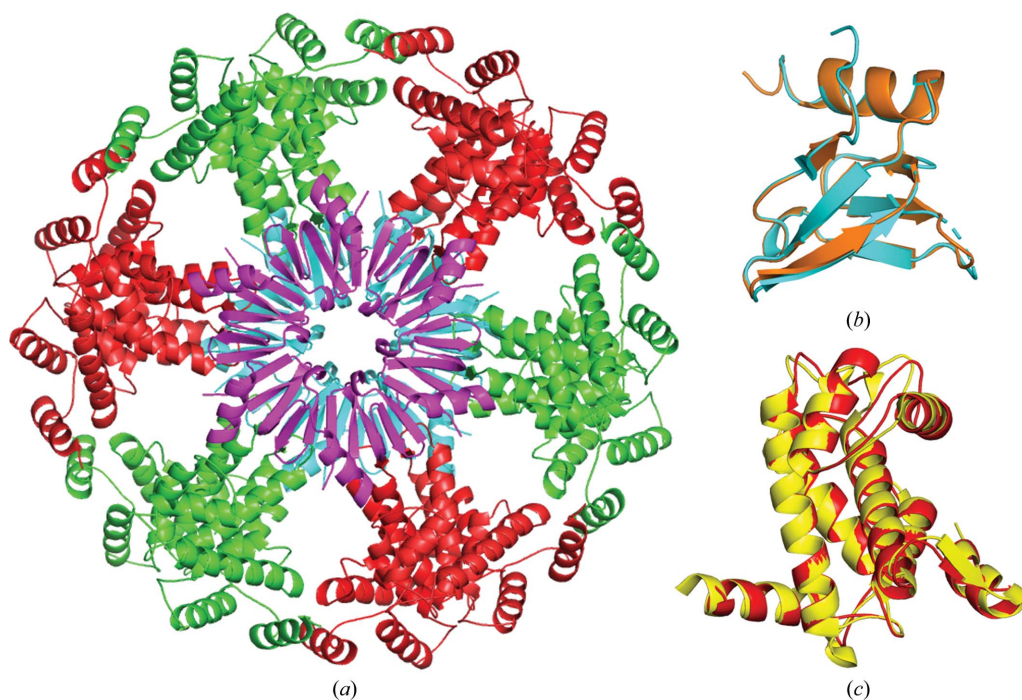
Although Hfq is a described contaminant that has been found to crystallize fortuitously instead of the target protein, a complex of parts of calpain I with Hfq has not been described previously. Hfq is part of the Sm-like family of proteins and is a toroid-shaped, highly conserved, homohexameric protein (Fortas *et al.*, 2015; Brennan & Link, 2007; Schulz & Barabas, 2014). Each 8–11 kDa subunit (depending on the host organism) comprises five coiled antiparallel β-strands and an N-terminal α-helix with an unstructured C-terminus. The expression levels of Hfq depend on the cellular growth rate and phase, with an estimate of 5000–10 000 oligomers per cell for log-phase *E. coli* in M9 medium, with 80–90% primarily situated in the cytoplasm (Kajitani *et al.*, 1994).

The Hfq–PEF(S) interactions show a 1:1 stoichiometry, and are primarily hydrophilic between the fourth EF-hand in the PEF(S) monomer and the Hfq nucleotide-binding site. Applying crystallographic symmetry, the classic homohexameric toroidal structure of Hfq can be observed and is highly similar to the Hfq conformation observed in Hfq alone (Schulz & Barabas, 2014; Fig. 3). The homodimeric PEF(S)

Figure 1
Asymmetric unit of the calpain PEF(S)–Hfq chaperone model in cartoon representation [PEF(S), green and red; Hfq, cyan and magenta].

Figure 2
SDS–PAGE (10%, reducing conditions) of Mono Q HR 10/10 purification (GE Healthcare) of CAPN1C115S + CAPNS1ΔGR, 0–0.5 *M* NaCl gradient over 7 CV (lane *M*, molecular-weight markers labelled in kDa; lanes A4–A17, peak fractions).
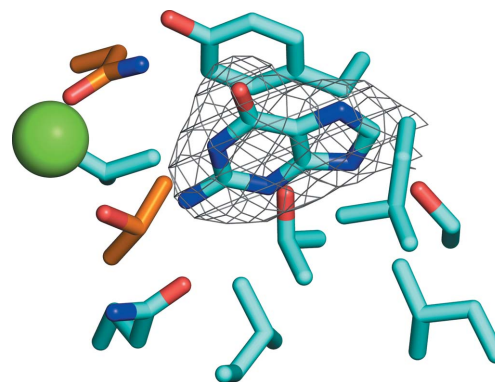
Figure 3
(a) Hfq–PEF(S) hexamer formed by crystallographic symmetry, with Hfq in cyan and magenta, and PEF(S) in green and red. (b) Hfq monomer from the Hfq–PEF(S) complex (cyan) aligned with an Hfq monomer from the native hexamer (orange; r.m.s.d. = 0.378 Å; Schulz & Barabas, 2014). (c) PEF(S) monomer from the Hfq–PEF(S) complex (red) aligned with the PEF(S) monomer from the homodimeric structure (yellow; Adams *et al.*, 2014).

surrounds the homohexamer of Hfq. The homodimeric PEF(S) structure is also highly analogous to previously observed PEF(S) complexes, suggesting little disruption of either homodimeric structure upon formation of the hetero-dimeric complex (Adams *et al.*, 2014). Hydrogen bonding from the backbone O atom of Gly153 in PEF(S) to the amino group of Gln52 in Hfq at a distance of 2.95 Å is observed in the complex. The interface areas between the PEF(S) dimer pairs (chains *A–B* and *C–D*) are 2093 and 2017 Å^2. The Hfq dimer pairs (chains *E–F* and *G–H*) share interface areas of between 668 and 717 Å^2 (Krissinel, 2015). The mixed PEF(S)–Hfq interface areas are smaller than each homodimer interface, at between 444 and 506 Å^2. The interface between PEF(S) and Hfq forms on the proximal face of the Hfq hexamer, which is reported to be the sRNA-binding site and has been shown to form a similar interface in an Hfq–catalase HPII complex (Yonekura *et al.*, 2013). Hfq has been reported to bind at least 20 proteins such as RNAse E, which are mostly involved in genetic processes, although the functional significance of these interactions remains unclear (Butland *et al.*, 2005).

Additional density for a small molecule was found in a hydrophobic pocket in the nucleotide-binding site of all Hfq monomers, making an aromatic stacking interaction with Tyr25 (Fig. 3). The pocket is lined with leucines and iso-leucines, further increasing the hydrophobicity of the pocket. The pocket is well known to accommodate nucleotides as ligands in Hfq. The ligand was modelled as guanine, giving the best fit to the electron density (Fig. 4). Additional density for the ribose or phosphate groups of a nucleoside or nucleotide could not be traced. Although guanine had not previously been described as a ligand for this pocket, it is not

unreasonable to speculate that Hfq is rather indiscriminate towards the base, given its function as an RNA chaperone.

We wondered whether Hfq might be involved as a chaperone to keep PEF(S) in solution while the large subunit PEF(L) is folded to be ready to form a complex with PEF(S). Indeed, the strength of the interaction between PEF(S) and PEF(L) ($\Delta G = -27.2$ kJ mol^{-1}) estimated using *PISA* (Krissinel, 2010) was found to be higher than that between PEF(S) and Hfq ($\Delta G = -12.6$ kJ mol^{-1}). Thus, the PEF(S)–Hfq oligomer may be an appropriate description of how PEF(S) is retained separately from the rest of the heterodimeric calpain I until the counterpart PEF(L), or domain IV, is ready to displace the chaperone and latch onto its PEF(S) and complete the folding during expression in this *E. coli* strain.

Figure 4
Guanine bound in the hydrophobic pocket between Hfq (cyan) and PEF(S) (orange). Chain residues were selected within 4 Å of the guanine ligand and Ca^{2+} ion (green sphere), and the map is contoured at 1.0σ.

While this structure was not the original goal of these crystallographic experiments, it does provide insight into the stress that the overexpression of proteins can have on the host bacteria, and the subsequent response to these processes. The primary focus of studies of Hfq has been on its RNA-binding capabilities. It has been shown to be involved in distinct metabolic pathways, including sugar transport, membrane remodelling and quorum sensing (Brennan & Link, 2007; Fortas *et al.*, 2015). Hfq protein–protein interactions have been studied to a lesser extent, not accounting for a chaperone function as a result of cellular stress. The structure presented here suggests that Hfq could play a greater role in the cellular stress response via protein-binding interactions than was previously thought (Schulz & Barabas, 2014).

References

Adams, S. E., Rizkallah, P. J., Miller, D. J., Robinson, E. J., Hallett, M. B. & Allemann, R. K. (2014). *J. Struct. Biol.* **187**, 236–241.
Adams, S. E., Robinson, E. J., Miller, D. J., Rizkallah, P. J., Hallett, M. B. & Allemann, R. K. (2015). *Chem. Sci.* **6**, 6865–6871.
Bartus, R. T., Elliott, P. J., Hayward, N. J., Dean, R. L., Harbeson, S., Straub, J. A., Li, Z. & Powers, J. C. (1995). *Neurol. Res.* **17**, 249–258.
Brennan, R. G. & Link, T. M. (2007). *Curr. Opin. Microbiol.* **10**, 125–133.
Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J. & Emili, A. (2005). *Nature*, **433**, 531–537.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Fortas, E., Piccirilli, F., Malabirade, A., Militello, V., Trépout, S., Marco, S., Taghbalout, A. & Arluison, V. (2015). *Biosci. Rep.* **35**, 1–9.
Goll, D. E., Thompson, V. F., Li, H., Wei, W. & Cong, J. (2003). *Physiol Rev.* **83**, 731–801.
Hallett, M. B. & Dewitt, S. (2007). *Trends Cell Biol.* **17**, 209–214.
Hata, S., Kitamura, F. & Sorimachi, H. (2013). *Genes Cells*, **18**, 753–763.
Hata, S., Ueno, M., Kitamura, F. & Sorimachi, H. (2012). *J. Biochem.* **151**, 417–422.
Hitomi, K., Uchiyama, Y., Ohkubo, I., Kunimatsu, M., Sasaki, M. & Maki, M. (1998). *Biochem. Biophys. Res. Commun.* **246**, 681–685.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Hosfield, C. M., Elce, J. S., Davies, P. L. & Jia, Z. (1999). *EMBO J.* **18**, 6880–6889.
Kajitani, M., Kato, A., Wada, A., Inokuchi, Y. & Ishihama, A. (1994). *J. Bacteriol.* **176**, 531–534.
Krissinel, E. (2010). *J. Comput. Chem.* **31**, 133–143.
Krissinel, E. (2015). *Nucleic Acids Res.* **43**, W314–W319.
Luo, T., Yue, R., Hu, H., Zhou, Z., Yiu, K. H., Zhang, S., Xu, L., Li, K. & Yu, Z. (2015). *Arch. Biochem. Biophys.* **586**, 1–9.
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.
Miller, D. J., Adams, S. E., Hallett, M. B. & Allemann, R. K. (2013). *Future Med. Chem.* **5**, 2057–2074.
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.
Ono, Y., Saido, T. C. & Sorimachi, H. (2016). *Nat. Rev. Drug Discov.* **15**, 854–876.
Pal, G. P., Elce, J. S. & Jia, Z. (2001). *J. Biol. Chem.* **276**, 47233–47238.
Pal, G. P., Veyra, T. D., Elce, J. S. & Jia, Z. (2003). *Structure*, **11**, 1521–1526.
Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., Uski, V., von Delft, F., Waterman, D., Wilson, K., Winn, M. & Wojdyr, M. (2018). *Acta Cryst.* D**74**, 68–84.
Santella, L., Kyozuka, K., De Riso, L. & Carafoli, E. (1998). *Cell Calcium*, **23**, 123–130.
Schulz, E. C. & Barabas, O. (2014). *Acta Cryst.* F**70**, 1492–1497.
Strobl, S., Fernandez-Catalan, C., Braun, M., Huber, R., Masumoto, H., Nakagawa, K., Irie, A., Sorimachi, H., Bourenkow, G., Bartunik, H., Suzuki, K. & Bode, W. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 588–592.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wang, C., Shi, D., Song, X., Chen, Y., Wang, L. & Zhang, X. (2016). *Neurochem. Int.* **97**, 15–25.
Yamashima, T. (2004). *Cell Calcium*, **36**, 285–293.
Yonekura, K., Watanabe, M., Kageyama, Y., Hirata, K., Yamamoto, M. & Maki-Yonekura, S. (2013). *PLoS One*, **8**, e78216.