STRUCTURAL BIOLOGY COMMUNICATIONS

ISSN 2053-230X

Expression and crystallographic studies of the D1D2 domains of C4.4A, a homologous protein to the urokinase receptor

Shanli Chen,[a]‡ Lin Lin,[b]‡ Cai Yuan,[c]* Henrik Gårdsvoll,[d,e] Mette C. Kriegbaum,[d,e] Michael Ploug[d,e] and Mingdong Huang[a]*

[a]College of Chemistry, Fuzhou Univerisity, Fuzhou, Fujian 350108, People's Republic of China, [b]Division of Hemostasis and Thrombosis, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02115, USA, [c]College of Bioscience and Biotechnology, Fuzhou Univerisity, Fuzhou, Fujian 350108, People's Republic of China, [d]The Finsen Laboratory, Rigshospitalet, DK-2200 Copenhagen, Denmark, and [e]Biotech Research and Innovation Centre, Copenhagen University, DK-2200 Copenhagen, Denmark. *Correspondence e-mail: cyuan@fzu.edu.cn, hmd_lab@fzu.edu.cn

Received 26 April 2017
Accepted 2 July 2017

Edited by R. L. Stanfield, The Scripps Research Institute, USA

‡ These authors made equal contributions.

Keywords: C4.4A; LU domains; glycosylation; prognostic cancer biomarkers.

C4.4A is a glycosylphosphatidylinositol-anchored membrane protein comprised of two LU domains (Ly6/uPAR-like domains) and an extensively O-glycosylated C-terminal Ser/Thr/Pro-rich region. C4.4A is a novel biomarker for squamous epithelial differentiation. Its expression is dysregulated under various pathological conditions and it is a robust biomarker for poor prognosis in various malignant conditions such as pulmonary adenocarcinoma. To facilitate crystallization, the two LU domains were excised from intact C4.4A by limited proteolysis, purified and crystallized by the sitting-drop vapour-diffusion method. The crystals diffracted to 2.7 Å resolution and belonged to space group $C222_1$, with unit-cell parameters a = 55.49, b = 119.63, c = 168.54 Å. The statistics indicated good quality of the data, which form a solid basis for the determination of the C4.4A structure.

1. Introduction

C4.4A was initially discovered and defined as a metastasis-associated antigen in a murine pancreatic tumour cell line (Rösel *et al.*, 1998). It was named C4.4A because it was recognized by specific monoclonal antibody C4.4 (Rösel *et al.*, 1998). Under normal homeostatic conditions C4.4A has a very strict expression pattern, which is primarily limited to stratified squamous epithelium such as the skin (the stratum granulosum), oesophagus, vagina, rectum, oral cavity, cornea and the cuboidial epithelium of human placenta cells (Kriegbaum, Jacobsen *et al.*, 2011; Hansen *et al.*, 2004). The high expression of C4.4A in skin keratinocytes is maintained during the re-epithelization of incisional wounds in mice (Hansen *et al.*, 2004) and in tissue-engineered human skin cultures (B. Jacobsen *et al.*, submitted) as well as in chronic human wounds (Rösel *et al.*, 1998; Hansen *et al.*, 2004). A comprehensive study of invasive skin lesions shows that C4.4A expression is maintained in stratum spinosum during hyperplasia, but more importantly C4.4A expression reappears in deep invasive lesions whether these are malignant or benign (Kriegbaum *et al.*, 2015). Several studies have reported that C4.4A is highly expressed in various solid cancers of, for example, the breast (Miyake *et al.*, 2015), bladder (Smith *et al.*, 2001; Kriegbaum *et al.*, 2016), kidney (Konishi *et al.*, 2010), colon (Paret *et al.*, 2007; Konishi *et al.*, 2010) and oesophagus (Ohtsuka *et al.*, 2013; Hansen *et al.*, 2008). In the case of human pulmonary cancer, C4.4A expression has been the focus of intense studies. Normal healthy lung epithelium is devoid of C4.4A expression, but during early squamous

Table 1
Macromolecule-production information.

Source organism	*Homo sapiens*
DNA source	cDNA from human C4.4A and suPAR D3
Cloning vector	pMT/V5-His
Expression vector	pMT/V5-His
Expression host	*Drosophila* S2 cells
Complete amino-acid sequence of the construct produced†	LECYSCVQKADDGCSPNKMKTVKCAPGVDVCTEA-VGAVETIHGQFSLAVRGCGSGLPGKNDRGLDL-HGLLAFIQLQQCAQDRCNAKLNLTSRALDPAG-NESAYPPNGVECYSCVGLSREACQGTSPPVVS-CYNASDHVYKGCFDGNVTLTAANVTVSLPVRG-CVQDEFCTRDGVTGPGFTLSGSCCQGSRCNSD-LRNKTYFSPRIPPLVRLPPPEPTTVASTTSVT-TSTSAPVRPTSTTKPMPAPTSQTPRQGVEHEA-SRDEEPRLTGGAAGHQDDDDK*LELENLPQNGR-QCYSCKGNSTHGCSSEETFLIDCRGPMNQCLV-ATGTHEPKNQSYMVRGCATASMCQHAHLGDAF-SMNHIDVSCCTKSGCNHPDLDVQYRSG

† The underlined sequences denote the chymotrypsin and enterokinase recognition sites, respectively, and the italicized sequence indicates the uPAR-D3 fusion tag.

metaplasia (not malignant *per se*) pronounced C4.4A expression appears, which is excellently aligned with its use as a biomarker for squamous differentiation. From a cancer patient perspective it is nonetheless more relevant that C4.4A expression is found in some of the precursor lesions of pulmonary adenocarcinomas, an epithelium subtype that is not expected to express this biomarker (Jacobsen *et al.*, 2012). Accordingly, high expression levels of C4.4A constitute a robust biomarker predicting a poor prognosis for patients with pulmonary adenocarcinoma but not squamous cell carcinoma (Hansen *et al.*, 2007; Jacobsen & Ploug, 2008; Jacobsen *et al.*, 2013). Based on these findings, there is a great interest in studying C4.4A in various pathological conditions and new experimental tools are being developed to accomplish this, such as C4.4A-deficient mouse models (Kriegbaum *et al.*, 2016).

Homology considerations predict that C4.4A is a glycolipid-anchored protein belonging to the Ly6/uPAR/α-neurotoxin (LU) family (Rösel *et al.*, 1998). The extracellular region of C4.4A includes two such LU domains (D1 and D2) followed by a C-terminal region, which is rich in serine, threonine and proline (STP-rich region; Hansen *et al.*, 2004). Only a very few genes encoding proteins with multiple LU domains are found in the human genome and they are all found in a small gene cluster on chromosome 19q13 (Kriegbaum, Persson *et al.*, 2011; Kjaergaard *et al.*, 2008). Whereas the urokinase receptor (uPAR) has three LU domains, all of which participate in the dynamic assembly of a high-affinity binding site for its primary ligand uPA, C4.4A has only two LU domains and thus resembles its close structural homologue Haldisin, which is an epithelial biomarker in the stratum granulosum (Gårdsvoll *et al.*, 2013). Circumstantial evidence suggests that C4.4A plays a role in adhesion, migration and invasion similar to that of uPAR (Smith *et al.*, 2001; Thuma *et al.*, 2013; Kriegbaum *et al.*, 2015). However, how C4.4A interacts with its putative ligands to perform its functions and how the two LU domains are arranged remain unknown. Protein structure–function relationships in uPAR (Llinas *et al.*, 2005) have played an essential role in the successful clinical translation of a small nonamer peptide antagonist (Ploug *et al.*, 2001) for non-invasive imaging of uPAR expression by positron emission tomography in cancer patients (Ploug, 2013; Persson *et al.*, 2015). To set a similar stage for research on C4.4A, we now report the purification, crystallization and preliminary X-ray study of the C4.4A D1 and D2 domains (C4.4A-D1D2).

2. Materials and methods

2.1. Macromolecule production

Recombinant C4.4A-ent-uPAR-D3 fusion protein was produced in *Drosophila* S2 cells as a secreted protein and purified according to a previously published protocol (Gårdsvoll *et al.*, 2007). The production procedure is summarized in Table 1. Briefly, recombinant proteins were expressed using Drosophila-SFM at a constant temperature of 298 K and 115 rev min^{-1} in a shaker incubator. Expression of the protein was induced by the addition of 0.5 m*M* $CuSO_4$, and the conditioned medium was harvested after one week of induction. After centrifugation and clarification by filtration, the supernatant was applied onto an anti-uPAR monoclonal antibody R2-coupled immunoaffinity chromatography column. The column was washed with at least ten column volumes of washing buffer (10 m*M* NaH_2PO_4, 140 m*M* NaCl pH 7.4) and the target protein was then eluted with 0.1 *M* acetic acid, 0.5 *M* NaCl pH 2.5 and the collected fractions were immediately neutralized with 1 *M* Tris base pH 11. To remove the highly glycosylated STP-rich region, purified C4.4A-ent-uPAR D3 fusion protein was mixed with immobilized chymotrypsin (Princeton Separations) at 37°C for 4–6 h to cleave the exposed scissile peptide bonds Tyr200-Phe201 and Phe201-Ser202 between C4.4A domain 2 and the STP-rich region (Hansen *et al.*, 2004). The digested material was loaded onto a mAb R2 immunoaffinity chromatography column, where the liberated uPAR-D3 and the uncleaved fusion protein were absorbed. The flowthrough contained the target C4.4A-D1D2 protein, which was further purified *via* size-exclusion chromatography (SEC) on a Superdex 75 column mounted on an ÄKTA FPLC system (GE Healthcare). The purified protein was analyzed on 12% polyacrylamide gels followed by Coomassie Blue staining. The target protein was concentrated to 10 mg ml^{-1} using an Amicon Ultra centrifugal concentrator and finally stored at 193 K. Protein for crystallization was analysed by SEC on a Superdex 75 column.

2.2. Crystallization

In order to obtain crystals of the target protein, the purified C4.4A protein was concentrated to 10 mg ml^{-1} and then crystallized by the sitting-drop vapour-diffusion method. The crystallization conditions were screened in 96-well crystal-growth plates with a crystallization robot (Art Robbins Instruments) using the following commercial screening kits: Index, PEGRx HT, SaltRX, Crystal Screen and others form Hampton Research. After optimization, well formed crystals of C4.4A-D1D2 were obtained using 22.5% polyethylene

Table 2
Crystallization of C4.4A-D1D2.

Method	Sitting-drop vapour diffusion
Plate type	24-well
Temperature (K)	293
Protein concentration (mg ml^{-1})	10
Buffer composition of protein solution	20 m*M* Tris pH 7.5, 100 m*M* NaCl
Composition of reservoir solution	22.5%(*w*/*v*) PEG 4000, 0.1 *M* citric acid pH 3.6
Volume and ratio of drop	1:1
Volume of reservoir (ml)	1

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	X29, NSLS
Wavelength (Å)	0.9796
Temperature (K)	100
Detector	ADSC Quantum 315 CCD
Crystal-to-detector distance (mm)	400
Rotation range per image (°)	1
Total rotation range (°)	360
Exposure time per image (s)	1
Space group	$C222_1$
a, *b*, *c* (Å)	55.49, 119.63, 168.54
α, β, γ (°)	90, 90, 90
Resolution range (Å)	43.22–2.70 (2.78–2.70)
No. of unique reflections	15804 (1549)
Completeness (%)	99.73 (99.17)
Multiplicity	13.5 (8.9)
$\langle I/\sigma(I)\rangle$	29.74 (2.68)
Overall *B* factor from Wilson plot (Å^2)	76.61
$R_{\rm merge}$ (%)	5.718 (77.21)
$R_{\rm r.i.m.}$ (%)	1.606 (26.68)
$CC_{1/2}$	1.000 (0.897)

glycol 4000, 0.1 *M* citric acid pH 3.6 as precipitant in a 24-well plate. Crystallization information is shown in Table 2.

2.3. Data collection and processing

Prior to X-ray data collection, the crystals were picked up in a cryoloop (0.1–0.2 mm; Hampton Research), immediately soaked in mother liquor containing an additional 25%(*v*/*v*) glycerol as a cryoprotectant for 10–20 s and then flash-cooled in liquid nitrogen. Diffraction experiments were conducted under a cryogenic condition (100 K) on beamline X29 at Brookhaven National Synchrotron Light Source (NSLS). A total of 360 images were collected at a crystal-to-detector distance of 400 mm with 1 s exposure for every 1° oscillation frame. Diffraction data sets were processed using the automated data-processing pipeline *xia*2 (Winter, 2010) with options that run *XDS* (Kabsch, 2010). Data-collection and processing statistics are summarized in Table 3.

3. Results and discussion

Recombinant C4.4A-ent-uPAR-D3 protein with a C-terminal uPAR D3 domain as a purification tag (Gårdsvoll *et al.*, 2007) was expressed in *Drosophila* S2 cells with a yield of 20 mg l^{-1} and purified by immunoaffinity chromatography. Crystallization of the full-length C4.4A protein, which was generated by enterokinase cleavage of the C4.4A-ent-uPAR-D3 fusion protein, failed. This is most likely to be owing to highly heterogeneous *O*-glycosylation of the STP-rich region. C4.4A-D1D2 was also generated by limited proteolysis using immobilized chymotrypsin owing to the presence of an exposed sequence between the C4.4A D2 domain and the STP-rich region (Hansen *et al.*, 2004). C4.4A-D1D2 was purified to homogeneity by immunoaffinity chromatography combined with gel filtration (Fig. 1*a*). The protein is monomeric as judged by its elution volume (11.22 ml) from a gel-filtration column (Fig. 1*b*). The purified C4.4A-D1D2 protein was concentrated to 10 mg ml^{-1} for initial crystallization screening. Plate-like crystals (Fig. 2*a*) appeared in condition C4 of the PEGRx HT screen. Data from the best C4.4A-D1D2 protein crystal (Fig. 2*b*) were obtained to 2.7 Å resolution (Fig. 3) on the X29 beamline at NSLS. The crystal lattice belonged to space group $C222_1$, with unit-cell parameters *a* = 55.49, *b* = 119.63, *c* = 168.54 Å (Table 3). The most

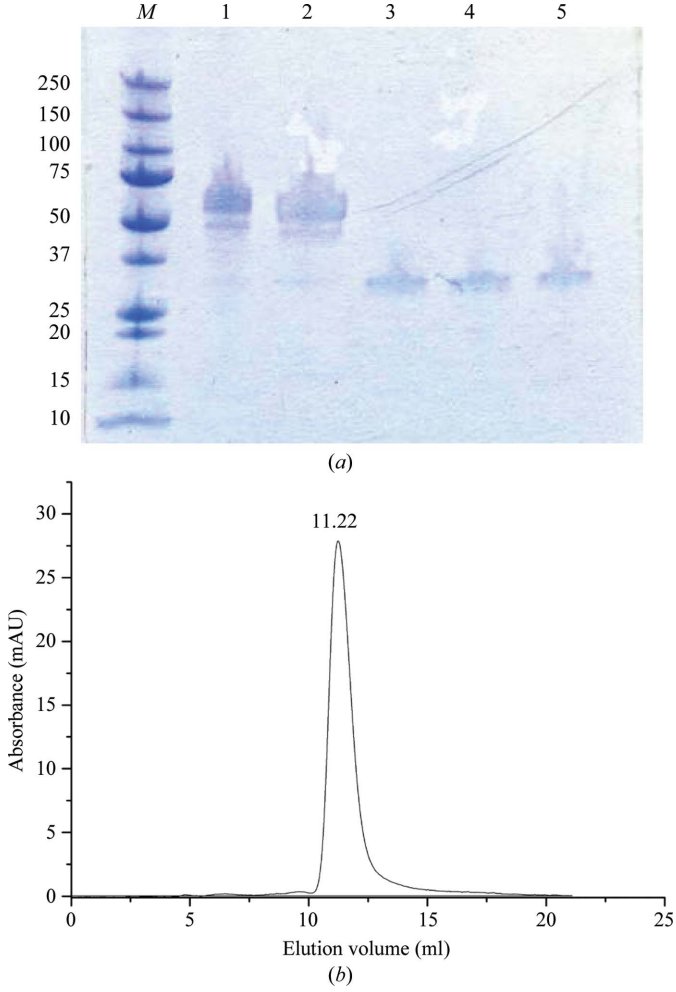
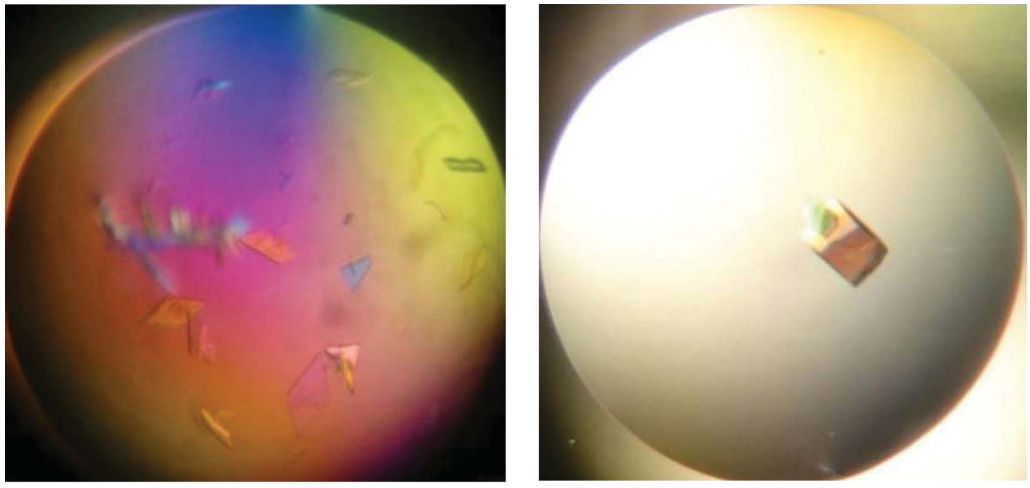


Figure 1
Purification of target C4.4A-D1D2 protein. (*a*) SDS–PAGE analysis of samples taken at different purification steps. Lane *M*, molecular-mass marker (labelled in kDa); lanes 1 and 2, C4.4A-ent-uPAR-D3, elution fractions from mAb R2 immunoaffinity chromatography; lanes 3–5, C4.4A D1D2, elution fractions from SEC, with an approximate molecular weight of 32 kDa. (*b*) Analysis of C4.4A-D1D2 on a Superdex 75 column. C4.4A-D1D2 eluted with a retention volume of 11.22 ml, which corresponds to a monomer.



(*a*) (*b*)

Figure 2
Crystals of C4.4A-D1D2. (*a*) Crystals from preliminary screening in a 96-well plate. (*b*) Crystals grown under the optimal condition (22.5% PEG 4000, 0.1 *M* citric acid pH 3.6) in a 24-well plate.

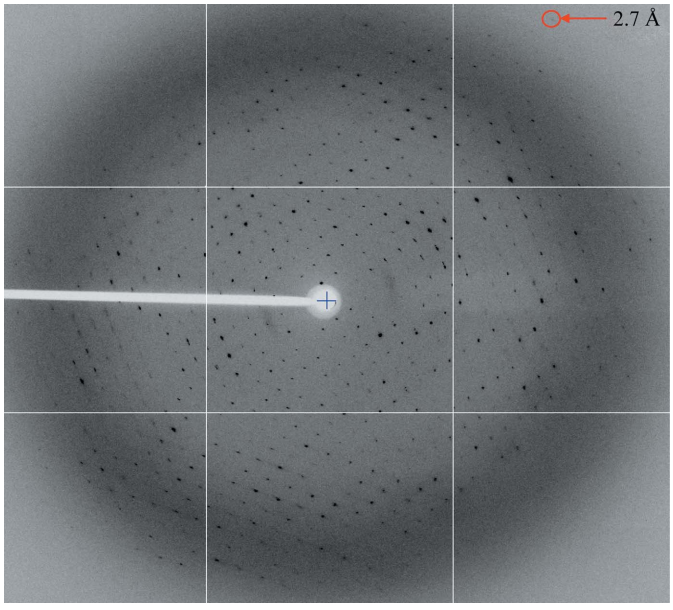


Figure 3
Representative X-ray diffraction image obtained from a crystal grown in the optimized crystal-growth condition.

probable Matthews coefficient was 2.33 Å^3 Da^{-1} and corresponded to two protein molecules per asymmetric unit, with a solvent content of 47.28%.

C4.4A shares relatively low sequence identity (28%) with its closest homologue, uPAR. Further studies are in progress to determine the structure of C4.4A-D1D2 using uPAR as a molecular-replacement template and elucidate the structure–function relationship of this protein.

Acknowledgements

We thank the staff of beamline X29 at the National Synchrotron Light Source for assistance during data collection. We thank Gitte Juhl Funch and Haldis Egholm Mønsted for excellent technical assistance.

Funding information

This work was supported by grants from the Natural Science Foundation of China (31670739, 31370737 and 31570745).

References

Gårdsvoll, H., Hansen, L. V., Jorgensen, T. J. D. & Ploug, M. (2007). *Protein Expr. Purif.* **52**, 384–394.
Gårdsvoll, H., Kriegbaum, M. C., Hertz, E. P., Alpízar-Alpízar, W. & Ploug, M. (2013). *J. Histochem. Cytochem.* **61**, 802–813.
Hansen, L. V., Gårdsvoll, H., Nielsen, B. S., Lund, L. R., Danø, K., Jensen, O. N. & Ploug, M. (2004). *Biochem. J.* **380**, 845–857.
Hansen, L. V., Laerum, O. D., Illemann, M., Nielsen, B. S. & Ploug, M. (2008). *Int. J. Cancer*, **122**, 734–741.
Hansen, L. V., Skov, B. G., Ploug, M. & Pappot, H. (2007). *Lung Cancer*, **58**, 260–266.
Jacobsen, B., Muley, T., Meister, M., Dienemann, H., Christensen, I. J., Santoni-Rugiu, E., Laerum, O. D. & Ploug, M. (2013). *J. Thorac. Oncol.* **8**, 152–160.
Jacobsen, B. & Ploug, M. (2008). *Curr. Med. Chem.* **15**, 2559–2573.
Jacobsen, B., Santoni-Rugiu, E., Illemann, M., Kriegbaum, M. C., Laerum, O. D. & Ploug, M. (2012). *Int. J. Cancer*, **130**, 2734–2739.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kjaergaard, M., Hansen, L. V., Jacobsen, B., Gardsvoll, H. & Ploug, M. (2008). *Front. Biosci.* **13**, 5441–5461.
Konishi, K., Yamamoto, H., Mimori, K., Takemasa, I., Mizushima, T., Ikeda, M., Sekimoto, M., Matsuura, N., Takao, T., Doki, Y. & Mori, M. (2010). *Cancer Sci.* **101**, 2269–2277.
Kriegbaum, M. C., Clausen, O. P., Lærum, O. D. & Ploug, M. (2015). *J. Histochem. Cytochem.* **63**, 142–154.
Kriegbaum, M. C. *et al.* (2016). *Sci. Rep.* **6**, 25833.
Kriegbaum, M. C., Jacobsen, B., Hald, A. & Ploug, M. (2011). *J. Histochem. Cytochem.* **59**, 188–201.
Kriegbaum, M. C., Persson, M., Haldager, L., Alpízar-Alpízar, W., Jacobsen, B., Gårdsvoll, H., Kjaer, A. & Ploug, M. (2011). *Curr. Drug Targets*, **12**, 1711–1728.
Llinas, P., Le Du, M. H.., Gårdsvoll, H., Danø, K., Ploug, M., Gilquin, B., Stura, E. A. & Ménez, A. (2005). *EMBO J.* **24**, 1655–1663.

Miyake, T., Ito, T., Yanai, A., Inoue, N., Miyagawa, Y., Murase, K., Imamura, M., Ichii, S., Takatsuka, Y., Nishizaki, T., Hirota, S., Ohtsuka, M., Yamamoto, H., Noguchi, S. & Miyoshi, Y. (2015). *Breast Cancer*, **22**, 366–373.
Ohtsuka, M. *et al.* (2013). *Int. J. Oncol.* **43**, 439–446.
Paret, C., Hildebrand, D., Weitz, J., Kopp-Schneider, A., Kuhn, A., Beer, A., Hautmann, R. & Zöller, M. (2007). *Br. J. Cancer*, **97**, 1146–1156.
Persson, M. *et al.* (2015). *Theranostics*, **5**, 1303–1316.
Ploug, M. (2013). *Theranostics*, **3**, 467–476.
Ploug, M., Østergaard, S., Gårdsvoll, H., Kovalski, K., Holst-Hansen, C., Holm, A., Ossowski, L. & Dano, K. (2001). *Biochemistry*, **40**, 12157–12168.
Rösel, M., Claas, C., Seiter, S., Herlevsen, M. & Zöller, M. (1998). *Oncogene*, **17**, 1989–2002.
Smith, B. A., Kennedy, W. J., Harnden, P., Selby, P. J., Trejdosiewicz, L. K. & Southgate, J. (2001). *Cancer Res.* **61**, 1678–1685.
Thuma, F., Ngora, H. & Zöller, M. (2013). *Mol. Oncol.* **7**, 917–928.
Winter, G. (2010). *J. Appl. Cryst.* **43**, 186–190.