Expression, purification, and crystallization of a plant polyketide cyclase from *Cannabis sativa*

Xinmei Yang[a], Takashi Matsui[a], Takahiro Mori[b], Futoshi Taura[c], Hiroshi Noguchi[d], Ikuro Abe[b]* and Hiroyuki Morita[a]*

[a]Institute of Natural Medicine, University of Toyama, 2630-Sugitani, Toyama, Toyama, 930-0194, Japan

[b]Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

[c]Graduate School of Medicine and Pharmaceutical Sciences, University of Toyama, 2630-Sugitani, Toyama, Toyama, 930-0194, Japan

[d]School of Pharmaceutical Sciences, University of Shizuoka, 52-1 Yada, Suruga, Shizuoka, 422-8526, Japan

Correspondence email: abei@mol.f.u-tokyo.ac.jp; hmorita@inm.u-toyama.ac.jp

Keywords: Cannabinoids; *Cannabis sativa*; Cyclase; Olivetolic acid.

Synopsis

Olivetolic acid cyclase (OAC) from *Cannabis sativa* L. has been overexpressed in *E. coli*, purified, and crystallized. Diffraction data have been collected to a resolution of 1.40 Å.

Abstract

Plant polyketides are a structurally diverse family of natural products. In the biosynthesis of plant polyketides, the construction of the carbocyclic scaffolds is a key step for diversifying the polyketide structure. Olivetolic acid cyclase (OAC) from *Cannabis sativa* L. is the only known plant polyketide cyclase that catalyzes the C2/C7 intramolecular aldol cyclization of linear pentyl tetra-β-ketide CoA to generate olivetolic acid in the biosynthesis of cannabinoid. The enzyme is also thought to belong to the dimeric α+β barrel (DABB) protein family. However, because of lack of the functional analysis of the other plant DABB proteins and low sequence identity with the functionally and structually characterized bacterial DABB proteins, the catalytic mechanism of OAC has remained unclear. To clarify the intimate catalytic mechanism of OAC, the enzyme was overexpressed in *Escherichia coli* and crystallized in a vapour-diffusion method. The crystals diffracted X-rays to 1.40 Å resolutions and belonged to space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b =$ 47.3 Å, c = 176.0 Å. Further crystallographic analysis will provide valuable insights into the structure–function relationship and catalytic mechanism of OAC.

1. Introduction

Plant polyketides, such as flavonoids, stilbenes, and phloroglucinols, are one of the largest and most important families of natural products, with remarkable structural diversity and biological activities. In the biosynthesis of plant polyketides, the construction of the carbocyclic scaffolds is a key step for diversifying the polyketide structures. This process is generally catalyzed by a type III polyketide synthase (PKS), by carbon chain elongation and subsequent cyclization of the highly reactive poly-β-keto intermediate (Schröder, 1999; Austin & Noel, 2003; Abe & Morita, 2010). Recent studies have suggested that the biosyntheses of plant polyketides, such as the anthranoid (Abe et al., 2005; Abdel-Rahman et al., 2013) and cannabinoid (Taura et al., 2009; Gagne et al., 2012), produced by *Aloe arborescens* and *Cannabis sativa*, respectively, require additional enzymes for proper folding cyclization of the linear poly-β-keto intermediate to generate the final products, as in the cases of bacterial polyketide biosyntheses.

The recently identified olivetolic acid cyclase (OAC) from *C. sativa* is a novel polyketide cyclase that is proposed to be involved in the biosynthesis of cannabinoid. The enzyme is the only known plant polyketide cyclase, and is thought to accept the linear pentyl tetra-β-ketide CoA, produced by a type-III PKS tetraketide synthase (TKS), as the substrate, and perform the C2–C7 aldol cyclization, thioester bond cleavage, and aromatization reactions to generate olivetolic acid (OA), without requiring any co-factors (Fig. 1, Gagne et al., 2012). OAC shares 30-48% identity with functionally unidentified plant dimeric $\alpha+\beta$ barrel (DABB) proteins, such as the structurally characterized heat stable protein (AtHS1) from *Arabidopsis thaliana* (48% identity, Bingman et al., 2004; Lytle et al., 2004), boiling stable protein (SP1) from *Populus tremula* (38% identity, Dgany et al., 2004), and At5g22580 from *A. thaliana* (32% identity, Cornilescu et al., 2004). OAC also shares low sequence identity (less than 20%) with bacterial DABB proteins, such as the structurally characterized bacterial polyketide cyclase, tetracenomycin F2 cyclase (Tcm I) from *Streptomyces glaucescens* (17% identity, Thompson et al., 2004), and the functionally distinct, structurally characterized ActVA-Orf6 monooxygenase from *S. coelicolor* (15% identity, Sciara et al., 2003) and 4-methylmuconolactone methylisomerase (MLMI) from *Pseudomonas reinekei* (13% identity, Marin et al., 2009).

On the basis of these findings and a comparison of the homology model of OAC with the three-dimensional structures of the plant and bacterial DABB proteins, OAC has been proposed to be a homodimeric protein consisting of 12.2 kDa subunits and to possess a hydrophobic tunnel as the active site cavity in each monomer (Gagne et al., 2012), as in the cases of the other structurally characterized DABB proteins. Furthermore, site-directed mutagenesis studies have suggested that three His residues (His5, His57, and His78) play crucial roles in the OA-forming activity (Gagne et al., 2012). However, the catalytic mechanism underlying the substrate and product specificities, the aldol cyclization and aromatization reactions, and the thioester bond cleavage, as well as the catalytic role of the three His residues in the OA-forming activity, have remained unclear. Therefore, to further clarify the intimate structural details of the OAC catalyzed reaction, we expressed glutathione S-transferase (GST)-fused recombinant OAC in *Escherichia coli*, removed the GST portion, and obtained good-quality crystals of the recombinant OAC.

2. Materials and methods

2.1. Expression and purification

The cDNA encoding full-length OAC, with *Sma*I/*Sal*I sites just before the initiation codon and after the stop codon, respectively, was purchased from Eurofins Genomics (Table 1). The cDNA was digested with *Sma*I/*Sal*I and was ligated into the *Sma*I/*Sal*I sites of the modified pQE-80L vector (QIAGEN), for expression as a fusion protein with GST at the N-terminus. A PreScission Protease (GE Healthcare) cleavage site was introduced between GST and OAC. The constructed expression plasmid was transformed into *E. coli* M15 cells (QIAGEN), and the cells harboring the plasmid were cultured to an OD_{600} of 0.6 in Luria broth (LB) medium, containing 100 μg ml^{-1} ampicillin, at 310 K. Isopropyl β-D-1-thiogalactopyranoside was added to a final concentration of 1 m*M* to induce protein expression, and the culture was incubated at 290 K for a further 20 h.

All of the following procedures were performed at 277 K. The *E. coli* cells were harvested by centrifugation at 5,000 g for 20 min, and then resuspended in 50 m*M* Tris-HCl buffer, pH 8.0, containing 200 m*M* NaCl, 5%(v/v) glycerol and 2 m*M* DTT (buffer A). The cells were disrupted by sonication, and the lysate was centrifuged at 6,000 g for 10 min. The supernatant was loaded on a COSMOGEL GST-Accept Resin column (Nacalai Tesque) equilibrated with buffer A. After the column was washed with 20 m*M* HEPES-NaOH buffer, pH 7.5, containing 100 m*M* NaCl, 5%(v/v) glycerol and 2 m*M* DTT (buffer B), the GST-tag was cleaved on the column by PreScission Protease (GE Healthcare) overnight, and OAC was eluted with buffer B. The resultant protein thus contains three additional N-terminal residues (GPG), derived

from the PreScission Protease recognition sequence. The protein solution was diluted ten-fold with 20 m*M* HEPES-NaOH buffer, pH 7.5, containing 2 m*M* DTT (buffer C), and then applied to a Resource Q column (GE Healthcare). The column was washed with buffer C containing 10 m*M* NaCl, and the protein was subsequently eluted at 50 m*M* NaCl, using a linear gradient of 10-200 m*M* NaCl. The protein was further purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare), and was concentrated to 17 mg ml^{-1} in 20 m*M* HEPES-NaOH buffer, pH 7.5, containing 25 m*M* NaCl.

A dynamic light-scattering (DLS) analysis was performed, using a DynaPro-MSXTC molecular-sizing instrument (Protein Solutions). After centrifugation through a 0.22 μm Ultrafree-MC filter (Millipore) to remove particulate material from the protein solution, the solution properties of the purified protein were monitored. Data were acquired as 50 scattering measurements at 293 K. Data from five sets of measurements were analyzed using the DYNAMICS software package (Protein Solutions) and averaged.

2.2. Crystallization

Initial crystallization conditions were found by screening with the commercial JCSG+ Suite screening kit (QIAGEN). The crystallizations were performed at 278 K, using the sitting-drop vapour-diffusion method with a 96-well plate. All crystallization drops were prepared by mixing 1 μl of protein solution (17 mg ml^{-1}) with an equal volume of reservoir solution, and were equilibrated against 50 μl reservoir solution. Clusters of crystals were observed three days later, in the crystallization condition consisting of 100 m*M* Bicine, pH 8.5, and 20%(w/v) PEG 6000. Further crystallization was attempted using Additive Screen HT (Hampton Research) at various pH values, together with the use of 20%–30%(w/v) PEG 6000 as a precipitant. Diffraction-quality crystals were finally obtained at 278 K, in 100 m*M* Tris-HCl, pH 8.8, 25%(w/v) PEG 6000, and 100 m*M* sodium malonate, using the sitting-drop vapour-diffusion method (Table 2).

2.3. Data collection and processing

Single crystals were transferred into a crystallization solution containing 10%(v/v) glycerol as a cryoprotectant, picked up with a nylon loop, and then flash-cooled at 100 K in a nitrogen-gas stream. X-ray diffraction data sets were collected on Beamline NW-12A at the Photon Factory (PF; wavelength 1.0000 Å) using an ADSC Q210r detector, with a distance of 117.1 mm between the crystal and the detector. A total of 360 frames were recorded, with a 0.5º oscillation angle and 1 s exposure time. The data were indexed, integrated, and scaled with the *XDS* program package (Kabsch, 2010). The Matthews volume (V_M) and the solvent content were calculated with *Xtriage* (Adams et al., 2010).

3. Results and discussion

The recombinant OAC was overexpressed in *E. coli*, as a fusion protein with a GST-tag at the N-terminus. After cleavage of the GST-tag, the purified OAC migrated as a single band with a molecular weight of 10 kDa upon 20%(w/v) SDS-PAGE, in agreement with the calculated molecular weight of 12.2 kDa (Fig. 2a). In contrast, the gel-filtration experiment by size exclusion chromatography indicated a molecular weight of 25 kDa (Fig. 2b). In addition to the gel-filtration analysis, the DLS analysis after the size-exclusion chromatography revealed a monomodal distribution, with a polydispersity value of 9.5% and a molecular weight estimate of 23 kDa, which allowed us to confirm that the recombinant OAC is a homodimeric protein, and also suggested that OAC is a member of the DABB protein family. The typical yield of the purified recombinant OAC was about 1.5 mg per liter of culture.

The OAC crystals appeared reproducibly in the optimized crystallization solution within a few days, with various sizes up to approximate dimensions of 0.15 × 0.15 × 0.12 mm. Among them, only the crystals with approximate dimensions of 0.07 × 0.07 × 0.06 mm gave good quality diffraction up to 1.40 Å resolution (Fig. 3). The detailed data collection

statistics are summarized in Table 3. The preliminary crystallographic analysis indicated that the crystals belonged to the space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 47.3$ Å, $c = 176.0$ Å. With two monomers in the asymmetric unit, the Matthews volume (V_M) was calculated to be 2.5 Å^3 Da^{-1} and the estimated solvent content was 50.7%, which is in the range normally observed for protein crystals. Structure determination by the molecular-replacement method is currently under way, using the *Phaser* (McCoy et al., 2007) and *Molrep* (Vagin & Teplyakov, 2010) programs with the crystal structures of the plant DABB proteins, AtHS1 (PDB entry 1q53), At5g22580 (PDB entry 1rjj), and SP1 (PDB entry 1si9), and the other bacterial homologues, such as TcmI (PDB entry 1tuw), ActVA-Orf6 (PDB entry 1lq9), and MLMI (PDB entry 2ifx) as search models. Simultaneously, we are also attempting to crystallize OAC complexed with its product and product analogues, together with the expression, purification, and crystallization of selenomethionine-labeled OAC. These structural analyses will provide valuable insights into not only the structure–function relationship and catalytic mechanism of OAC, but also the functional diversity of the DABB proteins.

Table 1

Macromolecule production information

Source organism	*Cannabis sativa* L.
DNA source	Chemical synthesis
Cloning vector	pEX-A
Expression vector	pQE-80L
Expression host	M15 (QIAGEN)
Complete amino acid sequence of the construct produced†	*MRGSHHHHHHGSMSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYE RDEGDKWRNKKFELGL EFPNLPYYIDGDVKLTQSMAIIRYIADKHN MLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDF ETLKVDFLSKLP EMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLD AFPKL VCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK* *SD*LEVLFQGPGMAVKHLIVLKFKDEITEAQKEEFFKTYVNLVNIIPAMKDVYWGKDVTQK NKEEGYTHIVEVTFESVETIQDYIIHPAHVGFGDVYRSFWEKLLIFDYTPRK

†The amino-acid sequence of glutathione S-transferase (GST) tag is shown in italics. PreScision protease recognition site (LEVLFQGP) is underlined. The peptide bond between the Q and G residues is cleaved by PreScision protease.

Table 2

Crystallization

Method	sitting-drop vapour-diffusion method
Plate type	96 well CrystalQuick, 288 square wells, flat bottom (greiner bio-one)
Temperature (K)	278
Protein concentration	17 mg ml^{-1}
Buffer composition of protein solution	20 m*M* HEPES-NaOH buffer, pH 7.5, and 25 m*M* NaCl
Composition of reservoir solution	100 m*M* Tris-HCl, pH 8.8, 25%(w/v) PEG 6000, and 100 m*M* sodium malonate
Volume and ratio of drop (μl)	1:1
Volume of reservoir (μl)	50

Table 3

Data collection and processing

Values for the outer shell are given in parentheses.

Diffraction source	NW-12A, PF
Wavelength (Å)	1.0000
Temperature (K)	100
Detector	ADSC Q210r

Crystal-detector distance (mm)	117.1
Rotation range per image (°)	0.5
Total rotation range (°)	180
Exposure time per image (s)	1
Space group	$P3_121$ or $P3_221$
a, b, c (Å)	47.3, 47.3, 176.0
Mosaicity (°)	0.193
Resolution range (Å)	50.0 - 1.40 (1.48 - 1.40)
Total No. of reflections	475,525 (74,586)
No. of unique reflections	45,736 (7,287)
Completeness (%)	98.9 (98.9)
Redundancy	10.4 (10.2)
$\langle I/\sigma(I)\rangle$	24.2 (12.3)
R_{merge}	0.074 (0.155)
Overall *B* factor from Wilson plot (Å^2)	10.19
No. molecules per AU	2
V_M (Å^3 Da^{-1})	2.5
V_{solv} (%)	50.7

$R_{merge} = \sum_{hkl}\sum_{i}|I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_{i} I_i(hkl)$, where I(*hkl*) is the intensity of reflection *hkl*, $\sum_{hkl}$ is the sum over all reflections, and $\sum_{i}$ is the sum over *i* measurements of reflection *hkl*.

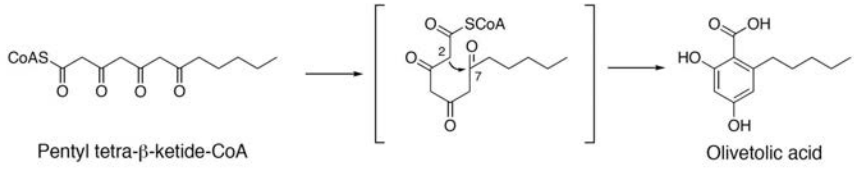

Figure 1

Proposed mechanism for the formation of olivetolic acid by OAC.

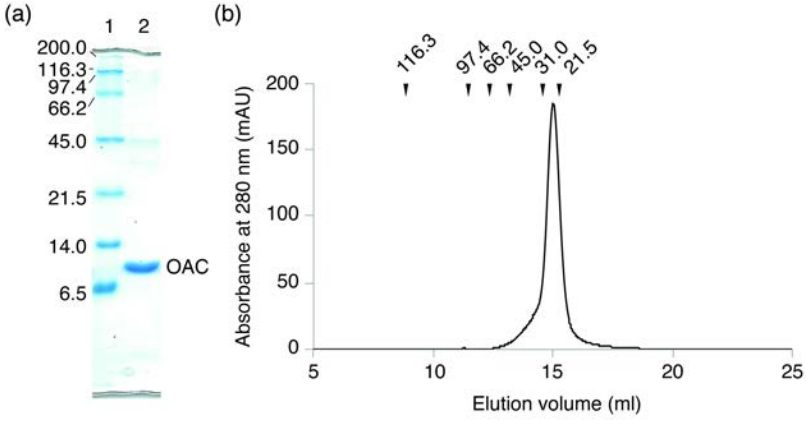

Figure 2

Analysis of molecular weight of OAC. (a) SDS-PAGE gel is shown. Lane 1, molecular weight marker; lane 2, peak fraction. Molecular weight markers are labelled in kDa. (b) Chromatogram of size-exclusion chromatography is shown. Peak represents OAC. Molecular weights (kDa) of marker proteins are indicated on the top of the chromatogram.

7_original.jpg

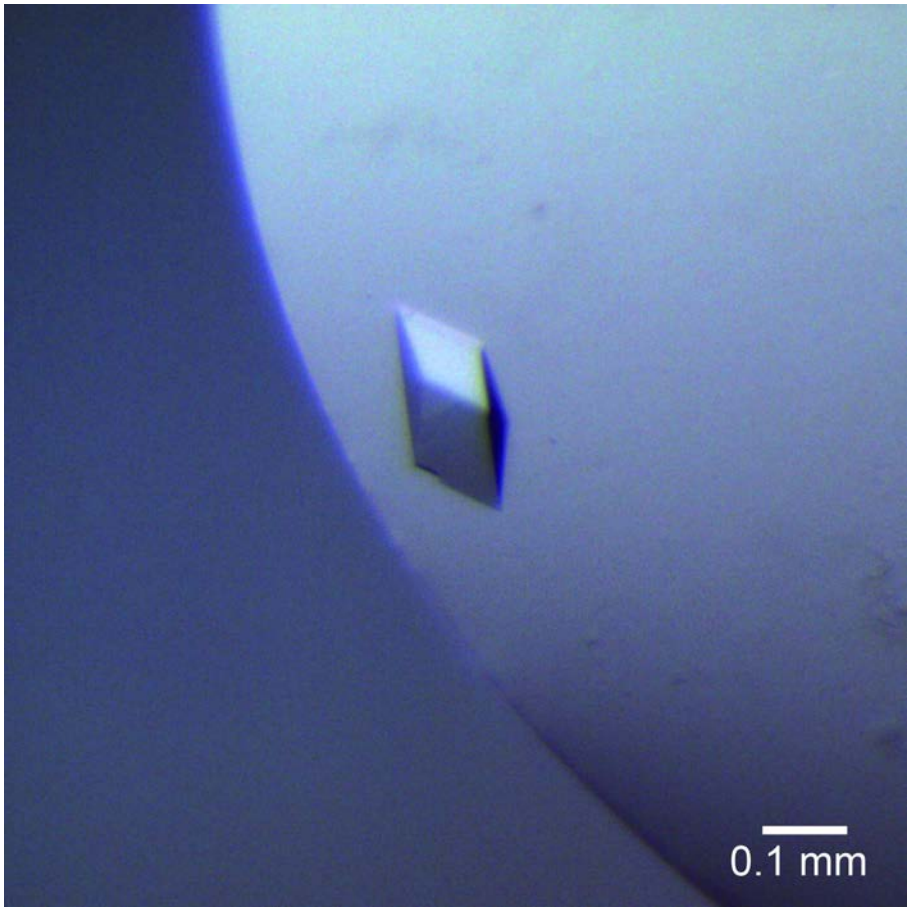

Figure 3

Crystals of OAC. The dimensions of the crystals were approximately 0.07 × 0.07 × 0.06 mm.

Acknowledgements

This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (to H.N., I.A., and H.M.), by a grant from Japan Foundation for Applied Enzymology (to H.M.), and by a grant from NAGASE Science Technology Foundation (to H.M.). The authors declare that they have no competing financial interests. We thank the beamline staff at the Photon Factory for their assistance with X-ray diffraction data collection.

References

Abe, I. & Morita, H. (2010). *Nat. Prod. Rep.* 27, 809–838.

Abe, I., Oruro, S., Utsumi, Y., Sano, Y. & Noguchi, H. (2005). *J. Am. Chem. Soc.* 127, 12709–12716.

Abdel-Rahman, I.A.M., Beuerle, T., Ernst, L., Abdel-Baky, A. M., Desoky, E. E-D. K., Ahmed, A. S. & Beerhues, L. (2013). *Phytochemistry* 88, 15–24.

Adams, O. D., Afonine, P. V., Bunkóczi, Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J.

S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr. D. Biol. Crystallogr.* 66, 213–221.

Austin, M. B. & Noel, J. P. (2003). *Nat. Prod. Rep.* 20, 79–110.

Bingman, C. A, Johnson, K. A, Peterson, F. C., Frederick, R. O., Zhao, Q., Thao, S., Fox, B. G., Volkman, B. F., Jeon, W. B., Smith, D. W., Newman, C. S., Ulrich, E. L., Hereman, A., Sussman, M. R., Markley, J. L. & Phillips, G. N. (2004). *Proteins* 57, 218–220.

Cornilescu, G., Cornilescu, C. C., Zhao, Q., Frederick, R. O., Peterson, F. C., Thao, S. & Markley, J. L. (2004). *J. Biomol. NMR*. 29, 387–390.

Dgany, O., Gonzalez, A., Sofer, O., Wang, W., Zolotnitsky, G., Wolf, A., Shoham, Y., Altman, A., Wolf, S. G., Shoseyov, O. & Almog, O. (2004). *J. Biol. Chem.* 279, 51516–51523.

Gagne, S. J., Stout, J. M., Liu, E., Boubakir, Z., Clark, S. M. & Page, J. E. (2012). *Proc. Natl. Acad. Sci. U. S. A.* 109, 12811–12816.

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

Lytle, B., Peterson, F., Kjer, K., Frederick, R., Zhao, Q., Thao, S., Bingman, C., Johnson, K., Phillips Jr., G. & Volkman, B. (2004). *J. Biomol. NMR* 28, 397–400.

Marin, M., Heinz, D. W., Pieper, D. H. & Klink, B. U. (2009). *J. Biol. Chem.* 284, 32709–32716.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Crystallogr.* 40, 658–674.

Sciara, G., Kendrew, S. G., Miele, A. E., Marsh, N. G., Federici, L., Malatesta, F., Schimperna, G., Savino, C. & Vallone, B. (2003). *EMBO J.* 22, 205–215.

Schröder, J. (1999). *Comprehensive Natural Products Chemistry* , Vol. 2, edited by D. E. Cane, pp. 749–771. Oxford: Elsevier.

Taura, F., Tanaka, S., Taguchi, C., Fukamizu, T., Tanaka, H., Shoyama, Y. & Morimoto, S. (2009). *FEBS Lett.* 583, 2061–2066.

Thompson, T. B., Katayama, K., Watanabe, K., Hutchinson, C. R. & Rayment, I. (2004). *J. Biol. Chem.* 279, 37956–37963.

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