Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Purification, crystallization and preliminary X-ray diffraction analysis of a hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT) from *Coffea canephora* involved in chlorogenic acid biosynthesis

Laura A. Lallemand,[a] James G. McCarthy,[b] Sean McSweeney[a] and Andrew A. McCarthy[c,d]*

[a]European Synchrotron Radiation Source, 6 Rue Jules Horowitz, BP 181, 38043 Grenoble, France, [b]Nestlé Research and Development, 101 Avenue Gustave Eiffel, Notre-Dame D'Oe, BP 49416, 37097 Tours, France, [c]Unit of Virus Host-Cell Interactions, UJF–EMBL–CNRS, UMR 5233, 6 Rue Jules Horowitz, 38042 Grenoble, France, and [d]European Molecular Biology Laboratory, 6 Rue Jules Horowitz, BP 181, 38042 Grenoble, France

Correspondence e-mail: andrewmc@embl.fr

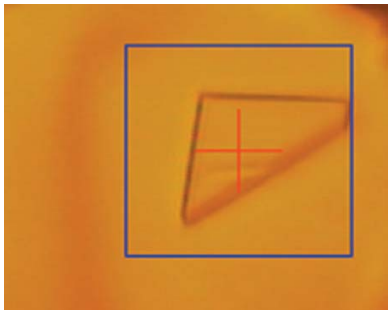
Received 14 March 2012
Accepted 27 April 2012

Chlorogenic acids (CGAs) are a group of soluble phenolic compounds that are produced by a variety of plants, including *Coffea canephora* (robusta coffee). The last step in CGA biosynthesis is generally catalysed by a specific hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT), but it can also be catalysed by the more widely distributed hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT). Here, the cloning and overexpression of HCT from *C. canephora* in *Escherichia coli* as well as its purification and crystallization are presented. Crystals were obtained by the sitting-drop technique at 293 K and X-ray diffraction data were collected on the microfocus beamline ID23-2 at the ESRF. The HCT crystals diffracted to better than 3.0 Å resolution, belonged to space group $P4_22_12$ with unit-cell parameters $a = b = 116.1$, $c = 158.9$ Å and contained two molecules in the asymmetric unit. The structure was solved by molecular replacement and is currently under refinement. Such structural data are needed to decipher the molecular basis of the substrate specifities of this key enzyme, which belongs to the large plant acyl-CoA-dependent BAHD acyltransferase superfamily.

1. Introduction

In addition to caffeine and its alkaloid precursors, *Coffea canephora* (robusta coffee) contains significant amounts of phenolic secondary metabolites known as chlorogenic acids (CGAs). This family of esters is formed between certain hydroxycinnamic acids and quinic acid or its derivative shikimic acid, with the most common ester being 5-caffeoylquinic acid (5-CQA). The other main groups of CGAs identified in green coffee beans include other CQA isomers, as well as dicaffeoylquinic and feruloylquinic acid isomers (Clifford & Wight, 1976). There is much evidence that this complex and diverse class of ester conjugates fulfil various physiological roles in plants. For example, they have been shown to be involved in the protection of plants from abiotic stresses such as UV damage (Burchard *et al.*, 2000). In addition, elevated CGA levels in tomato have been shown to increase resistance to bacterial infections (Niggeweg *et al.*, 2004). When obtained from plant-derived foods, they can also have a beneficial antioxidant effect on human health (Scalbert & Williamson, 2000). Indeed, some studies have shown that dietary CGAs can play a role in the prevention of important degenerative human conditions (Shahidi & Chandrasekara, 2009).

Some plants specifically accumulate chlorogenic acids containing only one type of hydroxyacid, such as quinic acid in coffee. This is consistent with the observation that the quinic acid precursor is usually present at higher levels than shikimic acid in plants (Boudet, 1973). In plants, both quinate and shikimate esters are synthesized from hydroxycinnamoyl-CoA thioesters by an acyltransfer reaction. The more widely distributed chlorogenic acid, 5-CQA, is synthesized from a caffeoyl-CoA thioester and quinic acid by a specific hydroxycinnamoyl-CoA quinate hydroxycinnamoyltransferase (HQT; EC 2.3.1.99) that is found in several plant species (Fig. 1). In addition, *C. canephora* also encodes the closely related hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HCT; EC



2.3.1.133), which has a preference for shikimic acid over quinic acid (Fig. 1). Two genes encoding hydroxycinnamoyltransferases have been isolated from *C. canephora* (Lepelley *et al.*, 2007). From sequence comparisons, they have been characterized as HQT and HCT, which differ in their substrate specificity, preferring quinic acid and shikimic acid, respectively (Sonnante *et al.*, 2010; Hoffmann *et al.*, 2003). A sequence alignment showed that they belong to the plant CoA-dependent acyltransferase BAHD superfamily, the name of which is derived from the first four biochemically characterized members (D'Auria, 2006; Tuominen *et al.*, 2011). As for the other BAHD enzymes, HCT can work in the reverse direction and catalyze the cleavage of chlorogenic acids (Fig. 1), as well as producing intermediates for lignin biosynthesis (Humphreys & Chapple, 2002).

The first representative structure of the BAHD superfamily was that of vinorine synthase (Ma *et al.*, 2005). Another member, anthocyanin malonyltransferase from *Dendranthema morifolium*, was subsequently solved in complex with malonyl-CoA (Unno *et al.*, 2007). More recently, structures of trichothecene 3-*O*-acetyltransferase from *Fusarium sporotrichioides* and *F. graminearum* complexed with either CoA or trichothecene mycotoxins have been described (Garvey *et al.*, 2008). All three enzymes show a similar fold to the previously described PapA5, a phthiocerol dimycocerosyl transferase from *Mycobacterium tuberculosis* (Buglino *et al.*, 2004). However, the enzyme described here is particularly interesting because it represents the first member to be crystallized which can accommodate an aromatic acyl-CoA substrate and either quinic acid or shikimic acid.

2. Materials and methods

2.1. Expression and purification

The *C. canephora* HCT gene (DDBJ/GenBank/EBI accession No. EF137954; Lepelley *et al.*, 2007) was subcloned into the expression vector pProEX_HTb (Invitrogen) using the *Bam*HI and *Xba*I restriction sites. A synthetic gene encoding HCT with codons optimized for *Escherichia coli* expression was subsequently obtained from Geneart (Regensburg, Germany) and cloned into pProEX_HTb (Invitrogen). All further experiments were carried out using the latter plasmid and an optimized expression protocol. The full-length protein was overexpressed with an N-terminal hexahistidine tag in *E. coli* BL21 Star (DE3) pLysS cells (Invitrogen) overnight at 291 K after induction with 1 m*M* IPTG. The cultures were centrifuged at 5000*g* for 20 min and the bacterial pellets were resuspended in a minimal volume of buffer consisting of 50 m*M* Tris–HCl pH 8.0, 500 m*M* NaCl, 10% glycerol, 5 m*M* β-mercaptoethanol (buffer *A*). The pellet stock was flash-cooled and stored at 193 K until purification.

Prior to cell lysis, lysozyme (200 µg ml^{-1}), DNaseI (20 µg ml^{-1}) and EDTA-free Complete Protease Inhibitor Cocktail tablets (Roche) were added to the thawing cells and stirred for 15 min at 277 K. The resuspended cells were lysed using a French pressure-cell press (two cycles at 69 MPa). The lysate was then centrifuged at 50 000*g* for 30 min at 277 K to remove the cell debris. A 5 ml HisTrap HP column (GE Healthcare) was equilibrated with 10% buffer *B* (buffer *A* supplemented with 500 m*M* imidazole pH 8.0). The soluble fraction of the cell lysate was loaded and an imidazole gradient from 10 to 100% buffer *B* was applied to elute the His-tagged protein. The fractions containing the majority of the recombinant protein were pooled, incubated with TEV protease and dialyzed overnight at 277 K in 50 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl, 10% glycerol, 1 m*M* DTT, 0.5 m*M* EDTA. The native HCT, which contained an additional glycine and alanine at the N-terminus after TEV cleavage, was subsequently recovered by subtractive chromatography using the same nickel column. The flowthrough was typically concentrated to 5–10 mg ml^{-1} and injected onto a Superdex 200 10/300 GL column (GE Healthcare) for size-exclusion chromatography with 20 m*M* Tris–HCl pH 7.5, 150 m*M* NaCl, 5 m*M* β-mercaptoethanol. The peak fractions were collected, run on a 12% SDS–polyacrylamide gel by electrophoresis and subsequently stained with Coomassie Blue

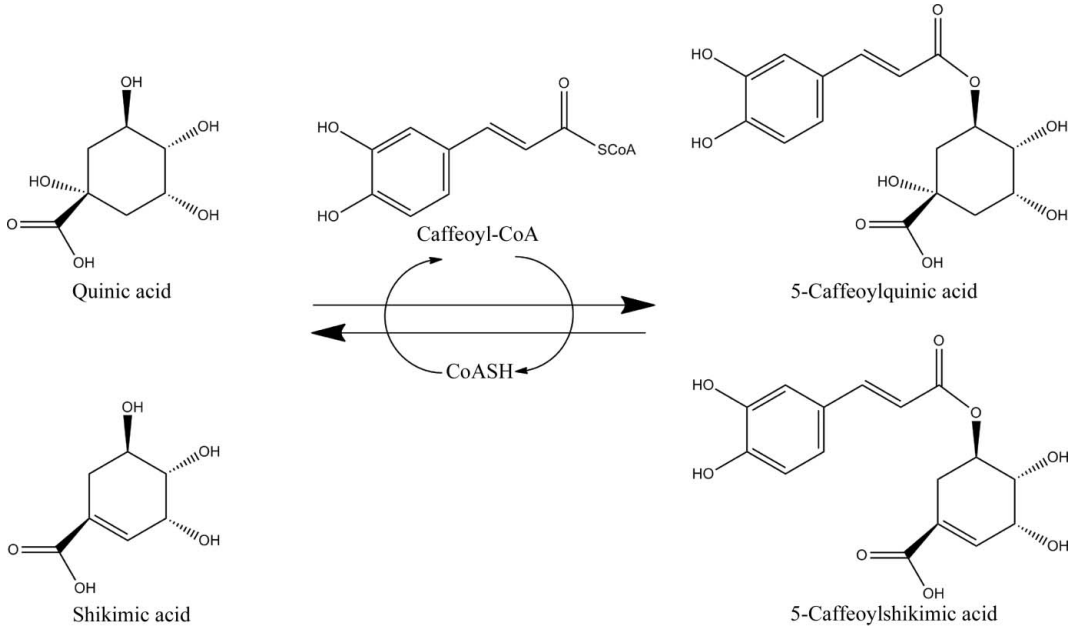


Figure 1
HCT can catalyse the transfer of activated hydroxycinnamoyl moieties such as caffeoyl-CoA to shikimic acid or quinic acid.

Table 1
Crystal data and data-collection statistics.

Values in parentheses are for the outermost shell.

X-ray source	ID23-2
Wavelength (Å)	0.873
No. of images	40
Oscillation range (°)	1.0
Unit-cell parameters (Å)	$a = b = 116.12$, $c = 158.93$
Space group	$P4_22_12$
No. of molecules in asymmetric unit	2
Matthews coefficient (Å^3 Da^{-1})	2.76
Solvent content (%)	55.5
Resolution range (Å)	45–3.0
Unique reflections	22073 (3143)
Observed reflections [$I/\sigma(I) > 1$]	72116 (10301)
Completeness (%)	99.0 (98.4)
$R_{\rm merge}$†	16.6 (56.6)
Mean $I/\sigma(I)$	8.5 (3.1)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$ calculated for the whole data set.

(Fig. 2*a*). The yield was approximately 2 mg of protein per litre of LB culture as determined using the Bradford assay (Bradford, 1976).

2.2. Crystallization and preliminary X-ray data

Sitting drops consisting of 100 nl HCT at 20 mg ml^{-1} and 100 nl precipitant solution were dispensed using a Cartesian crystallization robot into 96-well Greiner CrystalQuick plates. HCT crystals appeared after several weeks at 293 K in condition No. 62 of The Classics Suite (Qiagen) consisting of 1.6 *M* magnesium sulfate, 0.1 *M* MES pH 6.5. For X-ray measurements, the HCT crystals were fished directly from the Greiner plates and transferred into mother-liquor solution containing 20% glycerol before being flash-cooled in liquid nitrogen (Fig. 3). All X-ray diffraction data were collected on the highly automated microfocus beamline ID23-2 (Flot *et al.*, 2010) at the ESRF, Grenoble, France. HCT crystallized in space group $P4_22_12$ with two molecules per asymmetric unit, and a 3.0 Å resolution data set was collected. The data were processed and scaled using *XDS* (Kabsch, 2010) and *SCALA* (Evans, 2006). A summary of the data statistics is given in Table 1.

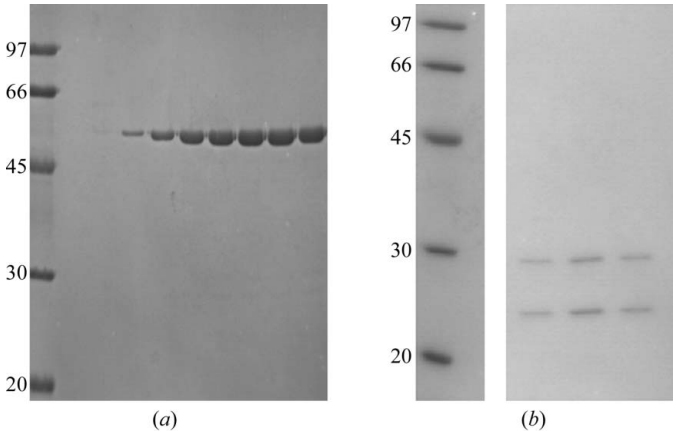


Figure 2
SDS–PAGE (12%) analysis of (*a*) purified HCT fractions after gel filtration and (*b*) a limited proteolysis digest of 0.5 mg ml^{-1} HCT with a 1:1000(*w*:*w*) trypsin dilution. Molecular-weight markers are labelled in kDa.

2.3. Proteolysis data

In our attempts to reproduce the initial crystals, we discovered that HCT (~48 kDa) was being cleaved over time at 293 K into two stable fragments (Fig. 2*b*). This degradation pattern could be reproduced by incubating 0.5 mg ml^{-1} HCT with a 1:1000(*w*:*w*) dilution of diverse proteases (trypsin, chymotrypsin and subtilisin) on ice. Electrospray ionization mass-spectrometric analysis of the two bands showed that they have a similar size (24 528 and 23 906 Da), with the sum being comparable to the mass of native HCT (48 464 Da). Subsequent N-terminal sequencing localized the protease-cleavage site in the central part of the enzyme, corresponding to a large crossover loop connecting the N- and C-terminal domains in BAHD structures. This must be a local disorder because the protein remains insensitive to further digestion. The digested protein has a similar gel-filtration profile and is still active *in vitro* (data not shown).

2.4. Thermofluor assays

Thermofluor assays were carried out on a freshly purified protein sample diluted to a final concentration of 10 µ*M*. Each condition was

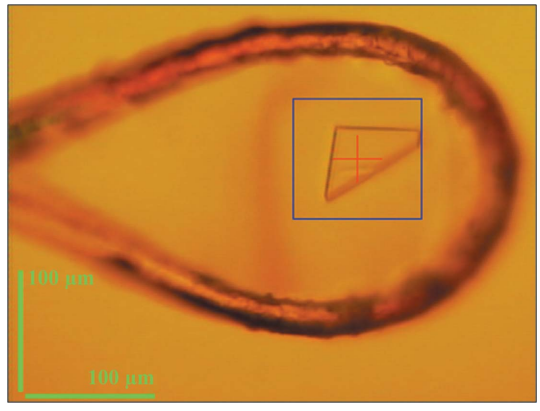


Figure 3
A crystal of HCT flash-cooled in a nylon loop mounted on an MD2M goniometer in a nitrogen stream at 100 K.

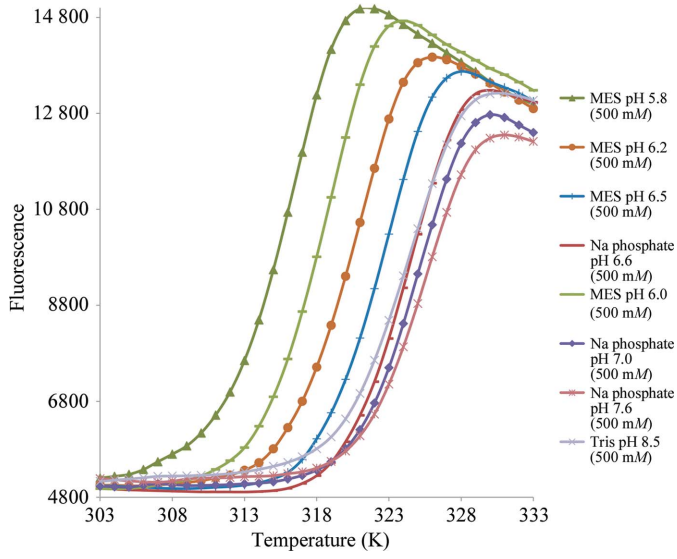


Figure 4
Thermofluor assay-buffer screen illustrating the improved stability in sodium phosphate buffer and at higher pH values.

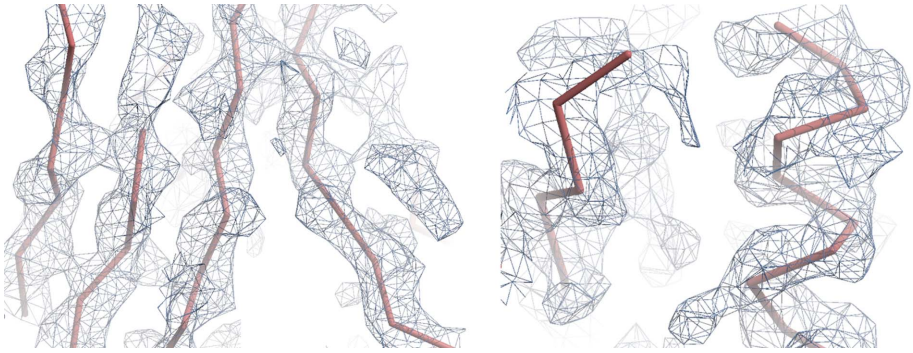


Figure 5
A *Coot* (Emsley & Cowtan, 2004) screenshot showing the best regions of a $2F_o - F_c$ map calculated with the molecular-replacement solution.

composed of 25 µl 2× test buffer solution or water as a control and 1.25 µl 200× SYPRO Orange preparation (Invitrogen) and was made up to a final volume of 50 µl per well. A real-time quantitative MX3005P PCR instrument (Stratagene) was used with standard 96-well PCR plates and optical seals. The plate was heated from 298 to 362 K with a heating rate of 1 K min^{-1} and the fluorescence intensity was measured using an SYBR filter (excitation and emission wavelengths of 492 and 516 nm, respectively). The raw data were analysed using *Microsoft Excel* by calculating the temperature value corresponding to the average of the maximum and minimum fluorescence intensity values.

3. Results and discussion

The native HCT gene results in a relatively low protein yield (typically 1–2 mg per litre of culture). We therefore decided to investigate whether this could be improved using a codon-optimized HCT synthetic gene. Although this did not result in a significantly improved yield, we decided to continue using the synthetic gene. A Thermofluor assay (Ericsson *et al.*, 2006) clearly illustrated that the protein is more stable in phosphate buffers or at higher pH ranges (Fig. 4). We therefore now routinely use sodium phosphate buffer for the initial purification steps and biochemical characterizations. HCT in the concentration range 5–20 mg ml^{-1} elutes from a gel-filtration column as a single peak corresponding to an apparent molecular weight of 45–50 kDa or approximately the calculated molecular weight of the monomeric protein. A similar molecular weight was obtained from dynamic light-scattering experiments prior to crystallization (normally at 10–20 mg ml^{-1}). This indicates that HCT is monomeric in solution, similar to other structural homologues characterized to date (Ma *et al.*, 2005; Unno *et al.*, 2007; Garvey *et al.*, 2008). Proteolysis experiments clearly showed the presence of a large disordered loop between the N- and C-terminal domains. This particular degradation pattern may explain why several batches of protein and crystallization screens were required before a single condition was identified to produce crystals of HCT. The single plate-like crystals grew to approximate dimensions of 0.05 × 0.05 × 0.01 mm (Fig. 3) after two months.

An X-ray data set was collected to better than 3.0 Å resolution from one of these crystals on the microfocus beamline ID23-2 at the ESRF (Flot *et al.*, 2010). The overall X-ray diffraction data are only of medium quality (Table 1). This is probably owing to the anisotropy observed in the diffraction pattern, with diffraction limits of 3.0 Å along the *a** and *b** directions and 2.8 Å along *c**, a feature that is often observed in plate-like crystals. Nevertheless, the data at low resolution are reasonable (R_{merge} = 4.2% for data between 50 and 8 Å). An analysis of the systematic absences suggested that the space group was $P4_22_12$. This corresponds to a Matthews coefficient of 2.8 Å^3 Da^{-1} and a solvent content of 56% with two molecules per asymmetric unit. Molecular-replacement calculations were carried out with *Phaser* (McCoy *et al.*, 2007). The search model consisted of chain *A* of vinorine synthase (PDB entry 2bgh; Ma *et al.*, 2005), which has a sequence identity of 23%. We used *Phaser* (McCoy *et al.*, 2007) to successfully place two molecules in the asymmetric unit. The first molecule resulted in a translation-function *Z* score (TFZ) of 4.8 and a log-likelihood gain (LLG) of 28. The second and last search resulted in a final TFZ of 8.7 and an LLG of 73. A single round of refinement (ten cycles of unrestrained refinement) in *REFMAC*5 (Murshudov *et al.*, 2011) using this molecular-replacement-derived model consisting of two unmodified vinorine synthase molecules resulted in an *R* and R_{free} of 47.6% and 52.8%, respectively. Initial weighted electron-density maps with Fourier coefficients $2F_o - F_c$ and *REFMAC*5 phases were of interpretable quality (Fig. 5), except for some regions such as the central part of the protein (amino acids 205–225). Further refinement and model-building cycles are currently in progress.

Despite extensive efforts, we have never reproduced these crystals again. In an effort to produce more reproducible crystals, we have recently mutated the trypsin-sensitive lysine residues to alanines in the crossover loop. This resulted in a more protease-resistant HCT that we can now reproducibly crystallize (manuscript in preparation). It therefore seems likely that proteolysis of the crossover loop hinders the crystallization of HCT. We hope that these results will enable us to perform a more systematic and detailed structural analysis.

We gratefully acknowledge the use of the Quality Control Platform at the Institut de Biologie Structurale, the High Throughput Crystallization Laboratory at the EMBL Grenoble and the EMBL-Grenoble/ESRF Joint Structural Biology Group for access to ESRF beamline ID23-2.

References

Boudet, A. (1973). *Phytochemistry*, **12**, 363–370.
Buglino, J., Onwueme, K. C., Ferreras, J. A., Quadri, L. E. & Lima, C. D. (2004). *J. Biol. Chem.* **279**, 30634–30642.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Burchard, P., Bilger, W. & Weissenböck, G. (2000). *Plant Cell Environ.* **23**, 1373–1380.
Clifford, M. N. & Wight, J. (1976). *J. Sci. Food Agric.* **27**, 73–84.
D'Auria, J. C. (2006). *Curr. Opin. Plant Biol.* **9**, 331–340.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.
Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S. & Mitchell, E. (2010). *J. Synchrotron Rad.* **17**, 107–118.
Garvey, G. S., McCormick, S. P. & Rayment, I. (2008). *J. Biol. Chem.* **283**, 1660–1669.
Hoffmann, L., Maury, S., Martz, F., Geoffroy, P. & Legrand, M. (2003). *J. Biol. Chem.* **278**, 95–103.
Humphreys, J. M. & Chapple, C. (2002). *Curr. Opin. Plant Biol.* **5**, 224–229.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Lepelley, M., Cheminade, G., Tremillon, N., Simkin, A., Caillet, V. & McCarthy, J. (2007). *Plant Sci.* **172**, 978–996.
Ma, X., Koepke, J., Panjikar, S., Fritzsch, G. & Stöckigt, J. (2005). *J. Biol. Chem.* **280**, 13576–13583.
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.
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.
Niggeweg, R., Michael, A. J. & Martin, C. (2004). *Nature Biotechnol.* **22**, 746–754.
Scalbert, A. & Williamson, G. (2000). *J. Nutr.* **130**, 2073S–2085S.
Shahidi, F. & Chandrasekara, A. (2009). *Phytochem. Rev.* **9**, 147–170.
Sonnante, G., D'Amore, R., Blanco, E., Pierri, C. L., De Palma, M., Luo, J., Tucci, M. & Martin, C. (2010). *Plant Physiol.* **153**, 1224–1238.
Tuominen, L. K., Johnson, V. E. & Tsai, C.-J. (2011). *BMC Genomics*, **12**, 236.
Unno, H., Ichimaida, F., Suzuki, H., Takahashi, S., Tanaka, Y., Saito, A., Nishino, T., Kusunoki, M. & Nakayama, T. (2007). *J. Biol. Chem.* **282**, 15812–15822.