Combining secondary-structure and protein solvent-accessibility predictions in methionine substitution for anomalous dispersion

Hsin-Yi Wu^a and Yi-Sheng Cheng^{a,b*}

^aInstitute of Plant Biology, National Taiwan University, Taipei 10617, Taiwan, and

^bDepartment of Life Science, National Taiwan University, Taipei 10617, Taiwan

Correspondence e-mail: chengys@ntu.edu.tw

Received 24 December 2013

Accepted 27 January 2014

In X-ray crystallographic analysis, the single-wavelength and multi-wavelength anomalous diffraction (SAD and MAD) methods have been widely used in order to solve the phase problem. Selenium-labelled methionine has been shown to be very effective for anomalous dispersion phasing, and at least one selenomethionine is required for every 100 amino acids. Some proteins, such as the *Arabidopsis thaliana* thylakoid lumen protein *AtTLP18.3*, can be over-expressed in an *Escherichia coli* system and high-quality protein crystals can be obtained. However, *AtTLP18.3* contains no methionine residues, and site-directed mutagenesis was required in order to introduce methionine residues into the protein. A criterion for the mutated residues is that they should avoid affecting the structure and function. In this study, several leucine and isoleucine residues were selected for methionine substitution by combining secondary-structure and solvent-accessibility predictions. From the secondary-structure prediction, mutated residues were first determined in the coil or loop regions at the junction of two secondary structures. Since leucine and isoleucine residues are hydrophobic and are normally buried within the protein core, these residues should have a higher solvent-accessibility prediction so that they would be partially buried or exposed in the protein. In addition, five residues (Leu107, Leu202, Ile133, Leu128 and Ile159) of *AtTLP18.3* were mutated to methionine residues. After overexpression and purification, only two single-mutant lines, L128M and I159M, could be crystallized. Finally, a double-mutation line of truncated *AtTLP18.3* with L128M and I159M mutations was constructed. The structure of the double mutant *AtTLP18.3* protein was resolved using the single-wavelength anomalous diffraction method at 2.6 Å resolution. The results indicated that a combination of secondary-structure and solvent-accessibility prediction for methionine substitution is a useful method in SAD and MAD phasing.

1. Introduction

The structures of biological macromolecules often provide a powerful insight into their molecular function and biological action. X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are widely used to determine protein structures. Recently, X-ray crystallography has been rapidly developing. Using a miniscule protein crystal, high-energy synchrotron radiation allows the collection of diffraction data. To determine a novel protein structure using X-ray crystallographic analysis, the anomalous diffraction methods multi-wavelength anomalous diffraction (MAD) and single-wavelength anomalous diffraction (SAD) have been used to solve the phase problem. Heavy atoms are required as anomalous diffraction centres, and selenium has been demonstrated to be very effective for phasing. Moreover, recombinant selenomethionyl protein can be overexpressed in *Escherichia coli* or other expression systems, and thus may provide a broad solution for anomalous diffraction. Usually, the phasing power of the heavy-atom markers is the major problem; for every 100 amino acids, at least one selenomethionine is needed for successful phasing. Some proteins can be used to obtain protein crystals in optimal crystal-growth conditions after protein purification, but contain no methionine residues. In order to solve the crystal structure using the SAD or MAD phasing method, the introduction of methionine residues into proteins that lack methionine is important.

AtTLP18.3 is an auxiliary protein that exists in the thylakoid lumen of the *Arabidopsis* chloroplast (Sirpio *et al.*, 2007; Mulo *et al.*, 2008), and contains a putative chloroplast transit peptide, a central domain of unknown function (DUF477; amino acids 83–235) and a C-terminal transmembrane region (amino acids 260–282). In order to reveal the molecular structure of the DUF477 domain, a truncated form of *AtTLP18.3* with only the DUF477 domain was constructed. After crystallization, crystals of recombinant *AtTLP18.3* protein could be grown in 2–3 d and diffraction data could be obtained using the synchrotron source at the National Synchrotron Radiation Research Center. However, there are no methionine residues in the recombinant *AtTLP18.3* protein. The truncated *AtTLP18.3* protein, which has 177 amino acids, requires at least two methionines for anomalous dispersion phasing. In this article, we describe how we chose the sites to introduce two methionine residues. Methods combining prediction of secondary structure and protein solvent-accessible surface area for methionine substitution were performed for efficient anomalous dispersion phasing.

2. Materials and methods

2.1. Bioinformatics analysis

The point mutations were selected based on the secondary structure and solvent accessibility (SA). Secondary-structure prediction was performed with the *PSIPRED* software (<http://bioinf.cs.ucl.ac.uk/psipred/>; McGuffin *et al.*, 2000). Solvent accessibility (SA) is the ratio of the solvent-accessible surface area of a residue *X* in the native state to that of the residue in an extended tripeptide (Ala-*X*-Ala or Gly-*X*-Gly; Gromiha *et al.*, 1999). The solvent-accessible surface areas of specific residues were predicted with the *Protein Solvent Accessible Surface Area Predictor* (ASAP; <http://ccb.imb.uq.edu.au/ASAP/>; Yuan *et al.*, 2006) and *SABLE* (<http://sable.cchmc.org>). The solvent accessibility of each residue of the solved protein structure was calculated using the web-based program *GETAREA* (<http://curie.utmb.edu/getarea.html>) and *ENDscript*. Protein atomic coordinates must be supplied in PDB format and uploaded to the site for calculation (Fraczkiewicz & Braun, 1998). The *jFATCAT* (rigid) method was used to compare the structures. The *jFATCAT* structure-alignment server is available at the RCSB PDB website (<http://source.rcsb.org/jfatcatserver/>; Prlic *et al.*, 2010).

2.2. Cloning and site-directed mutagenesis

The wild-type *AtTLP18.3* gene was amplified from *Arabidopsis* cDNA (*At1g54780*) by polymerase chain reaction (PCR) and cloned into a pGEX-6P-1 expression vector (GE Healthcare) using *EcoRI* and *XhoI* restriction sites. Site-directed mutagenesis was performed to introduce methionines into *AtTLP18.3*. Mutations were generated by whole-plasmid synthesis PCR (Weiner *et al.*, 1994) using the designated primers listed in Table 1 with the pGEX6P1-*AtTLP18.3* plasmid as a template. The conditions of PCR for site-directed mutagenesis consisted of a 5 min denaturation at 95°C followed by 16 cycles of 30 s at 95°C, 60 s at 55°C and 5.5 min at 72°C. After amplification, the reactions were digested with *DpnI* for 1 h at 37°C to remove template plasmid DNA. All constructs were verified by DNA sequencing. The wild-type and mutants of *AtTLP18.3* were overexpressed in *Escherichia coli* strain BL21 (DE3).

2.3. Protein expression and purification

The wild-type protein was grown at 37°C in Luria–Bertani medium containing 100 µg ml^{−1} ampicillin to an OD₆₀₀ of around 0.6, and

Table 1

Primers used for cloning and site-directed mutagenesis.

		Substitution	Primer
<i>AtTLP-EcoRI</i>			5'-CGACGCGAATTCGCCTCTGAGTTCAATATC-3'
<i>AtTLP-XhoI</i>			5'-CGATGCGCTCGAGTTAACTGAACTGCTCTCGCT-T-3'
M01F	L107M		5'-CGATGACGCTGGTGTATGAGTCGAGTGACG-AAGTC-3'
M01R	L107M		5'-GACTTCGTCACCTCGACTCATAACACCAGCGTC-ATCG-3'
M04F	L128M		5'-GATCTTGAATACAGAAAGAAAATGAGACTCA-ATTTCATCACTG-3'
M04R	L128M		5'-CAGTGATGAAATTGAGTCTCATTCTTCTTCTGT-ATTCAAGATC-3'
M03F	I133M		5'-GAGACTCAATTTTCATGACTGTCCGAAAGCTC-3'
M03R	I133M		5'-GAGCTTTTCGACAGTCATGAAATTGAGTCTC-3'
M05F	I159M		5'-GAAAAATGGTATCCTTCTATGGAAGAAGGTA-ACAATAAGG-3'
M05R	I159M		5'-CCTATTGTTACCTTCTTCCATAGAAGGATACC-ATTTTTC-3'
M02F	L202M		5'-CTACCGTCTCCGAAAATATGCCCGTACTAGCA-ACGG-3'
M02R	L202M		5'-CCGTTGCTAGTACGGGCATATTTTCGGAGACG-GTAG-3'

expression was induced by adding 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37°C for 4 h. The protein containing the double mutation was grown in M9 medium containing 100 µg ml^{−1} ampicillin. When the OD₆₀₀ reached 0.6–0.8, selenomethionine at 100 µg ml^{−1} was supplied into the M9 medium, which was incubated at 28°C for a further 1 h. Finally, the recombinant protein was induced with 1 mM IPTG at 28°C for 12–16 h. The cells were harvested by centrifugation at 6000g for 15 min. The cell pellet was resuspended in binding buffer (50 mM Tris–HCl pH 7.0, 150 mM NaCl) and lysed using a sonicator (Misonix model XL 3000). After centrifugation at 12 000g for 30 min at 4°C, the glutathione-S-transferase (GST) fusion protein was purified using a GSTrap FF column on an ÄKTAprime FPLC system (GE Healthcare). The GST tags were then cleaved by PreScission Protease (GE Healthcare). The purified protein was concentrated and dialyzed into a buffer containing phosphate-buffered saline (PBS). The selenomethionine-derivatized protein was purified using the same protocol as for the wild-type protein, except for the addition of 1 mM tris(2-carboxyethyl)phosphine.

2.4. Crystallization and data collection

Initial crystallization conditions were screened using the Crystal Screen and Crystal Screen 2 reagent kits (Hampton Research). The crystallization experiments were conducted using the hanging-drop vapour-diffusion method at 23°C, with 2 µl drops consisting of 1 µl 10 mg ml^{−1} protein and 1 µl reservoir solution suspended over 0.5 ml reservoir solution. After 3 d, crystals were observed in three different conditions: Crystal Screen conditions No. 10 [30% (w/v) polyethylene glycol 4000, 0.2 M ammonium acetate, 0.1 M trisodium citrate dihydrate pH 5.6], No. 22 [30% (w/v) polyethylene glycol 8000, 0.2 M ammonium sulfate, 0.1 M sodium cacodylate pH 6.5] and No. 28 [18% (w/v) polyethylene glycol 8000, 0.2 M calcium acetate hydrate, 0.1 M sodium cacodylate pH 6.5]. These conditions were optimized by varying the precipitant (PEG 4000–8000), the pH (4.0–9.0) and the protein concentration. Finally, purified native *AtTLP18.3* protein at 7.5 mg ml^{−1} in 1× PBS buffer was used for crystallization at 23°C. Most crystals of wild-type *AtTLP18.3* protein were grown in 0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate pH 6.5, 25–30% (w/v) polyethylene glycol 4000 within two weeks. Crystals of SeMet-containing protein were grown under similar conditions with a

Table 2
Crystallization.

	<i>A/TLP18.3</i>	<i>A/TLP18.3</i> L128M/I159M
Method	Hanging-drop vapour diffusion	Hanging-drop vapour diffusion
Plate type	24-well plates	24-well plates
Temperature (K)	293	293
Protein concentration (mg ml ⁻¹)	7.5	7.5
Buffer composition of protein solution	1× PBS buffer	1× PBS buffer, 2 mM tris(2-carboxyethyl)phosphine
Composition of reservoir solution	0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate pH 6.5, 25–30%(w/v) polyethylene glycol 4000	0.2 M sodium acetate trihydrate, 0.1 M sodium cacodylate pH 6.0, 25–30%(w/v) polyethylene glycol 4000, 2 mM tris(2-carboxyethyl)phosphine
Volume and ratio of drop	10 µl (1:1)	10 µl (1:1)
Volume of reservoir (ml)	0.5	0.5

slightly lower pH (6.0) and with the addition of 2 mM tris-(2-carboxyethyl)phosphine to all solutions (see Table 2).

Crystals were rapidly swept through mother liquor containing 25%(v/v) glycerol as a cryoprotectant and flash-cooled in liquid

nitrogen at 100 K. Diffraction data were collected using synchrotron radiation on the SPXF beamline BL13B1 equipped with CCD detectors (Q315 and Q210; Area Detector Systems Corporation) at the National Synchrotron Radiation Research Center in Taiwan. All diffraction intensities were integrated and scaled using the *HKL*-2000 software package (Otwinowski & Minor, 1997). The statistics of the

Figure 1
SDS-PAGE analysis of purified *A/TLP18.3* protein. Lane 1, low-molecular-weight protein marker (labelled in kDa). Lane 2, purified wild-type *A/TLP18.3* protein. Lane 3, purified mutant *A/TLP18.3*.

Figure 2
Crystal of wild-type *A/TLP18.3* protein.

(a)

(b)

Figure 3
(a) Crystal and (b) diffraction pattern of mutant *A/TLP18.3*.

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

	<i>At</i> TLP18.3	<i>At</i> TLP18.3 L128M/I159M
Beamline at NSRRC	13B1	13B1
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 46.87$, $b = 49.89$, $c = 76.51$	$a = 46.98$, $b = 49.84$, $c = 76.67$
Wavelength (Å)	0.97315	0.9788
Resolution range (Å)	30.0–1.6	30.0–2.6
Observed reflections	343890	57179
Unique reflections	24243	5797
Completeness (%)	99.3 (100.0)	97.8 (99.6)
$\langle I/\sigma(I) \rangle$	58.6 (8.7)	30.8 (38.4)
$R_{\text{merge}}^{\dagger}$ (%)	3.5 (31.6)	4.6 (6.2)

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the i th measurement and $\langle I(hkl) \rangle$ is the weighted average of all measured reflections.

preliminary crystallographic data for the wild-type and selenomethionine-substituted *At*TLP18.3 crystals are shown in Table 3.

2.5. Phosphatase activity assay

p-Nitrophenyl phosphate (*p*NPP; New England Biolabs) was used at a series of concentrations to determine the phosphatase activity of *At*TLP18.3. The incubated solution consisted of 15 µg purified *At*TLP18.3, 50 mM sodium acetate buffer pH 4.0 and the substrate to give a total volume of 0.1 ml. Incubation was performed at 37°C for 30 min. The reaction was terminated by the addition of 0.25 ml 500 mM NaOH. The released *p*-nitrophenol was measured at 405 nm using a GeneQuant 1300 spectrophotometer (GE Healthcare). The enzyme activity was calculated from the Beer–Lambert law as follows,

$$\text{enzyme activity} = \frac{V (\mu\text{l}) \times \text{OD}_{405\text{ nm}} (\text{cm}^{-1})}{\varepsilon \times \text{incubation time (min)} \times \text{enzyme } (\mu\text{g})}$$

and enzyme turnover number = enzyme activity × enzyme molecular weight (Da), where ε is the molar extinction coefficient ($\text{M}^{-1} \text{cm}^{-1}$).

3. Results and discussion

3.1. Residue selection for selenomethionine substitution

In this study, crystals of selenomethionine-substituted protein were used to collect data for phasing. After removing the GST tag, the amino-acid sequence of the truncated *At*TLP18.3 protein consists of 177 residues; however, it does not contain any methionine residues. Substitution of leucine or isoleucine by methionine can be performed without a significant conformational change. Therefore, these two residues are suitable substitution candidates for methionine (Bordo & Argos, 1991; Ohmura *et al.*, 2001). In the truncated *At*TLP18.3 protein sequence, there are 15 leucine residues and nine isoleucine residues. Initially, we randomly chose two leucines at positions 107 (Leu107) and 202 (Leu202), and one isoleucine at position 133 (Ile133) for methionine substitution. The protein yields of these three mutants were as high as that of wild-type *At*TLP18.3, but none of them could be crystallized.

We then selected other residues according to secondary-structure and solvent-accessibility (SA) predictions. Solvent accessibility shows different distribution ranges of protein residues, such as completely buried (0–20%), partially buried (20–50%) and exposed (>50%). For partially buried mutations, information concerning the selected residues is very important in predicting changes in stability; however, this is not the case for buried mutations. The stability changes caused by partially buried mutations in all secondary structures were mainly the result of hydrophobicity (Gromiha *et al.*, 1999). Based on the

Figure 4

Comparison of the predicted and observed secondary structure and solvent accessibility of the *At*TLP18.3 protein. The predicted secondary structure (line 1) was predicted using *PSIPRED*: pink barrels represent α -helices, while yellow arrows signify β -sheets. All mutations are shown in red circles. The secondary structure (line 2) and solvent accessibility (line 3) of the solved structure were determined using *ENDscript*. For the secondary structure, α -helices and β -strands are shown as cylinders and arrows, respectively. The solvent accessibility of the solved structure (line 3) is shown in different colours, with white indicating buried residues (<0.1), light blue indicating intermediate (0.1–0.4), blue indicating accessible (>0.4) and blue with red borders indicating highly exposed (>1). The predicted area exposed to the solvent (line 4) is represented by the relative solvent accessibility expressed as a percentage (black to white).

secondary-structure prediction, we thus chose residues that were located at the junction of a coil region and another secondary structure (α -helix or β -sheet). We suggested that when the residues at the junction position were substituted by other amino acids, the overall structure would undergo no significant change. Therefore, we mutated two other residues, leucine at position 128 (L128M, SA 30.6%, partially buried, at the junction of the coil region and the β -sheet) and isoleucine at position 159 (I159M, SA 49.4%, partially buried, at the junction of the coil region and the α -helix). In addition to the wild-type protein, these two point-mutant proteins, L128M and

I159M, can be expressed and crystallized (see Figs. 1, 2 and 3). In contrast, the predictions of the prior three mutations are as follows: L107M (SA 8%, protein interior, at the junction of the coil region and the α -helix), L202M (SA 0.3%, protein interior, at the junction of the coil region and the α -helix) and I133M (SA 0%, protein interior, in the β -sheet). These three residues were predicted to be completely buried, and the crystallization results showed that partially buried mutations were a better choice for methionine substitution than fully buried mutations. (All mutations are shown in Figs. 4 and 5.) Finally, we solved the *Ar*TLP18.3 structure by the SAD method using the

Figure 5
(a) Ribbon drawing showing the locations of the five mutated residues: L107M, L202M, I133M, L128M and I159M. The view from a 90° rotation shows that two residues, Leu128 and Ile159, are on the surface protein. (b) Different anomalous Fourier map contoured at 1.0σ (blue) and 4.5σ (red). (c) Structural superposition of wild-type and mutant *Ar*TLP18.3 protein. The proteins are shown as ribbons, with the wild-type protein in red and the double mutant in yellow. Generated from PDB entries 3pvh and 3ptj.

Table 4

Steady-state kinetics data for the wild type and variants.

All assays were performed and analyzed as described in the text. The kinetic parameters listed are the average of three separate measurements. In the calculation of the relative k_{cat}/K_M ratio, the k_{cat}/K_M of the wild type is taken as 1 in each case.

Variant	K_M (mM)	k_{cat} (s ⁻¹)	k_{cat}/K_M (mM ⁻¹ s ⁻¹)	Relative ratio	Reference
Wild type	42.56 ± 7.41	9.48 ± 0.85	4.49	1	Wu <i>et al.</i> (2011)
L107M	43.15 ± 8.60	9.09 ± 0.94	4.74	0.946065	This study
L202M	42.51 ± 7.40	9.09 ± 0.82	4.67	0.960561	This study
I133M	47.29 ± 10.13	8.84 ± 0.79	5.35	0.838852	This study
L128M	43.93 ± 9.32	9.04 ± 1.00	4.86	0.923088	This study
I159M	42.47 ± 7.40	9.41 ± 1.08	4.51	0.9944	This study
L128M/I159M	43.51 ± 6.99	9.29 ± 0.77	4.67	0.9593	This study

double mutant L128M and I159M (PDB entries 3pvh and 3ptj); as a result, we showed that the *A*/TLP18.3 protein is an acid phosphatase (Wu *et al.*, 2011).

3.2. Comparison of mutant and wild-type proteins

Using the *jFATCAT* structure-alignment server found on the RCSB PDB website, we also performed pairwise structure alignments between the double-mutation and wild-type crystal structures. The *jFATCAT*-defined *P*-value provides a probability measurement of two structures being identified as similar. A smaller *P*-value indicates that the two corresponding structures have a more statistically significant similarity. Overall, the structure of the double mutant is very similar to that of the wild type, with a *P*-value of 0.00 and a C^α root-mean-square deviation (r.m.s.d.) of 0.34 Å.

Furthermore, although not all of the mutants could be crystallized under similar conditions as wild-type *A*/TLP18.3, we analyzed these mutants to determine whether they had an effect on the phosphatase activity of the protein. Table 4 compares the kinetics parameters of the *p*NPP phosphatase activity of the wild type and all other mutants. The k_{cat}/K_M for all mutants are only slightly different from that of wild-type *A*/TLP18.3. These results show that the mutation of leucine or isoleucine to methionine did not affect the overall structure and the phosphatase activity of the *A*/TLP18.3 protein.

4. Conclusion

In order to use selenomethionine-substituted proteins for the anomalous dispersion phasing method, we summarize a strategy for introducing methionine residues into protein sequences without methionine. The strategy is as follows. (i) Detect all of the leucine and isoleucine residues in the protein sequence. (ii) Choose the leucine or isoleucine residues that are partially buried or exposed according to the solvent-accessibility prediction. All residues are then classified into three major types: completely buried (SA 0–20%), partially buried (SA 20–50%) and exposed (SA > 50%). After prediction, we found that more than half of the solvent accessibilities of leucines and isoleucines are below 10%; this may be because leucine and isoleucine residues are hydrophobic and are normally buried inside the protein core. We compared the the predicted solvent accessibility (from the sequence) and that of the solved structure (from the PDB file), and the results usually showed a higher predicted accessibility

Table 5

Comparison of secondary structure and solvent accessibility of the wild type and variants.

The secondary structure where the mutated residues are located is shown in bold type. The predicted data are shown in parentheses.

Residue	Secondary structure	Solvent accessibility (%)	Location	Crystallization	Phosphatase activity
Leu107	Coil-α-helix (coil- α -helix)	0 (8)	Interior (interior)	—	+
Leu202	α -Helix (α -helix-coil)	0 (0.3)	Interior (interior)	—	+
Ile133	β-Sheet (β -sheet)	0 (0)	Interior (interior)	—	+
Leu128	Coil-β-sheet (coil- β -sheet)	8.2 (30.6)	Interior (partially buried)	+	+
Ile159	Coil-α-helix (α -helix-coil)	77.3 (49.4)	Exposed (partially buried)	+	+

than was found in the solved structure (see Table 5). In order to choose suitable leucine or isoleucine residues, we suggest that those with a higher predicted solvent accessibility are better choices. (iii) Use the secondary-structure prediction and choose a residue at the junction of a coil region with another secondary structure. Each step can be completed on the bioinformatics tool website. We successfully combined the secondary-structure and protein solvent-accessible surface area predictions for methionine substitution to use the SAD phasing method, and solved the structure of the *A*/TLP18.3 protein. This process can also be used to solve many other protein crystal structures that lack methionine residues.

H-YW acknowledges Ming-Hao Liu for proofreading the manuscript. This work was supported by the National Science Council, Taiwan (grant No. NSC.98-2313-B-002-059-MY2 to Y-SC). Part of this research was carried out at the National Synchrotron Radiation Research Center, a national user facility supported by the National Science Council of Taiwan. The Synchrotron Radiation Protein Crystallography Facility is supported by the National Research Program for Genomic Medicine.

References

- Bordo, D. & Argos, P. (1991). *J. Mol. Biol.* **217**, 721–729.
- Fraczkiewicz, R. & Braun, W. (1998). *J. Comput. Chem.* **19**, 319–333.
- Gromiha, M. M., Oobatake, M., Kono, H., Uedaira, H. & Sarai, A. (1999). *Protein Eng.* **12**, 549–555.
- McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). *Bioinformatics*, **16**, 404–405.
- Mulo, P., Sirpiö, S., Suorsa, M. & Aro, E. M. (2008). *Photosynth. Res.* **98**, 489–501.
- Ohmura, T., Ueda, T., Hashimoto, Y. & Imoto, T. (2001). *Protein Eng.* **14**, 421–425.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Prlic, A., Bliven, S., Rose, P. W., Bluhm, W. F., Bizon, C., Godzik, A. & Bourne, P. E. (2010). *Bioinformatics*, **26**, 2983–2985.
- Sirpio, S., Allahverdiyeva, Y., Suorsa, M., Paakkarinen, V., Vainonen, J., Battchikova, N. & Aro, E. (2007). *Photosynth. Res.* **91**, 208.
- Weiner, M. P., Costa, G. L., Schoettlin, W., Cline, J., Mathur, E. & Bauer, J. C. (1994). *Gene*, **151**, 119–123.
- Wu, H.-Y., Liu, M.-S., Lin, T.-P. & Cheng, Y.-S. (2011). *Plant Physiol.* **157**, 1015–1025.
- Yuan, Z., Zhang, F., Davis, M. J., Bodén, M. & Teasdale, R. D. (2006). *J. Proteome Res.* **5**, 1063–1070.