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

Crystallization and preliminary crystallographic studies of Hyp-1, a St John's wort protein implicated in the biosynthesis of hypericin

Humberto Fernandes,[a] Malgorzata Konieczna,[a] Robert Kolodziejczyk,[b] Grzegorz Bujacz,[a,c] Michal Sikorski[a] and Mariusz Jaskolski[a,b]*

[a]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, [b]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland, and [c]Institute of Technical Biochemistry, Technical University of Lodz, Poland

Correspondence e-mail: mariuszj@amu.edu.pl

Received 20 March 2008
Accepted 3 April 2008

According to a debated hypothesis, the biosynthesis from emodin of the medicinally important natural compound hypericin is catalyzed in St John's wort (*Hypericum perforatum*) by the phenolic oxidative-coupling protein Hyp-1. Recombinant St John's wort Hyp-1 has been overexpressed in *Escherichia coli* and obtained in single-crystal form. The crystals belong to the orthorhombic system, space group $P2_12_12_1$, with unit-cell parameters a = 37.5, b = 76.7, c = 119.8 Å, contain two protein molecules in the asymmetric unit and diffract X-rays to 1.73 Å resolution.

1. Introduction

St John's wort (*Hypericum perforatum* L.) is a herbaceous perennial plant that has received considerable interest in North America and Europe owing to its medicinal properties. In addition to its application in folk medicine for various ailments, which dates back thousands of years, several recent clinical studies have demonstrated the effectiveness of St John's wort as a phytomedicinal treatment for depression (Briskin & Gawienowski, 2001). Morphologically, St John's wort is characterized by the presence of several types of secretory structures, including translucent glands, secretory canals and the most important structure in this plant, dark glands, which do not have a cavity and are mainly visible on leaves and flowers (Onelli *et al.*, 2002; Zobayed *et al.*, 2006). The occurrence of these metabolically active multicellular structures is unique and characteristic of some species of the genus *Hypericum* (Kosuth *et al.*, 2006).

In addition to diverse physiological functions during the life cycle of the plant, secondary metabolites in medicinal plants also have an important role as a source of active pharmaceuticals (Kosuth *et al.*, 2006). The main bioactive compounds of *H. perforatum* belong to the hypericin group (hypericin and pseudohypericin; Fig. 1) and are thought to function as defensive allelochemicals in the plant (Zobayed *et al.*, 2006). Hypericin, a photosensitive red-coloured naphthodianthrone, is the major medicinal compound of *H. perforatum*. It has been successfully tested in numerous studies for the treatment of mild to moderate depression (Deltito & Beyer, 1998), as well as in antitumour and antiviral (HIV, hepatitis C virus) tests (Kubin *et al.*, 2005 and references therein). Screening studies for inhibitory effects of hypericin on various pharmaceutically important enzymes, such as MAO (monoaminoxidase), PKC (protein kinase C), dopamine-β-hydroxylase, reverse transcriptase, telomerase and CYP (cytochrome P450), have yielded results supporting the therapeutic potential of hypericin. Studies of the effects of hypericin on GABA-activated (γ-aminobutyric acid) currents and NMDA (*N*-methyl-D-aspartate) receptors also indicate therapeutic promise in stroke (Kubin *et al.*, 2005).

Bais *et al.* (2003) hypothesized that in *H. perforatum* hypericin is synthesized from the emodin precursor (Fig. 1*a*) in an enzymatic dimerization reaction. Emodin is an anthraquinone which has been suggested to act as a feeding deterrent, to be an allelopathic compound, to have antimicrobial activity and to participate in seed germination and dispersal (Izhaki, 2002).

The final localization of hypericin in the dark glands is well accepted (Briskin & Gawienowski, 2001), but the site of its synthesis is unknown, with the dark glands being a possibility (Onelli *et al.*, 2002) that is supported by the presence of emodin at a high concentration in the dark glands and its absence in other tissues (Zobayed *et al.*, 2006).

Bais and coworkers were able to isolate a gene, *hyp-1*, the product of which seems to be responsible for the direct conversion of emodin to hypericin in dark-grown cell cultures of *H. perforatum*. The cDNA of *hyp-1* has 725 bp with an open reading frame of 480 nucleotides (including the stop codon) and codes for a protein of 159 amino-acid residues with a molecular weight of 17.8 kDa and a pI of 5.54. Following the mRNA encoding the Hyp-1 protein, Kosuth *et al.* (2006) expected to find elevated transcription levels in the dark glands. However, the results were negative; in addition, Kosuth and coworkers found the highest level of transcription of the *hyp-1* gene, but only traces of hypericin, in the roots, where there are no dark glands at all. These findings may have several explanations: (i) the sites of biosynthesis and accumulation of hypericin are different, (ii) the level of steady-state mRNA is not correlated with the level of the encoded polypeptide, (iii) the conditions of the *in vitro* experiment differ from those *in vivo* or (iv) it is questionable whether the *hyp-1* gene is involved in hypericin biosynthesis (Kosuth *et al.*, 2006).

A *BLAST* search using the deduced amino-acid sequence of Hyp-1 detected similarity to the Bet v 1/PR-10 family of proteins that are found in a wide variety of plant species (Bais *et al.*, 2003). Protein-sequence comparisons revealed significant homology between Hyp-1 and a major apple allergen (Mal d 1), with 45.3% sequence identity and 57.2% sequence similarity (Fig. 2), followed by Pru av 1, a major cherry allergen (37.5% sequence identity and 50.6% sequence similarity). PR-10, or class 10 pathogenesis-related proteins, are only found in plants (Colditz *et al.*, 2007; Handschuh *et al.*, 2007) and include potent pollen and fruit allergens such as Bet v 1 (birch), Mal d 1 and Cor a 1.04 (hazel). Although they are ubiquitous and typically encoded by multigene families, their true physiological role is unknown. One theory, invalidated by a number of negative examples, viewed PR-10 proteins as ribonucleases (Moiseyev *et al.*, 1994) involved in plant defence responses. A recent hypothesis postulates a hydrophobic ligand-binding role for the PR-10 proteins. It is based on the structural similarity between PR-10 proteins and cytokinin-specific binding proteins (CSBPs; Pasternak *et al.*, 2006) and on structural reports on ligand complexes of PR-10 proteins (Fernandes *et al.*, 2008; Markovic Housley *et al.*, 2003). Apart from Bet v 1 (Gajhede *et al.*, 1996; Schweimer *et al.*, 1999; Mirza *et al.*, 2000; Markovic Housley *et al.*, 2003; Spangfort *et al.*, 2003; Holm *et al.*, 2004), the best structurally characterized PR-10 members are a group of yellow lupin homologues (Biesiadka *et al.*, 2002; Pasternak *et al.*, 2005; Fernandes *et al.*, 2008). These structures have established that the classic PR-10 fold consists of a seven-stranded antiparallel β-sheet wrapped around a long C-terminal α-helix. These two main structural elements, together with two accessory α-helices, enclose a large hydrophobic cavity with a volume of up to 4500 Å^3.

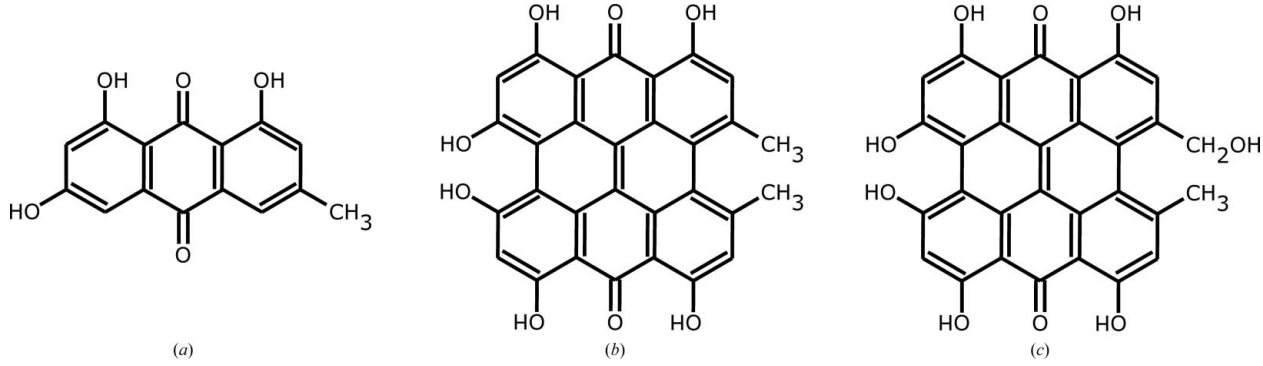

Figure 1
Structures of (*a*) emodin, (*b*) hypericin and (*c*) pseudohypericin.

```
HpHyp-1        MAAYTIVKEEESPTAPHRLFKALVLERHQVLVKAQPHVFKSGEIIEGDGGVGTVTKITFV 60
Mal d 1        MGVYTFENEYTSEIPPPRLFKAFVLDADNLIPKIAPQAIKHAEILEGDGGPGTIKKITFG 60
Bet v 1        MGVFNYETETTSVIPAARLFKAFILDGDNLFPKVAPQAISSVENIEGNGGPGTIKKISFP 60
LlPR10-2B      MGVFTFQDEYTSTIAPAKLYKALVTDADIIIPK-AVETIQSVEIVEGNGGPGTIKKLTFI 59
VrCSBP         M-VKEFNTQTELSVRLEALWAVLSKDFITVVPKVLPHIVKDVQLIEGDGGVGTILIFNFL 59

HpHyp-1        DGHPLTYMLHKFDEIDAANFYCKYTIFEGDVLRDNIEKVVYEVKLEAV-GGGSKGKITV 119
Mal d 1        EGSQYGYVKHKIDSVDEANYSYAYTLIEGDALTDTIEKVSYETKLVAS-GSGSIIKSIS 119
Bet v 1        EGFPFKYVKDRVDEVDHTNFKYNYSVIEGGPIGDTLEKISNEIKIVATPDGGSILKISN 119
LlPR-10.2B     EGGESKYVLHKIEAIDEANLGYNYSIVGGVGLPDTIEKISFETKLVEGANGGSIGKVTI 118
VrCSBP         PEVSPSYQREEITEFDESSHEIGLQVIEGGYLNQGLSYYKTTFKLSEIEEDKTLVNVKI 118

HpHyp-1        SYHPKPGCTVNEEEVKIGEKKAYEFYKQVEEYLAANPEVFA 159
Mal d 1        HYHTKGDVEIKEEHVKAGKEKAHGLFKLIESYLKGHPDAYN 159    (45.3 / 57.2%)
Bet v 1        KYHTKGDHEVKAEQVKASKEMGETLLRAVESYLLAHSDAYN 160    (35.6 / 51.9%)
LlPR-10.2B     KIETKGDAQPNEEEGKAAKARGDAFFKAIESYLSAHPD-YN 158    (39.4 / 56.9%)
VrCSBP         SYDHDSDIEEKVTPTKTSQS-TLMYLRRLERYLSNGS-A-- 155    (20.6 / 36.2%)
```

Figure 2
Sequence alignment of *H. perforatum* Hyp-1 (GenBank accession No. AAN65449; Benson *et al.*, 2007), Mal d 1 (apple; AAC26136), Bet v 1 (birch; P15494), LlPR-10.2B (lupin; AAF77634) and VrCSBP (mung bean; BAA74451) generated using *ClustalW* (Chenna *et al.*, 2003). Conserved amino acids are highlighted. At the end of the aligned sequences, the degree of identity/similarity to the *H. perforatum* Hyp-1 sequence is shown in parentheses. Note the high degree of conservation in a glycine-rich region extending between residues 45 and 54 in Hyp-1, which corresponds to a structurally invariant loop in PR-10 proteins.

The main objectives of the present project are to confirm the structural classification of *H. perforatum* Hyp-1 in the PR-10 folding class and to establish the structural determinants of its interactions with emodin and hypericin, with the ultimate goal being the elucidation of its role in hypericin biosynthesis. As the first step, we have established reproducible crystallization conditions for Hyp-1 in ligand-free form and demonstrate that the crystals diffract X-rays to at least 1.73 Å resolution.

2. Material and methods

2.1. Cloning, overexpression and purification

RNA was purified from freshly ground *H. perforatum* leaves using a commercially available kit (Qiagen) and following the recommendations of the supplier. cDNA was synthesized for 1 h at 315 K using 200 units of SuperScript II Reverse Transcriptase (Invitrogen), 2 pmol gene-specific primer and ∼5 mg total RNA in 20 µl PCR buffer. Standard PCR was performed immediately after first-strand cDNA synthesis using the following primers: promoter proximal primer 5′-CACCATGGCGGCGTACACTAT and promoter distal primer 5′-TTAAGCGAAAACTTCAGGATTACG. The PCR product was cloned into pGEM T-Easy Vector (Promega) and positive candidate clones were sequenced. The *hyp-1* coding sequence of 477 bp was PCR-amplified from the pGEM vector and introduced into a pET151/D vector (Invitrogen) for expression. Recombinant *H. perforatum* Hyp-1 with an N-terminal His-tag fusion was produced in *Escherichia coli* strain BL21star(DE3) cells (Invitrogen) using the T7 promoter/T7 RNA polymerase system (Studier *et al.*, 1990). 2 ml of an overnight culture (30 ml) grown at 310 K in LB media in the presence of 100 µg ml^{-1} ampicillin was used as an inoculum for the expression culture (250 ml). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 m*M* at an OD_{600} of ∼1. After 5 h growth at 303 K, the cells were harvested by centrifugation at 4300*g* and 277 K for 10 min and the pellet was frozen at 253 K. Cells were resuspended in lysis buffer (20 m*M* Tris–HCl pH 7.5, 5% glycerol) and disrupted by sonication (Ultrasonic Processor, Model CV33, Fisher; 5 × 20 s bursts at 80% power). Cell debris was pelleted by centrifugation at 15 555*g* and 277 K for 1 h. The soluble fraction was passed through a DE52 cellulose column equilibrated with 20 m*M* sodium phosphate buffer pH 7.5 with the addition of 5% glycerol. Fractionation was carried out by stepwise elution with NaCl. The fractions eluted with 100 and 200 m*M* NaCl were further purified by affinity chromatography using IMAC-Sepharose (GE Healthcare). Fractionation was carried out by stepwise elution with imidazole. The fractions eluted with 50 and 100 m*M* imidazole were pooled and the His tag was cleaved during overnight digestion at 277 K using His-tagged TEV protease in 10 m*M* HEPES buffer pH 7.5. The TEV protease, the cleaved His tag and any unprocessed material were removed by a second run on the IMAC-Sepharose column. The recombinant protein has a hexapeptide N-terminal extension (GIDPFW–) as a cloning artifact. The final yield of recombinant Hyp-1 protein after purification was ∼30 mg per litre of liquid culture.

2.2. Crystallization

Prior to crystallization, Hyp-1 was concentrated to 6 mg ml^{-1} in 3 m*M* sodium citrate buffer pH 6.3. Screening for crystallization conditions was performed in 24-well Linbro plates using hanging-drop vapour diffusion against 0.5 ml well solution at 292 K by mixing 1.5 µl protein solution with 1.5 µl well solution. Crystal Screen I and II and PEG/Ion Screen solutions from Hampton Research were used to find initial crystallization conditions, which were then optimized using Additive Screen, also from Hampton Research. Small crystals appeared after 2 d in a drop where the reservoir consisted of 0.2 *M* $MgCl_2$, 0.1 *M* Tris–HCl pH 8.5 and 30%(*w*/*v*) PEG 4000. Crystal size could be improved by using 0.4 *M* NaCl instead of 0.2 *M* $MgCl_2$. The crystals reached their final dimensions of up to 0.10 × 0.02 × 0.02 mm in about two weeks (Fig. 3).

2.3. X-ray data collection

Crystals were briefly soaked in a cryoprotectant solution obtained by mixing the well solution with 50%(*v*/*v*) PEG 400 in a 1:1 ratio and were then flash-cooled at 100 K in a stream of cold N_2 gas (Teng,

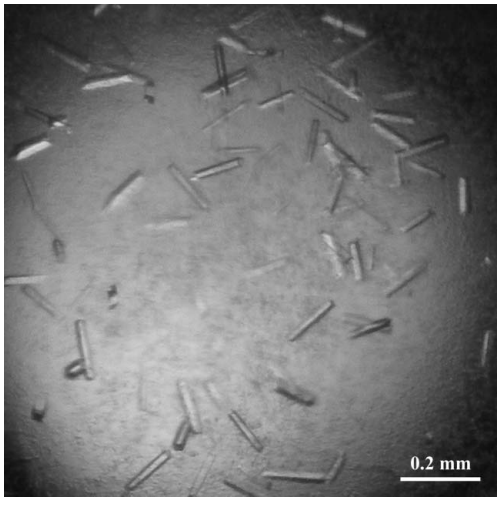

Figure 3
Single crystals of *H. perforatum* Hyp-1 protein. The average dimensions of the crystals are 0.10 × 0.02 × 0.02 mm.

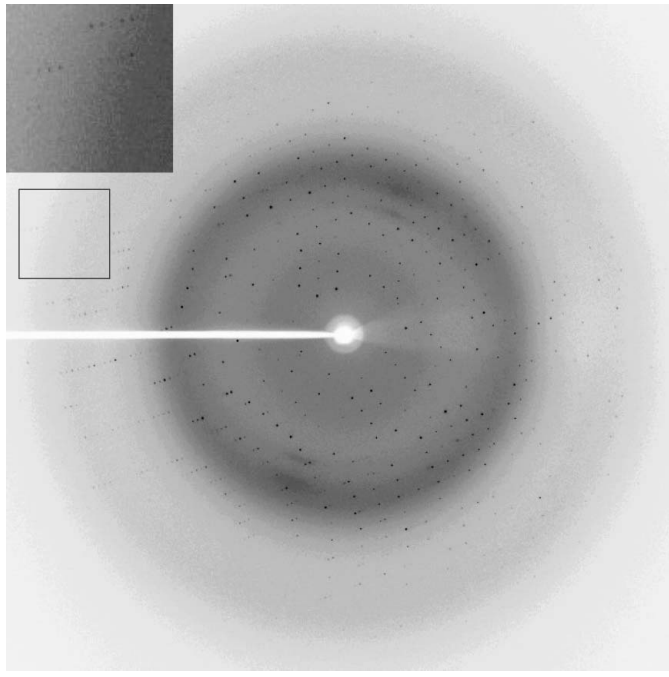

Figure 4
An X-ray diffraction pattern recorded for a single crystal of *H. perforatum* Hyp-1 protein using synchrotron radiation (0.5° oscillation). The edge of the detector (framed, inset) corresponds to a resolution of 1.73 Å.

Table 1
Summary of crystal data and data-collection statistics.

Values in parentheses are for the last resolution shell.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 37.5, b = 76.7, c = 119.8
Resolution limits (Å)	30.0–1.73 (1.79–1.73)
Radiation source	BESSY 14.1
Wavelength (Å)	0.91841
Temperature (K)	100
Mosaicity (°)	0.43
No. of measured reflections	247614
No. of unique reflections	36647
Redundancy	6.76
$R_{\rm merge}$†	0.070 (0.496)
Completeness (%)	99.1 (91.6)
$\langle I/\sigma(I)\rangle$	24.9 (2.1)

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl.

1990). X-ray diffraction data extending to 1.73 Å resolution were measured in 360 0.5° oscillation steps on the BESSY 14.1 beamline (Berlin, Germany) using a MAR 225 Mosaic CCD detector (Fig. 4). The data were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). The final statistics are reported in Table 1.

3. Results and discussion

After IMAC-Sepharose affinity purification the recombinant Hyp-1 protein was essentially pure, migrating with an apparent molecular weight of approximately 18 kDa on SDS–PAGE. Crystallization optimization experiments indicated sensitivity to the presence of metal cations and the types of PEG and pH buffer used. The crystals grew as thin needles, but diffracted X-rays to 1.73 Å resolution. They were orthorhombic, with systematic absences indicating space group $P2_12_12_1$. The unit-cell volume was most consistent with the presence of two protein molecules in the asymmetric unit, corresponding to a Matthews coefficient of 2.34 Å^3 Da^{-1} and a solvent content of 47.4% (Matthews, 1968). Size-exclusion chromatography suggested that analogously to classic PR-10 proteins, Hyp-1 was monomeric in solution. Analysis of the self-rotation function was inconclusive as no clear peaks not related to crystal symmetry could be observed. Thus, the nature of the assembly in the asymmetric unit has yet to be determined.

Attempts to solve the structure by molecular replacement using a number of programs and several search models derived from PR-10 coordinates available in the PDB (Berman *et al.*, 2000) have been unsuccessful. This failure may indicate that *H. perforatum* Hyp-1 does not have a PR-10 fold or that its structural distortion is sufficiently great to prevent recognition by vector methods. The latter possibility is quite likely as it has precedent with other PR-10-like structures (Pasternak *et al.*, 2006, 2008). The problem lies in the C-terminal α-helix of PR-10 proteins, which shows very low sequence conservation and a high degree of structural deformation. Work is in progress towards solving the *H. perforatum* Hyp-1 structure by the isomorphous replacement method.

This work was supported in part by grants from the Ministry of Science and Higher Education to MJ (No. N N204 2584 33) and to MS (Nos. 6 P04B 004 21 and 2 P04A 053 27). HF was supported by a Marie Curie Fellowship from the European Union.

References

Bais, H. P., Vepachedu, R., Lawrence, C. B., Stermitz, F. R. & Vivanco, J. M. (2003). *J. Biol. Chem.* **278**, 32413–32422.
Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. (2007). *Nucleic Acids Res.* **35**, 21–25.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Biesiadka, J., Bujacz, G., Sikorski, M. M. & Jaskolski, M. (2002). *J. Mol. Biol.* **319**, 1223–1234.
Briskin, D. P. & Gawienowski, M. C. (2001). *Plant Phys. Biochem.* **39**, 1075–1081.
Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, D. D. (2003). *Nucleic Acids Res.* **31**, 3497–3500.
Colditz, F., Niehaus, K. & Krajinski, F. (2007). *Planta*, **226**, 57–71.
Deltito, J. & Beyer, D. (1998). *J. Affect. Disord.* **51**, 345–351.
Fernandes, H., Pasternak, O., Bujacz, G., Bujacz, A., Sikorski, M. M. & Jaskolski, M. (2008). *J. Mol. Biol.* doi:10.1016/j.jmb.2008.03.027.
Gajhede, M., Osmark, P., Poulsen, F. M., Ipsen, H., Larsen, J. N., Joost van Neerven, R. J., Schou, C., Lowenstein, H. & Spangfort, M. D. (1996). *Nature Struct. Biol.* **3**, 1040–1045.
Handschuh, L., Femiak, I., Kasperska, A., Figlerowicz, M. & Sikorski, M. M. (2007). *Acta Biochim. Pol.* **54**, 783–796.
Holm, J., Gajhede, M., Ferreras, M., Henriksen, A., Ipsen, H., Larsen, J. N., Lund, L., Jacobi, H., Millner, A., Wurtzen, P. A. & Spangfort, M. D. (2004). *J. Immunol.* **173**, 5258–5267.
Izhaki, I. (2002). *New Phytol.* **155**, 205–217.
Kosuth, J., Katkovcinova, Z., Olexova, P. & Cellarova, E. (2006). *Plant Cell Rep.* **26**, 211–217.
Kubin, A., Wierrani, F., Burner, U., Alth, G. & Grunberger, W. (2005). *Curr. Pharm. Des.* **11**, 233–253.
Markovic Housley, Z., Degano, M., Lamba, D., von Roepenack-Lahaye, E., Clemens, S., Susani, M., Ferreira, F., Scheiner, O. & Breiteneder, H. (2003). *J. Mol. Biol.* **325**, 123–133.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mirza, O., Henriksen, A., Ipsen, H., Larsen, J. N., Wissenbach, M., Spangfort, M. D. & Gajhede, M. (2000). *J. Immunol.* **165**, 331–338.
Moiseyev, G. P., Beintema, J. J., Fedoreyeva, L. I. & Yakovlev, G. I. (1994). *Planta*, **193**, 470–472.
Onelli, E., Rivetta, A., Giorgi, A., Bignami, M., Cocucci, M. & Patrignani, G. (2002). *Flora*, **197**, 92–102.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pasternak, O., Biesiadka, J., Dolot, R., Handschuh, L., Bujacz, G., Sikorski, M. M. & Jaskolski, M. (2005). *Acta Cryst.* D**61**, 99–107.
Pasternak, O., Bujacz, A., Biesiadka, J., Bujacz, G., Sikorski, M. & Jaskolski, M. (2008). *Acta Cryst.* D**64**, 595–606.
Pasternak, O., Bujacz, G. D., Fujimoto, Y., Hashimoto, Y., Jelen, F., Otlewski, J., Sikorski, M. M. & Jaskolski, M. (2006). *Plant Cell*, **18**, 2622–2634.
Schweimer, K., Sticht, H., Boehm, M. & Roesch, P. (1999). *Appl. Magn. Reson.* **17**, 449–464.
Spangfort, M. D., Mirza, O., Ipsen, H., Van Neerven, R. J., Gajhede, M. & Larsen, J. N. (2003). *J. Immunol.* **171**, 3084–3090.
Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). *Methods Enzymol.* **185**, 60–89.
Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Zobayed, S. M., Afreen, F., Goto, E. & Kozai, T. (2006). *Ann. Bot.* **98**, 793–804.