Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The purification, crystallization and preliminary structural characterization of glucose-1-phosphate thymidylyltransferase (RmlA), the first enzyme of the dTDP-L-rhamnose synthesis pathway from *Pseudomonas aeruginosa*

Wulf Blankenfeldt,[a] Marie-France Giraud,[a] Gordon Leonard,[b] Rahim Rahim,[c] Carole Creuzenet,[c] Joseph S. Lam[c] and James H. Naismith[a]*

[a]Centre for Biomedical Sciences, North Haugh, The University, St Andrews, Fife KY16 9ST, Scotland, [b]Joint Structural Biology Group, ESRF, F38043, Grenoble CEDEX, France, and [c]Department of Microbiology, University of Guelph, Ontario, N1G 2W1, Canada

Correspondence e-mail: naismith@st-and.ac.uk

Received 9 May 2000
Accepted 19 July 2000

Glucose-1-phosphate thymidylyltransferase (RmlA; E.C. 2.7.7.24) is the first of four enzymes involved in the biosynthesis of dTDP-L-rhamnose, the precursor of L-rhamnose, a key component of the cell wall of many pathogenic bacteria. RmlA catalyses the condensation of thymidine triphosphate (dTTP) and α-D-glucose-1-phosphate (G1P), yielding dTDP-D-glucose. RmlA from *Pseudomonas aeruginosa* has been overexpressed and purified. Crystals of the enzyme have been grown using the sitting-drop vapour-diffusion technique with PEG 6000 and lithium sulfate as precipitant. Several diffraction data sets of single frozen crystals were collected to a resolution of 1.66 Å. Crystals belonged to space group $P1$, with unit-cell parameters a = 71.5, b = 73.1, c = 134.7 Å, α = 89.9, β = 80.9, γ = 81.1°. The asymmetric unit contains eight monomers in the form of two RmlA tetramers with a solvent content of 51%. Selenomethionine-labelled protein has been obtained and crystallized.

1. Introduction

L-Rhamnose is a 6-deoxyhexose that is found in the cell wall of many pathogenic microorganisms. In Gram-negative bacteria it is one of the important residues of the O-antigen of lipopolysaccharide, a factor that is a key determinant for the virulence of these species. In *P. aeruginosa*, this sugar is a constituent of the core oligosaccharide and serves as the receptor for O-antigen polymer (Rahim *et al.*, 2000). Mycobacteria, on the other hand, utilize L-rhamnose in the arabinogalactan (AG) that attaches the lipid mycolic acid layer to the peptidoglycan layer (McNeil *et al.*, 1990). Inhibitors of the synthesis of AG stop cell growth and some, such as ethambutol, are effective drugs in the treatment of tuberculosis (see, for example, Deng *et al.*, 1995). As L-rhamnose is not found in mammals, inhibition of its biosynthetic pathway presents a target for the development of novel antibiotics.

Four enzymes, glucose-1-phosphate thymidylyltransferase (RmlA), dTDP-D-glucose 4,6-dehydratase (RmlB), dTDP-6-deoxy-D-*xylo*-4-hexulose (RmlC) and dTDP-6-deoxy-L-*lyxo* 4-hexulose-4-reductase (RmlD), are required for the synthesis of dTDP-L-rhamnose from α-D-glucose-1-phosphate and dTTP. Significantly, these proteins are highly conserved amongst microorganisms (see, for example, Ma *et al.*, 1997; Graninger *et al.*, 1999) and therefore conclusions drawn from the structure of a protein from one species will have strong implications for the corresponding enzyme from other organisms.

After having successfully crystallized RmlB, RmlC and RmlD (Giraud, Gordon *et al.*, 1999; Giraud, McMiken *et al.*, 1999; Allard *et al.*, 2000), we here describe the overexpression, purification and crystallization of the pathway's first enzyme, RmlA, from *P. aeruginosa*. This protein is not only involved in the biosynthesis of L-rhamnose but also in the pathways leading to other 6-deoxysugars such as L-talose or D-fucose, as these share common intermediates in the conversion of D-glucose to their end-products (Nakano *et al.*, 2000; Yoshida *et al.*, 1999). The enzyme is highly homologous to other bacterial sugar nucleotide transferases (*e.g.* glucose-1-phosphate uridylyltransferase). The sugar nucleotide transferases catalyse the first step in all sugar-nucleotide chemistry and are of key importance in biology and biotechnology. As RmlA shows no sequence relationship to any protein structure currently deposited in the PDB (Sussman *et al.*, 1998), it is expected to contain a novel fold. The currently favoured mechanism is that in which the oxygen of G1P carries out a nucleophilic attack on the nucleotide triphosphate (NTP) α-phosphate group. This is an ordered sequential bi-bi mechanism and is consistent with the known

inversion of stereochemistry at the NTP α-phosphate group (Sheu & Frey, 1978; Fig. 1).

P. aeruginosa is a Gram-negative bacterium that colonizes many children with cystic fibrosis, where it is a significant cause of morbidity and mortality. In addition, it is an opportunistic pathogen that can cause a wide variety of infections, particularly in victims of severe burns and in patients who are for any reason immunosuppressed. This makes *P. aeruginosa* one of the most prevalent pathogens in hospital-acquired infections. Owing to its high resistance to antibiotics it is a particularly dangerous pathogen and any approach towards its control is highly sought.

2. RmlA overexpression and purification

The open reading frame of the gene encoding RmlA from *P. aeruginosa* was amplified using PCR with primers that incorporated a 5′ *Nco*I and a 3′ *Bam*HI site to facilitate cloning into a modified pET23a(+) vector (Newton & Mangroo, 1999). The plasmid also contained a sequence coding for a 6×His-tag on the N-terminus of RmlA to allow an easy purification on metal-chelating columns. Expression involves the IPTG (isopropyl-β-D-thiogalactoside) inducible T7 promotor and ribosome-binding sites conferred by the vector. The sequence of the amplified and cloned gene was confirmed to be identical to the chromosomal copy (except for the N-terminal 6×His-tag).

In order to overexpress RmlA, *Escherichia coli* BL21(λDE3) cells transformed with the plasmid were grown at 310 K in Luria–Bertani medium containing 100 μg ml^{-1} ampicillin until the OD_{600} reached 0.6–0.8. Expression of the protein was then induced by addition of 1 m*M* IPTG. After a further 3 h of culture, cells were harvested by centrifugation (20 min, 6000*g*, 277 K).

The cell pellet was resuspended in a lysis buffer containing 20 m*M* Tris–HCl pH 8.5, 100 m*M* NaCl, 2 m*M* DTT, 5 m*M* PMSF, 100 μg ml^{-1} hen egg-white lysozyme. After 30 min incubation at room temperature, the viscosity of the mixture was decreased by the addition of DNAse I (20 μg ml^{-1}) and by sonication (five cycles of 1 min interrupted by 1 min periods on ice). The suspension was centrifuged at 20 000*g* and 277 K for 20 min and the supernatant was brought to 20% ammonium sulfate saturation. After incubation on ice for 1 h, a second centrifugation (20 min, 20 000*g*, 277 K) was carried out and the supernatant was dialysed against two changes of 1 l 20 m*M* Tris–HCl pH 7.0, 20 m*M* imidazole, 500 m*M* NaCl. The filtered protein solution was passed through a POROS-MC column which had been pre-loaded with nickel sulfate. Proteins were eluted with a 20–500 m*M* imidazole gradient. A protein with a molecular weight corresponding to RmlA (∼34 kDa) was found in a peak eluting at approximately 200 m*M* imidazole. Fractions corresponding to this peak were pooled, concentrated with a 10 kDa cutoff Amicon membrane and dialysed against two changes of 1 l 20 m*M* Tris–HCl pH 8.5 at 277 K, containing 10 m*M* EDTA in the first change to remove contaminating nickel ions. For further purification, the protein was applied to a POROS-HQ anion-exchange column on a BioCAD 700E Workstation. Elution was achieved with a 50–1000 m*M* NaCl gradient. RmlA eluted at a salt concentration of 200 m*M*. Pooled fractions were brought to a protein concentration of approximately 4 mg ml^{-1} as determined by Bradford assay (Bradford, 1976) using bovine serum albumin as standard and were dialysed against two changes of 50 m*M* Tris–HCl pH 7.5 at 277 K. Prior to crystallization experiments, DTT was added to 4 m*M* and the solution was filtered through a 0.2 μm membrane. This procedure typically yielded 30 mg of pure protein per litre of bacterial culture. Small aliquots of the purified protein could be stored at 255 K without deterioration for several months without the addition of cryoprotecting agents.

Selenomethionine labelling of *P. aeruginosa* RmlA could not be achieved in *met*$^-$ B834(λDE3) *E. coli* cells. Under all conditions tested, the protein formed inclusion bodies. Selenomethionine-enriched protein was therefore produced by inhibition of the methionine biosynthesis pathway in *E. coli* BL21(λDE3) (Doublie, 1997). Briefly, cells were grown in M9 medium (64 g l^{-1} $Na_2HPO_4{\cdot}7H_2O$, 15 g l^{-1} KH_2PO_4, 2.5 g l^{-1} NaCl, 5 g l^{-1} NH_4Cl, 1 m*M* $MgSO_4$, 0.4% glucose, 0.1 m*M* $CaCl_2$) at 310 K until the OD_{600} reached 0.6. At this stage, the amino acids lysine, phenylalanine and threonine were added to final concentrations of 100 mg l^{-1} and isoleucine, leucine and valine to 50 mg l^{-1}. Seleno-L-methionine was added to a concentration of 60 mg l^{-1}. The temperature was lowered to 303 K and the culture left to shake for a further 15 min before protein overexpression was induced with 1 m*M* IPTG. After 6 h, cells were harvested and lysed as described above. 13 mg of pure protein per litre of culture could be isolated.

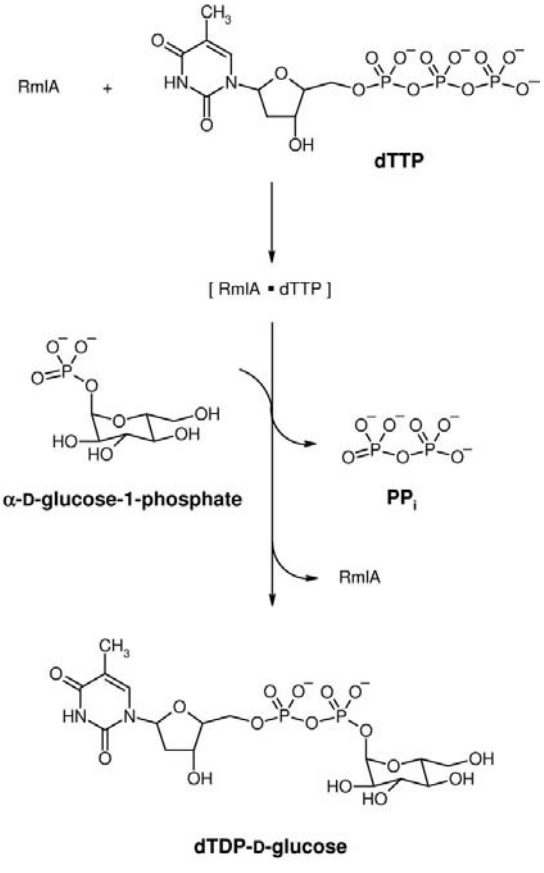

Figure 1
The reaction catalysed by RmlA.

3. Protein analysis

Following the two HPLC steps, the protein appeared to be pure as judged by an SDS silver nitrate stained gel (single band at an apparent molecular weight of 34 kDa), the calculated molecular weight based on sequence being 33 773 Da. A single peak with a molecular weight of 33 803 Da was found in the MALDI mass spectrum. Dynamic light-scattering results (DynaPro 801) indicated the native protein to be monodisperse, with a molecular weight in the range 106–122 kDa, indicative of a trimeric or tetrameric protein. N-terminal sequencing was per-

Table 1
Data-collection statistics for non-Se-labelled RmlA crystals.

Values in parentheses refer to the highest shell.

Crystal grown in the presence of	Glucose-1-phosphate (G1P)	Glucose-1-phosphate	dTDP-glucose	G1P/thymidine/AlF_3	dTMP
X-ray source	Rotating anode	ID14EH1	ID14EH1	ID14EH1	BM14
Wavelength (Å)	1.542	0.934	0.934	0.934	0.978
Resolution (Å)	36.27–2.27 (2.40–2.27)	40.49–1.90 (1.99–1.90)	40.49–1.77 (1.86–1.77)	36.27–1.87 (1.97–1.87)	32.12–1.66 (1.74–1.66)
Unit cell (Å, °)	a = 71.6,	a = 71.7	a = 71.6	a = 71.3	a = 71.5
	b = 73.5	b = 73.7	b = 73.4	b = 73.1	b = 73.1
	c = 134.2	c = 134.5	c = 134.3	c = 133.7	c = 134.7
	α = 89.9	α = 90.0	α = 89.9	α = 90.0	α = 89.9
	β = 80.8	β = 80.9	β = 80.6	β = 81.4	β = 80.9
	γ = 80.8	γ = 80.9	γ = 80.9	γ = 81.6	γ = 81.1
V_M (Å^3 Da^{-1})	2.55	2.56	2.54	2.52	2.54
Total measurements	305802	377690	483036	333697	1141819
Unique reflections	116157	202988	242450	183207	297160
$I/\sigma(I)$	6.8 (4.8)	9.2 (1.9)	7.1 (2.4)	6.9 (2.4)	10.9 (2.0)
Average redundancy	2.6 (2.3)	1.9 (1.9)	2.0 (1.3)	1.8 (1.3)	3.8 (2.0)
Completeness (%)	94.2 (84.2)	96.1 (95.6)	93.9 (77.4)	84.7 (66.8)	94.2 (81.2)
R_{merge}† (%)	6.5 (15.2)	5.4 (28.6)	7.2 (30.2)	6.7 (29.1)	5.3 (33.2)

† $R_{merge} = \sum\sum I(h)_j - \langle I(h)\rangle / \sum\sum I(h)_j$, where $I(h)$ is the measured diffraction intensity and the summation includes all observations.

formed and confirmed the protein to be RmlA.

The efficiency of the selenomethionine-labelling procedure was scrutinized by MALDI mass spectrometry. A shift of +304 Da was found for the intact labelled protein, corresponding well to the predicted additional mass of 282 Da (six methionine residues per chain). In an important and useful second check, sulfur methionine-containing fragments were undetectable in the MALDI mass spectrum of a tryptic protein digest.

4. RmlA crystallization

Initial crystallization conditions were obtained from Hampton Research Crystal Screens I and II (Jancarik & Kim, 1991; Cudney *et al*., 1994) plus NaCl, PEG 6000, PEG 6000/lithium sulfate and MPD grids. The sitting-drop vapour-diffusion method (Ducruix & Giegé, 1992) with 4 µl of protein sample and 4 µl of precipitant at 293 K was used throughout. Crystals appeared in 27 of the initial 192 conditions, in some cases 10 min after setup. Most promising were results from the PEG 6000/lithium sulfate grid and hence these conditions were further optimized. Plate-type crystals of approximate dimensions 0.3 × 0.3 × 0.05 mm (Fig. 2) were obtained after 1–7 d using 9–12%(*w*/*v*) PEG 6000, 0.5 *M* lithium sulfate and 0.1 *M* citrate/NaOH pH 4.6 as precipitating solution. The initial very high mosaicity of these crystals could be greatly reduced by the addition of 1–2 µl 10–50 m*M* G1P, dTMP or dTDP-glucose to the protein prior to crystallization.

5. Data collection

A 2.2 Å resolution data set from a single flash-frozen crystal grown in the presence of G1P was collected in-house at 110 K using a Nonius/MacScience DIP2000 imaging-plate detector system. Data were recorded as 245 non-overlapping 20 min 1° oscillations. Cryoprotection was achieved by washing the crystal in mother liquor supplemented with 16%(*v*/*v*) PEG 600 for 10–15 s. The oscillation images were indexed and integrated with the program *MOSFLM* (Leslie, 1992) and scaled with the *CCP*4 program *SCALA* (Collaborative Computational Project, Number 4, 1994). Higher resolution data sets of crystals grown in the presence of G1P, dTMP, dTDP-glucose or thymidine/glucose-1-phosphate/AlF_3 were measured at the ESRF, Grenoble on beamlines ID14EH1 and BM14 (Table 1). All crystals were triclinic, with approximate unit-cell parameters a = 71, b = 73, c = 134 Å, α = 89.9, β = 81, γ = 81°. All attempts to index or reduce the data in a higher space group failed; a native Patterson map shows no non-origin peak. A majority of the crystals were actually twinned. This could only be detected from the diffraction pattern. Trial and error was used to locate single crystals for analysis. In some cases, flash-annealing with the crystal remaining in the loop (Yeh & Hol, 1998) helped to achieve a less mosaic diffraction pattern.

In addition to the above data sets, a three-wavelength MAD experiment with a selenomethionine-labelled crystal that was grown in the presence of G1P was carried out at beamline BM14 of the ESRF, Grenoble (Table 2). The crystal-to-detector distance was adjusted so that the outer rim of the detector area corresponded to a resolution of 2.8 Å. 730 non-overlapping 0.5° oscillations were recorded at each of three wavelengths. The three wavelengths were chosen from an EXAFS scan of the crystal to correspond to the maximum of f'' (peak), the maximum modulus of f' (inflection) and the minimum modulus of f' (remote). These data were indexed and integrated with *DENZO* and scaled with *SCALEPACK* (Otwinowski & Minor, 1996).

6. Preliminary structural characterization

At the beginning of this study, it was not clear whether native RmlA is a trimeric or a tetrameric protein. A self-rotation search of the TMP data set with *REPLACE* (Tong & Rossmann, 1997) reveals three major (>30σ) plus several minor (~10–15σ) twofold axes (Fig. 3). In addition, a 60° and a 120° rotation axis (15.4σ and 14.6σ) are found lying parallel to the major 180° rotation axes at φ = 8° and ψ = 98° (data not shown). The interpretation of these results was greatly aided by the determination of Se-atom positions with the program *SOLVE* 1.17 (Terwilliger & Berendzen, 1999). 24 sites were found, which could easily be grouped into eight equivalent clusters of three atoms. The clusters fall into two sets of four, indicating that RmlA is a tetramer and that the unit cell of the *P*1 crystal form contains two tetrameric molecules. The rotation superimposing the two tetramers can be described as either as a 60, a 120 or a 180° rotation axis, depending on which monomer is used as a reference point. This explains the existence of major and minor twofold axes in the κ = 180° self-rotation search. Firstly, there are inter- and intra-tetramer 180° axes lying

Figure 2
Photograph of RmlA crystals. These crystals have been grown in the presence of dTMP.

Table 2
Data-collection statistics for MAD experiment on BM14 (ESRF, Grenoble).

Values in parentheses refer to the highest shell.

Position	Peak	Inflection	Remote
Wavelength (Å)	0.9790	0.9791	0.8855
Resolution (Å)	30.0–2.8 (2.87–2.80)		
Unit cell (Å, °)	a = 71.6, b = 73.9, c = 133.8, α = 89.8, β = 80.3, γ = 80.2		
V_M (Å^3 Da^{-1})	2.54		
Total measurements	221219	220585	224654
Unique reflections	121843	121982	121112
$I/\sigma(I)$	31.7 (9.0)	31.6 (8.8)	29.7 (9.8)
Average redundancy	1.8 (1.0)	1.8 (1.0)	1.9 (1.7)
Completeness (%)	93.0 (63.2)	93.0 (63.2)	92.7 (69.4)
R_{merge}† (%)	2.2 (8.9)	2.3 (9.5)	2.4 (8.9)

† $R_{merge} = \sum\sum I(h)_j - \langle I(h)\rangle / \sum\sum I(h)_j$, where $I(h)$ is the measured diffraction intensity and the summation includes all observations.

parallel to each other. Secondly, in the case of exclusively intra-tetramer axes only two pairs of two monomers are superimposed, while in the other case the intramolecular vectors of eight protein chains contribute to the peak in the self-rotation function, leading to a stronger signal.

The asymmetric unit of the crystal contains approximately 2400 amino-acid residues and has a solvent content of 51%, corresponding to a V_M of 2.54 Å^3 Da^{-1} (Matthews, 1968).

Phasing calculations based on the Se-atom positions are under way.

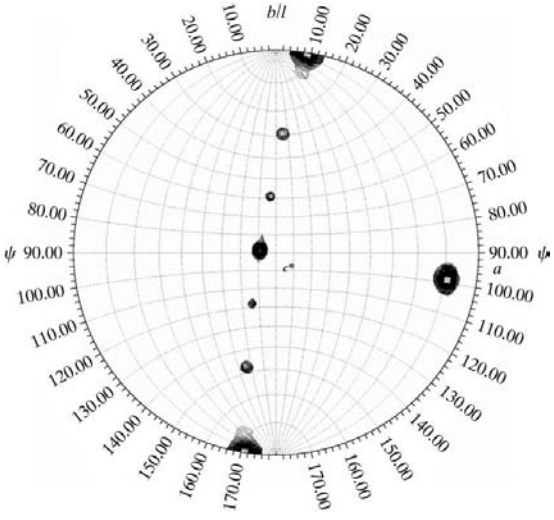

Figure 3
$\kappa = 180°$ section of the self-rotation search in the TMP data set. The search was performed with *REPLACE* (Tong & Rossmann, 1997). Search angle: polar XYK; orthogonalization AXABZ.

This project is supported by a grant from the Wellcome Trust to JHN (056851) and a grant from the Canadian Bacterial Diseases Network to JSL. RR and CC are recipients of a studentship and a fellowship, respectively, from the Canadian Cystic Fibrosis Foundation. The use of beamlines at the ESRF, Grenoble, France is gratefully acknowledged.

References

Allard, S. T., Giraud, M. F., Whitfield, C., Messner, P. & Naismith, J. H. (2000). *Acta Cryst.* D**56**, 222–225.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–252.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. & McPherson, A. (1994). *Acta Cryst.* D**50**, 414–423.
Deng, L., Mikusova, K., Robuck, K. G., Scherman, M., Brennan, P. J. & McNeil, M. R. (1995). *Antimicrob. Agents Chemother.* **39**, 694–701.
Doublie, S. (1997). *Methods Enzymol.* **276**, 523–530.
Ducruix, A. & Giegé, R. (1992). *Crystallization of Nucleic Acids and Proteins, a Practical Approach.* Oxford: IRL Press.
Giraud, M. F., Gordon, F. M., Whitfield, C., Messner, P., McMahon, S. A. & Naismith, J. H. (1999). *Acta Cryst.* D**55**, 706–708.
Giraud, M. F., McMiken, H. J., Leonard, G. A., Messner, P., Whitfield, C. & Naismith, J. H. (1999). *Acta Cryst.* D**55**, 2043–2046.
Graninger, M., Nidetzky, B., Heinrichs, D. E., Whitfield, C. & Messner, P. J. (1999). *J. Biol. Chem.* **274**, 25069–25077.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Leslie, A. G. W. (1992). *CCP4/ESF–EACMB Newslett. Protein Crystallogr.* **26**.
Ma, Y., Mills, J. A., Belisle, J. T., Vissa, V., Howell, M., Bowlin, K., Scherman, M. S. & McNeil, M. (1997). *Microbiology*, **143**, 937–945.
McNeil, M., Daffé, M. & Brennan, P. J. (1990). *J. Biol. Chem.* **265**, 18200–18206.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Nakano, Y., Suzuki, N., Yoshida, Y., Nezu, T., Yamashita, Y. & Koga, T. (2000). *J. Biol. Chem.* **275**, 6806–6812.
Newton, D. T. & Mangroo, D. (1999). *Biochem. J.* **339**, 63–69.
Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.
Rahim, R., Burrows, L. L., Monteiro, M. A., Perry, M. B. & Lam, J. S. (2000). Submitted.
Sheu, K. F. & Frey, P. A. (1978). *J. Biol. Chem.* **253**, 3378–3380.
Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. E. (1998). *Acta Cryst.* D**54**, 1078–1084.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–61.
Tong, L. & Rossmann, M. G. (1997). *Methods Enzymol.* **276**, 594–611.
Yeh, J. I. & Hol, W. G. (1998). *Acta Cryst.* D**54**, 479–480.
Yoshida, Y., Nakano, Y., Nezu, T., Yamashita, Y. & Koga, T. (1999). *J. Biol. Chem.* **274**, 16933–16939.