electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a Nudix protein from *Pyrobaculum aerophilum* reveals a dimer with two intersubunit β-sheets

Shuishu Wang, Cameron Mura, Michael R. Sawaya, Duilio Cascio and David Eisenberg

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of a Nudix protein from *Pyrobaculum aerophilum* reveals a dimer with two intersubunit β-sheets

Shuishu Wang, Cameron Mura, Michael R. Sawaya, Duilio Cascio and David Eisenberg*

UCLA-DOE Laboratory of Structure Biology, 611 Charles E. Young Drive East, 201 Boyer Hall, Los Angeles, CA 90095-1570, USA

Correspondence e-mail: david@mbi.ucla.edu

Received 28 September 2001
Accepted 21 January 2002

PDB References: 1jrk, r1jrksf; 1k26, r1k26sf; 1k2e, 1k2esf.

Nudix proteins, formerly called MutT homolog proteins, are a large family of proteins that play an important role in reducing the accumulation of potentially toxic compounds inside the cell. They hydrolyze a wide variety of substrates that are mainly composed of a nucleoside diphosphate linked to some other moiety *X* and thus are called Nudix hydrolases. Here, the crystal structure of a Nudix hydrolase from the hyperthermophilic archaeon *Pyrobaculum aerophilum* is reported. The structure was determined by the single-wavelength anomalous scattering method with data collected at the peak anomalous wavelength of an iridium-derivatized crystal. It reveals an extensive dimer interface, with each subunit contributing two strands to the β-sheet of the other subunit. Individual subunits consist of a mixed highly twisted and curved β-sheet of 11 β-strands and two α-helices, forming an α-β-α sandwich. The conserved Nudix box signature motif, which contains the essential catalytic residues, is located at the first α-helix and the β-strand and loop preceding it. The unusually short connections between secondary-structural elements, together with the dimer form of the structure, are likely to contribute to the thermostability of the *P. aerophilum* Nudix protein.

1. Introduction

Nudix proteins form a large family of proteins that are found in all organisms. They hydrolyze a wide variety of substrates that contain a nucleoside diphosphate linked to some other moiety, *X*. Therefore, these enzymes are called Nudix hydrolases (Bessman *et al.*, 1996). Their functions are mainly to reduce the level of these potentially toxic compounds and the accumulation of metabolic intermediates (Safrany *et al.*, 1998; O'Handley *et al.*, 2001). The best-studied Nudix enzyme is the MutT protein from *Escherichia coli* (Shimokawa *et al.*, 2000; Abeygunawardana *et al.*, 1995; Taddei *et al.*, 1997; Wagner *et al.*, 1997; Bhatnagar *et al.*, 1991; Frick *et al.*, 1995; Porter *et al.*, 1996). It reduces the rate of AT to CG transversion several thousand-fold by hydrolyzing the mutagenic nucleotide 8-oxo-dGTP to give 8-oxo-dGMP, thus preventing it from being incorporated into DNA (Bessman *et al.*, 1996). Its catalytic activity is reported to depend on the ligation of an essential metal ion. Structural and mutagenesis studies have identified the metal ligands as part of a signature motif conserved among the Nudix family, GX_5EX_7RE*UX*EE*X*G*U*, where *U* is a hydrophobic residue and *X* is any amino acid (Harris *et al.*, 2000; Shimokawa *et al.*, 2000; Lin *et al.*, 1996, 1997). Furthermore, the structure of this motif appears to be conserved among three members of the Nudix family for which structures have been determined: *E. coli* MutT, diadenosine

tetraphosphate hydrolase ($A_{P4}A$ hydrolase) from *Lupinus angustifolius* (Swarbrick *et al.*, 2000) and an ADP-ribose pyrophosphatase (ADPRase) from *E. coli* (Gabelli *et al.*, 2001). All three structures contain an α-β-α sandwich in which the conserved residues in the Nudix motif are located in helix α1 and in the loop preceding it.

Beyond the conserved catalytic core structure, Nudix proteins demonstrate variations in peripheral structure and oligomerization state, perhaps contributing to the wide variety of substrate specificities observed in this family. For example, the $A_{P4}A$ hydrolase has a structure similar to that of the *E. coli* MutT, but with one extra β-strand and two more α-helices. On a larger scale, the structures of MutT and $A_{P4}A$ hydrolase are monomeric, whereas ADPRase is a dimer with its N-terminal domain swapping between the subunits. Furthermore, in *E. coli* MutT the entire adenosine moiety of the substrate analog AMPCPP binds in the active-site cleft behind helix α1 (Lin *et al.*, 1997); however, in ADPRase, the terminal ribose of the substrate binds in the cleft, while the adenine moiety binds to the N-terminal domain.

Here, we report the crystal structure of a Nudix protein from the hyperthermophilic archaeon *P. aerophilum* (PA) determined with single-wavelength anomalous scattering data. The PA Nudix protein was found to be a dimer, with each subunit having a similar fold to the *E. coli* MutT. Unlike the structure of ADPRase, the PA Nudix protein dimerizes through two intersubunit β-sheets that lie in the center of the polypeptide sequence. The active-site cleft has a distinctly different structure from those of other Nudix hydrolases of known structure. In this paper, the structure is analyzed in terms of the catalytic function and the thermostability.

2. Experimental procedures

2.1. Cloning, expression and purification of the PA Nudix

The open reading frame (ORF) that encodes the putative Nudix protein reported here was found *via* a BLAST search of the PA genome with the sequence of the *E. coli* MutT protein. A phosmid clone containing the PA *mutT* ORF was kindly provided by the laboratory of Dr Jeffrey Miller (UCLA). The following primers were utilized for polymerase chain reaction (PCR) amplification of the gene with Deep Vent DNA polymerase (New England Biolabs): CATATGATCGTTAC-CAGCGGCGTTTTA (sense) and AAGCTTTGAAA-TTTTTCCCAGTCTATATAG (antisense). Amplified DNA was subcloned into the pCR-Blunt vector (Invitrogen) to give an intermediary vector that was then digested with the restriction enzymes *Nde*I and *Hin*dIII. The fragment containing the *mutT* gene was purified by gel extraction and ligated into a bacterial expression vector [pET22b(+), Novagen] that had been linearized by double digestion with the same enzymes. The resulting recombinant plasmid [designated pET22b(+)-mutT] adds a C-terminal 6×His tag to the expressed protein. The plasmid was transformed into chemically competent NovaBlue *E. coli* cells (Novagen) for screening and the DNA sequence of the *mutT* gene was confirmed by sequencing (Davis Sequencing, Inc.) A point mutant of PA Nudix (M16L) in which a Leu was substituted for Met16 was prepared by overlap-extension PCR in order to avoid heterogeneity resulting from a second translation-initiation site at this position.

Recombinant PA Nudix was overexpressed by transformation of the pET22b(+)-mutT plasmid into chemically competent BL21(DE3) *E. coli* cells. Single colonies were used to inoculate Luria–Bertani broth supplemented with 100 µg ml^{-1} ampicillin and the cultures were grown at 310 K to an OD_{600} of about 0.8. Isopropyl-β-D-thiogalactoside (IPTG) at 1 m*M* was then added to induce overexpression and cells were grown for a further 3–4 h. Cells were harvested by centrifugation for 10 min at 8000*g* and stored at 253 K.

Thawed cell pellets were resuspended at room temperature in a high ionic strength buffer containing 20 m*M* HEPES pH 7.8, 0.5 *M* NaCl, 0.5 m*M* PMSF, with or without 0.5%(*v*/*v*) Triton X-100. Cells were lysed by lysozyme treatment (∼0.5 mg ml^{-1} for 30 min at room temperature) followed by two or three passes through a French press operating at ∼7.6 MPa or by sonication on ice for 5 × 1 min with 1 min intervals. Lysed cells were maintained on ice and cell debris was cleared by centrifugation at ∼37 000*g* for 30 min. The supernatant of the cell lysate was either directly loaded onto an Ni^{2+}-charged HiTrap chelating column (Pharmacia) or pretreated by heating to ∼348–353 K for 10 min and clearing the denatured insoluble *E. coli* proteins by centrifugation at 37 000*g* for 35 min (>80% purity was achieved with this heat treatment). The final full-length protein was >99% pure as estimated by several independent techniques (SDS–PAGE, MALDI–TOF and electrospray mass spectrometry).

2.2. Crystallization and data collection

The purified protein was exchanged by dialysis or by three cycles of dilution and concentration into a dilute buffer (10 m*M* Tris, 5 m*M* EDTA pH 8.0 or 5 m*M* HEPES, 10 m*M* NaCl pH 7.5) and then concentrated to above 10.0 mg ml^{-1} at 277 K in a Centriprep ultrafiltration device. Crystallization experiments were carried out using the hanging-drop vapor-diffusion method. In each drop, 3–5 µl of protein solution was mixed with an equal volume of well solution. Hampton Crystal Screen kits I and II were used for the initial screening of crystallization conditions. Crystals were obtained under several conditions. The best crystals were obtained from drops set up with well solutions containing 5–8% PEG 4000 or PEG MME (polyethylene glycol monomethyl ether) 2000, 5% 2-propanol, 50 m*M* $(NH_4)_2SO_4$ and 100 m*M* NaOAc pH 4.8 (native-1 crystals) or well solutions consisting of 100 m*M* MES pH 6.2 and 15% 2-methyl-2,4-pentanediol (MPD) (native-2 crystals) at room temperature. Crystals were soaked for 5 min in the well solution but with 30% of glycerol or with 50% of MPD for native-2 crystals and flash-frozen in a stream of N_2 gas at 100 K. An iridium derivative was obtained by soaking a crystal which was obtained under similar conditions as native-1 crystals in the well solution containing 20 m*M* $IrCl_3$ for 8 h. The crystal was then similarly flash-frozen. Diffraction

data from Ir-derivatized and native-1 crystals were collected at beamline X8C, NSLS, Brookhaven National Laboratory and those from the native-2 crystal were collected on an in-house Rigaku FRD generator with an R-AXIS IV++ detector. For the iridium-derivative crystal, a single set of data was collected at the peak anomalous wavelength of iridium. The diffraction data were indexed, integrated and scaled with the programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1996). Data-processing statistics of the iridium-derivative data and two native data sets are listed in Table 1. The iridium and native-1 crystals are not isomorphous, despite their close similarity in unit-cell parameters. Thus, the native data were not used for initial phasing. After solving the structures, it became clear that the crystals have different packing interfaces along the crystal *z* axis. The two crystals pack in the same way in the *xy* plane, but the packing between layers is shifted along the *y* axis by about 18.6 Å.

2.3. Phasing and structure determination

Initial phases were determined for the data collected at the peak anomalous wavelength from the iridium-derivative crystal. Heavy-atom positions were identified with the programs *XTALVIEW* (McRee, 1992) and *SHELXD* (Uson & Sheldrick, 1999) and then refined with the program *MLPHARE* (Otwinowski, 1991; Collaborative Computational Project, Number 4, 1994). Density modification, *i.e.* solvent flattening, histogram matching, multi-resolution modification and twofold NCS averaging, was carried out with the program *DM* (Cowtan, 1994).

The automatic refinement procedure *ARP*/*wARP* (Perrakis *et al.*, 1999) was then used for automatic model building. The initial model and the electron-density maps were displayed with the graphics program *O* (Jones *et al.*, 1991) and the model was manually rebuilt to fix the side chains and to add residues that have clear electron density. The model was then refined using *CNS* (Brünger *et al.*, 1998) with NCS restraints. σ_A-weighted $2F_o - F_c$ and $F_o - F_c$ electron-density maps were calculated from the refined model and the model was manually fixed using the program *O*. This cycle was repeated until the refinement converged. This refined model was used as a model for molecular replacement of the native-1 data with the program *AMoRe* (Navaza, 1994) or of the native-2 data with the program *EPMR* (Kissinger *et al.*, 1999) and refinement was similarly carried out with the programs *CNS* and *O*.

The final refinement statistics are listed in Table 2. Over 91% of residues in the native-1 structure and over 93% in the Ir-derivative structure fall into the most favored regions in the Ramachandran plot. Some glycerol or MPD molecules from the cryogenic solution were found in the refined structures (see Table 2). Positive $F_o - F_c$ density was found near a cluster of four His side chains, His19, His31, His89 and His91, in the active-site cleft. A water molecule modelled into this site was not sufficient to account for this density. Since the protein was purified from an Ni-chelating column, an Ni^{2+} ion was placed into the density. The hydroxyl group of Tyr84 side chain is also in a good position to function as a ligand.

In the Ir-derivative structure all the Ir atoms were found to bind in surface cavities on the edge of bulk-solvent channels. Three of them, Ir1, Ir2 and Ir3, were found at the crystal-packing interface, while Ir4 was found near the Nudix box residues of subunit *A*. The strongest iridium peak, Ir1, was found at a packing interface of three molecules and surrounded by five water molecules, which take five positions of an octahedral coordination. The interactions between the

Table 1
Data-collection statistics for PA Nudix.

Values in parentheses are for the last resolution bin (1.92–1.85 Å for the Ir-derivative data and 1.86–1.80 Å for the native-1 data).

	Ir derivative†	Native-1†	Native-2‡
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Unit-cell parameters			
a (Å)	52.42	52.71	52.71
b (Å)	71.66	72.61	72.61
c (Å)	85.76	85.18	85.18
β (°)			99.96
Resolution (Å)	100–1.85	100–1.8	90–2.4
Unique reflections	25455	30654	23707
Observed reflections	348337	398012	84048
Completeness (%)	89.3 (52.0)	98.8 (96.4)	98.2 (97.0)
Anomalous completeness§ (%)	87.6 (69.8)		
$I/\sigma(I)$	18.7 (3.2)	18.4 (3.2)	14.1 (3.0)
R_{sym}¶	0.076 (0.470)	0.080 (0.443)	0.093 (0.465)
No. of subunits per a.u.	2	2	4
X-ray wavelength (Å)	1.1053	1.1	1.5418

† Data for the Ir derivative and native-1 were collected at a synchrotron. The unit-cell parameters for the two data sets are similar. However, the two crystals have different packing and therefore are not isomorphous. ‡ Native-2 crystals were obtained from drops equilibrated against 100 m*M* MES pH 6.2 and 15% MPD. Data were collected on an R-AXIS IV++ detector. § Anomalous data were processed to 2.0 Å, with the last bin from 2.07–2.0 Å. ¶ $R_{sym} = \sum |(I_{hkl} - \langle I_{hkl}\rangle)|/\sum I_{hkl}$, where $\langle I_{hkl}\rangle$ is the average of I_{hkl} over all symmetry equivalents.

Table 2
Statistics of atomic refinement of PA Nudix.

	Ir derivative	Native-1	Native-2
Resolution range for refinement (Å)	20–1.85	20–1.8	20–2.4
No. of protein atoms	2374	2416	4732
Average *B* factor, protein atoms† (Å^2)	26.1	19.65	31.9
No. of other molecules‡	2 SO_4 (25.2, 36.6), 4 Ir (34.7), 2 Ni (74.4), 3 glycerol (37.1), 5 acetate (46.2), 149 H_2O (34.4)	2 SO_4 (21.1, 42.6), 2 Ni (66.1), 6 glycerol (38.8), 2 acetate (30.9), 244 H_2O (29.6)	6 MPD (64.8), 173 H_2O (30.9)
R.m.s.d.			
Bond length (Å)	0.017	0.019	0.013
Bond angle (Å)	1.8	1.8	1.8
R_{free}§	0.218	0.219	0.274
R_{work}§	0.182	0.183	0.190

† The Ir-derivative crystal has a higher mosaicity (about 1.2°) than the native crystal (about 0.7°). ‡ Average *B* factors are given in parentheses. § *R* factors were calculated using data in the resolution range for refinement without a σ cutoff. R_{free} was calculated with a set of data (8%) never used in the refinement. R_{work} was calculated against the data used in the refinement.

protein atoms and Ir1 and Ir4 seem to be mediated by water. Ir2 and Ir3 are approximately related by the twofold NCS, each bound near a sulfate ion, as well as the side chains of Asp65, Asp66 and Asn67 of a symmetry-related molecule.

3. Results

3.1. Protein expression and purification

The PA Nudix gene was cloned and overexpressed in *E. coli* with a His tag at the C-terminus. It was found that the wild-type protein was expressed as a mixture of the full-length and a truncated form that lacks the first 15 N-terminal residues. Apparently, the Met16 codon acts as a second start site for translation of this truncated protein. The DNA sequence ~10 bp upstream of the Met16 codon corresponds to a bacterial ribosome-binding site, thus exacerbating the problem. Similar problems have been reported with other proteins (Matsumiya *et al.*, 2001). In order to circumvent the alternative start site, a point mutant (M16L) was made in which a Leu was substituted for Met16. This mutant was expressed as a single full-length polypeptide and was purified to >99% pure. It differs from the wild-type sequence only in the M16L mutation and 14 amino acids appended to the C-terminus (the 6×His tag and an eight amino-acid linker). This M16L mutant was used for all the studies reported here.

3.2. Phasing with single-wavelength anomalous scattering data

The crystal structure was determined using single-wavelength anomalous scattering data collected at a wavelength near the peak of fluorescence at the L_{III} edge of iridium. An anomalous difference Patterson map revealed a single well occupied Ir site and a few much weaker potential sites. *SHELXD* (Uson & Sheldrick, 1999) was used to evaluate the heavy-atom sites and how well they correlate with the Patterson map. The two top sites were input to *MLPHARE* (Otwinowski, 1991; Collaborative Computational Project, Number 4, 1994) for refinement and calculating phases. The anomalous difference Fourier map calculated from these phases gave four unique peaks above the 5.5σ contour level. The positions of the two new peaks also matched the third and fifth sites predicted by *SHELXD*. These four sites were then input to *MLPHARE* for refinement and phase calculation. The four Ir atoms (designated Ir1, Ir2, Ir3 and Ir4) have occupancies of 0.74, 0.50, 0.36 and 0.39 and temperature factors of 57, 59, 64 and 67 Å^2, respectively, after refinement with *MLPHARE*. The atom Ir1, which has the highest occupancy and lowest *B* factor, gave the strongest peak in the Patterson map.

Density modification with solvent flattening, histogram matching and multi-resolution modification using *DM* (Cowtan, 1994) was carried out to improve the phases and to identify the correct absolute configuration. The Fourier map after *DM* showed enough secondary-structure elements to locate the twofold non-crystallographic symmetry (NCS) axis. Twofold NCS averaging with *DM* significantly improved the quality of the electron-density map (Fig. 1) and the entire map was continuous and could be traced easily. Automatic model building with *ARP*/*wARP* (Perrakis *et al.*, 1999) at a resolution of 1.85 Å built more than 80% of the model. The map calculated at this resolution with phases from the auto-built model showed clear side chains for more than 80% of the residues and a complete model was built in a single round of manual rebuilding.

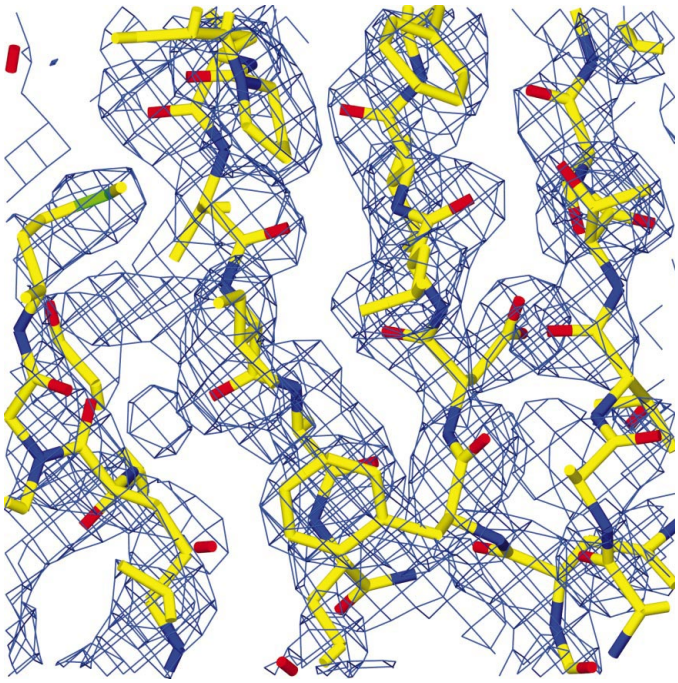

Figure 1
A section of the initial electron-density map superimposed on the final coordinates. The map was calculated at 2.5 Å with phases after twofold NCS averaging and contoured at 1.3σ. The figure was generated with the program *O* and rendered with *POV-RAY*.

3.3. Overall structure

The refined crystal structure of the PA Nudix protein shows that it forms a dimer. Each subunit also contributes two β-strands to the β-sheet of the other subunit to form an extensive dimer interface. The dimer is held together mainly by the two β-sheets to which both subunits contribute strands. Hydrogen-bonding interactions occur between strand β5b of one subunit and β6 of the other subunit (Figs. 2 and 3). There are eight main-chain hydrogen bonds between two subunits in the dimer. In addition, a patch of residues with hydrophobic side chains, Ile2, Ile77, Leu94, Pro38, Pro74, Tyr61, Tyr96, Phe92 and Met73, located at the center of the dimer interface, also contribute to the binding energy of the dimer. The molecular surface of each monomer, calculated with the program *GRASP* (Nicholls *et al.*, 1991), is about 7500 Å^2 and the total buried surface in the dimer interface is about 2000 Å^2. The dimer formation in solution was confirmed by analytical ultracentrifugation studies (data not shown).

Each subunit (referred to as *A* and *B*) is composed of a mixed β-sheet and two α-helices (Fig. 2). The β-sheet is highly

curved and twisted and wraps around the helices. The topology of the protein is shown in Fig. 3. Most of the strands are connected by tight turns or short loops, with the longest loop consisting of only six residues. The β-sheet can be divided into a central sheet and a subsheet, which are connected through strand $\beta2$. The N-terminal part of strand $\beta2$ forms hydrogen bonds with the C-terminal part of $\beta1$, but then curves and makes a subsheet with strands $\beta3$, $\beta8$ and $\beta9$. Strands $\beta1$ and $\beta4$ to $\beta7$, together with $\beta5a'$ and $\beta5b'$ from the other subunit, form the central β-sheet.

The two subunits are similar to each other. However, subunit *A* has a long C-terminal helix that extends away from the molecular surface, while in subunit *B* the C-terminal residues turn back and form a more compact structure, making the helix shorter (Fig. 2). The residues in the linker to the His tag are ordered in the crystal structure, although the His tag itself is not. In the case of subunit *A*, these residues extend the helix and make contacts with nearby molecules in the crystal packing. In one crystal form in the $P2_1$ space group (native-2) there are four molecules per asymmetric unit, composed of a pair of dimers, each dimer similar to native-1. Both subunits in the dimer have long straight C-terminal helices which pack against the C-terminal helices of the other dimer to form a tetramer with 222 symmetry. However, the interactions between the helices mainly involve the residues of the linker to the His tag. Therefore, the tetramer is not physiologically relevant.

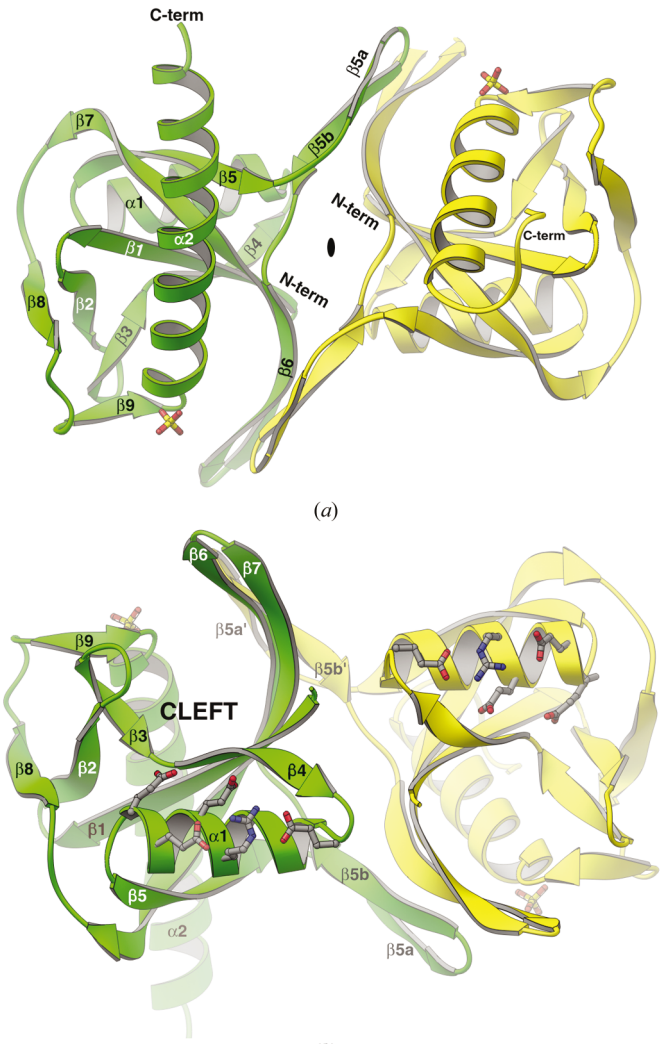

Figure 2
Ribbon diagrams of the PA Nudix structure. (*a*) shows the dimer with the twofold NCS axis perpendicular to the paper plane. Each subunit is composed of a long highly twisted central β-sheet and a small subsheet connected through strand $\beta2$, sandwiched between two α-helices. Each subunit contributes two β-strands ($\beta5a$ and $\beta5b$) to the β-sheet of the other subunit to form the dimer interface. Subunit *A* (on the left) has a long C-terminal helix $\alpha2$ with the eight amino-acid linker at the C-terminal end extending the helix, while subunit *B* has a shorter helix $\alpha2$ and its C-terminus and the linker fold back to form a more compact structure. (*b*) is a near 180° rotation of (*a*) along a horizontal axis. A cleft formed by the β-strands connects helix $\alpha1$ at the front to the N-terminal end of helix $\alpha2$ at the back. The conserved arginine and glutamate side chains, as well as two sulfate ions found at the N-terminal end of $\alpha2$ of each subunit, are shown as sticks. This figure was generated using *RIBBONS* (Carson, 1997).

3.4. The conserved residues and the active-site cleft

The Nudix box sequence signature motif, $G^{30}X_5E^{36}X_7R^{44}E^{45}UXE^{48}E^{49}XG^{51}U$ (the superscripts refer to the residue number in the PA Nudix sequence) spans the strand $\beta4$, helix $\alpha1$ and the N-terminal part of strand $\beta5$ (Figs. 4 and 5). The conserved residues are at the same relative positions in three-dimensional space as those in the *E. coli* MutT

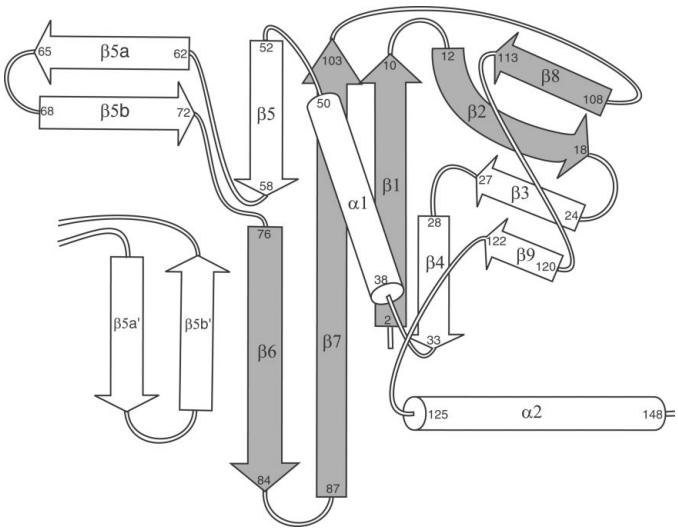

Figure 3
The topology diagram of the PA Nudix structure. The strand $\beta1$ is connected by a type II β-turn to $\beta2$, which forms hydrogen bonds with the C-terminal part of $\beta1$ and then makes a curve to form hydrogen bonds to $\beta3$, making a subsheet with strands $\beta8$ and $\beta9$. A proline makes a kink at the end of $\beta3$, allowing $\beta4$ to form hydrogen bonds to the N-terminal half of $\beta1$. A small loop connects $\beta4$ to helix $\alpha1$. Strands $\beta5$ and $\beta6$ together form an almost continuous strand antiparallel to the longest strand $\beta7$. Strands $\beta5a$ and $\beta5b$ go to the other subunit and form part of its central β-sheet, while strands $\beta5a'$ and $\beta5b'$ are from the other subunit. The last helix $\alpha2$ is connected to $\beta9$ by a type I β-turn. There are more β-strands in this protein structure than in the *E. coli* MutT. The shaded β-strands are the strands that are present in the *E. coli* MutT structure.

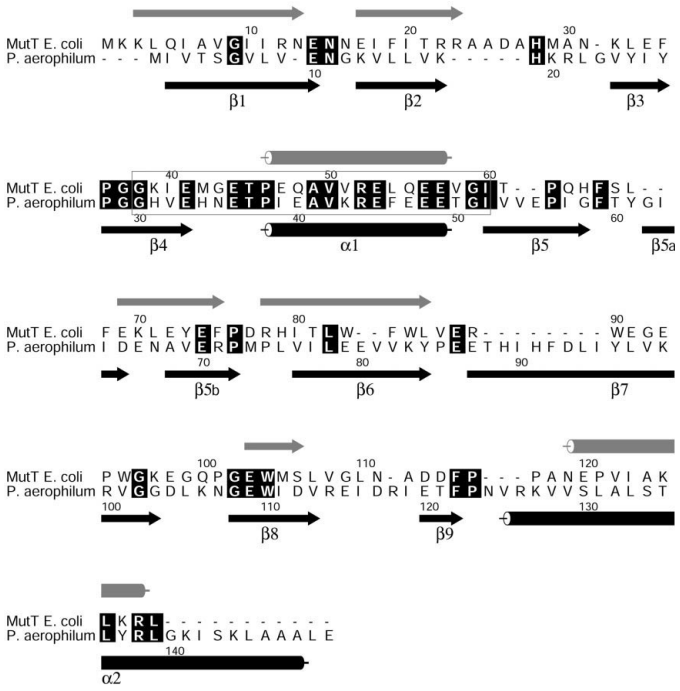

Figure 4
Sequence alignment between the *E. coli* MutT and PA Nudix. The sequences were aligned with *CLUSTALW* (Thompson *et al.*, 1994) and the figure was generated using *ALSCRIPT* (Barton, 1993). The sequence of the PA Nudix protein also includes an eight amino-acid linker at the C-terminal end. Identical residues are highlighted in black. Helices and β-strands of the two structures are marked with cylinders and arrows, respectively. The two sequences align well only around the Nudix box signature motif (boxed residues). The PA enzyme has more β-strands and much shorter loops.

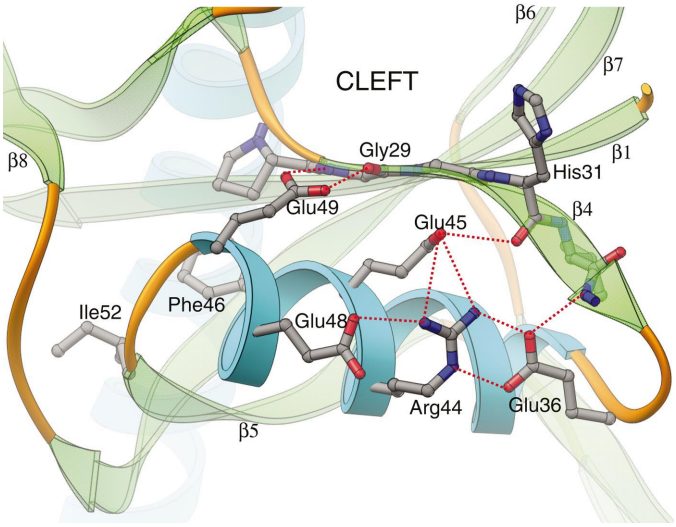

Figure 5
Conserved Nudix box residues. The structure around the active site is shown in ribbon representation, with residues in strand β4 and conserved side chains of the catalytic helix shown as sticks. The semi-conserved hydrophobic side chains, Phe46 and Ile52, are also shown. All the conserved side chains are on one side of helix α1, while the side chains of Phe46 and Ile52 are buried and face the central β-sheet. Hydrogen bonds from the conserved side chains are shown as red dashed lines.

structure (Fig. 6). These residues probably play similar roles throughout the Nudix enzyme family. All side chains of the conserved residues are on the same side of helix α1. The semi-conserved hydrophobic residues, Phe46 and Ile52, interact with the residues on the central β-sheet and are probably important for anchoring the catalytic helix to the central β-sheet. The conserved glutamate side chains are involved in hydrogen bonds with nearby side-chain and main-chain atoms (Fig. 5) and are in conformations that are not capable of coordinating a metal ion. It is possible that when in solution or at high temperatures the side chains can rearrange to bind the metal ion. However, the carbonyl O atom of the conserved Gly30 is involved in a hydrogen bond to the strand β1 and is not available for coordinating the metal ion. The carbonyl O atom of the corresponding glycine in *E. coli* MutT was found to be a ligand to the enzyme-bound metal ion (Lin *et al.*, 1997). However, in both the $A_{P4}A$ hydrolase and the ADPRase the corresponding glycine was also found to have a conformation incapable of coordinating the metal ion. It is likely that it is the small size of this conserved glycine residue that is important for the catalytic function of the Nudix enzymes. The side chain of the conserved Arg44 in the PA Nudix is well ordered and its guanidinium group interacts with the side chains of Glu45, Glu36 and Glu48 (Fig. 5). The putative function of this conserved arginine in the *E. coli* MutT is to orient the side chain of Glu45, which is proposed to function as a general base (Harris *et al.*, 2000; Lin *et al.*, 1997).

There is a deep narrow cleft connecting the catalytic helix α1 at the front to the N-terminus of helix α2 at the back (Fig. 2*b*). Strands β1, β4 and β7 form the floor of the cleft. Strand β6 and part of β7, together with β5a′ and β5b′ from the other subunit, form one wall of the cleft, while β2-loop-β3 and β9 to the N-terminal end of α2 form the other wall. In the NMR solution structure of the *E. coli* MutT, the nucleoside group of the substrate analog AMPCPP was found to bind in the corresponding cleft (Fig. 6). It is possible that in the PA Nudix protein this cleft also functions in the recognition of the substrate. The loop between β2 and β3 and the turn between β6 and β7 tilt toward the center of the cleft, partially covering the top of the cleft. The positions of these residues differ most greatly between the two subunits in the dimer (Fig. 7), thus indicating their potential flexibility for accommodating the binding of substrates.

3.5. Sulfate-binding sites

For the Ir derivative and native-1 crystals, which were obtained in the presence of 50 m*M* $(NH_4)_2SO_4$, a sulfate ion was found at the N-terminus of the last helix (α2) of each subunit, sitting at one end of the active-site cleft (Figs. 2*b* and 6). The O atoms of the sulfate ion form hydrogen bonds with the main-chain N atoms of Asn124, Val125, Arg126 and Lys127. In molecule *A*, both side chains of Arg126 and Lys127 are well ordered and form hydrogen bonds with the O atoms of the sulfate. However, in subunit *B* these two residues have partially disordered side chains. The sulfate bound to subunit *B* has a higher *B* factor (Table 2), possibly indicating a lower

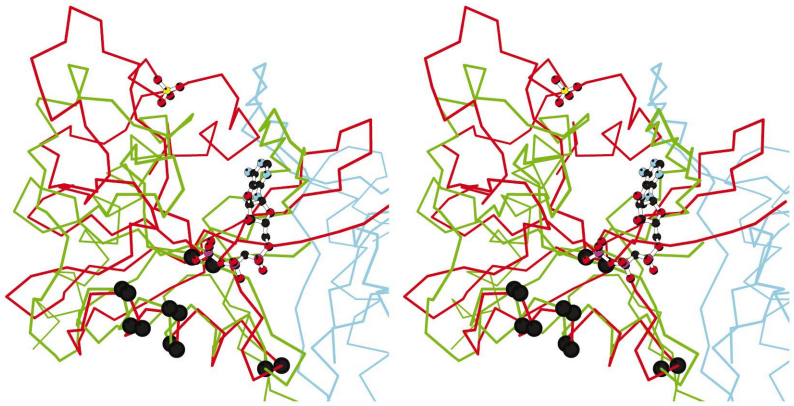

Figure 6
Structure superposition between the subunit *A* of PA Nudix and the *E. coli* MutT. The *E. coli* MutT structure (PDB code 1tum; Lin *et al.*, 1997) is shown in red and its bound AMPCPP shown in ball-and-stick representation. The two structures were aligned based on β1, α1 and part of β7 of the PA structure. They have similar folds and the conserved residues in the Nudix motif (shown as black dots) are at similar positions. However, away from helix α1 the two structures are quite different. The active-site cleft is much narrower and deeper in the PA Nudix structure. The adenosine group of the AMPCPP in the *E. coli* MutT structure binds in the active-site cleft. A sulfate ion was found at the N-terminus of α2 of the PA protein (shown in ball-and-stick), indicating a possible binding site for a phosphate group. The figure was generated with *MOLSCRIPT* (Kraulis, 1991).

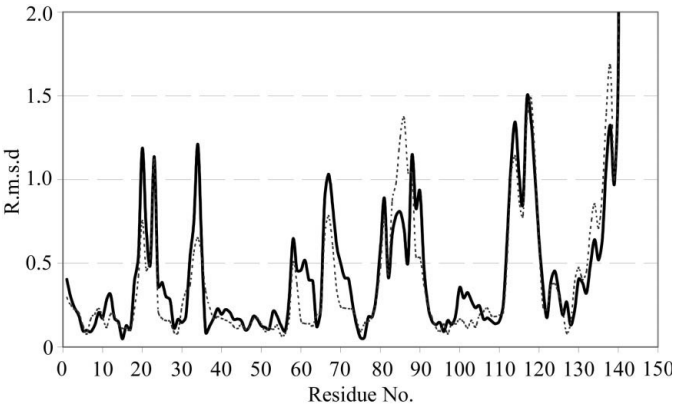

Figure 7
R.m.s.d. in Å of the main-chain atomic positions for the superimposed subunits of the dimer. The solid line is from the structure of the iridium-derivative crystal and the dashed line is from the native crystal. Higher r.m.s.d. was observed for residues 18–25, 79–90 and 64–69, which form the walls of the active-site cleft, indicating their flexibility. The peak around residue 32 is a consequence of partial disorder of the loop (higher *B* factors), while peaks around residues 110–120 and after 135 arises from differences in the crystal contacts of the two subunits. A plot of the r.m.s.d. atomic positions between the dimers of the two different crystals (not shown) also gives the same peaks at residues 18–25, 79–90 and 64–69, but with a smaller scale.

occupancy. Asn124 has main-chain dihedral angles in the left-handed helix region of the Ramachandran plot ($\varphi \simeq 70^\circ$ and $\psi \simeq 0^\circ$). This is necessary for the optimal interaction of its backbone amide group with the sulfate ion. In one crystal form ($P2_1$ space group with two molecules in one asymmetric unit, data not shown), molecule *B* has a very low occupancy of sulfate and the electron-density maps indicated that there are two conformations for Asn124, with one conformation having regular main-chain dihedral angles ($\varphi \simeq -75^\circ$ and $\psi \simeq -25^\circ$). We note that in *E. coli* MutT the amino-terminal Asn residue of helix 2 (N119) interacts with the 6-oxo group of dGTP (Lin *et al.*, 1997).

4. Discussion

4.1. Comparison with other Nudix hydrolases of known three-dimensional structure

The PA Nudix protein follows the same fold as the *E. coli* MutT. However, the two structures are quite different. Fig. 6 shows the superposition of the C^α traces of *E. coli* MutT onto that of the subunit *A* of PA Nudix. Only β1, α1 and part of β7 can be aligned. The r.m.s.d. over the 30 C^α atoms aligned is 1.48 Å. When structures are aligned based on these secondary-structure elements, most of the long β-strands are close in space in the two structures. However, the short β-strands and α2 are shifted relative to each other. The β-strands of the PA Nudix enzyme are highly curved compared with those in the MutT and the $A_{P4}A$ hydrolase. As a result, the active-site cleft is narrower and deeper in the PA Nudix structure. This indicates that the PA Nudix hydrolase is likely to have a different substrate from those of MutT or $A_{P4}A$ hydrolase. In the crystal structure of ADPRase complexed with ADP-ribose (Gabelli *et al.*, 2001), the terminal ribose group binds in the active-site cleft, while the adenine moiety interacts with the N-terminal domain, which is involved in dimer formation through domain swapping. The PA Nudix does not have this N-terminal domain and thus is unlikely to have the same ADP-sugar hydrolase activity.

In the PA Nudix hydrolase structure, the N-terminal end of the long helix α2 is located at the other end of the active-site cleft opposite the catalytic helix and a sulfate ion was found binding at the N-terminal end of α2 (Figs. 2*b* and 6). The distance between the Nudix box residues to the sulfate-binding site is about 20 Å. This sulfate-binding site could potentially be a binding site for a phosphate group on the substrate, although the distance is large. It is possible that this Nudix enzyme hydrolyzes substrates that have a phosphate group away from the scissile bond that can bind to the N-terminal end of the helix α2. An attempt to identify potential substrates for the PA Nudix protein showed that it is inactive against 14 typical substrates for known Nudix enzymes (ADP-ribose, ADP-mannose, ADP-glucose, GDP-mannose, GDP-glucose, UDP-mannose, UDP-glucose, Ap_2A, Ap_3A, Ap_4A, NADH, deamino-NADH, NAD and FAD) (Yang, Wang and Mura, unpublished results). These results suggest that PA Nudix is potentially a novel Nudix enzyme. Alternatively, PA Nudix may be inactive because of the M16L mutation or because the active form lacks residues 1–15.

4.2. Thermostability

P. aerophilum is a hyperthermophile that grows optimally at 373 K (Volkl *et al.*, 1996). The dimeric form of PA Nudix might

contribute to the thermostability of this enzyme. Intersubunit interactions have been proposed to be a major stabilization mechanism for hyperthermophilic proteins (Vieille & Zeikus, 2001). In addition, ion pairs and hydrophobic interactions between subunits make the dimer more resistant to dissociation. Two ion pairs were found between the subunits, Glu79 of one subunit to Arg71 of the other. Another factor that may contribute to the thermostability of this protein is that PA Nudix has very few residues in loops. Thompson & Eisenberg (1999) compared the sequences of about 20 complete genomes and found that thermophilic proteins generally have loop deletions relative to their mesophilic homologs. Sequence alignment between *E. coli* MutT and PA Nudix showed a deletion at loop 1 of the PA Nudix sequence (Fig. 4). In addition, the loops are shortened by incorporating more residues in the secondary-structure elements. About 75% of residues are in β-strands and α-helices, compared with only about 49% in the *E. coli* MutT protein (Fig. 4). More β-strands are present in the PA Nudix structure and the β-strands are mostly connected by tight β-turns.

We would like to thank Dr Hanjing Yang for lots of helpful discussion on Nudix hydrolases and their functions.

References

Abeygunawardana, C., Weber, D. J., Gittis, A. G., Frick, D. N., Lin, J., Miller, A.-F., Bessman, M. J. & Mildvan, A. S. (1995). *Biochemistry*, **34**, 14997–15005.

Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.

Bessman, M. J., Frick, D. N. & O'Handley, S. F. (1996). *J. Biol. Chem.* **271**, 25059–25062.

Bhatnagar, S. K., Bullions, L. C. & Bessman, M. J. (1991). *J. Biol. Chem.* **266**, 9095–9054.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54,** 905–921.

Carson, M. (1997). *Methods Enzymol.* **277**, 493–505.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.

Frick, D. N., Weber, D. J., Abeygunawardana, C., Gittis, A. G., Bessman, M. J. & Mildvan, A. S. (1995). *Biochemistry*, **34**, 5577–5586.

Gabelli, S. B., Bianchet, M. A., Bessman, M. J. & Amzel, L. M. (2001). *Nature Struct. Biol.* **8**, 467–472.

Harris, T. K., Wu, G., Nassiah, M. A. & Mildvan, A. S. (2000). *Biochemistry*, **39**, 1655–1674.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kissinger, C., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Lin, J., Abeygunawardana, C., Frick, D. N., Bessman, M. J. & Mildvan, A. S. (1996). *Biochemistry*, **35**, 6715–6726.

Lin, J., Abeygunawardana, C., Frick, D. N., Bessman, M. J. & Mildvan, A. S. (1997). *Biochemistry*, **36**, 1199–1211.

McRee, D. E. (1992). *J. Mol. Graph.* **10**, 44–46.

Matsumiya, S., Ishino, Y. & Morikawa, K. (2001). *Protein Sci.* **10**, 17–23.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Nicholls, A., Sharp, K. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.

O'Handley, S. F., Dunn, C. A. & Bessman, M. J. (2001). *J. Biol. Chem.* **276**, 5421–5426.

Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, p. 80. Warrington: Daresbury Laboratory.

Otwinowski, Z. & Minor, W. (1996). *Methods Enzymol.* **276**, 307–326.

Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Porter, D. W., Nelson, V. C., Fivash, M. J. & Kasprzak, K. S. (1996). *Chem. Res. Toxicol.* **9**, 1375–1381.

Safrany, S. T., Caffrey, J. J., Yang, X., Bembenek, M., Moyer, M. B., Burkhart, W. A. & Shears, S. B. (1998). *EMBO J.* **17**, 6599–6607.

Shimokawa, H., Fujii, Y., Furuichi, M., Sekiguchi, M. & Nakabeppu, Y. (2000). *Nucleic Acids Res.* **28**, 3240–3249.

Swarbrick, J. D., Bashtannyk, T., Maksel, D., Zhang, X.-R., Blackburn, G. M., Gayler, K. R. & Gooley, P. R. (2000). *J. Mol. Biol.* **302**, 1165–1177.

Taddei, F., Hayakawa, H., Bouton, M.-F., Cirinesi, A.-M., Matic, I., Sekiguchi, M. & Radman (1997). *Science*, **278**, 128–130.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.

Thompson, M. J. & Eisenberg, D. (1999). *J. Mol. Biol.* **290**, 595–604.

Uson, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.

Vieille, C. & Zeikus, G. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 1–43.

Volkl, P., Markiewicz, P., Baikalov, C., Fitz-Gibbon, S., Stetter, K. O. & Miller, J. H. (1996). *Nucleic Acids Res.* **24**, 4373–4378.

Wagner, J., Kamiya, H. & Fuchs, R. P. P. (1997). *J. Mol. Biol.* **265**, 302–309.