Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of an aliphatic amidase from *Geobacillus pallidus* RAPc8

Serah W. Kimani,[a] Vinod B. Agarkar,[b] Donald A. Cowan,[b] Muhammed F.-R. Sayed[b] and B. Trevor Sewell[b,c]*

[a]Department of Molecular and Cell Biology, University of Cape Town, Rondebosch, South Africa, [b]Advanced Research Centre for Applied Microbiology, Department of Biotechnology, University of the Western Cape, Bellville, South Africa, and [c]Electron Microscope Unit, University of Cape Town, Rondebosch, South Africa

Correspondence e-mail: trevor.sewell@uct.ac.za

Received 21 May 2007
Accepted 4 August 2007

PDB Reference: aliphatic amidase, 2plq, r2plqsf.

The amidase from *Geobacillus pallidus* RAPc8, a moderate thermophile, is a member of the nitrilase superfamily and catalyzes the conversion of amides to the corresponding carboxylic acids and ammonia. It shows both amide-hydrolysis and acyl-transfer activities and also exhibits stereoselectivity for some enantiomeric substrates, thus making it a potentially important industrial catalyst. The crystal structure of *G. pallidus* RAPc8 amidase at a resolution of 1.9 Å was solved by molecular replacement from a crystal belonging to the primitive cubic space group $P4_232$. *G. pallidus* RAPc8 amidase is homohexameric in solution and its monomers have the typical nitrilase-superfamily α-β-β-α fold. Association in the hexamer preserves the eight-layered α-β-β-α:α-β-β-α structure across an interface which is conserved in the known members of the superfamily. The extended carboxy-terminal tail contributes to this conserved interface by interlocking the monomers. Analysis of the small active site of the *G. pallidus* RAPc8 amidase suggests that access of a water molecule to the catalytic triad (Cys, Glu, Lys) side chains would be impeded by the formation of the acyl intermediate. It is proposed that another active-site residue, Glu142, the position of which is conserved in the homologues, acts as a general base to catalyse the hydrolysis of this intermediate. The small size of the substrate-binding pocket also explains the specificity of this enzyme for short aliphatic amides and its asymmetry explains its enantioselectivity.

1. Introduction

Amidases belong to the hydrolase family (acylamide amidohydrolase subclass) and are specific for linear amides (EC 3.5.1) and cyclic amides (EC 3.5.2), catalyzing the hydrolysis of carboxylic acid amides to their corresponding free carboxylic acids and ammonia. They occur in both prokaryotes and eukaryotes. In many cases they occur in the same operon as nitrile hydratases (NHases), presumably to enable the hydrolysis of nitriles in a bi-enzymatic pathway (Fournand & Arnaud, 2001). The gene coding for the amidase from *Geobacillus pallidus* RAPc8 occurs in such an operon (Cameron *et al.*, 2005). Although the physiological role of amidases is still not clear, it has been suggested that they are involved in the metabolism of aldoximes (Kato *et al.*, 1999, 2000; reviewed by Cameron *et al.*, 2003), forming the third stage in the three-step pathway that involves aldoxime dehydratase, NHase and amidase.

Bacterial aliphatic amidases are the most extensively characterized, particularly as a consequence of their potential in the large-scale production of acrylic as well as other acidic products in industry (Hughes *et al.*, 1998). Amidases exhibit a

wide range of substrate specificities and some exhibit stereoselectivity (Mayaux *et al.*, 1990, 1991; Hashimoto *et al.*, 1991; Hirrlinger *et al.*, 1996), a property that can be exploited to allow the production of enantiopure acids that would be difficult to produce by other methods.

Although attempts to group amidases have been reported in a number of studies (Chebrou *et al.*, 1996; Pace & Brenner, 2001; Fournand & Arnaud, 2001), their classification has not been definitively formulated (Pertsovich *et al.*, 2005). However, based on the amino-acid sequence and structural similarities, bacterial aliphatic amidases can now be broadly divided into two groups. Group 1 consists of amidases that are structurally related to the nitrilase-superfamily enzymes, while group 2 mainly represents amidases belonging to the amidase-signature family. The latter are structurally unrelated to nitrilases (Pertsovich *et al.*, 2005; Banerjee *et al.*, 2002). Unlike signature amidases, which are either homodimeric or homo-octameric in their active form, most of the characterized nitrilase-related aliphatic amidases form homotetrameric and homohexameric complexes in solution (Pertsovich *et al.*, 2005).

Recent biochemical studies have shown that *G. pallidus* RAPc8 amidase exhibits both amide-hydrolysis and acyltransfer activities, having high specificity for short-chain aliphatic amides such as acrylamide, propionamide and acetamide (Makhongela *et al.*, 2007). This enzyme is moderately active on substituted short-chain and mid-length aliphatic amides and has no activity at all on long-chain aliphatic amides, aromatic substrates or urea amides. *G. pallidus* RAPc8 amidase exhibits substantial levels of chiral selectivity for D-enantiomers, particularly D-lactamide, with no activity on L-lactamide. This enzyme therefore has potential uses in the synthesis of chiral acids.

Based on primary structure comparisons, conserved structural motifs and biochemical characterization (Agarkar *et al.*, 2006; Makhongela *et al.*, 2007), *G. pallidus* RAPc8 amidase has been confirmed to be a member of the nitrilase superfamily of enzymes. Although it has an average of 19% sequence identity with enzymes of the nitrilase superfamily whose crystal structures had been determined at the time this work was performed (PDB codes 1j31, 1f89, 1erz and 1ems; Sakai *et al.*, 2004; Kumaran *et al.*, 2003; Nakai *et al.*, 2000; Pace *et al.*, 2000), a predicted secondary structure of *G. pallidus* RAPc8 amidase showed high structural homology with these enzymes (Agarkar *et al.*, 2006). Two additional amidase structures from *Helicobacter pylori* (PDB codes 2dyu, 2e2k and 2e2l) and from *Pseudomonas aeruginosa* (PDB code 2uxy), which have identities of 29% and 80%, respectively, to the *G. pallidus* enzyme and which were solved independently, have recently been reported (Hung *et al.*, 2007; Andrade *et al.*, 2007). Both of these enzymes have the extended carboxy-terminal sequence relative to the early crystal structures that is frequently found in the nitrilase branch of the superfamily.

A conserved Cys, Glu, Lys (CEK) catalytic triad is present in all the members of the nitrilase superfamily. Structural and kinetic evidence suggests that the Glu acts as a general base catalyst, increasing the nucleophilicity of the Cys so that it can attack the nitrile, carbamyl or amide group of the substrate. The proposed role of the lysine is to stabilize the tetrahedral intermediate so that the acyl intermediate can form with the loss of ammonia (Nakai *et al.*, 2000; Brenner, 2002). The active site of the *H. pylori* amidase (Hung *et al.*, 2007) was visualized both in the absence and presence of a formamide substrate by mutating the active-site Cys166 to serine. The proximity of another conserved glutamate which made contact with the formamide was recognized and it was proposed that its role was to maintain the side-chain geometry of the CEK triad as well as to facilitate the docking of the substrate into the correct site. In the mutant, the amide C atom is seen to be located 2.02 Å from the position of the Cys166 S atom in the wild-type enzyme and the orientation of the amide group in the active site prior to the nuclephilic attack is defined by interactions with both Glu59 and Glu141. The main-chain amide of the residue following the active-site cysteine is located so that it could stabilize the first tetrahedral intermediate prior to the formation of the acyl intermediate. The preparation protocol of the *P. aeruginosa* amidase (Andrade *et al.*, 2007) fortuitously allowed the visualization of the acyltransfer intermediate of acetamide onto hydroxylamine as a distorted tetrahedral intermediate, an acetohydroxamic acid derivative, with an occupancy of 0.6, giving substantial insight into the reaction mechanism.

The nitrilase branch of the superfamily has been shown to form homo-oligomeric spirals or helices (Sewell *et al.*, 2003; Thuku *et al.*, 2007). Enzymes from this branch have not thus far been crystallized, although homology modelling has been used to fit the low-resolution envelope determined by electron microscopy. An interesting property of the nitrilase from *Rhodococcus rhodochrous* J1 is that it is inactive as a dimer and only attains activity when the dimers associate to form spirals. Expanding the database of homologous structures is important both to characterize the interfaces and thus fully understand the spirals as well as to determine the mechanism of activation by spiral formation. The one-start spiral is constructed by association of the subunits across two interfaces: the 'A surface' and the 'C surface' (Sewell *et al.*, 2003). Mutation of residues in these surfaces leads to inactivation of the enzyme, providing further evidence of a connection between these interfaces and the active site (Sewell *et al.*, 2005). In addition, *G. pallidus* RAPc8 amidase, in common with the newly determined amidase structures, has an extended C-terminus that was absent in the earlier nitrilase-superfamily crystal structures but is present in most nitrilases (Agarkar *et al.*, 2006).

Analyses of both gel-filtration and native PAGE profiles indicated that *G. pallidus* RAPc8 amidase is a homohexamer in solution (Makhongela *et al.*, 2007). Electron microscopy showed that the complex had *D*3 symmetry, a finding confirmed by crystal packing after a *G. pallidus* RAPc8 amidase molecular-replacement solution was obtained using a mixed polyalanine structure of the hypothetical protein PH0642 from *Pyrococcus horikoshii* (PDB code 1j31; Sakai *et al.*, 2004) as a search probe as reported by Agarkar *et al.* (2006).

Table 1
Crystallographic information.

Values in parentheses are for the outer shell.

Data-collection statistics	
Wavelength (Å)	1.5418
Rotation range (°)	48
Space group	$P4_232$
Unit-cell parameters (Å, °)	$a = b = c = 130.39$, $\alpha = \beta = \gamma = 90.00$
Resolution range (Å)	46.1–1.90 (1.97–1.90)
Total observations	240879
Unique observations	30300
Completeness (%)	99.30 (94.80)
Redundancy	7.90 (3.90)
Signal-to-noise ratio [$I/\sigma(I)$]	18.27 (3.23)
R_{sym}† (%)	11.0 (31.5)
χ^2	0.97 (0.87)
Wilson plot average B factor (Å^2)	12.70
Matthews coefficient (Å^3 Da^{-1})	2.41
Solvent content (%)	49
Refinement statistics	
No. of non-H atoms	
Protein	2673
Water	304
No. of reflections	
Working set	28752
Test set (free)	1501
R factor‡/R_{free}§ (%)	14.50/17.50
R.m.s. deviations from ideality	
Bond lengths (Å)	0.01
Bond angles (°)	1.31
Average B value (Å^2)	10.56

† $R_{sym} = \sum|I - \langle I\rangle|/\sum\langle I\rangle$, in which I is a measured intensity and $\langle I\rangle$ is the average intensity from multiple measurements of symmetry-related reflections. ‡ R factor = $\sum||F_o| - |F_c||/\sum|F_o|$, where $|F_o|$ and $|F_c|$ are observed and calculated structure-factor amplitudes, respectively. § R_{free} was calculated using a random set containing 5% of observations which were omitted during refinement.

This paper reports the refinement of the *G. pallidus* RAPc8 amidase crystal structure and presents a detailed analysis of the structure and its implications for the mechanism, specificity and subunit association. In particular, analysis of the active site suggests a possible further catalytic role for the conserved glutamate residue previously implicated by Hung *et al.* (2007).

2. Materials and methods

2.1. Structure determination and refinement

The structure was solved by molecular replacement in the primitive cubic space group $P4_232$ using a model based on PDB entry 1j31, with non-identical residues (78%) replaced by alanine as described by Agarkar *et al.* (2006). Rigid-body refinement was performed with *Phaser* (Collaborative Computational Project, Number 4, 1994; Read, 2001) and an initial electron-density map was generated using *REFMAC*5 (Collaborative Computational Project, Number 4, 1994; Murshudov *et al.*, 1997). We were unable to interpret this map, which had a phasing figure of merit of 0.19, and we employed a number of strategies to improve it. The first interpretable map (which had a phasing figure of merit of 0.31) was generated with phases calculated from the average of all four known homologues. The initial model (1j31 substituted with *G. pallidus* RAPc8 amidase side chains using *SCWRL*; Canutescu *et al.*, 2003) was automatically rebuilt using *RESOLVE* (Terwilliger, 2004*a*,*b*) within *PHENIX* (Adams *et al.*, 2002). 20 cycles involving prime-and-switch phasing, solvent flattening (assuming a 50% solvent content) and automatic model building were executed. As a result, the figure of merit of the phases increased to 0.55 and the map was readily interpretable. Alternating cycles (23 in all) of manual model building using *O* (Jones *et al.*, 1990) and restrained refinement using *REFMAC*5 were repeated until 340 of the 348 expressed amino acids had been located and the R factor in the resolution range 46.1–1.90 Å was reduced to 14.5% (Table 1).

2.2. Model validation

The accuracy, precision and correctness of the refined *G. pallidus* RAPc8 amidase model was assessed using various validation tools, including *PROCHECK* (Collaborative Computational Project, Number 4, 1994; Laskowski *et al.*, 1993), *WHATCHECK* (Hooft *et al.*, 1996) and *MOLPROBITY* (Lovell *et al.*, 2003).

2.3. Structure analysis

The online *CASTp* (Computed Atlas of Surface Topography of proteins) server (Binkowski *et al.*, 2003) was used to analyze the pockets and cavities on the surface and interior of the complete *G. pallidus* RAPc8 amidase model. The *BAVERAGE* program (Collaborative Computational Project, Number 4, 1994) was used to calculate average B factors for the final model for both the main chain and the side chains. The topology diagrams were generated with *TOPS* (Westhead *et al.*, 1999) and modified using *TOPDRAW* (Bond, 2003). Figures were drawn with *PyMOL* (DeLano, 2004) unless otherwise stated in the figure legend.

2.4. Molecular modelling

The models of the acyl intermediates were created with *Accelrys Discovery Studio* 1.7 and the energies of the structures were minimized using *CHARMM* (Brooks *et al.*, 1983) and the Accelrys CHARMm forcefield.

3. Results and discussion

3.1. Model rebuilding and refinement

The final model accounted for residues 1–340 of the 348 in the translated sequence. No electron density was observed for the eight C-terminal amino acids. The reflection data to 1.9 Å resolution allowed the location of 304 water molecules in the final structure. Overall, the refined *G. pallidus* RAPc8 amidase model was of high quality, showing good agreement with the diffraction data and minimal errors. The refinement statistics are shown in Table 1.

3.2. The monomer fold

The monomer of *G. pallidus* RAPc8 amidase is an α/β globular protein that consists of 11 α-helices and 14 β-strands

interconnected by external loops. A cartoon representation of a *G. pallidus* RAPc8 amidase subunit and the topology of the β-sheets and α-helices are shown in Fig. 1. Helices α1, α2 and α3 form one outer helical layer, while helices α4, α5 and α6 form the other. β-strands β1, β2, β3, β4, β5 and β14 form one β-sheet, while β-strands β6, β7, β8, β9, β12 and β13 form the other sheet. The two six-stranded β-sheets are sandwiched between the two helical layers, forming a four-layer α-β-β-α sandwich architecture (Nakai *et al.*, 2000). The two β-sheet layers have a similar alternation of parallel and antiparallel β-strands; the three central β-strands of each β-sheet (β1, β2, β3 and β7, β8, β9) form a parallel β-sheet, while the rest of the strands are arranged in an antiparallel configuration. The interface between the two β-sheets is tightly packed with hydrophobic residues. Helices α8, α9, α10 and α11 are found on the extended C-terminal region, while β-strands β10 and β11 are on a long external loop that links β9 and β12. Four residues form a 3_{10}-helix (shown in cyan in Fig. 1) at the active site between β7 and α5. The two α/β halves are connected by two loops, one at each end: a 20-residue loop between β5 and β6 on one end and an eight-residue loop between β13 and β14 on the other end. A structure-similarity search of the DALI database (Holm & Sander, 1993) with *G. pallidus* RAPc8 amidase identified the four nitrilase-superfamily structures (PDB codes 1j31, 1f89, 1erz and 1ems) with high structural similarity (average Z score of 32.0). The structural similarity with other reported α/β four-layer sandwich proteins was poor.

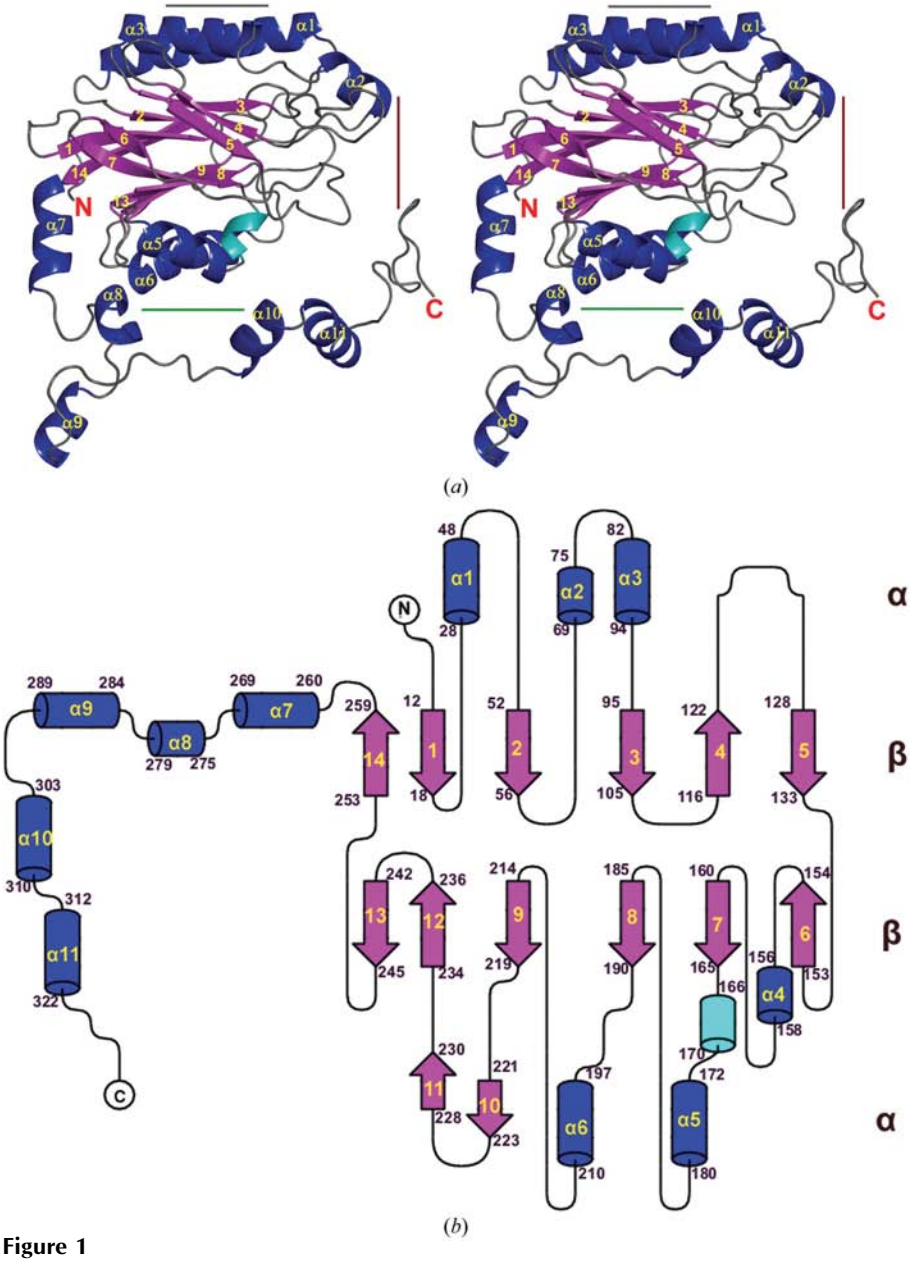

Figure 1
The fold of *G. pallidus* RAPc8 amidase. (*a*) Stereoview of a cartoon representation of a *G. pallidus* RAPc8 amidase monomer. The bars indicate the locations of the interacting surfaces seen in the spiral-forming homologues. The green bar indicates the 'A surface', the red bar the 'C surface' and the grey bar the 'D surface', following the nomenclature of Sewell *et al.* (2005). (*b*) α-Helix and β-sheet topology of *G. pallidus* RAPc8 amidase. β-Sheets (labelled 1–14) are shown as purple arrows, while α-helices are shown as blue cylinders. The cyan cylinder in the topology diagram is the 3_{10}-helix on which the active-site Cys166 resides. The secondary-structure elements are numbered in sequence from the N-terminus. The topology diagram was generated using the program *TOPS* (Westhead *et al.*, 1999) and manually redrawn in *TOPDRAW* (Bond, 2003) for simplification.

3.3. The hexamer

As suggested earlier by gel-filtration chromatography, native PAGE and electron microscopy (Makhongela *et al.*, 2007), *G. pallidus* RAPc8 amidase exists in solution as a homohexamer (Fig. 2) having *D*3 point-group symmetry. It measures 85 Å parallel to the threefold axis and 94 Å at the widest point perpendicular to the threefold axis (Fig. 2*a*). The monomers associate with one another across two major interfaces located approximately at right angles, as can be clearly seen from the cylindrical projection of density around the threefold axis (Fig. 2*b*). The subunit interfaces 1–2, 3–4 and 5–6 correspond to the 'A surface' in the nomenclature of Sewell *et al.* (2003). The C-terminal sequence comprises a series of short α-helices which also contribute to this interface by forming an interlock (Fig. 2*c*). Interfaces 2–3, 4–5 and 6–1 are on the opposite side of the monomer to the 'C surface' seen in the spiral structures (Sewell *et al.*, 2003; Thuku *et al.*, 2007) and primarily involve interactions between residues in α7 and the N-terminal loop (Fig. 2*d*). A key difference between the spiral structures visualized to date and the amidase is

that the C-terminus is positioned on the inside of the spiral whereas it on the outside of the hexamer. Association across the 'A surface', which seems to be conserved in the nitrilase superfamily, results in an eight-layered α-β-β-α:α-β-β-α sandwich structure.

3.4. Interfaces

The two types of interacting surfaces are located on opposite sides of the hexamer along each of the three twofold axes in such a way that interface 1–2 is opposite 4–5, 3–4 is opposite 6–1 and 5–6 is opposite 2–3. The 2–3, 4–5 and 6–1 surfaces are formed by helices α7 and the N-terminal loops of the two interacting monomers. The electrostatic interactions which contribute to this interface are listed in Table 2 and illustrated in Fig. 3(*c*).

As in the homologous structures, the conserved 1–2, 3–4 and 5–6 interfaces ('A surfaces') in *G. pallidus* RAPc8 amidase are formed by helices α5 (residues 172–186) and α6 (residues 197–210) from interacting monomers, as well as the C-terminal tails (residues 273–340) that form the interlock described above. The helices α5 and α6 pack against the corresponding elements in the twofold-related monomer as shown in Fig. 3(*a*) in such a way that their hydrophobic residues (especially Met202, Met203 and Met207) intermesh. In addition, Glu173

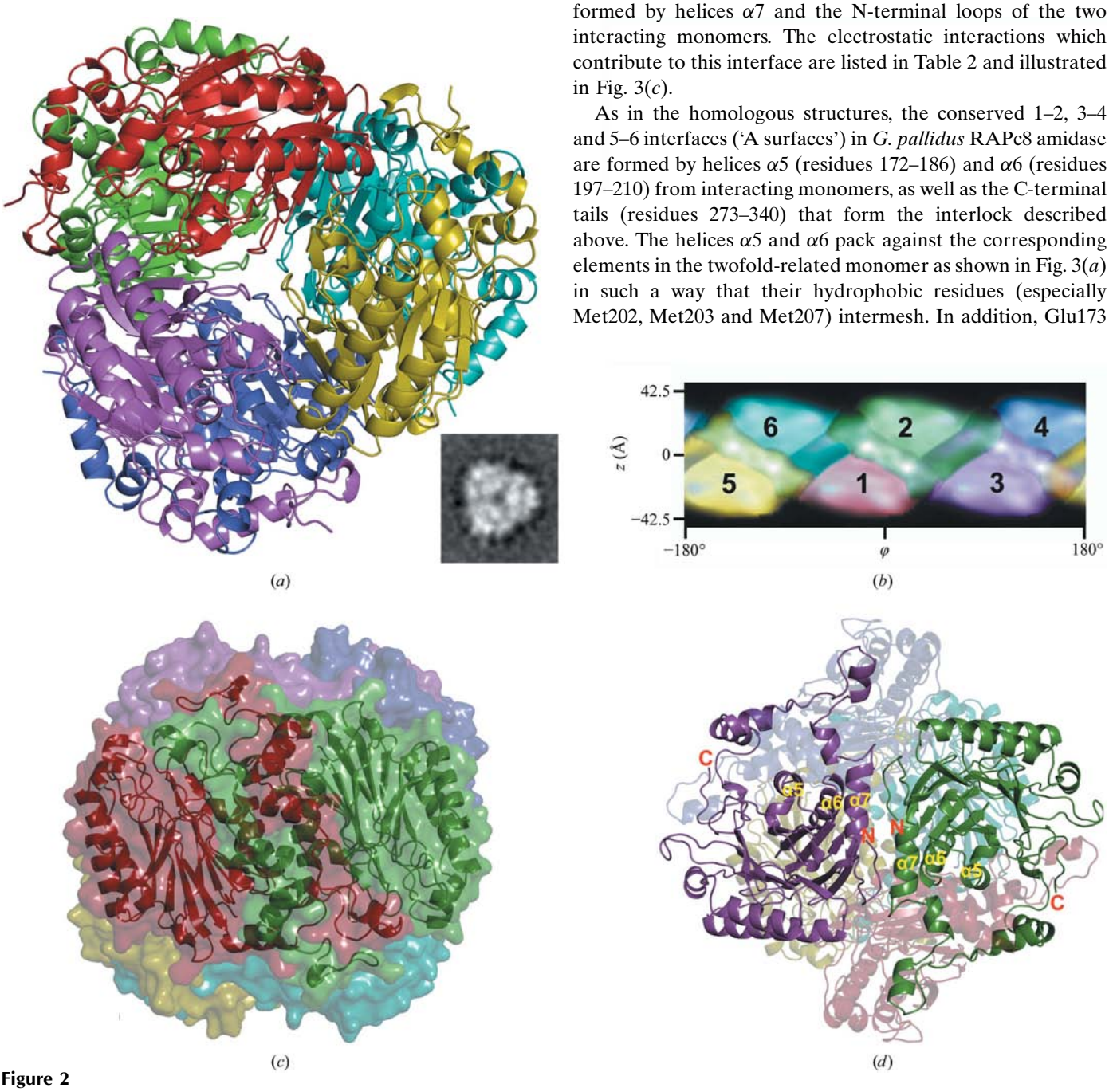

Figure 2
Views of *G. pallidus* RAPc8 amidase homohexameric structure. (*a*) A cartoon representation of the biological complex viewed down the crystallographic threefold axis. The hexamer is composed of a trimer of dimers. The three twofold axes are perpendicular to the threefold axis, giving rise to a complex with *D*3 point-group symmetry having a projection that closely resembles that observed by negative-stain electron microscopy (Makhongela *et al.*, 2007; insert). (*b*) A cylindrical projection of the density of the hexamer, with the cylinder axis and the threefold axis aligned. The density of each monomer was projected separately, coloured and then combined to form the composite image. The monomers are labelled and coloured in a manner consistent with (*a*). The value of the projected density is higher in lighter areas. The conserved 'A surfaces' link 1–2, 3–4 and 5–6. A second twofold interacting surface links 1–6, 2–3 and 4–5. (*c*) A transparent surface representation of the hexamer viewed down the twofold axis linking monomers 1 and 2 in the *G. pallidus* RAPc8 amidase hexameric complex. The monomers 1 (red) and 2 (green) are also shown as cartoons which together with the surface clearly illustrate the interlock formed by the C-terminal sequences in the hexamer as well as the contact of the tails with the core structure. (*d*) Cartoon image of the second twofold interface. Interaction between monomer 2 (green) and monomer 3 (magenta) viewed down a twofold axis. Helices α7 and the N-terminal loops are labelled. The conserved dimer interface ('A surface') involves helices α5 and α6, which are labelled.

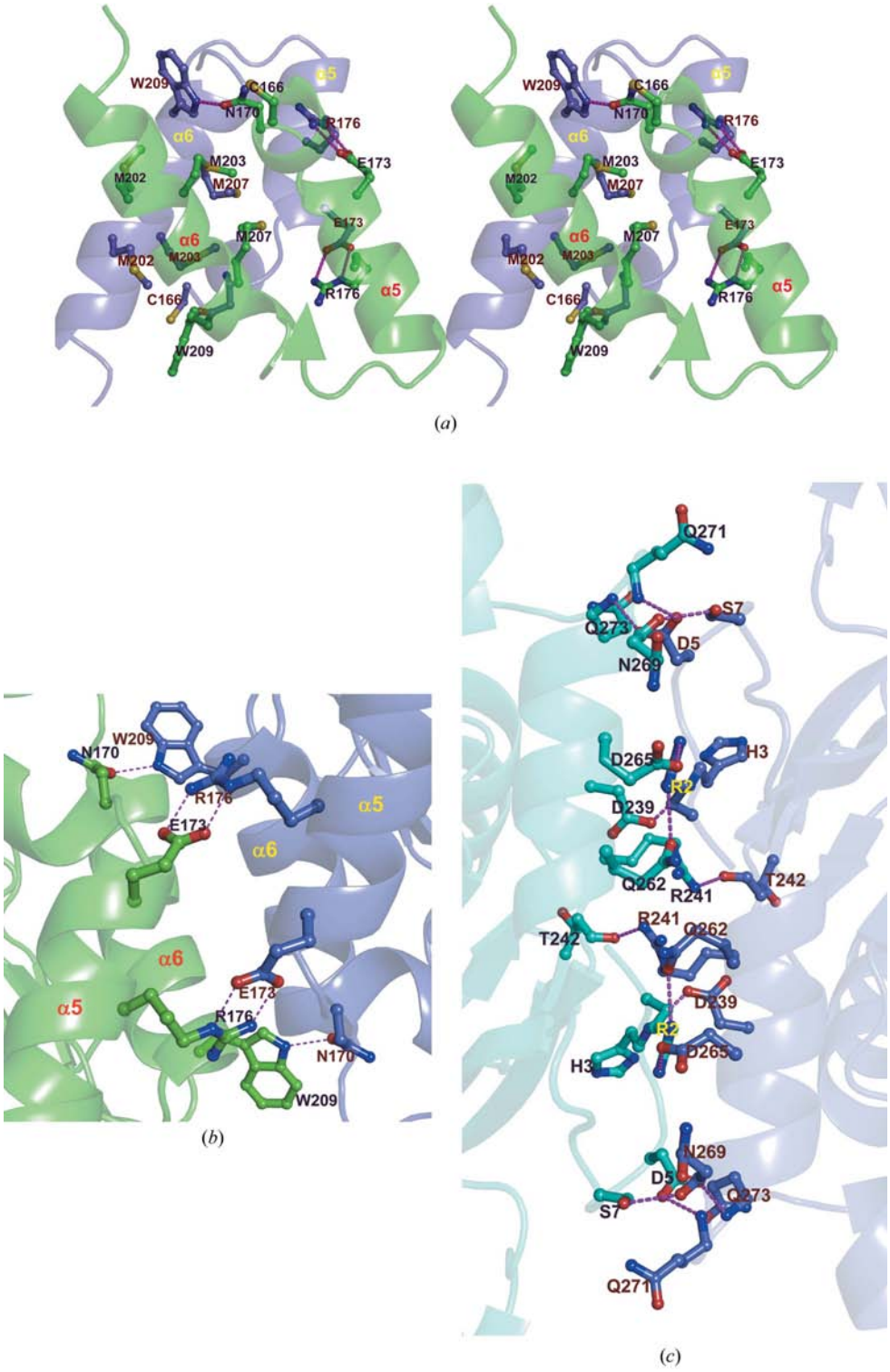

Figure 3
Details of the interactions at the interfaces. (*a*) A stereo diagram of the 'A surface' interactions viewed perpendicular to the twofold axis. Helices from monomer 1 are coloured green, while those from monomer 2 are coloured blue. The figure shows the methionine packing in α6 and the connection between the interface and the active site through the hydrogen bond between Trp209 and Asn170, which stabilizes the 3_{10}-helix on which Cys166 is located. (*b*) The details of the 'A surface' interaction between helices α5 viewed parallel to the twofold axis. Arg176 and Glu173 form salt bridges linking the subunits. Trp209 and Asn170 contribute a hydrogen bond to the interface. (*c*) Interactions of helices α7 and the N-terminal loops in monomers 2 (blue) and 3 (cyan) on the second twofold-related interface. This interface is stabilized mainly by electrostatic interactions, with two salt bridges between Asp265 and Arg2 of both subunits.

and Arg176 form salt bridges (Fig. 3*b*) between the paired subunits. Hydrogen bonds also exist between $N^{\varepsilon 1}$ of Trp209 located in α6 and $O^{\delta 1}$ of Asn170 located in the short segment of 3_{10}-helix supporting the active-site Cys166 in the symmetry-related subunit. This interaction links this interface to the structure of the active site and possibly provides a reason why the formation of this interface is obligatory for activity. The electrostatic interactions contributing to the 'A surface' are listed in Table 3.

3.5. The extended C-terminal sequence

Residues 272–340 of the amidase form an extended C-terminal tail relative to the non-amidase homologue structures. The tail sequence is folded into four helices, α8, α9, α10 and α11, which together interlock near the 'A surface' with the equivalent construct from the twofold-related monomer. The tails are located on the outside of the hexamer. Several electrostatic interactions (Table 4) occur between the interlocked tails, most notably two salt bridges between Arg295 and Glu320. Electrostatic contacts also exist between the tails and the core structures.

The interlocking of the C-terminal tails could play roles in strengthening the 'A surface' and in linking the monomers in the process of dimer formation. The latter role was first proposed by Nakai *et al.* (2000) in the structure of *N*-carbamyl-D-amino-acid amidohydrolase (DCase; PDB code 1erz). However, the C-terminal tails in *G. pallidus* RAPc8 amidase are much longer, a feature that is also found in other amidases and the helical nitrilases, cyanide hydratases and cyanide dihydratases. However, the lack of sequence homology

Table 2
Electrostatic interactions and hydrogen bonds at the twofold-interacting surface formed by the α7 and N-terminal loops.

Subunit 2	Subunit 3	Distance (Å)
Arg2 $N^{\eta1}$	Gln262 $O^{\varepsilon1}$	2.94
Arg2 $N^{\eta2}$	Asp265 $O^{\delta2}$	2.95
His3 N	Asp239 $O^{\delta2}$	2.94
Asp5 $O^{\delta1}$	Gln273 N	2.71
Asp5 $O^{\delta2}$	Gln271 N	2.85
Ser7 O^{γ}	Asn269 O	2.98
Ser9 O^{γ}	Asp265 $O^{\delta2}$	2.67
Arg241 $N^{\eta1}$	Thr242 O	2.98

Table 3
Electrostatic interactions at the 'A surface'.

Subunit 1	Subunit 2	Distance (Å)
Arg176 $N^{\eta2}$	Glu173 $O^{\varepsilon1}$	2.89
Arg176 N^{ε}	Glu173 $O^{\varepsilon2}$	2.80
Trp209 $N^{\varepsilon1}$	Asn170 $O^{\delta1}$	2.88

Table 4
Electrostatic interactions and hydrogen bonds in the C-terminal interlock.

Subunit 1	Subunit 2	Distance (Å)
Lys113 N^{ζ}	Glu292 $O^{\varepsilon2}$	2.72
Ala114 N	Ser290 O^{γ}	2.93
Trp138 $N^{\varepsilon1}$	Gln273 $O^{\varepsilon1}$	2.97
Tyr145 O^{η}	Glu292 $O^{\varepsilon1}$	2.61
Asp167 $O^{\delta2}$	Lys278 N^{ζ}	2.70
Asp177 $O^{\delta1}$	Phe304 N	2.80
Ala270 O	Arg324 $N^{\eta1}$	2.85
Gln273 $O^{\varepsilon1}$	Tyr192 O^{η}	2.69
Arg295 $N^{\eta2}$	Glu320 $O^{\varepsilon2}$	2.68
Ala298 O	Trp308 $N^{\varepsilon1}$	2.88
Pro140 O	Thr285 $O^{\gamma1}$	2.65
Asn269 $O^{\delta1}$	Thr329 $O^{\gamma1}$	2.70

and even predicted secondary-structural homology in the C-terminal tail of these non-amidase species suggests the possibility of high structural variability in this region.

Further interesting crystal-packing interactions occur at the inter-hexamer interface. Lys36 and Glu82 form a pair of symmetrically related salt bridges with the corresponding residues on a neighbouring hexamer. This interaction is possibly analogous to that occurring at the 'D surface' in the helical nitrilases (Sewell *et al.*, 2005). These interactions may also account for the tendency of the hexamers to form the long rows visualized by electron microscopy (Makhongela *et al.*, 2007).

3.6. The active-site environment

G. pallidus RAPc8 amidase shares structural homology with the members of the nitrilase superfamily. One of the conserved motifs in the nitrilase enzymes is the cysteine, glutamate and lysine (CEK) catalytic triad (Novo *et al.*, 1995; Pace & Brenner, 2001) that is involved in the hydrolysis of the carbon–nitrogen bond. In *G. pallidus* RAPc8 amidase, the conserved CEK catalytic triad is comprised of Glu59, Lys134 and Cys166. The arrangement of these side chains is similar to that of the other nitrilase-superfamily structural homologues (Fig. 4*a*).

A buried cleft corresponding to the binding pocket was identified and characterized in each subunit of the *G. pallidus* RAPc8 amidase structure by *CASTp* (Binkowski *et al.*, 2003). The pocket is formed by residues in the loops between strand β2 and the small helix α2, strands β5 and β6, strand β7 and helix α5, as well as strand β8 and helix α6, and is located near the edge of the dimer interface. This active-site location is similar to all α/β four-layer sandwich structures including nitrilase-superfamily enzymes.

The *G. pallidus* RAPc8 amidase active-site cleft is lined by residues Glu59, Tyr60, Trp138, Glu142, Cys166, Gly191, Tyr192 and Met193 as depicted in Fig. 4(*b*). A hydrogen-bonding network exists that probably plays a role in maintaining a highly constrained configuration of the residues in the substrate-binding site. The catalytic residue Cys166 is located on a β-strand–turn–helix structural motif formed by strand β7 and helix α5. This motif is termed a nucleophilic elbow (Kumaran *et al.*, 2003) and is found in many α/β four-layer sandwich hydrolases (Schrag & Cygler, 1997). As for the catalytic cysteines in the other CN-hydrolase structures, Cys166 has a rare but allowed combination of peptide-bond dihedral angles that falls in the recently described type II′ turn conformation (Lovell *et al.*, 2003) region of the Ramachandran plot. The carbonyl O atom of Cys166 forms hydrogen bonds with the N^{ε} (3.02 Å) and $N^{\eta2}$ (2.83 Å) atoms of Arg188; this hydrogen-bonding network may play a role in maintaining the orientation of the Cys166 side chain in the active site.

3.7. Cys166 modification

Unlike other nitrilase-superfamily structures in which a number of bound water molecules believed to participate in catalysis are present in the active sites, no waters were found in the active-site pocket of *G. pallidus* RAPc8 amidase. However, both $2F_{\rm obs} - F_{\rm calc}$ and positive difference electron-density maps had density linked to Cys166, indicating that it was modified, possibly by oxidation, in the crystal (Fig. 5). This density very closely resembles that seen in the case of the DJ-1 protein (Wilson *et al.*, 2004), which contains a Glu, His, Cys catalytic triad. Attempts to fit this density either as a sulfinic acid or a sulfenic acid did not refine unambiguously. However, refinement did suggest that the density was produced by more than one atom. We therefore suggest that the Cys166 oxidative modification is a mixture of species, following the example of Wilson *et al.* (2004). The importance of this modification is uncertain, but any form of oxidation would decrease the nucleophilicity of the S^{γ} atom in the modified cysteine residue; this may explain why *G. pallidus* RAPc8 amidase enzyme progressively lost activity during purification in the absence of reducing agents (Makhongela *et al.*, 2007). We reject the possibility of another covalent adduct arising from the mother liquor, which contained citrate and acetate, as citrate is too large and acetate is known not to be an inhibitor of the enzyme. An artefact of the type described by Andrade *et al.*

(2007) is also unlikely as no substrates were used in the purification process.

3.8. Structural basis of substrate specificity

Unlike other nitrilase-superfamily structures, where the active-site clefts are reported to be large and deep (Kumaran *et al.*, 2003; Nakai *et al.*, 2000; Sakai *et al.*, 2004), the *G. pallidus* RAPc8 amidase binding pocket is relatively shallow, with a volume of 53.59 Å^3. For comparison, the DCase substrate-binding pocket has a volume of 1134.7 Å^3, making it over 20 times larger than the *G. pallidus* RAPc8 amidase pocket. The small size of the active-site pocket may explain the specificity of *G. pallidus* RAPc8 amidase for short aliphatic amides as observed by Makhongela *et al.* (2007). The bulky side chain of Trp138 covers the binding pocket, such that analysis by *CASTp* (Binkowski *et al.*, 2003) could not locate any openings of sufficient size for the entry of substrate. We conclude that access to the active site is dependent on the movement of residues in the loop Pro137–Gly143. In this loop, Trp138 and Ile141 are packed against Tyr227 and Met193. Two small openings to the outside environment pass Tyr192 on one side of Trp138 and Tyr67 on the other.

The cavity itself is highly asymmetric. Useful insight can be obtained by modelling the acyl intermediates of some of the substrates studied by Makhongela *et al.* (2007). A small pocket is available near Lys134 to accommodate the carbonyl of the acyl intermediate. This location is filled by one of the O atoms of the putative sulfinic acid and orientates the plane of the acyl carbonyl. The bulk of the substrate extends up towards Trp138. The clash with Trp138 limits the length of the unbranched substrate to four C atoms and would certainly not allow aromatic substrates.

The geometry of the active site provides a satisfactory explanation for the observed enantioselectivity of *G. pallidus* RAPc8 amidase. It is clear that the hydroxyl group of the acyl intermediate formed from L-lactamide would clash with the carboxyl of Glu142, whereas no clash would occur in the case of D-lactamide.

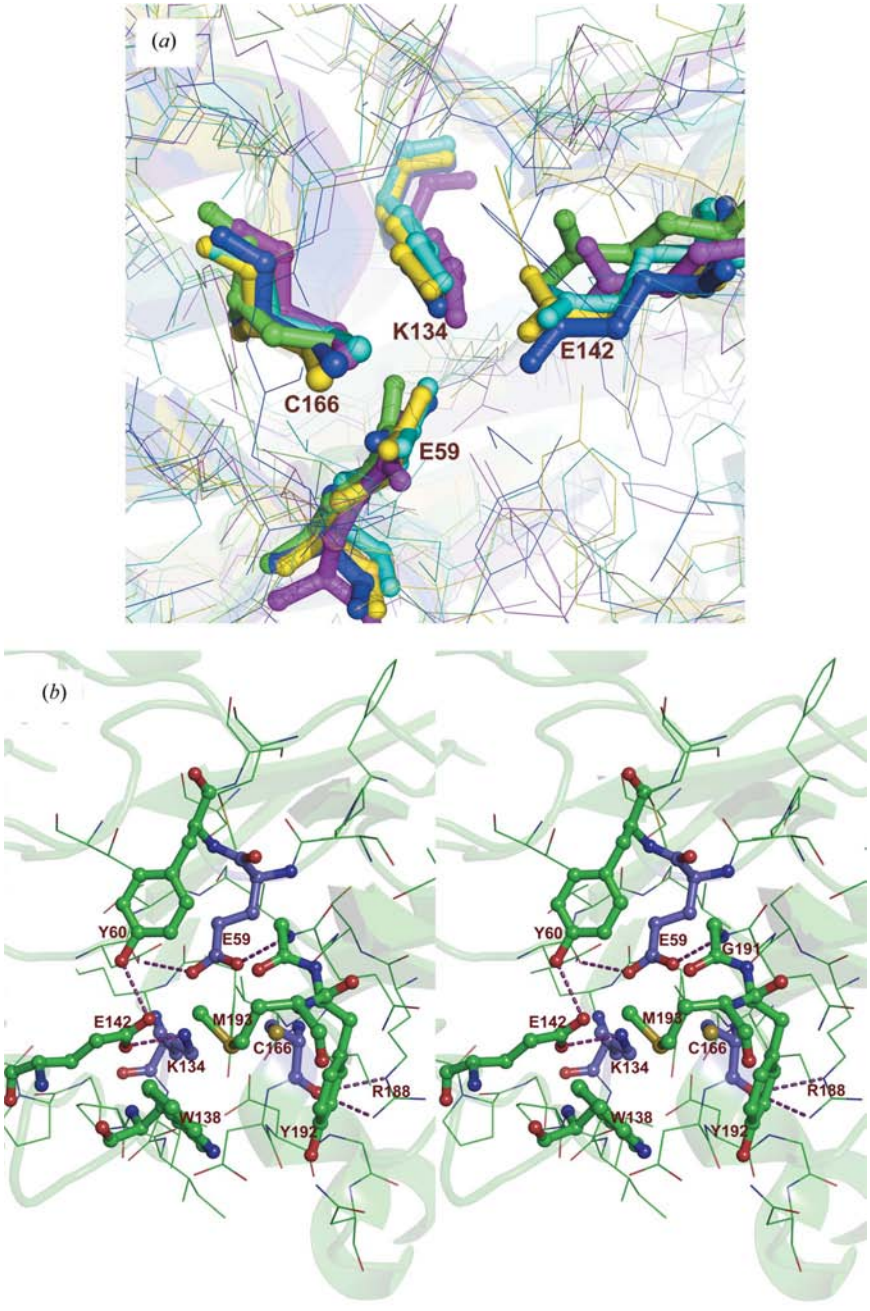

Figure 4
(*a*) The arrangement of the established catalytic triad residues and Glu142 in *G. pallidus* RAPc8 amidase compared (Cohen, 1997) with homologous nitrilase-superfamily structures (PDB codes 1j31, 1f89, 1erz and 1ems). 1j31 is in yellow, 1ems in green, 1erz in cyan, 1f89 (modified as described in Appendix *A*) in magenta and *G. pallidus* RAPc8 amidase in blue. The catalytic triad residues (Glu59, Lys134, Cys166) and Glu142 have similar positions to those of the corresponding residues in the four non-amidase structures, indicating their role in driving catalysis in the superfamily. (*b*) Stereoview of the active site, showing the network of probable electrostatic interactions. The catalytic triad residues are shown in blue.

3.9. Implications for the reaction mechanism

The small number of exposed residues in the active site of *G. pallidus* RAPc8 amidase severely limits the choice of catalytic residues. The conventional wisdom (Nakai *et al.*, 2000; Brenner, 2002), shown in Fig. 6, that Cys166 acts as a nucleophile, Glu59 acts as a general base catalyst and that Lys134 stabilizes the tetrahedral intermediate state. This seems to be confirmed by the placement of these residues which coincide closely with those of the other nitrilase homologues (Fig. 4*b*).

The acyl intermediate (Fig. 6) formed by these interactions must then be hydrolysed. The geometry of the active site in the case of the *G. pallidus* RAPc8 amidase suggests that Glu59 is not a candidate for the general base catalyst as it is inaccessible to the aqueous environment, being buried behind the acyl intermediate (Fig. 7).

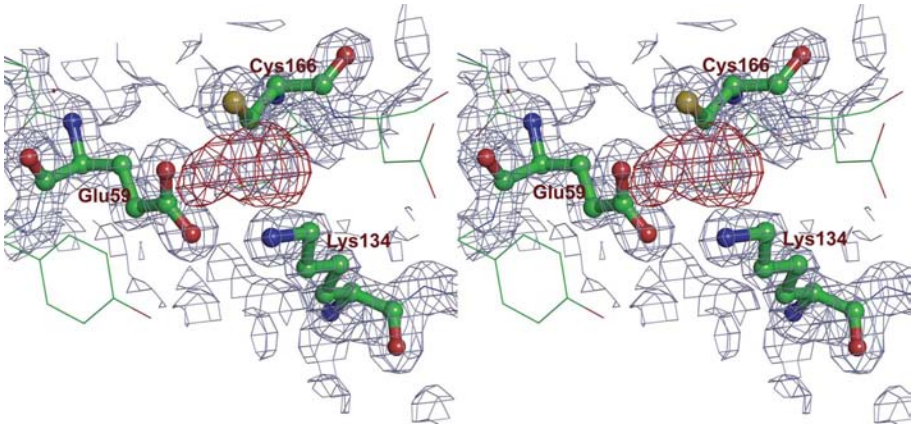

Figure 5
A stereoview of the electron-density maps around the active-site residues Glu59, Lys134 and Cys166. The $2F_{obs} - F_{calc}$ map contoured at 1.3 is shown in blue, while the positive $F_{obs} - F_{calc}$ difference map contoured at 3.0σ is shown in red. The positive difference electron density around the S^{γ} atom of Cys166 suggests oxidative modification, probably to a mixture of species including sulfinic acid.

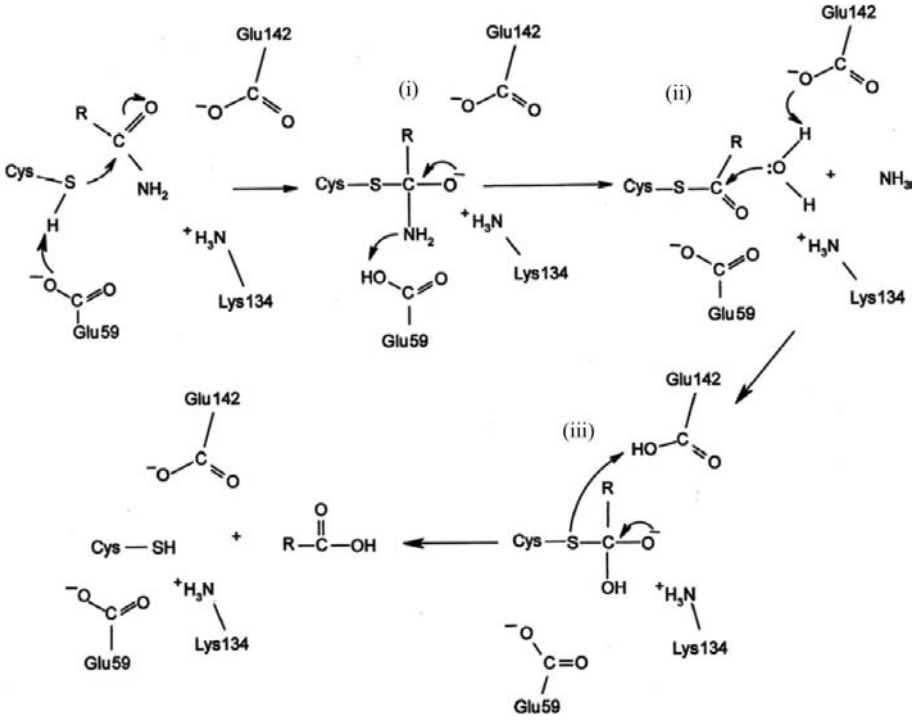

Figure 6
A putative mechanism for the *G. pallidus* RAPc8 amidase, illustrating the proposed role of Glu142. Cys166 acts as a nucleophile, Glu59 acts as a general base catalyst and Lys134 stabilizes the tetrahedral intermediate (i). It is proposed that Glu142 acts as a general base catalyst, enhancing the nucleophilicity of the water which hydrolyses the acyl intermediate (ii). The positioning of Glu142 at a distance of 3.7 Å from the acyl carbon is in accordance with the necessary stereoelectronic considerations and is suggested because Glu59 is buried beneath the acyl group. The source of the proton in step (iii) is speculative, as the distance between Cys166 S^{γ} and the carboxylate O atom of Glu142 is 5.3 Å.

A second glutamate residue (Glu142) is present in the active site. It has its carboxyl O atom located 5.30 Å from the sulfur of Cys166 and 3.72 Å from a carboxyl O atom of Glu59, which in turn has its other carboxyl O atom located 3.46 Å from the S atom of Cys166. The work of Hung *et al.* (2007) has demonstrated its role in positioning the substrate. The tetrahedral intermediate (Fig. 6) is possibly stabilized by Lys134 and the main-chain amino group of residue 167 (Hung *et al.*, 2007; Andrade *et al.*, 2007). This leads to the formation of an acyl intermediate, with the production of ammonia.

Modelling the acyl intermediate in the *G. pallidus* RAPc8 amidase structure gives insight into the mechanism of its hydrolysis and suggests a hypothesis for a further role for Glu142. The additional density that we have interpreted as oxidation of Cys166 is located in such a way that it blocks access of both Glu59 and Lys134 to the outside environment. If the acyl intermediate is modelled so as to partially fill this density, then it too would block access to these residues. Glu142 is located further towards the outside than the acyl carbon, at a distance of about 3.7 Å from its putative position, making it a readily available candidate for the general base catalyst. The geometry of the modelled active site would also allow a collinear trajectory of the lone pair of electrons in the sp^3 orbital of the water and the sp^2 orbital of the acyl intermediate, which would facilitate hydrolysis.

Examination of the sequences of large numbers of nitrilases and nitrilase homologues, aligned on the basis of structure and guided by the structure of the amidase, clearly shows that this residue is conserved. It is also relevant that despite differing secondary structure in the homologues, with the exception of the putative CN hydrolase from yeast (Kumaran *et al.*, 2003) the position of this residue is conserved. A re-examination of this structure suggests (see Appendix *A*) that it can be reinterpreted so as to remove this exception. Glu142 is located in a loop that is directed into the helix, forming the 'C surface' identified by Sewell *et al.* (2003). A role for this residue in the catalytic mechanism would explain why dimer association is necessary for activity (Nagasawa *et al.*, 2000),

suggesting a model in which the formation of the 'C surface' locates the residue at the correct position for it to play a role in catalysis. No such obvious connection is suggested in the case of the other catalytic residues.

4. Conclusions

The crystal structure of *G. pallidus* RAPc8 amidase has revealed high structural homology to solved structures of other nitrilase-superfamily enzymes, especially the other two recently published structures of amidases from different species. The similarity to the amidase from *Pseudomonas aeruginosa* (Andrade *et al.*, 2007) is remarkable, with a C^{α} r.m.s. deviation of only 0.26 Å. The only significant difference is in the region of helix α3, which is involved in inter-hexamer crystal-packing interactions in both crystal forms. The amidase from *Helicobacter pylori* (Hung *et al.*, 2007) is slightly less similar, with a C^{α} r.m.s. deviation of 1.21 Å. In each of the three amidases the active site is visualized in a different state, giving insight into substrate binding, the formation of the acyl intermediate and its hydrolysis. The CEK catalytic triad also has a similar configuration to a carbamoylase (Chen *et al.*, 2003) and a putative nitrilase despite the fact that the enzymes catalyse different reactions; further studies will be necessary to understand what differentiates the catalytic mechanisms. Comparison of the interacting surfaces in the amidase with those of the spiral and helical members of the nitrilase family studied by us shows that the dyad interface observed across the members in the family is conserved. On the other hand, the interactions leading to the formation of the threefold axis occur on the opposite side of the monomer such that the C-terminus is exposed on the outside rather than the inside of the complex. *G. pallidus* RAPc8 amidase has a C-terminal sequence that is slightly longer than that of the helical members of the family and that forms an interlock by interacting with the opposing monomer across the dyad interface. However, lack of sequence conservation or even predicted secondary structure in this region suggests that caution is advised in general conclusions about this feature of the molecule. A pocket corresponding to the active-site cleft of *G. pallidus* RAPc8 amidase has been identified that is small enough to explain the specificity of this enzyme for short aliphatic amides and the enantioselectivity of the enzyme. We speculate that the structure suggests the involvement of Glu142 in catalysing the nucleophilic attack of water on the acyl intermediate because of its location in the constricted environment of the active-site pocket. We further speculate that association in the vicinity of the loop in which Glu142 is situated is necessary for positioning this residue and hence activity in the case of the nitrilase from *Rhodococcus rhodochrous* J1.

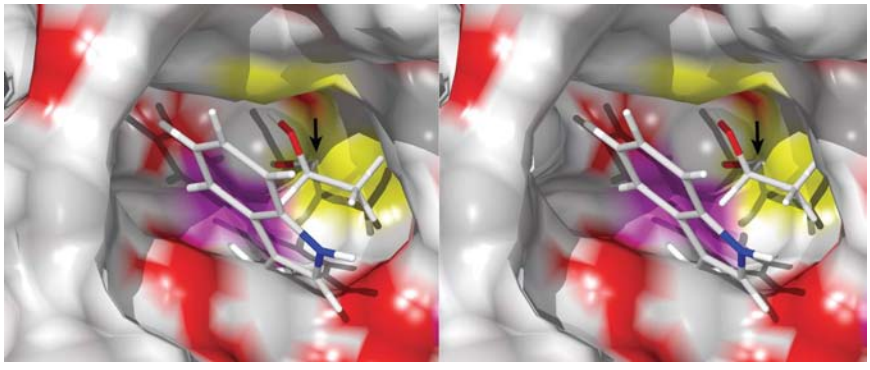

Figure 7
A stereo surface rendering of the active-site pocket lying behind Trp138. The surface is coloured in CPK colours corresponding to the exposed surface atoms. The two magenta patches directly behind the wire rendition of Trp138 represent the carboxyl O atoms of Glu142. The acyl intermediate formed after reaction with D-lactamide has been modelled and is also illustrated as a wire rendition. The location of the sp^2 acyl carbon is indicated by the black arrow. The yellow surface behind the acyl carbon represents the sulfur of Cys166. The acyl oxygen is located in a pocket near Lys134. Glu59 is located directly behind the acyl oxygen and is inaccessible to solvent. It can be seen clearly that the L-lactamide enantiomer would produce a clash between the hydroxyl and the carboxyl of Glu142. The image was drawn with *UCSF Chimera* (Pettersen *et al.*, 2004).

APPENDIX *A*

The role we propose for Glu142 depends in part on the interpretation of the location of Glu144 in the structure of the yeast putative CN-hydrolase protein (PDB code 1f89; Kumaran *et al.*, 2003) being questionable. The map of 1f89 was regenerated from the deposited coordinates and structure factors and the density in this region was examined in detail. It is worth noting that there is no density at the confidence level of the rest of the map to support the nonconsensus published location of Glu144 (both backbone and side chain). Indeed, clear density is not visible for any residues between 132 and 146. Positive difference density at the 2σ level is clearly visible at the consensus location of Glu144.

It is worth noting that Kumaran *et al.* (2003) did not position residues 132–141. The sequence in this region (KVHLFDVIDPNGISFHESETL) is very similar and identical in length to that of the Nit domain of NitFhit (PDB code 1ems; Pace *et al.*, 2000; KLHLFDLEIPGKVRLMESEFS). In 1ems this loop has been built as antiparallel β-sheet extending out into solution. The temperature factors of most of the atoms in this loop are high, indicating that it is flexible. If residues 132–146 in 1f89 are built according to the structure of 1ems, the end of the loop clashes with a symmetry-related molecule. This accounts for the low density, as the loop presumably flexes in a variety of ways to accommodate the crystal packing.

A re-interpretation of the density for residues 132–134 and 142–146 is possible at the 0.8σ level in both of the subunits related by noncrystallographic symmetry. Refinement of the re-interpreted model leads to density at the 1.2σ level for the side chain of Glu144 at the consensus location.

The authors wish to thank the Carnegie Corporation of New York for funding the Structural Biology Programme and the

National Research Foundation for funding this project. We would also like to thank Tom Terwilliger and Liang Tong for their kind assistance at various stages of the refinement. SK was funded by the Carnegie Corporation of New York and the University of Cape Town. VA was funded by the National Research Foundation and the University of the Western Cape.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.

Agarkar, V. B., Kimani, S. W., Cowan, D. A., Sayed, M. F.-R. & Sewell, B. T. (2006). *Acta Cryst.* F**62**, 1174–1178.

Andrade, J., Karmalik, A., Carrondo, M. A. & Frazão, C. (2007). *J. Biol. Chem.* **282**, 19598–19605.

Banerjee, A., Sharma, R. & Banerjee, U. C. (2002). *Appl. Microbiol. Biotechnol.* **60**, 33–44.

Binkowski, T. A., Naghibzadeg, S. & Liang, J. (2003). *Nucleic Acids Res.* **31**, 3352–3355.

Bond, C. S. (2003). *Bioinformatics*, **19**, 311–312.

Brenner, C. (2002). *Curr. Opin. Struct. Biol.* **12**, 775–782.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *J. Comput. Chem.* **4**, 187–217.

Cameron, R. A., Sayed, M. & Cowan, D. A. (2005). *Biochim. Biophys. Acta*, **1725**, 35–46.

Cameron, R., Tsekoa, T. & Cowan, D. A. (2003). *Adv. Appl. Microbiol.* **52**, 123–158.

Canutescu, A. A., Shelenkov, A. A., Dunbrack, R. L. Jr (2003). *Protein Sci.* **12**, 2001–2014.

Chebrou, H., Bigey, F., Arnaud, A. & Galzy, P. (1996). *Biochim. Biophys. Acta*, **1298**, 285–293.

Chen, C. Y., Chiu, W. C., Liu, J. S., Hsu, W. H. & Wang, W.-C. (2003). *J. Biol. Chem.* **278**, 26194–26201.

Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.

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

DeLano, W. L. (2004). *PyMOL.* DeLano Scientific, San Carlos CA, USA. http://www.pymol.org.

Fournand, D. & Arnaud, A. (2001). *J. Appl. Microbiol.* **9**, 381–393.

Hashimoto, Y., Nishiyama, M., Ikehata, O., Horinouchi, S. & Beppu, T. (1991). *Biochim. Biophys. Acta*, **1088**, 225–233.

Hirrlinger, B., Stolz, A. & Knackmuss, H. J. (1996). *J. Bacteriol.* **178**, 3501–3507.

Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.

Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.

Hung, C. L., Liu, J. H., Chiu, W. C., Huang, S. W., Hwang, J. K. & Wang, W.-C. (2007). *J. Biol. Chem.* **282**, 12220–12229.

Hughes, J., Armitage, Y. C. & Symes, K. C. (1998). *Antonie Van Leeuwenhoek*, **74**, 107–118.

Jones, T. A., Bergdoll, M. & Kjeldgaard, M. (1990). *Crystallographic and Modeling Methods in Molecular Design*, edited by C. Bugg & S. Ealick, pp. 189–195. Berlin: Springer-Verlag.

Kato, Y., Ooi, R. & Asano, Y. (2000). *Appl. Environ. Microbiol.* **66**, 2290–2296.

Kato, Y., Tsuda, T. & Asano, Y. (1999). *Eur. J. Biochem.* **263**, 662–670.

Kumaran, D., Eswaramoorthy, S., Gerchman, S. E., Kycia, H., Studier, F. W. & Swaminathan, S. (2003). *Proteins*, **52**, 283–291.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.

Makhongela, H. S., Glowacka, A. E., Agarkar, V. B., Sewell, B. T., Weber, B., Cowan, D. & Burton, S. G. (2007). *Appl. Microbiol. Biotechnol.* **75**, 801–811.

Mayaux, J. F., Cerbelaud, E., Soubrier, F., Faucher, D. & Petre, D. (1990). *J. Bacteriol.* **172**, 6764–6773.

Mayaux, J. F., Cerbelaud, E., Soubrier, F., Yeh, P., Blanche, F. & Petre, D. (1991). *J. Bacteriol.* **173**, 6694–6704.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Nagasawa, T., Wieser, M., Nakamura, T., Iwahara, H., Yoshida, T. & Gekko, K. (2000). *Eur. J. Biochem.* **267**, 138–144.

Nakai, T., Hasegawa, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Ueki, T., Nanba, H., Ikenaka, Y., Takahashi, S., Sato, M. & Tsukihara, T. (2000). *Structure*, **8**, 729–739.

Novo, C., Tata, R., Clemente, A. & Brown, P. R. (1995). *FEBS Lett.* **367**, 275–279.

Pace, H. C. & Brenner, C. (2001). *Genome Biol.* **2**, reviews0001.1–0001.9.

Pace, H. C., Hodawadekar, S. C., Draganescu, A., Huang, J., Bieganowski, P., Pekarsky, Y., Croce, C. M. & Brenner, C. (2000). *Curr. Biol.* **10**, 907–917.

Pertsovich, S. I., Guranda, D. T., Podchernyaev, D. A., Yanenko, A. S. & Svedas, V. K. (2005). *Biochemistry (Mosc.)*, **70**, 1280–1287.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.

Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.

Sakai, N., Tajika, Y., Yao, M., Watanabe, N. & Tanaka, I. (2004). *Proteins*, **57**, 869–873.

Schrag, J. D. & Cygler, M. (1997). *Methods Enzymol.* **284**, 85–107.

Sewell, B. T., Berman, M., Meyers, P. R., Jandhyala, D. & Benedik, M. J. (2003). *Structure*, **11**, 1413–1422.

Sewell, B. T., Thuku, R. N., Zhang, X. & Benedik, M. J. (2005). *Ann. NY Acad. Sci.* **1056**, 153–159.

Terwilliger, T. (2004*a*). *J. Synchrotron Rad.* **11**, 49–52.

Terwilliger, T. C. (2004*b*). *Acta Cryst.* D**60**, 2144–2149.

Thuku, R. N., Weber, B. W., Varsani, A. & Sewell, B. T. (2007). *FEBS J.* **274**, 2099–2018.

Westhead, D. R., Slidel, T. W., Flores, T. P. & Thornton, J. M. (1999). *Protein Sci.* **8**, 897–904.

Wilson, M. A., St Amour, C. V., Collins, J. L., Ringe, D. & Petsko, G. A. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 1531–1536.