doi:10.1016/j.jmb.2005.12.049 J. Mol. Biol. (2006) 357, 221–235

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Crystal Structure of a Glycyl Radical Enzyme from *Archaeoglobus fulgidus*

Lari Lehtiö[1,2], J. Günter Grossmann[3], Bashkim Kokona[4], Robert Fairman[4] and Adrian Goldman[1]*

[1]*Institute of Biotechnology University of Helsinki, PO Box 65, FIN-00014, Helsinki, Finland*

[2]*National Graduate School in Informational and Structural Biology Åbo Akademi University, Tykistökatu 6 FI-20520 Tuku, Finland*

[3]*Synchrotron Radiation Department, CCLRC Daresbury Laboratory, Daresbury, Warrington, WA4 4AD, UK*

[4]*Department of Biology, Haverford College, 370 Lancaster Ave, Haverford PA 19041, USA*

**Corresponding author*

We have solved the crystal structure of a PFL2 from *Archaeglobus fulgidus* at 2.9 Å resolution. Of the three previously solved enzyme structures of glycyl radical enzymes, pyruvate formate lyase (PFL), anaerobic ribonucleotide reductase and glycerol dehydratase (GD), the last one is clearly most similar to PFL2. We observed electron density in the active site of PFL2, which we modelled as glycerol. The orientation of the glycerol is different from that in GD, and changes in the active site indicate that the actual substrate of PFL2 is bigger than a glycerol molecule, but sequence and structural homology suggest that PFL2 may be a dehydratase. Crystal packing, solution X-ray scattering and ultracentrifugation experiments show that PFL2 is tetrameric, unlike other glycyl radical enzymes. *A. fulgidus* is a hyperthermophile and PFL2 appears to be stabilized by several factors including an increased number of ion pairs, differences in buried charges, a truncated N terminus, anchoring of loops and N terminus *via* salt-bridges, changes in the oligomeric interface and perhaps also the higher oligomerization state of the protein.

Keywords: pyruvate formate lyase; glycerol dehydratase; PFL2; glycyl radical; hyperthermophile

Introduction

The reaction catalysed by pyruvate formate lyase (PFL) was described in the 1940s[1,2] and has been studied extensively since it was established that it contained a glycyl radical.[3] Recently, several other enzymes have been identified in the pyruvate formate lyase sequence family: ketoacid formate lyase,[4] glycerol dehydratase (GD),[5] benzyl succinate synthetase[6] and *p*-hydroxyphenylacetate decarboxylase.[7] However, many of the proteins in the PFL family are still of unknown function[8] and the structures of only two of them have been solved: PFL[9,10] and GD.[11]

Enzymes in the PFL sequence family have a ten-stranded α/β-barrel consisting of two sets of five parallel α/β units assembled in an antiparallel manner.[9,10] This structure is common in the larger family of glycyl radical enzymes that also contain the anaerobic ribonucleotide reductases.[12,13] Ribonucleotide reductases and PFL-like enzymes share similarities in structure and mechanism, but lack significant sequence similarity.[8,9] A cysteine residue on a loop inside the barrel of glycyl radical enzymes is converted into a radical CysS· when substrate binds. In the absence of substrate, the radical resides on a glycine residue in a C-terminal loop, Gly734 (PFL numbering), which is converted to a glycyl radical when the enzyme is activated. Glycyl radical enzymes such as PFL display half-of-the-sites reactivity,[14] so that only one of the glycine residues of the dimer is in the radical state at a time. This radical is regenerated in each reaction cycle. The solved structures are of the inactive non-radical enzyme, and so the conformational changes, especially around the glycine loop, that must occur during activation cannot be visualised.[15] The changes are due to the sp^3 to sp^2 transition at the Gly734 C^{α}, suggesting that the conformational change during activation of one PFL monomer makes the glycine in the other monomer inacces-

Abbreviations used: PFL, pyruvate formate lyase; GD, glycerol dehydratase; SAXS, small-angle X-ray scattering; rmsd, root-mean-square deviation; SROV, square-roots of variance.

E-mail address of the corresponding author: adrian.goldman@helsinki.fi

sible to activating enzyme, resulting in half-of-the-sites reactivity. Activation of PFL-like enzymes, like other glycyl radical enzymes, is performed by a specific activating enzyme belonging to the radical-SAM superfamily.[16] These proteins generate radical species by reductive cleavage of *S*-adenosyl-methionine through an $[Fe\text{-}S]_4$ centre.[17–19]

When we started this work only the structure of *Escherichia coli* PFL had been solved, and in order to expand our knowledge of these PFL sequence family enzymes, we decided to solve the structure of the archaeal *Archaeoglobus fulgidus* PFL2. We chose this enzyme because *A. fulgidu*s is an anaerobic hyperthermophile and so the glycyl radical, once generated, might be more stable at ambient temperature in an oxygen-free atmosphere than the PFL glycyl radical, allowing direct structural studies on the activated enzyme. Recently O'Brien and co-workers[11] solved the structure of another member of the PFL-family, *Clostridium butyricum* GD. This GD and *A. fulgidus* PFL2 are more similar to each other than either is to *E. coli* PFL[8] even though both *C. butyricum* and *E. coli* are mesophilic bacteria. We describe here the structure of PFL2, which, as expected, is very similar to GD, and describe those structural features that explain its increased thermal stability.

Results

Overall structure

The structure of PFL2 resembles that of the other glycyl radical enzymes, in particular that of glycerol dehydratase (GD), with a root-mean-square deviation (rmsd) per C^{α} of 1.4 Å for 734 atoms out of 773 residues in PFL2 and 786 in GD (Figure 1(a)). The major differences are in the loops and helices surrounding the ten-stranded barrel, in particular residues 128–144, 456–475, 696–708. In addition, the long helices (168–242) found at the dimer interface of PFL are more bent in PFL2 than in GD (Figure 1(b)). The ten-stranded barrel of the PFL-family is very well conserved. Surprisingly, O'Brien *et al.*[11] reported that the rmsd per C^{α} between the barrel strands of GD and PFL was a remarkable 6.8 Å, despite the conserved core and claimed that, if the helices surrounding the barrel were taken into account, the difference would be even larger (>8 Å).[11] We, however, see a difference of 1.8 Å rmsd per C^{α} between PFL2 and PFL (571 C^{α} atoms) and 1.7 Å between GD and PFL (602 C^{α} atoms). When only the barrel strands of PFL are used in the alignment, as defined by Becker *et al.*,[10] the rmsd is 1.8 Å for both PFL2 and GD when compared to PFL. It is evident from the structural alignments that, as expected from the sequence alignments,[8] the N-terminal part is the least conserved. Despite that, in our view, an rmsd value this low for ~75% of the structures indicates significant similarity. In fact, GD and PFL are even more similar than PFL2 and PFL.

Oligomerization and crystal packing

Dynamic light-scattering data indicated that PFL2 would be a trimer,[8] and the self-rotation function calculated for the diffraction data indicated that the crystal contained a 3-fold symmetry element (peak at $\theta=55°$, $\phi=45°$, $\chi=120°$). Surprisingly, the asymmetric unit of the crystal contains a dimer, not a trimer. It has an unusual symmetry that generates the peak in the self-rotation function. The monomers make an angle of 120° with respect to each other, thus forming 2/3 of the trimer we expected to find. A monomer from the adjacent asymmetric unit fills in the gap left in the 2/3 trimer, producing an asymmetric trimer (Figure 2(a)). In addition, the pseudo-trimer axis is coincident with the body diagonal of the pseudo-cubic *C*222 cell (Table 1). The combination of pseudo-trimer and pseudo-cubic cell generates an approximate 3-fold along the body diagonal direction in the diffraction data.

The dimer found in the asymmetric unit has a very small interaction surface: only 220 $Å^2$ (Protein–Protein interaction server†). The packing in the crystal actually suggests that a tetramer is the biologically relevant unit. This tetramer is a dimer of dimers, where each dimer is similar to the GD and PFL dimer, burying 1430 $Å^2$ in the monomer–monomer interface. The D2 tetramer forms by packing two of these dimers together. All the subunits of the tetramer consist of a copy of a single monomer from an asymmetric unit, and so the monomers in a tetramer are identical by space group symmetry. The rmsd between A and B monomers, related by the non-crystallographic 3-fold symmetry, is 0.32 Å. Monomer B creates a tetramer similarly to monomer A. Each of the monomers in the tetramer interacts only with two other monomers and therefore the buried surface area between PFL-like dimers of the tetramer is 2×1120 $Å^2$ for one dimer, yielding a total buried surface area of 10,200 $Å^2$ per tetramer (Figure 2(b)).

The inconsistency between the DLS data[8] and the structure prompted us to re-characterize the oligomerization properties in solution by SAXS and analytical ultracentrifugation (see below). The interaction between the physiological monomers at room temperature is clearly not very strong, since we see always also a monomer state in gel-filtration experiments and at high pH of the native PAGE the oligomerization seems to be severely disturbed.[8]

Analytical ultracentrifugation and SAXS

The apparent mass of PFL2 varied with the protein concentration and centrifugation speed (Table 2A), which is due to the presence of different oligomeric states. The observed molecular mass at the lowest centrifugal force and highest protein concentration of 348 kDa correlates extremely well

† http://www.biochem.ucl.ac.uk/bsm/PP/server/

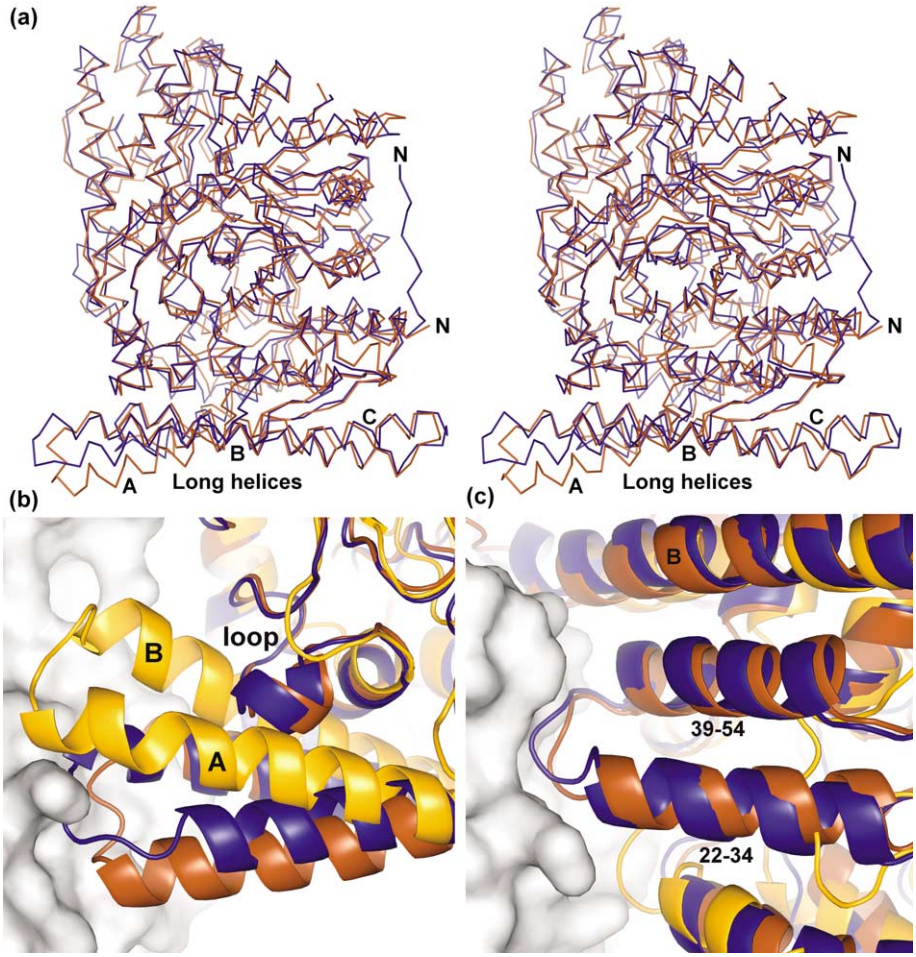

Figure 1. (a) A C^{α} trace comparing a monomer of PFL2 (brown) with a monomer of GD (blue). The N termini have been labelled to show the truncation in PFL2. The long helices that in PFL2 and GD are composed of three separate helices are also labelled: A (168–185), B (189–224) and C (226–243 in PFL2 numbering). (b) A more detailed view of the helices bending at the PFL-like dimer interface. In addition to (a), PFL is also shown in orange. The long helices A and B and loop (521–532) that are inserted between helices are labelled and shown as a cartoon like the rest of the protein in PFL2 and GD. The surface of the adjacent monomer of PFL2 is shown in grey. (c) Inserted sequence in PFL2 (22–54) and GD forming a helix-loop-helix motif at the dimer interface. Helices 22–34 and 39–54 are shown as a cartoon representation and labelled, as is helix B in (a) and (b). This helix-loop-helix motif is completely missing from *E. coli* PFL.

with the calculated mass of a tetramer (348, 668 Da). In considering two-state models, the best fit is to a 1–4 equilibrium scheme for all protein concentrations, as judged by the square-roots of variance (SROV) (Table 2B). A small amount of aggregation is predicted though, particularly at the highest protein concentration. For the 4 μM data, a three-state model, 1-4-8, shows an improvement in the fit. This feature may also be explained by a small amount of aggregate in the sample, as we observed by DLS.[8] Interestingly, global fits to all concentrations and all speeds (Table 2B, last column) suggest that the aggregation does not significantly impact the quality of the fits but reveals a slightly better fit to a 1-2-4 model than the simpler 1-4 model. Little dimer is predicted though, based on the best-fit dissociation constants of $K_{21}=3.2\times10^{-8}$ M and $K_{41}=1.6\times10^{-22}$ M^3. This indicates, as expected, that the tetramer is not extremely stable, but it is in the range for normal tetrameric protein.[20] A three-state 1-3-6 model also fits equally as well as any of our 1-4 schemes, but it predicts a significant amount of hexamer at the ultracentrifugation loading concentration. This is inconsistent with the X-ray scattering studies.

X-ray scattering studies of PFL2 in solution at pH 6.5 provided a scattering profile and distance distribution function that is consistent with a molecule of a radius of gyration of 47.2($\pm$0.2) Å and a maximum dimension of 140($\pm$5) Å. Indeed, it is the tetramer described above that fits the X-ray scattering curve best, yielding a goodness-of-fit value (χ) of 1.69. None of the other possible models fit the data appropriately including the tetramer ($\chi=2.92$) and trimers indicated in Figure 2(a) ($\chi=3.99$ and 4.37), a trimer obeying 3-fold symmetry ($\chi=3.96$), the dimer present in the asymmetric unit ($\chi=5.69$), a PFL-like dimer ($\chi=5.30$) and a monomer ($\chi=7.85$). It was also evident that, at pH 7.5, the radius of gyration was noticeably smaller (46.6($\pm$0.2) Å) and the minimum

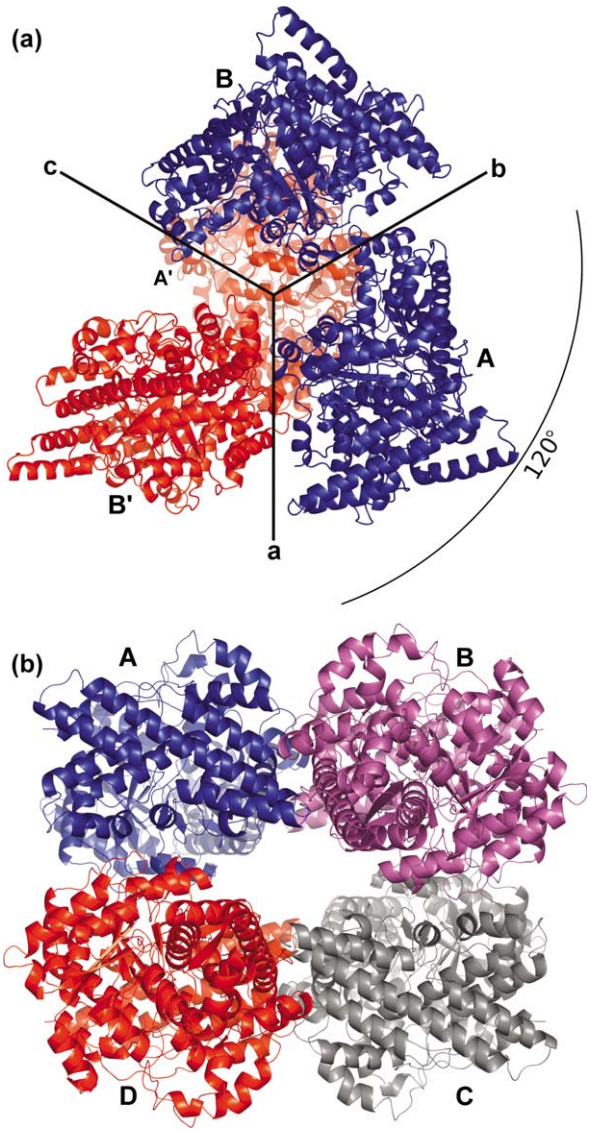

Figure 2. (a) A view of the false trimer viewed along the body diagonal of the pseudo-cubic unit cell. The two blue monomers (labelled A and B) are related by a non-crystallographic axis, as are the two red monomers (A′ and B′); A′ and B′ are related to A and B by crystallographic symmetry operators. The direction of the unit cell axes are indicated by *a*, *b* and *c*. As can be seen, A and B are related by an approximately 120° rotation, but B′ is not. (b) The biologically relevant tetramer of PFL2. The monomers are coloured differently. The view is along one of the 2-fold symmetry axes of the *D*2-tetramer, which is formed using the crystallographic symmetry operators. Monomers A + B and C + D form the "PFL-like" dimers; the A–D and B–C interfaces are unique to PFL2.

around $q = 0.09$ Å^{-1} was no longer as pronounced. These differences are with earlier obervations that the tetramer dissociates at high pH.[8]

PFL-like dimer interface

As expected,[8] the PFL-like dimer interface in PFL2 is like that in GD but different from that in PFL. The end of helix 168–185, the beginning of helix 189–224 and the loop connecting them contribute a lot to the interface. The helices in PFL2 and GD are 10 Å longer than in PFL, and a third helix (226–243) is added at the N terminus of helix 168–185 (helix C in Figure 1(a)). The dimer interface end of the helices is also in a different orientation. A movement of ~12 Å leaves room between the helices and the barrel core, which is filled by insertion of a loop (521–533; Figure 1(b)). In PFL2 and GD there is also another inserted sequence (22–54) forming a helix-loop-helix motif at the dimer interface (Figure 1(c)). This loop contributes to the oligomeric interface *via* hydrophobic interactions (C^{α}, C^{β} and C^{γ} of $Glu37^{A}$, $Pro38^{A}$, $Phe127^{B}$ and $Pro130^{B}$) as well as an ionic interaction between $Glu37^{A}$ and $Arg121^{B}$.

Because the oligomeric interface is different in PFL2 and GD from that in PFL, the orientation of the monomers with respect to each other is different. When the monomer As are aligned, and the Gly radicals superimpose with a difference of 0.65–1.05 Å, the B monomer of PFL2 is rotated 47° with respect to the monomer B of PFL. This rotation is along an axis running close to the mass centres of the monomers. In GD this rotation is very similar (rotation of 40°). The difference in the position of the radical glycine residues in monomer B, when A monomers are aligned, is 8.5 Å between GD and PFL, 7.6 Å between PFL2 and PFL, and 2.7 Å between PFL2 and GD.

Table 1. Summary of data processing and refinement

Resolution (Å)	20–2.9 (3.0–2.9)
Wavelength (Å)	1.00
Number of observations	384,799 (37,366)
Number of unique reflections	52,531 (5035)
Space group	*C*222
Unit-cell parameters (Å)	$a = 167.03$ $b = 174.17$ $c = 162.46$
Completeness (%)	99.6 (100)
R_{merge} (%)[a]	8.5 (43.5)
$I/\sigma(I)$	20.4 (5.2)
R-factor (%)[b]	19.9
R_{free} (%)[c]	24.5
Numbers of atoms (in asymmetric unit)	
Protein	12,222
Water	167
Other	32
rmsd	
Bond lengths (Å)	0.011
Bond angles (°)	1.54
Ramachandran plot	
Residues in most favoured regions (%)	88.1
Residues in additionally allowed regions (%)	11.0
Residues in generously allowed regions (%)	0.7

Values in parentheses are for the highest resolution shell.
[a] $R_{merge} = \Sigma_i |I_i - \langle I \rangle| / \Sigma \langle I \rangle$, where I is an individual intensity measurement and $\langle I \rangle$ is the average intensity for this reflection with summation over all data.
[b] *R*-factor is defined as $\Sigma ||F_{obs}| - |F_{calc}|| / \Sigma |F_{obs}|$, where F_{obs} and F_{calc} are observed and calculated structure-factor amplitudes, respectively.
[c] *R*-free is the *R* factor for the test set (5% of the data).

Table 2. Analytical ultracentrifugation

A. *The observed molecular at different concentrations and centrifugation speed*

Concentration	8000	10,000	12,000	14,000	Overall M_r
4 μM	348,000	295,000	247,000	184,000	278,000±26,000
2 μM	342,000	301,000	228,000	244,000	272,000±26,000
0.4 μM	254,000	245,000	190,000	179,000	208,000±18,000

B. *Analysis of the possible oligomeric states at different protein concentrations*

Model	4 μM	2 μM	0.4 μM	Global
Monomer	37.8	20.0	26.0	26.1
Dimer	20.8	10.9	10.0	13.0
Trimer	11.2	6.60	11.4	7.05
Tetramer	14.5	8.41	20.0	11.7
1⇒2	20.8	10.9	10.0	13.0
1⇒3	10.4	5.62	4.89	9.05
1⇒4	5.43	3.71	3.24	5.43
2⇒4	6.92	4.57	7.17	6.01
1⇒2⇒4	5.43	3.71	3.24	5.15
1⇒4⇒8	4.17	2.82	3.24	5.43
1⇒3⇒6	4.22	2.78	3.49	5.09

The square-root of variance $\times 10^{-3}$ is reported for each oligomeric model. Those with the smallest variance and thus fitting the data best, 1⇒4, 1⇒2⇒4, and 1⇒4⇒8, are shaded.

Fingerprints of thermostability

Nature appears to use several ways to make proteins thermostable.[21] We have analysed the sequence and structural properties of PFL2 in order to analyse what effects might contribute to its thermal stability.

The amino acid composition of PFL2 is consistent with the pattern seen in thermophilic enzymes. In PFL2, the (E+K/Q+H) ratio is 3.8, in the middle of the range for thermophiles (3.2–4.6),[22] and the changes in frequency of Lys, Glu, Gln, His and Arg are as seen in other thermophilic proteins.

GD is a dimer like PFL with a buried surface area of 1190 Å^2 per monomer. The corresponding interface in the PFL2 tetramer is bigger (1430 Å^2) and contains a striking feature presumably affecting the stability of the protein: there is an ionic network in the middle of the interaction surface. GD contains only a single ion pair at the central region involving Lys200 and Asp197 from the same monomer. In PFL2 there is a six-centered ionic network between Glu507(A), Arg530(A) and Glu191(B) and the respective residues in the other monomer (Figure 3(a)). This region includes residues at the PFL-like interface described above (Figure 1(b)).

The slightly smaller (1120 Å^2) AD interface in tetrameric PFL2 is formed by surface complementarity and by ion pairs (Figure 3(b)). The dimer–dimer interface has six short (<4.5 Å) ionic interactions across the dimer interface (Arg165(A)–Asp415(D)–Arg166(A) and Asp417(A)–Arg416(D) and the respective residues in the other monomer). Furthermore, this interface has many other polar residues and coordinated water molecules, making quite an extensive hydrogen-bonding network at the interface. The interface is both hydrophilic and hydrophobic. The latter is caused mainly by two specific residues per monomer: Tyr343 and Ile342. Tyr343 protrudes into the opposite subunit, thus contributing greatly to the surface complementarity by interacting with the surrounding hydrophobic residues (Ile324(A) Pro161(D), Tyr281(D) and Phe432(D)) (Figure 3(b)). In addition, the Tyr343(A) hydroxyl is hydrogen-bonded to Asn420(D). Symmetrically there is a second similar area at the interface, bridged by Ile342(A) and Ile342(D) which interact with each other.

Salt-bridges[21] can contribute to protein thermostability, and we have shown[23] that uncompensated buried charges may be involved in specifically destabilising proteins, as they occur less frequently in thermophilic proteins, even though the percentage of charged residues does not change. We therefore analysed the structures of the monomers of PFL2 and GD, seeking electrostatic differences that might explain why PFL2 would be more thermostable than GD.

The number of ionic interactions between atoms per monomer is significantly increased in PFL2 in comparison with GD. There are 73 ion pairs at a distance of 4.5 Å or less in PFL2 compared to just 63 in GD and merely 48 in PFL. The additional ionic interactions in PFL2 occur in specific positions as well as on the surface of the protein. The most important ones are at the N terminus and at the surface of the protein. The N terminus of PFL2 is anchored to the protein by two ionic triples (Asp4-Arg361-Glu7 and Glu231-Arg5-Asp302). While the latter interaction also exists in GD, the N terminus of GD is eight residues longer and is not stabilised by salt-bridges (Figure 1(a)). Finally, and probably most importantly, PFL2 has two additional eight-centre and one additional five-centre and ionic networks, all on the surface of the protein: Arg166-Glu533-Arg178-Glu174-Arg177-Glu173-Arg169-Glu245, Glu248-Lys241-Glu252-(Lys234-Glu237)-Arg249-Glu292-Lys288

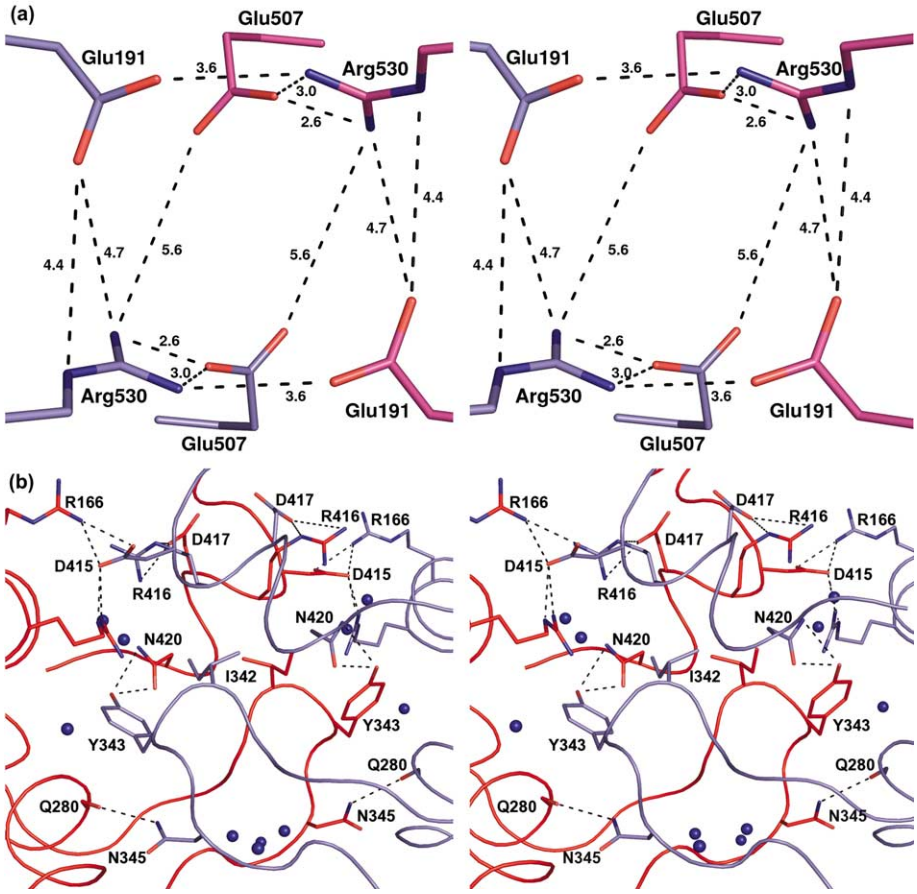

Figure 3. (a) Additional ionic network contributing to the stability of PFL2 at the contact surface between monomers forming the PFL-like dimer in PFL2. As in Figure 2(b), stick models of residues in monomers A and B are coloured in blue and magenta, respectively. (b) AD surface of PFL2 showing the ionic interactions and direct hydrogen bonds across the surface. Monomers are shown as ribbons and coloured as in Figure 2(b): monomer A in blue and B in red. Residues participating in the interactions are shown as stick models and labelled.

and Glu176-Lys200-Glu204-Lys48-Glu52. The one ionic network in GD that is absent from PFL2 (Lys441-Asp424-Lys438-Asp427-Arg287) is from a region of PFL2 that is buried in the AD interface of the tetramer, which is not present in GD as it is a dimer.

There is also a difference in the proportion of buried charged atoms. In PFL2, of the 210 charged groups (as defined in Materials and Methods), 17.6% (37) are totally buried and 11.4% (24) are partially buried, whereas in GD of the total 210 charged groups, 18.6% (39) are fully buried and 14.8% (31) are partially buried.

Of the 28 positions when there was a buried charge in one or other but not both of the structures, ten were in structurally similar positions in both proteins (Table 3A). These residues are thus unlikely to contribute to the increased stability of PFL2. Some of the differences, such as those at PFL2 Glu149, Asn261 and His265, were compensated by changes elsewhere, thus making it unclear which of the structures would be more stable. In eight positions (Table 3C), a charged residue in GD (Glu28, Lys323, Arg325, His332, Glu443, His446, His513 and His692) is replaced by a neutral residue in PFL2, eliminating the charge and presumably increasing the thermostability through improved hydrophobic packing in the protein.[23] At five positions, an uncharged residue in GD is replaced by a charged one in PFL2 (Arg365, His370, Asp424, His489 and Asp547) (Table 3D).

Finally, PFL2 contains a stabilising Asp residue, while GD contains a destabilising one. PFL2 Asp527 makes five charged neutral hydrogen bonds, three to the backbone amide groups (Thr511 and Ser529) and two to side-chain hydroxyl groups (Ser515 and Ser529) (Figure 4(a)). The buried Asp527 is thus probably stabilising, and it firmly anchors the 525–532 surface loop to the protein. This loop also participates in oligomerisation, as one residue in the loop, Arg530, is involved in the ionic interactions between monomers (Figure 3(a)). Conversely, GD Asp506 is buried in a hydrophobic pocket (Figure 4(b)). In order to hydrogen bond to both the side-chain hydroxyl and the main-chain carbonyl group of Ser448, the Asp506 carboxylate group must be protonated and, therefore, Asp506 is presumably destabilising (with respect, e.g. to the Asp506Asn variant[23]).

Table 3. Comparison of buried charged amino acids that do not participate in ion pairs in PFL2 and in GD

A. *Structurally similar positions*

Amino acid in PFL2	Amino acid in GD
Glu267	Glu277
Arg277	Arg287
Asp279	Asp289
Glu428	Glu435
Asp550	Asp558
Lys591	Lys601
Arg621	Arg631
Arg649	Arg659
His722	His733
Arg749	Arg760

B. *Unclear which one is more stable*

Glu149	Tyr157
Asn261	His271
His265	Asn275

C. *Charged residues in GD that are neutral in PFL2*

Amino acid in PFL2	Amino acid in GD
Ser20	Glu28
Pro316	Lys323
Phe318	Arg325
Tyr325	His332
Phe436	Glu443
Ser439	His446
Asn505	His513
Tyr682	His692

D. *Neutral residues in GD that are charged in PFL2*

Arg365	Pro372
His370	Ser377
Asp424	Ile431
Asp547	Asn555
His489	Tyr497

E. *Special cases*

Amino acid in PFL2	Location	Amino acid in GD	Location
Asp527	Four charged-neutral hydrogen bonds: Two backbone and 2 Ser	Ser535	Hydrogen bonded to Asp538
Ser498	Hydrogen bonded to Glu149	Asp506	Two hydrogen bonds to neutral residues, hydrophobic environment, possibly protonated

The Table is aligned according to the position in the structure and the buried charged residues are shaded.

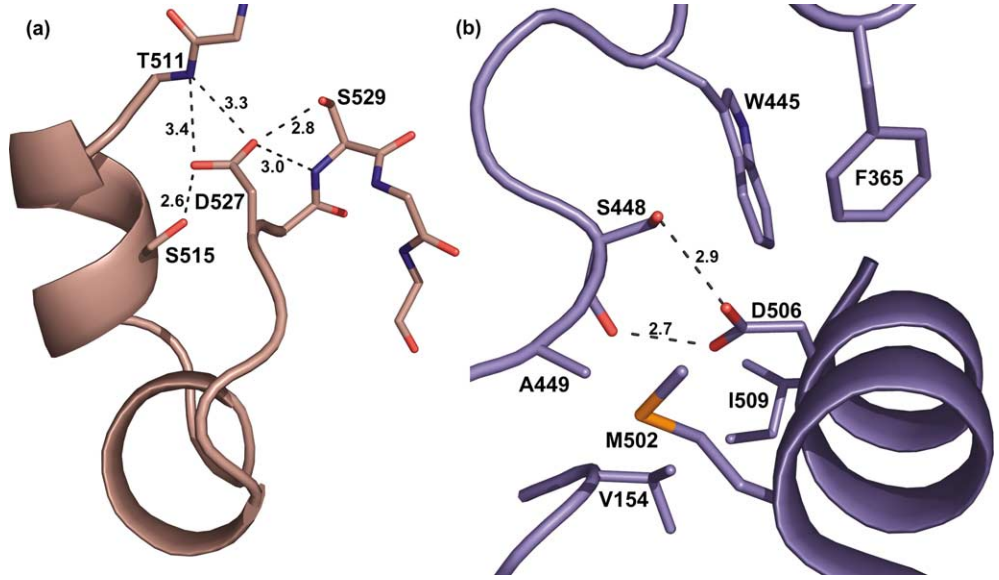

Figure 4. (a) Asp527 of PFL2. A buried charge that anchors loop 525–532 to the protein. Hydrogen bonds, including the long-distance interactions to the backbone amine of Thr511, are shown with distances in Å. (b) Destabilising, possibly protonated, Asp506 in GD. Hydrogen bonds to the Ser448 hydroxyl and carbonyl groups are shown. This, together with the hydrophobic environment, suggests that Asp506 is protonated. If it is not protonated, it forms a very unfavourable interaction with the carbonyl group of Ser448.

Active site

The proposed mechanism of GD starts with abstraction of a hydrogen atom from C1 of glycerol, thus creating a substrate radical.[11] The reaction is then believed to proceed *via* a tricyclic radical intermediate, which is stabilized by the active site histidine residues (His164 and His281). The histidine residues interact with the migrating C2 hydroxyl group. As His281 also hydrogen bonds to Lys323, the role of the histidine would be to act as a general acid and protonate the migrating hydroxyl group. This would lower the energetic barrier of the reaction.

Although the structures of PFL2 and GD are similar, there are differences in the active site (Figure 5(a)). There are two amino acid changes near the active site cysteine, which change the volume of the active site: PFL2 Met632 and Asn332 replace GD Ser642 and Asn332 and Tyr339. The effect of these two opposing changes is to change the shape and increase the volume of the active site cavity from 18.4 Å^3 in GD to 23.7 Å^3 in PFL2 (Figure 5(a)). This corresponds to an increase in substrate volume of about two non-hydrogen atoms.

We observed a 5.4σ peak in the ($F_o - F_c$) difference electron density map right next to the Cys-Met pair (Figure 5(a)). Water molecules and the isopropanol used in the crystallization mixture do not satisfy the density but glycerol, which was used in the cryo-

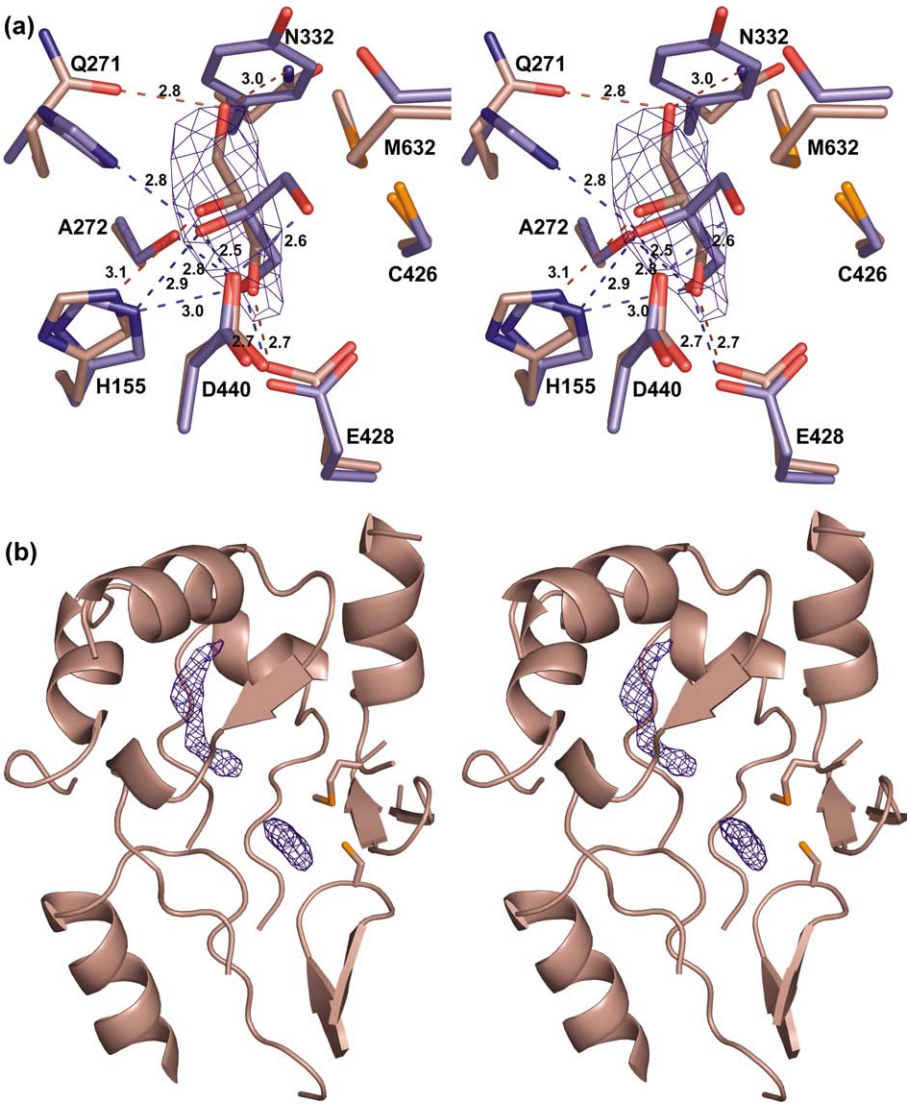

Figure 5. (a) Comparison of the active sites of PFL2 (brown) and GD (blue) with bound glycerol molecules, with the active site residues and glycerol shown as sticks. Hydrogen bonds of the glycerol molecules are shown as broken lines with distances in Å. Structures were aligned with C^{α} atoms of the residues shown in the Figure, resulting in an rmsd value of 0.65 Å. The final σ_A weighted $2F_o - F_c$ electron density map is contoured at 1σ around the glycerol molecule in PFL2. (b) Positive $F_o - F_c$ density, contoured at 3σ, observed in the electron density maps in the active site and in the cavity leading from the active site to the surface of the protein, before adding ligands to the model. The positive tube-like density was modelled as triethylene glycol and the active site density as glycerol. Part of the final PFL2 model is shown as a brown cartoon model, and the Cys426 and Met632 residues of the active site of PFL2 are shown as sticks to show that the orientation is similar to (a).

protectant solution, does. The glycerol probably did not enter the protein during the quick cryo-soak, which lasted only a few seconds, but became bound to the protein during purification as the lysis buffer also contained glycerol.[8] Glycerol may thus be a substrate or product analog of PFL2. The glycerol molecule adopts a somewhat different conformation in monomers A and B, mainly at the O3 hydroxyl group. However, as the temperature factors for the glycerol are significantly lower in monomer A than in monomer B (41.5 Å^2 *versus* 58.2 Å^2), we compared PFL2 monomer A to the GD active site with glycerol bound. In PFL2, the O3 of glycerol is hydrogen bonded to the Gln271 and Asn332 side-chains (Figure 5(a)), while C3 and O3 in GD bind differently; this region of the active site includes the three residues that are different in PFL and GD. The positions of the three substrate atoms, O1, O2 and C1, from which the hydrogen atom is abstracted, align well. However, the proposed essential His281 in GD is not conserved in PFL2, being replaced by Gln271. His164, presumed to be less critical, is conserved, suggesting that this may indeed be the general acid–base catalyst for the 1,2 hydroxyl shift (see Discussion).[11]

Tunnel to active site

A tunnel leads from the surface of PFL2 to the active site. This tunnel also contained continuous electron density (Figure 5(b)), which we modelled as triethylene glycol. At the current resolution (2.9 Å) we cannot rule out the possibility that there is an unknown co-substrate molecule bound to the enzyme, analogous to CoA in PFL. The tunnel is surrounded by hydrophobic residues, whereas in glycerol dehydratase, a similar tunnel is much more polar and is occupied by a continuous chain of water molecules, leading from the substrate binding Asp447 to the surface of the protein. In PFL2, the only polar residues at hydrogen bonding distance from the cavity are Asp440 (GD Asp447), Gln271 (GD His281) and His155 (GD His164) near the active site and Arg85 (GD Trp93) on the surface.

Radical generation and storage

The environment around the Gly752^{734} in PFL2 is similar to other PFL-like enzymes, as expected. (In what follows, the superscript indicates the corresponding residue in *E. coli* PFL.) As in PFL,[8] the amine group of the conserved PFL2 Asn724^{706} makes an unusual N–N hydrogen bond to the main-chain nitrogen of the radical glycine. (A small correction of the published model[11] makes this also possible for GD: a 180° rotation of the C^{δ}–C^{γ} bond of Gln735 is required.) There is, however, one intriguing difference that may affect the activation mechanism: the role of the conserved Arg771^{753} from the C-terminal helix to the glycine loop, which in PFL and GD makes a hydrogen bond to the backbone carbonyl group of the loop. In PFL2, Arg771^{753} is in a different orientation and does not make contact with the glycine loop (Figure 6(a)). Instead, it makes hydrogen bonds to the backbone carbonyl group of Pro661^{635} and the hydroxyl group of Thr663^{637}, which lie in a proline-rich loop, next to Arg771^{753}. Pro661^{635} is conserved, but Thr663^{637} is replaced by Arg in GD and His in PFL. The difference may explain why Arg771^{753} can adopt this different conformation in PFL2.

Even though PFL2 is more like GD than PFL, the Gly752^{734} conformation is like that of PFL, not GD. The major difference in the loop is due to a change in the conformation of Gly752^{734} and the preceding Ala751^{Ser733} (Figure 6(b)). The change is not due to substrate binding, since the same conformation is also seen in the substrate-free native GD structure,[11] but is due primarily to a change in the conformation of Ala751^{Ser733} (Figure 6(b)). In PFL2 and PFL the Ala751^{Ser733} carbonyl group hydrogen bonds to the side-chain of Ser690^{664}, but the corresponding residue in GD is leucine. The lack of the hydrogen bond allows the loop to adopt a different conformation and the leucine prevents the loop adopting the PFL2/PFL-type conformation by steric hindrance; the distance from the Ala751^{733} O conformation in PFL2 to the leucine side-chain would be 1.9 Å.

Discussion

Oligomeric structure

The crystal structure and solution X-ray scattering data all indicate that PFL2 is a tetramer. Analytical ultracentrifugation supports this conclusion. PFL2 thus superficially, resembles the oligomeric structure of the heterooctameric ($\beta_4\gamma_4$) PFL-family enzyme, *p*-hydroxyphenylacetate decarboxylase.[24] However, in that enzyme, the oligomerization is due to phosphorylation of the smaller subunit. The demonstration that PFL2 is a tetramer is surprising, because dynamic light-scattering (DLS)[8] had shown that there was a major peak at a relative molecular mass of 297,000. The approximate molecular mass for the smaller monomer peak is 95 kDa. This, in conjunction with the 3-fold axis visible in the rotation function (see above), led us originally to conclude that PFL2 was mainly a trimer. The agreement between the SAXS, analytical ultracentrifugation and the crystal packing rules out this result and proves that the biologically relevant oligomer is a tetramer. The precision of the DLS measurement is thus not very high, despite the claims of the manufacturer†). This may be because the low-angle light-scattering (LALS) is not measurable in the PD2000.

This tetrameric arrangement is novel among glycyl radical enzymes. The N-terminal region, including the long helix pair, makes most of the monomer–monomer contacts in the PFL-like

† http://www.precisiondetectors.com/

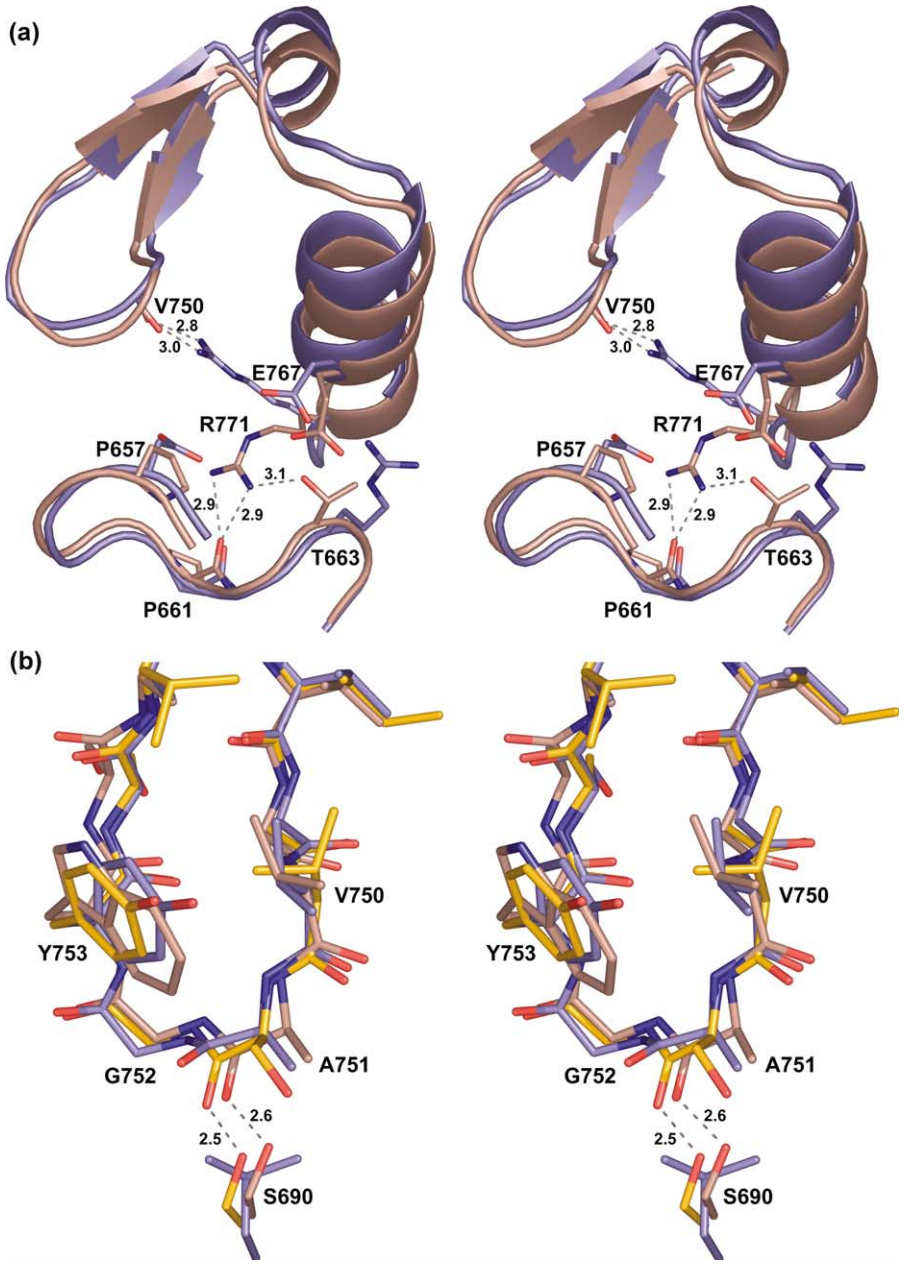

Figure 6. (a) Different conformations of Arg771, which participates in the complex formation with the activating enzyme. The backbone of the 656–665 and 746–773 regions is shown as a cartoon representation with PFL2 in grey and GD in blue. Key residues in the different interactions of Arg771 in PFL2 and GD are shown as sticks, with the hydrogen bonds of Arg771 shown as broken lines with distances in Å. The conformation in PFL is similar to the GD model shown here. (b) Comparison of glycine loops in the solved structures of PFL-family enzymes. Only C^{α} atoms of the loop were included in the alignment (11 atoms). The region from 748–755 (PFL2 numbering) is shown as sticks, except for the Arg749 side-chain, which was omitted for clarity. Carbon atoms of PFL2, GD and PFL are shown in brown, blue and orange, respectively. The hydrogen bond between Ser690 and the carbonyl group of the radical glycine is shown dotted for PFL and PFL2 with distances in Å.

dimer and is one of the least conserved regions within the PFL-family.[8] PFL2, GD and PFL all have different interfaces even though the overall fold is similar. Consequently, the allosteric mechanism by which PFL achieves glycyl radical activation cannot in detail be the same in PFL2 and GD. It would be interesting to see whether the other PFL-like enzymes, GD or PFL2, actually have half-of-the-sites activity like PFL.

Protein evolution

Our recent study[8] found only two PFL-like enzymes from archaea, *A. fulgidus* and *Methanothermobacter thermoautotrophicus* PFL2s. Since then, a third archael PFL2 has been found, from *Pyrococcus kodakaraensis* (*Thermococcus kodakaraensis*),[25] showing 35% identity with the *A. fulgidus* PFL2. The *P. kodakaraensis* enzyme is most homologous (over

50% identity) to the enzymes of sulphate-reducing delta-proteobacteria, *Geobacter sulfurreducens*[26] and *Desulfotalea psychrophila*.[27] This group of enzymes are most homologous to the enzymes from sulfate and sulfonate-reducing bacteria present in the BSS group (benzylsuccinate synthase) in our previous analysis.[8] At the sequence level, they differ from the benzylsuccinate synthases by having a glycine residue, instead of a leucine, adjacent to the active site cysteine (Cys426 in PFL2 and 418 in PFL). As PFL-like enzymes are quite uncommon in archaea but common in bacteria, the PFL-like genes may have been acquired by lateral gene transfer from bacteria, as occurs for other proteins.[28] For *A. fulgidus* PFL2, the gene transfer, if it occurred, appears to be ancient. First, the PFL2 protein sequence is consistent with the protein being from a hyperthermophile (the T_{opt} value for *A. fulgidus* growth is 83 °C) (see Results) and, second, the codon usage is similar to that of a hyperthermophile, showing increased frequency of AGR arginine codons, where R denotes purine. In PFL2, 98% of Arg codons (45/46) are AGR codons, meaning that the PFL2 gene comes from an organism growing at exteremely high temperatures;[22] the average value for *A. fulgidus* is 90.6%. Recent gene transfer from a mesophilic bacterium thus seems unlikely. This may also shed some light on the possible reaction catalysed by this enzyme (see below).

Enzyme activation and mechanism

We have not yet been able to show an enzymatic function for PFL2 because we have been unable to express its activating enzyme.[8] However, based on structural and sequence homology, it is not a PFL. It does not possess the cysteine pair found in PFL (Cys418 and Cys419), nor does it have the arginine residues 176 and 435 that are believed to be required for catalysis in PFL, by directing the thiyl radical addition to the keto group of pyruvate and by stabilization of the formate radical anion that results from the fragmentation of the oxyradical intermediate.[9] We also believe that it is unlikely to be a benzylsuccinate synthase[6] or *p*-hydroxyphenylacetate decarboxylase.[7] First, the PFL2 active site is too small to accommodate such substrates and, second, it contains neither arginine nor lysine residues, so the substrate cannot be negatively charged.

Could PFL2 be a dehydratase? PFL2 does not contain His281 that O'Brien and co-workers[11] have described as essential for catalysis of GD, but the evidence for this in their work is equivocal, as their His164, conserved in PFL2 as His155, is positioned between the O2 and O1 (Figure 5(a); see Figure 6 of O'Brien *et al.*[11]). This would be the natural location for a residue supposed to be involved in catalysing the 1,2 hydroxyl radical shift. It therefore seems to us likely that PFL2 is a dehydratase, and we suggest that in both GD and PFL2 His155 (GD His164) is the required general acid. This requires that His155 is protonated, which is possible as the pK_a value of histidine is 6.0 and it makes an ionic interaction with the adjacent Asp440.

Consistent with this view, the orientation of the C1 and C2 is similar in PFL2 and GD, placing O1 and O2 of glycerol in similar positions with similar hydrogen bonding patterns and C1 in an orientation that makes it possible for Cys426 to abstract a hydrogen atom from it (Figure 5(a)).

However, the PFL2 active site is larger and its geometry different from that in GD (Figure 5(a)), and so the true substrate is probably slightly bigger than glycerol, possibly containing eight non-hydrogen atoms. It may also be less hydrophilic as GD Ser642 is replaced by Met632 and the active site tunnel, which in PFL2 contains a PEG molecule, is less hydrophilic than the equivalent one in GD.

The conformation of Arg771 in PFL2 is very different than in both GD and PFL. This arginine is important in activation; the Arg-Lys mutation of the equivalent residue Arg782 in GD captures a complex of GD with the activating enzyme.[11] In PFL2, Arg771 appears to close the activation site by binding to the proline-rich loop. The absence of positively charged residues at the surface may enable this "closed" conformation seen in PFL2 (Figure 6(a)). This feature may be related to the increased thermostability of PFL2.

Thermostability

The higher oligomeric state in PFL2 is consistent with the fact that *A. fulgidus* is a hyperthermophile, as thermostable proteins often adopt more highly oligomerised forms than their mesophilic counterparts.[29–32] Consistent with this, all the mesophilic glycyl radical enzymes are dimers.[9,11,12] In PFL2, there is a monomer–tetramer equilibrium at room temperature as shown by SAXS and analytical ultracentrifugation. The tetramer is not very stable especially at high pH, which might be due to deprotonation of the surface His49 and His617.

Oligomerization, however, is not always important in protein thermostabilisation.[33] An increased number of salt-bridges is frequently found in thermophilic enzymes[21] and molecular simulations have suggested that salt-bridges would resist the effects of high temperature better than either hydrophobic or polar interactions.[34] In particular, ionic interactions at protein–protein interfaces are known to stabilize proteins[35] and this kind of stabilization mechanism is present in PFL2 as shown by the ionic interactions at the interface (Figure 3(a)). In contrast, one way to significantly destabilize a structure is to completely bury charged residues.[23] Both these electrostatic mechanisms seem to be important in stabilization of PFL2. The number of salt-bridges is increased to 73 (*versus* 63 in GD and 48 in PFL). Changes in the burial of charged residues in PFL2 in comparison with GD produce better interactions and better hydrophobic packing of the protein (Table 3), as shown most strikingly by the comparison of PFL2

Asp527 and GD Asp506 (Figure 4). The truncation of the N terminus and its anchoring to the rest of the protein (Figure 1) may also be important for its increased stability. Similarly, the anchoring of $Arg771^{753}$ to a region away from the glycine loop (Figure 6(a)), unlike in GD and PFL, may also be related to thermostability, as it may prevent unwanted activation of the protein at low temperatures.

The increased thermostability of PFL2 is thus not due to any single factor. An increased number of ionic interactions, improvement in burial of charged residues, truncation and anchoring of N terminus and loops, and the higher oligomeric state presumably all contribute.

Materials and Methods

Chemicals

All the chemicals used in the experiments were purchased from commercial vendors and used as supplied. The components in the cystallization mixture were from Sigma (PEG-8000, Hepes, DTT) and Fluka (isopropanol, DTT).

Protein purification and crystallization

A. fulgidus PFL2 was purified as described,[8] except that all the buffers were changed from Tris–HCl (pH 7.5) to Bis–Tris (pH 6.4) as our results showed that at higher pH the protein was in multiple oligomeric states.[8] In addition, we added a hydrophobic interaction chromatography step between the anion-exchange and gel-filtration chromatography, as follows. Ammonium sulphate was added to a final concentration of 0.5 M to the combined fractions from the anion-exchange column and the sample was loaded onto a phenyl Sepharose column (HiLoad 16/10 Phenyl Sepharose; Pharmacia). Protein was eluted from the column with a decreasing gradient of ammonium sulfate (0.5-0 M ammonium sulphate, 50 mM Bis–Tris (pH 6.4) and 1 mM DTT). PFL2 eluted at the very end of the gradient. The fractions were combined, concentrated with Centriprep (Amicon), applied to the gel-filtration column and purified further as described elsewhere.[8]

Initial crystallization conditions were found with a precipitant synergy screen.[36] The final crystallization conditions contained 16% (w/v) PEG 8000, 8% (v/v) isopropanol, 80 mM Hepes (pH 7.5), 160 mM ammonium sulphate and 1 mM DTT. The crystallization drop consisted of 1 μl of 20 mg/ml protein solution and 1 μl of precipitant solution. We equilibrated the drops against 0.5 ml of well solution using sitting-drop, vapour-diffusion with a Chryschem plate (Hampton Research). Plate-like crystals appeared overnight.

Structure solution

Before data collection, the crystal was dipped in well solution supplemented with 15% (v/v) glycerol and flash-cooled to 100 K in the nitrogen stream. Data were collected to 2.9 Å resolution at ESRF beamline ID29 at a wavelength of 1.00 Å. Crystals belong to orthorhombic space group *C*222 with unit cell axes of $a=167.03$ Å, $b=174.17$ Å and $c=162.46$ Å. Data were processed with XDS[37] and the structure solved by molecular replacement using CNS1.1.[38] The glycerol dehydratase monomer (PDB code 1R9D) was used as a search model.[11] After finding an initial solution, we used Bodil[39] to convert the molecular replacement solution into a rough model of PFL2 by draping the PFL2 sequence onto the GD skeleton. The sequences of PFL2 and GD were aligned in Bodil (sequence identity 32%) and a structure-based model of PFL2 was generated. We find that this makes refinement quicker, as the starting model for refinement is thus somewhat closer to the correct structure than it would otherwise be. No further energy minimization or regularization was done prior to refinement against diffraction data with CNS1.1 using the maximum likelihood target.

In the initial stages of refinement, NCS restraints with a weight of 500 were applied between monomers and only grouped *B*-factors were used. Later, when the model improved, the restraints were released. The final rounds of refinement were done with Refmac.[40] The test set, 5% of reflections, for the R_{free} calculation was preserved during the change of the refinement program and the final *R*-factors describing the error in the model are 19.9% and 24.5% for *R* and R_{free}, respectively (Table 1).

Solution X-ray scattering (SAXS)

X-ray scattering data were collected at station 2.1[41] of the Synchrotron Radiation Source (SRS), Daresbury Laboratory, UK. Two samples of PFL2 were prepared at 20 mg/ml and the buffer was changed to 20 mM Tris–HCl (pH 7.5), 100 mM NaCl and 1 mM DTT or 20 mM Bis–Tris (pH 6.5), 100 mM NaCl and 1 mM DTT. Experiments were performed in a standard cell, cooled to 4 °C, at sample concentrations between 0.5 and 20 mg/ml, according to published procedures.[42] This included the examination of the scattering data, which had been collected in individual time-frames of 60s, to ensure that no observable radiation damage, protein aggregation, or deposition on the cell windows, had occurred. A sample-to-detector distance of 1 m was used for concentrated samples only to cover the momentum transfer interval $0.03\ \text{Å}^{-1} \le q \le 0.76\ \text{Å}^{-1}$ (q is the modulus of the momentum transfer and defined as $q=4\pi \sin \Theta/\lambda$, where 2Θ is the scattering angle and λ is the X-ray wavelength of 1.54 Å). In view of the size of PFL2, additional measurements at a distance of 4.25 m with low concentration samples were performed, permitting q-values between $0.01\ \text{Å}^{-1}$ and $0.18\ \text{Å}^{-1}$ to be measured. The latter provided a sufficiently large scattering range for merging profiles collected at short and long detector distances. The q-range was calibrated using silver behenate powder and wet rat tail collagen (based on a diffraction spacing of 58.38 Å and 670 Å, respectively). The radius of gyration and the intraparticle distance distribution function $p(r)$ were calculated from the experimental scattering data using the program GNOM,[43] which also leads to a reliable estimate of the maximum particle dimension, the value of r at which $p(r)$ drops to zero. High-resolution models based on the PFL2 crystal structure were used for the simulation of scattering curves with the program CRYSOL.[44] This method takes into account the solvent effect by surrounding the protein with a hydration shell of thickness 3 Å and fitting its excess average scattering density. CRYSOL produces a goodness-of-fit value, χ, as a measure of the consistency between structural model and experimental data. A very reliable agreement between

experiment and simulation is generally obtained for χ-values smaller than 2.

Analytical ultracentifugation

Samples of *A. fulgidus* PFL2 were prepared at three different protein concentrations at pH 6.5. The final protein concentrations were 4 μM, 2 μM, and 0.4 μM based on a calculated molar extinction coefficient ε_{280} of 87,400 M^{-1} cm^{-1}. The solutions were prepared in a buffer containing 20 mM Tris–HCl (pH 6.5), 100 mM NaCl, and 1 mM DTT. Attempts to run sedimentation equilibrium analyses at higher pH were unsuccessful due to problems with protein oxidation. The samples were dialysed for 18 h at 4 °C against 2000 ml of buffer each. The following day, 110 μl of sample and 125 μl of buffer (blank) were loaded in six-sectored charcoal centerpieces and run using a Beckman ProteomeLab XL-A analytical ultracentrifuge equipped with an An60 Ti rotor. Samples were spun at 8000 rpm for one day and scans were taken every 4 h. The last two scans were compared with Winmatch until equilibrium was reached. A 22 h delay time was required for the samples to reach equilibrium. Afterwards, samples were spun at 10,000, 12,000, and 14,000 rpm at 4 °C. The data were analysed using WinNonlin V1.035. Winmatch and WinNonlin programs were obtained from the National Analytical Ultracentrifugation facility at the University of Connecticut. Parameters for analysis were calculated using Sednterp.[45]

Analysis of structural features

The volume of the active site cavities was calculated with VOIDOO[46] with a probe size of 1.0 Å. The volume accessible by the probe centre was used for comparisons because the cavity is not completely closed.

For the analysis of buried charges, we used a definition of charged groups similar to Kajander *et al*.;[23] O and C of Asp and Glu, N^{ζ} of Lys; all four atoms of the guanido group of Arg and $N^{\varepsilon 2}$ of histidine. The accessible surface area of these groups was compared to the same area in Gly-X-Gly peptide in vacuum. Calculations as well as analysis of salt-bridges were done with WHAT IF.[47] Structural alignments were done with the program O.[48] The polar and hydrophobic surface areas and the Figures illustrating the structures in this publication were made with Pymol†.

Protein Data Bank accession number

The experimental structure factors and coordinates of the refined PFL2 model have been deposited in the RCSB Protein Data Bank with accession code 2F3O.

Acknowledgements

This work was supported by grants from the Academy of Finland (1105157 and 178376 to A.G.), by the Sigrid Juselius foundation, by Biocentrum Helsinki and by a National Science Foundation grant (MCB-0211754 to R.F.). L.L. is a member of the Informational and Structural Biology Graduate School. Beamtime was made available under the European Union Improving Human Potential Programme (Access to Research Infrastructures) at the ESRF under contract no. HPRI-CT-1999-00022. We thank Dr Elspeth Gordon for providing support at the ESRF. We also thank Dr Veli-Pekka Jaakola and M. Sc. Tarja Parkkinen for help with the data collection, and Dr Roman Tuma, Professor Peter Schönheit and Dr Igor Fabrichniy for helpful discussions.

† http://pymol.sourceforge.net

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.12.049

References

1. Chantrenne, H. & Lipmann, F. (1950). Coenzyme A dependence and acetyl donor function of the pyruvate-formate exchange system. *J. Biol. Chem.* **187**, 757–767.
2. Utter, M. F., Werkman, C. H. & Lippman, F. (1944). Reversibility of the phosphoroclastic split of pyruvate. *J. Biol. Chem.* **154**, 723–724.
3. Wagner, A. F., Frey, M., Neugebauer, F. A., Schäfer, W. & Knappe, J. (1992). The free radical in pyruvate formate-lyase is located on glycine-734. *Proc. Natl Acad. Sci. USA*, **89**, 996–1000.
4. Heßlinger, C., Fairhurst, S. A. & Sawers, G. (1998). Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in *Escherichia coli* that degrades L-threonine to propionate. *Mol. Microbiol.* **27**, 477–492.
5. Raynaud, C., Sarcabal, P., Meynial-Salles, I., Croux, C. & Soucaille, P. (2003). Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proc. Natl Acad. Sci. USA*, **100**, 5010–5015.
6. Leuthner, B., Leutwein, C., Schulz, H., Horth, P., Haehnel, W., Schiltz, E. *et al*. (1998). Biochemical and genetic characterization of benzylsuccinate synthase from *Thauera aromatica*: a new glycyl radical enzyme catalysing the first step in anaerobic toluene metabolism. *Mol. Microbiol.* **28**, 615–628.
7. Selmer, T. & Andrei, P. I. (2001). *p*-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycyl radical enzyme catalysing the formation of *p*-cresol. *Eur. J. Biochem.* **268**, 1363–1372.
8. Lehtiö, L. & Goldman, A. (2004). The pyruvate formate lyase family: sequences, structures and activation. *Protein Eng. Des. Sel.* **17**, 545–552.
9. Leppänen, V. M., Merckel, M. C., Ollis, D. L., Wong, K. K., Kozarich, J. W. & Goldman, A. (1999). Pyruvate formate lyase is structurally homologous to type I ribonucleotide reductase. *Struct. Fold. Des.* **7**, 733–744.
10. Becker, A., Fritz-Wolf, K., Kabsch, W., Knappe, J., Schultz, S. & Volker Wagner, A. F. (1999). Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nature Struct. Biol.* **6**, 969–975.
11. O'Brien, J. R., Raynaud, C., Croux, C., Girbal, L., Soucaille, P. & Lanzilotta, W. N. (2004). Insight into the mechanism of the B12-independent glycerol

dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry,* **43**, 4635–4645.
12. Logan, D. T., Andersson, J., Sjöberg, B. M. & Nordlund, P. (1999). A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science*, **283**, 1499–1504.
13. Young, P., Andersson, J., Sahlin, M. & Sjöberg, B. M. (1996). Bacteriophage T4 anaerobic ribonucleotide reductase contains a stable glycyl radical at position 580. *J. Biol. Chem.* **271**, 20770–20775.
14. Unkrig, V., Neugebauer, F. A. & Knappe, J. (1989). The free radical of pyruvate formate-lyase. Characterization by EPR spectroscopy and involvement in catalysis as studied with the substrate-analogue hypophosphite. *Eur. J. Biochem.* **184**, 723–728.
15. Lehtiö, L., Leppänen, V. M., Kozarich, J. W. & Goldman, A. (2002). Structure of *Escherichia coli* pyruvate formate-lyase with pyruvate. *Acta Crystallog. sect. D*, **58**, 2209–2212.
16. Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F. & Miller, N. E. (2001). Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucl. Acids Res.* **29**, 1097–1106.
17. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T. & Drennan, C. L. (2004). Crystal structure of biotin synthase, an *S*-adenosylmethionine-dependent radical enzyme. *Science*, **303**, 76–79.
18. Hanzelmann, P. & Schindelin, H. (2004). Crystal structure of the *S*-adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans. *Proc. Natl Acad. Sci. USA*, **101**, 12870–12875.
19. Lepore, B. W., Ruzicka, F. J., Frey, P. A. & Ringe, D. (2005). The X-ray crystal structure of lysine-2,3-aminomutase from *Clostridium subterminale*. *Proc. Natl Acad. Sci. USA*, **102**, 13819–13824.
20. Tolan, D. R., Schuler, B., Beernink, P. T. & Jaenicke, R. (2003). Thermodynamic analysis of the dissociation of the aldolase tetramer substituted at one or both of the subunit interfaces. *Biol. Chem.* **384**, 1463–1471.
21. Vieille, C. & Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1–43.
22. Farias, S. T. & Bonato, M. C. (2003). Preferred amino acids and thermostability. *Genet. Mol. Res.* **2**, 383–393.
23. Kajander, T., Kahn, P. C., Passila, S. H., Cohen, D. C., Lehtiö, L., Adolfsen, W. *et al*. (2000). Buried charged surface in proteins. *Struct. Fold. Des.* **8**, 1203–1214.
24. Andrei, P. I., Pierik, A. J., Zauner, S., Andrei-Selmer, L. C. & Selmer, T. (2004). Subunit composition of the glycyl radical enzyme *p*-hydroxyphenylacetate decarboxylase. A small subunit, HpdC, is essential for catalytic activity. *Eur. J. Biochem.* **271**, 2225–2230.
25. Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. & Imanaka, T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* **15**, 352–363.
26. Methe, B. A., Nelson, K. E., Eisen, J. A., Paulsen, I. T., Nelson, W., Heidelberg, J. F. *et al*. (2003). Genome of *Geobacter sulfurreducens*: metal reduction in subsurface environments. *Science*, **302**, 1967–1969.
27. Rabus, R., Ruepp, A., Frickey, T., Rattei, T., Fartmann, B., Stark, M. *et al*. (2004). The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Environ. Microbiol.* **6**, 887–902.
28. Nelson, K. E., Clayton, R. A., Gill, S. R., Gwinn, M. L., Dodson, R. J., Haft, D. H. *et al*. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, **399**, 323–329.
29. Bräsen, C., Urbanke, C. & Schönheit, P. (2005). A novel octameric AMP-forming acetyl-CoA synthetase from the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *FEBS Letters*, **579**, 477–482.
30. Walden, H., Taylor, G. L., Lorentzen, E., Pohl, E., Lilie, H., Schramm, A. *et al*. (2004). Structure and function of a regulated archaeal triosephosphate isomerase adapted to high temperature. *J. Mol. Biol.* **342**, 861–875.
31. Eijsink, V. G., Bjørk, A., Gåseidnes, S., Sirevåg, R., Synstad, B., van den Burg, B. & Vriend, G. (2004). Rational engineering of enzyme stability. *J. Biotechnol.* **113**, 105–120.
32. Thoma, R., Hennig, M., Sterner, R. & Kirschner, K. (2000). Structure and function of mutationally generated monomers of dimeric phosphoribosylanthranilate isomerase from *Thermotoga maritima*. *Struct. Fold. Des.* **8**, 265–276.
33. Nordberg, K. E., Abou-Hachem, M., Holst, O., Danson, M. J. & Hough, D. W. (2002). *Rhodothermus marinus*: a thermophilic bacterium producing dimeric and hexameric citrate synthase isoenzymes. *Extremophiles*, **6**, 51–56.
34. Thomas, A. S. & Elcock, A. H. (2004). Molecular simulations suggest protein salt bridges are uniquely suited to life at high temperatures. *J. Am. Chem. Soc.* **126**, 2208–2214.
35. Bell, G. S., Russell, R. J., Connaris, H., Hough, D. W., Danson, M. J. & Taylor, G. L. (2002). Stepwise adaptations of citrate synthase to survival at life's extremes. From psychrophile to hyperthermophile. *Eur. J. Biochem.* **269**, 6250–6260.
36. Majeed, S., Ofek, G., Belachew, A., Huang, C. C., Zhou, T. & Kwong, P. D. (2003). Enhancing protein crystallization through precipitant synergy. *Structure (Camb)*, **11**, 1061–1070.
37. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallog.* **26**, 795–800.
38. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
39. Lehtonen, J. V., Still, D. J., Rantanen, V. V., Ekholm, J., Björklund, D., Iftikhar, Z. *et al*. (2004). BODIL: a molecular modeling environment for structure-function analysis and drug design. *J. Comput. Aided Mol. Des.* **18**, 401–419.
40. Murshudov, G. N. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
41. Towns-Andrews, E., Berry, A., Bordas, J., Mant, G. R., Murray, P. K., Roberts, K. *et al*. (1989). Time-resolved X-ray diffraction station: X-ray optics, detectors, and data acquisition. *Intl. Conf. Synch. Rad. Instr.* **60**, 2346–2349.

42. Grossmann, J. G., Hall, J. F., Kanbi, L. D. & Hasnain, S. S. (2002). The N-terminal extension of rusticyanin is not responsible for its acid stability. *Biochemistry*, **41**, 3613–3619.
43. Semenyuk, A. V. & Svergun, D. I. (1991). GNOM–a program package for small-angle scattering data processing. *J. Appl. Crystallog.* **24**, 537–540.
44. Svergun, D., Barberato, C. & Koch, M. H. J. (1995). CRYSOL–a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallog.* **28**, 768–773.
45. Laue, T., Shaw, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992). Analytical ultracentrifugation. In *Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. & Horton, J. C., eds), pp. 90–125, Royal Society of Chemistry, Cambridge, UK.
46. Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallog. sect. D*, **50**, 178–185.
47. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52–56.
48. Jones, T. A. (1974). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallog.* **11**, 268–272.

Edited by M. Guss

(Received 18 October 2005; received in revised form 5 December 2005; accepted 7 December 2005)
Available online 3 January 2006