PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Crystal structure of DFA0005 complexed with α-ketoglutarate: A novel member of the ICL/PEPM superfamily from alkali-tolerant *Deinococcus ficus*

Cheng-Jen Liao,[1] Ko-Hsin Chin,[1,2] Chao-Hsiung Lin,[3] Peter Shi-Fong Tsai,[3,4] Ping-Chiang Lyu,[5] Chiu-Chung Young,[6] Andrew H.-J. Wang,[7] and Shan-Ho Chou[1,2]*

[1] Institute of Biochemistry, National Chung-Hsing University, Taichung, Taiwan, Republic of China
[2] National Chung Hsing University Biotechnology Center, National Chung-Hsing University, Taichung, Taiwan, Republic of China
[3] Department of Life Science and Institute of Genomics Science, National Yang-Ming University, Pei-To, Taiwan, Republic of China
[4] Molecular and Genomics Medicine, National Health Research Institute, Miao-Li, Taiwan, Republic of China
[5] Department of Life Science, National Tsing-Hwa University, Hsin-Chu, Taiwan, Republic of China
[6] Department of Soil and Environmental Sciences, National Chung-Hsing University, Taichung, Taiwan, Republic of China
[7] Institute of Biological Chemistry Academia Sinica, Nankang, Taipei, Taiwan, Republic of China

ABSTRACT

The crystal structure of the DFA0005 protein complexed with α-ketoglutarate (AKG) from an alkali-tolerant bacterium Deinococcus ficus has been determined to a resolution of 1.62 Å. The monomer forms an incomplete α7/β8 barrel with a protruding α8 helix that interacts extensively with another subunit to form a stable dimer of two complete α8/β8 barrels. The dimer is further stabilized by four glycerol molecules situated at the interface. One unique AKG ligand binding pocket per subunit is detected. Fold match using the DALI and SSE servers identifies DFA0005 as belonging to the isocitrate lyase/phosphoenolpyruvate mutase (ICL/PEPM) superfamily. However, further detailed structural and sequence comparison with other members in this superfamily and with other families containing AKG ligand indicate that DFA0005 protein exhibits considerable distinguishing features of its own and can be considered a novel member in this ICL/PEPM superfamily.

Proteins 2008; 73:362–371.

***Key words:** Deinococcus ficus; organic waste left-over decomposition; alkali-tolerant; ICL/PEPM superfamily; α-ketoglutarate ligand; AKG.*

INTRODUCTION

The pale-pink alkali-tolerant *Deinococcus ficus* (DF strain CC-FR2-10T) is a novel bacterium isolated from the rhizosphere of the sacred tree Ficus religiosa L.[1] It is found to possess high degrading power for food and organic waste at pH 10 and is being applied to industrial usage. It contains one chromosome of 2,801,970 bp, and three megaplasmids of 642,681, 395,118, and 311,736 bp, respectively. The total genome sequence has been recently completed (Lin *et al.*, manuscript in preparation), and a structural genomics project is followed to study the structures of proteins encoded in this genome and to study the alkali adaptation mechanism exhibited by this alkali-tolerant bacterium compared with their non-alkaline counterparts.[2,3]

The crystal structure of DFA0005 gene product from this bacterium complexed with AKG ligand is solved to a resolution of 1.62 Å using the MAD approach of X-ray crystallography. Fold matching search using the secondary structure matching (SSM)[4] or Dali[5] program identifies DFA0005 as a member of the ICL/PEPM superfamily, including enzymes such as ICL (isocitrate lyase, involved in the glyoxylate cycle),[6] MICL (2-methylisocitrate lyase, involved in the propionate catabolism),[7–9] PEPM (phosphoenolpyruvate mutase, involved in the unusual C—P bond

The coordinates and structural factors of DFA0005 have been deposited in the Protein Data Bank with an ID of 2ZE3.
Grant sponsor: National Science Council, Taiwan, ROC; Grant numbers: 94-2113-M005-003 and 95-2113-M005-018; Grant sponsor: Ministry of Education; Grant sponsor: National Research Program for Genomic Medicine, Taiwan, ROC.
*Correspondence to: Shan-Ho Chou, Institute of Biochemistry National Chung-Hsing University, Taichung, 40227, Taiwan, Republic of China. E-mail: shchou@nchu.edu.tw.
Received 7 December 2007; Revised 15 February 2008; Accepted 6 March 2008
Published online 23 April 2008 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22071

formation present in a variety of antibiotic, antifungal, and herbicidal compounds such as fosfoymcin, amioethyl phosphonate conjugate etc.),[10–13] carboxy-PEPM (carboxyphosphoenolpyruvate mutase, involved in the bialaphos synthesis),[14–16] PEPH (phosphonopyruvate hydrolase, involved in the formation of pyruvate),[17] and MOBL (3-methyl-2-oxobutanoate hydroxymethyltransferase, involved in the formation of ketopantoate).[18,19] Although diverse in functions, they all bind substrates with a conserved α-hydroxy- or α-keto-carboxylate moiety.[20,21] Thus, they all contain conserved amino acid residues for interacting with the common moiety in a wide variety of substrates, including citrate, 2-methylisocitrate, phosphonopyruvate, 3-methyl-oxobutanoate, and so forth. It is believed that the functional diversity within the ICL/PEPM superfamily is even greater than currently known, since many uncharacterized proteins with potentially different activities have been obtained from sequence analysis.[9,20]

In this respect, we report the discovery of a novel member in this superfamily from the alkali-tolerant bacterium *Deinococcus ficus*. It encodes a protein (DFA0005) that retains a similar core structure with those of the ICL/PEPM superfamily, but exhibits many distinguishing features of its own: (a) it has a deeply penetrated AKG binding pocket, not observed in the ICL/PEPM superfamily or any other enzyme/AKG complexes published to date[8,9,11,13,17]; (b) the ligand pocket contains many characteristic polar residues to bind the tail carboxylate moiety of AKG in an orientation different from all other tail moieties of the substrates currently observed in the ICL/PEPM superfamily; (c) it does not contain any metal ion as usually required for the head HOOC—CO moiety binding[7,8,13]; (d) the binding pocket does not encompass a characteristic long β4-α4 loop necessary for gating the substrate binding and product release commonly exploited in this superfamily[9,13]; and (e) it exists as a dimer, not a tetramer or pentamer as usually observed in the MICL/PEPM superfamily members.[10]

METHODS

Cloning and purification

The DFA0005 gene fragment was PCR amplified directly from the local alkali-tolerant *Deinococcus ficus* (DF strain CC-FR2-10T) with the forward 5′- TACTT CCAATCCAATGCTATGACCCACGCCGATCACGCCCGC and backward 5′-TTATCCACTTCCAATGTTACGCGGT GGGGCGGTGGAAGAGGT primers. A ligation-independent cloning (LIC) approach[22,23] was carried out to obtain the desired construct. The final construct codes for a N-terminal His_6 tag, a 17 amino acid linker, and a DFA0005 target protein (275 aa) under the control of a T7 promoter. Over-expression of the Hig_6-tag target protein was induced by the addition of 1 m*M* IPTG at 291 K for 24 h. The target protein was purified by immobilized metal affinity chromatography (IMAC) on a nickel column (Sigma). The His_6 tag and linker were cleaved from DFA0005 by TEV (Tobacco Etch Virus) protease at 291 K for 24 h. For crystallization, DFA0005 was further purified on an anion exchanger column (AKTA, Pharmacia). The final target protein has purity greater than 99% and contains only an extra tripeptide (SNA) at the N-terminal end. Se-Met labeled DFA0005 was produced using a non-auxotroph *E. coli* strain C41 as host in the absence of methionine but with ample amounts of Se-met (100 mg/L). The M9 medium consists of 1 g of NH_4Cl, 3 g of KH_2PO_4, and 6 g of Na_2HPO_4 supplemented with 20% (W/V) of glucose, 0.3% (W/V) of $MgSO_4$, and 10 mg of $FeSO_4$ in 1 L of double-distilled water. The induction was conducted at 18°C for 24 h by the addition of 1 mL of 0.5 m*M* IPTG. Purification of the Se-Met labeled DFA0005 was performed using the protocols as established for the native protein.

Crystallization

For crystallization, the native protein was concentrated to 14 mg/mL in 20 m*M* Tris pH 8.0 and 80 m*M* NaCl using an Amicon Ultra-10 (Millipore). Screening for crystallization condition was performed using sitting-drop vapor diffusion in 96-well plates (Hampton Research) at 295 K by mixing 0.5 μL protein solution with 0.5 μL reagent solution. Initial screens including the Hampton sparse matrix Crystal Screens 1 and 2, a systematic PEG-pH screen and a PEG/Ion screen were performed using the Gilson C240 crystallization workstation. Oval-like crystals appeared in one day from a reservoir solution comprising 0.1*M* Na Hepes pH 7.5, lithium sulfate monohydrate 1.5 *M*. Crystals suitable for diffraction experiments were grown by mixing 2.0 μL protein solution with 2.0 μL reagent solution at room temperature and reached dimensions of 0.5 × 0.2 × 0.1 mm after 1 day. Se-Met labeled DFA0005 was crystallized in a similar way, except 1.3 *M* of lithium sulfate monohydrate was used. Although crystals did not appear sharp, the diffraction pattern turned out to be good to 1.68 Å.

Data collection

Crystals were flash-cooled at 100 K under a stream of cold nitrogen. X-ray diffraction data were collected using the National Synchrotron Radiation Research Center (NSRRC) beamline 13B1 in Taiwan. A two-wavelength MAD data set up to 1.62 Å resolution was obtained. The data were indexed and integrated using HKL2000 processing software,[24] giving data set that are ∼99% complete with R_{merge} of 5.9%/6.6% for all intensities and 38%/46% for the last resolution range. The refinement of selenium atom positions, phase calculation, and density modification were performed using the program SOLVE/

Table I
Statistics of Data Collection and Structural Refinement of DFA0005

Space group	$P6_322$	
Unit-cell parameters	$a = b = 108.84$ and $c = 114.12$ Å	
Mosaicity (°)	0.45	
Wavelength (Å)	High remote	Inflection
	0.96478	0.97950
Resolution range (Å)	50 – 1.65 (1.70–1.65)[a]	50 – 1.62 (1.68–1.62)[a]
Data cutoff (σF)	2.0	2.0
Number of measured reflection	765362 (71696)[a]	786950 (68502)[a]
Number of unique reflection	189632 (17924)[a]	190285 (18027)[a]
Redundancy	4.3 (4.0)[a]	4.1 (3.8)[a]
Completeness (%)	99.1 (93.5)[a]	99.3 (94.5)[a]
R_{merge}(%)	5.9 (38)[a]	6.6 (46)[a]
Mean I/σ (I)	14.3 (2.5)[a]	15.6 (3.1)[a]
R_{free} test set size (%)	5	
R/R_{free} (%)	18.4/20.1	
Solvent content (%)	53	
Matthew coefficient, Å^3/Da	2.1	
Model content		
Protein residues/Water	274/337	
Glycerol/α-ketoglutarate	2/1	
Deviation from ideal geometry		
Bond lengths (Å)	0.011	
Bond angles (°)	1.345	
Chiral	0.098	
Ramachandran plot (%)		
Residues in most favored regions	95.7	
Residues in additional allowed regions	2.8	

[a]Values in parenthesis are for the outmost shell, while the preceding values refer to all data.

RESOLVE.[25] The model was manually adjusted using the XtalView/Xfit package. CNS[26] was then used for refinement to a final R_{cryst} of 18.4% and R_{free} of 20.1%, respectively. The crystals belong to the $P6_322$ space group. The data and refinement statistics were summarized in Table I.

RESULTS

Overall quality

DFA0005 is one of the gene products encoded by the mega-plasmid A of this bacterium. Its structure [Fig. 1(a)] has been determined to a resolution of 1.62 Å using the MAD approach. The data collection, model, and refinement statistics are summarized in Table I. The electron density map is well interpretable from residue 4 to the end of residue 277. Two additional electron density regions are found to which the glycerol molecule can fit very well [Fig. 1(b)]. It is reasonable since glycerol was employed as the cryoprotectant in this experiment. However, an unexpected well-defined electron density region is also recognized, to which only AKG molecule can fit [Figs. 1(b) and 3(b)]. The final model of DFA0005 [Fig. 1(a)] thus comprises one protein molecule of 274 residue long, 337 water, two glycerol, and one AKG molecules. The Matthew's coefficient for DFA0005 is 2.1 A^3/Da and the estimated solvent content is 53%. Examination of the Ramachandran plot calculated from the PdbViewer[27] shows that 95.7% of the residues of nonglycine and nonproline residues are in the most favored regions, with 2.9% of the residues in the additional allowed regions. Four residues (1.5%) are found to deviate from the allowed torsional angles. The side chains of Glu87 are found in the mouth of the water channel, and participate in the H-bonding network leading to the AKG binding pocket [Fig. 4(b)]. Gln60 and Thr61, opposite to Glu87, also locate in the mouth of the water channel. The side chain of Thr61 participates in H-bonding with one water molecule in the channel, and its amino group forms a H-bond with the carbonyl group of Gly59. The participation of these amino acid residues in the H-bonding network may lead to the deviation of their backbone torsion angles from the ideal values. Finally, the side chain of Glu168 in the top of α5 helix forms a H-bond with the carbonyl group of Gly163 in the β5-α5 loop. The deviation of backbone torsion angles of Ala164 in the Gly163Ala164Thr165 bulge may be necessary to reverse the chain direction from β5 to α5.

Dimeric structure

DFA0005 packs in the crystal as a dimer [Fig. 1(b)], with each monomer adopting a modified version of α/β barrel.[10] In this modified α/β barrel, only seven helices

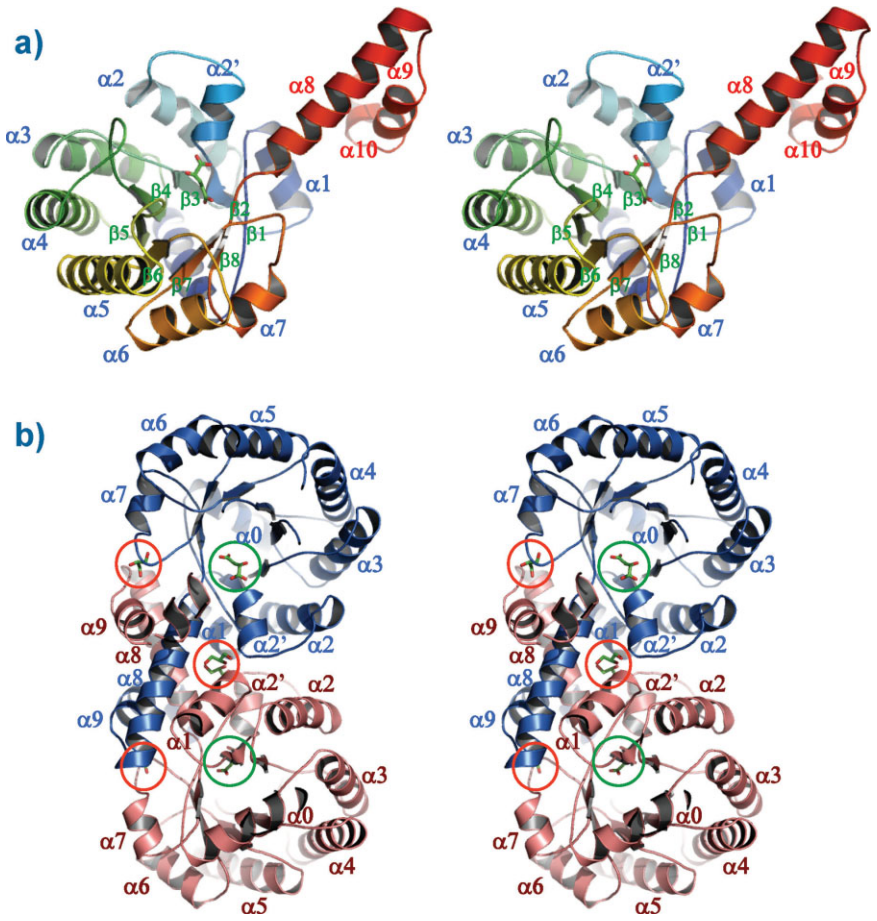

Figure 1

*(**a**) The stereo monomer structure of DFA0005 drawn in ribbon and color-coded from blue (N-terminal) to red (C-terminal). The monomer forms an incomplete β8/α7 barrel, with the C-terminal α8/α9/α10 pointing toward the next subunit to form two complete β8/α8 barrels. (**b**) The stereo dimeric structure drawn in ribbon. One subunit is colored in blue, while the other in violet. Four glycerol molecules situated at the dimeric interface are drawn in ball-and-stick and circled in red. Two AKG molecules located in the barrels are circled in green. Figures 1, 3, and 4 were drawn by the PyMol program (DeLano Scientific LLC, 2004).*

were found to embrace the central β barrel. The eighth helix protrudes from the barrel [Fig. 1(a)], which, along with the two ensuing helices (α9 and α10), mediate the helical swapping with another subunit [Fig. 1(b)]. The strayed helix of one monomer thus forms the eighth helix of the second monomer, and vice versa. This helical swapping beautifully mediates the tight dimeric association to form two complete α/β barrels. The N-terminal end of the barrel is blocked by a helix α0 that traverses the bottom of the barrel [Fig. 1(a,b)].

The dimeric state of DFA0005 in solution was ascertained both by gel filtration chromatography and analytical ultracentrifuge studies (data not shown). The protein–protein interaction server (PPIS) was used to analyze the dimer interface.[28,29] Roughly 20% (2603 Å^2) of solvent-accessible surface area of each subunit is buried upon dimer formation. The majority of contacting residues is contributed from those located in helices α1/α2′ and helices α8/α9/α10 protruding toward the second subunit. Interactions are extensive, including a total of 17 polar contacts within a 3.8 Å distance limit and 46 nonpolar contacts within a 5.0 Å distance limit between protein atoms. Although no salt bridge is found within a 4.0 Å distance limit, six H-bonds are detected, including the Ala27H^{N}(A)-Ala50C═O(B) (2.92 Å), Arg219NH1(A)-Glu254O$^{\varepsilon 1}$(B) (2.70 Å), and Gln244N$^{\varepsilon 2}$(A)-Glu268O$^{\varepsilon 1}$(B) (3.08 Å) atom pairs, and vice versa.

Other than these interactions, four glycerol molecules [circled in red in Fig. 1(b)] situated between the two subunits are found which greatly improve the dimer stability. Two are located in the cavities formed between the β1-α1 loops and the α8′ helix, forming five additional H-bonds between the O^1-Asp25O$^{\delta}$(B), O^3-Thr239O$^{\gamma}$(A), O^2(A)-O^2(B) atom pairs, while another two are located in the cavities formed between the β7-α7 loops and the α8/α9 helices, forming 10 more H-bonds between

the O^1-Ser215O^γ(A), O^1-Pro216C=O(A), O^3-Asp260O^γ(B), O^2-Asp260O^γ(B), O^2—H_2O—Ser215C=O(A) atom pairs. These four glycerol molecules thus form extensive H-bond networks between the DFA0005 subunits and contribute a total of fifteen H-bonds during dimer formation.

Comparison with other ICL/PEPM superfamily proteins

A structural fold and homolog search performed by the DALI[5] and SSM[4,30] servers using the final model DFA0005 coordinates generates a number of structural homologues, all belonged to the ICL/PEPM superfamily. A multiple structural alignment was carried out for some representing structures using the MUSTANG program,[31] with their resulting sequence alignments displayed in Figure 2(a) and structural superimposition in Figure 3(a). The pairwise sequence identity percentages of DFA0005 with 2QIW, IM1B, 2HJP, 1O5Q, 1XG4, and 1M3U are 33.0, 28.4, 31.3, 31.5, 30.6, and 26.7% respectively, while the rmsd values of superimposition of DFA0005 structure with those of 2QIW, IM1B, 2HJP, 1O5Q, 1XG4, and 1M3U are 1.30, 1.28, 1.38, 1.30, 1.27, and 1.70 Å, for 212, 196, 200, 201, 209, and 153 Cαs, respectively. These data indicate that DFA0005 is likely a member of the ICL/PEPM superfamily, and is more close to 2QIW but more distant from 1M3U. However, several significant differences in the loop positions for DFA0005 are detected after the alignment of these protein structures [Fig. 3(a)]; (a) the α2-α2′ loop in DFA0005 exhibits a distinct twist from all other loops in the ICL/PEPM superfamily members, although their loop length are similar; (b) the β4-α4 loop in DFA0005 is rather short and cannot exhibit the conformational flexibility necessary for gating the ligand entrance/exit characteristic of the PEPM/PEPH/MICL/ICL family. The most significant difference of DFA0005 with other ICL/PEPM superfamily members is the unique AKG ligand binding pocket present in the DFA0005 that will be described in the following paragraph.

Unique AKG ligand and water channel

Besides the glycerol molecules, one AKG molecule per subunit is found to locate within the barrel [circled in green in Fig. 1(b)], which is clearly identified from the $F_o - F_c$ electron density map as shown in Figure 3(b). It must be a native ligand for DFA0005 since no AKG was added extrinsically during cultivation, protein purification, or crystallization. The head moiety (HOOC-CO) of AKG is situated at the top of the barrel, and the C3-C6 tail moiety (CH_2CH_2COOH) penetrating toward the bottom of the barrel, forming a well-established binding pocket [Fig. 3(c)]. Interestingly, most head moiety (HOOC—CO) of ligands in the ICL/PEPM superfamily are found to superimpose okay [circled in red in Fig. 3(d)], except that of the ketopantoate of MOBL (colored in gray). However, their tail moieties are found to deviate significantly toward three different directions; upward for isocitrate, sulfopyruvate, and phosphoenolpyruvate, horizontal for ketopantoate, and downward for AKG. A well formed water channel, an extensive H-bonding network, and a number of unique amino acid residues around the AKG ligand are associated with the unique orientation of the AKG ligand. Such results clearly indicate that DFA0005 is a special protein in the ICL/PEPM superfamily and the unique AKG binding in DFA0005 is not an artifact. The hydrophilic water channel, comprising at least fourteen well visible water molecules, is found situated on top of the mouth of the β-barrel [Fig. 4(a)]. Close-up of the channel [Fig. 4(b)] shows an extensive H-bond network for the water molecules. Two deeply buried water molecules are found in the bottom of the AKG binding pocket. The detailed interactions between the water molecules, the AKG, and the surrounding amino acid residues in the pocket are shown in the Ligplot scheme of [Fig. 4(c)]; while the head AKG moiety interacts with three conserved amino acid residues (Asp82, Glu112, and Arg151, marked in solid red circles in Fig. 2) either directly or indirectly, the tail carboxylate moiety interacts with several polar side chains located in the bottom of the binding pocket. Thus Thr42 in strand β2, Asn80 in β3, Asn110 in β4, Asn208 in β7, and Arg228 in β8 all form one indirect H-bond through the deeply buried water molecules, and Ser230 in β8 forms one direct H-bond, with the tail carboxylate oxygen atoms. Residues in strands β5 and β6 contribute no H-bond since they are too far away from the tail carboxylate oxygen atoms. Importantly, most hydrophobic residues situated at the bottom of the β-barrel in other ICL/PEPM superfamily have been replaced by polar residues to interact with the tail carboxylate moiety and the two deeply buried water molecules in DFA0005. Thus Leu/Ile/Val residues in position 80 of β3 strand, His/Val/Cys residues in position 110 of β4 strand, Leu/Val/Ile residues in position 208 of β7 strand, Met/Ile/Leu residues in position 228 of β8 strand, and Leu/Ile/Ala residues in position 230 of β8 strand have been replaced with polar residues of Asn, Asn, Asn, Arg, and Ser, respectively [marked by solid blue circles in Fig. 2 and boxed in Fig. 4(c)] to accommodate the very polar tail carboxylate moiety of AKG.. The well-established H-bond interaction network in this ligand binding pocket can explain the good visibility of the AKG ligand in the DFA0005 $F_o - F_c$ electron density map [Fig. 3(b)]. Furthermore, we also find one critical residue that may partially account for the fact that this pocket in DFA0005 is not suitable for ligand binding in other ICL/PEPM superfamily members. The critical position 41 (boxed in reverse in Fig. 2) is originally occupied by very bulky residues such as Trp/Tyr/Leu. They have now been

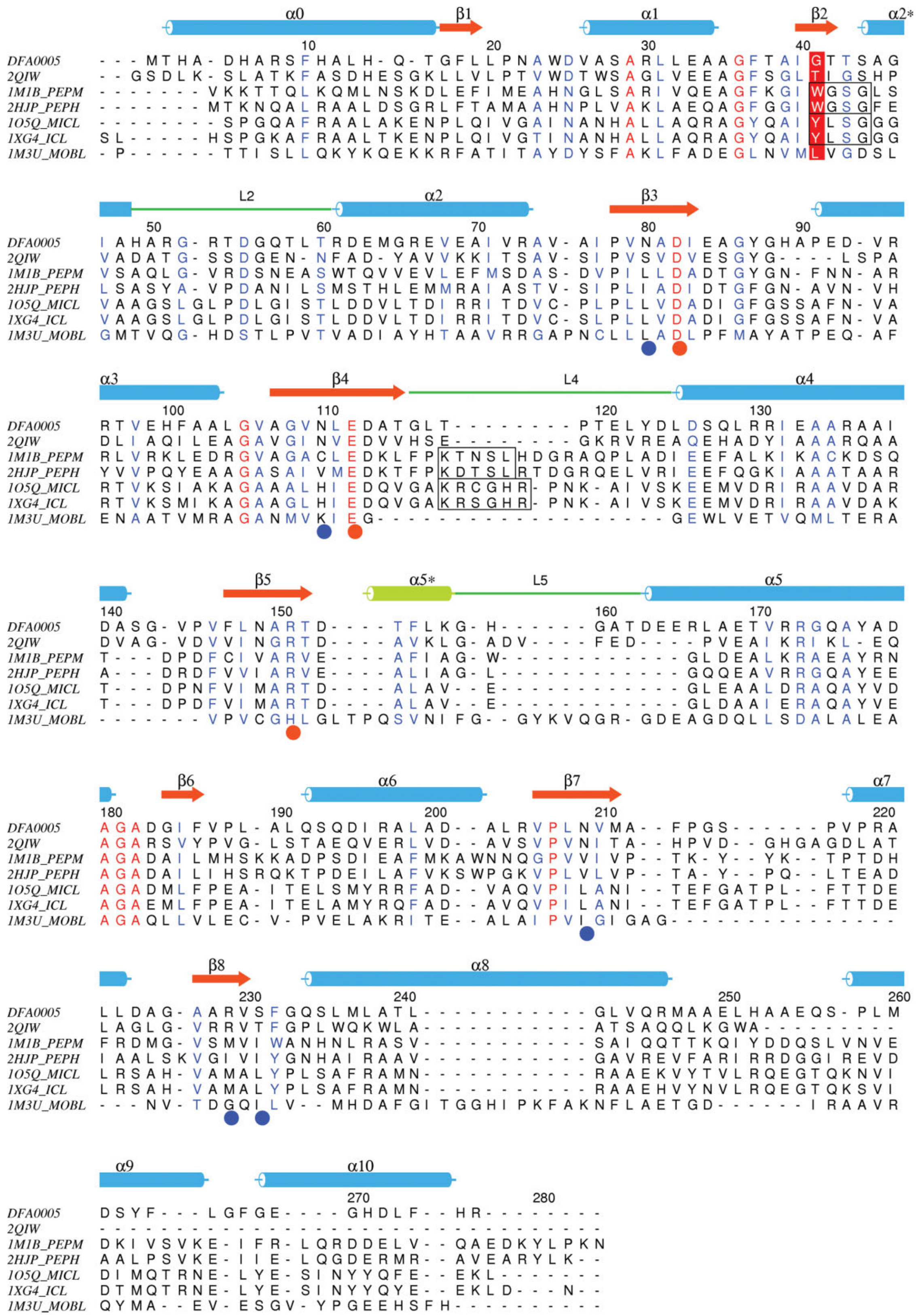

Figure 2

Multiple structure and sequence alignments of DFA0005 with other typical proteins in the ICL/PEPM superfamily generated by the MUSTANG[31] and formatted by the Alscript.[36] MICL, ICL, PEPH, PEPM stand for 2′-methylisocitrate lyase, isocitrate lyase, phosphoenolpyruvate hydrolase, and phosphoenolpyruvate mutase, respectively. Identical residues are shown in red letters and highly conserved residues in blue letters. The secondary structure elements for α-helices (horizontal blue cylinders) and β-strands (horizontal red arrows) are also shown on the top along with their annotations. Conserved amino acid residues for binding with the similar head moiety are marked by solid red circles, while those for binding with the tail moiety of AKG in DFA0005 are marked by solid blue circles. The critical residues partially determining the binding orientation of ligands in the ICL/PEPM superfamily are shown in white letters with a red background. The two signature amino acid patches conserved for the typical ICL/MICL or PEPM/PEPH families in the β2-α2′ and L4 regions are boxed. The α2′-α2 (L2), β4-α4 (L4), and β5-α5 (L5) loops are also annotated and connected by green lines.

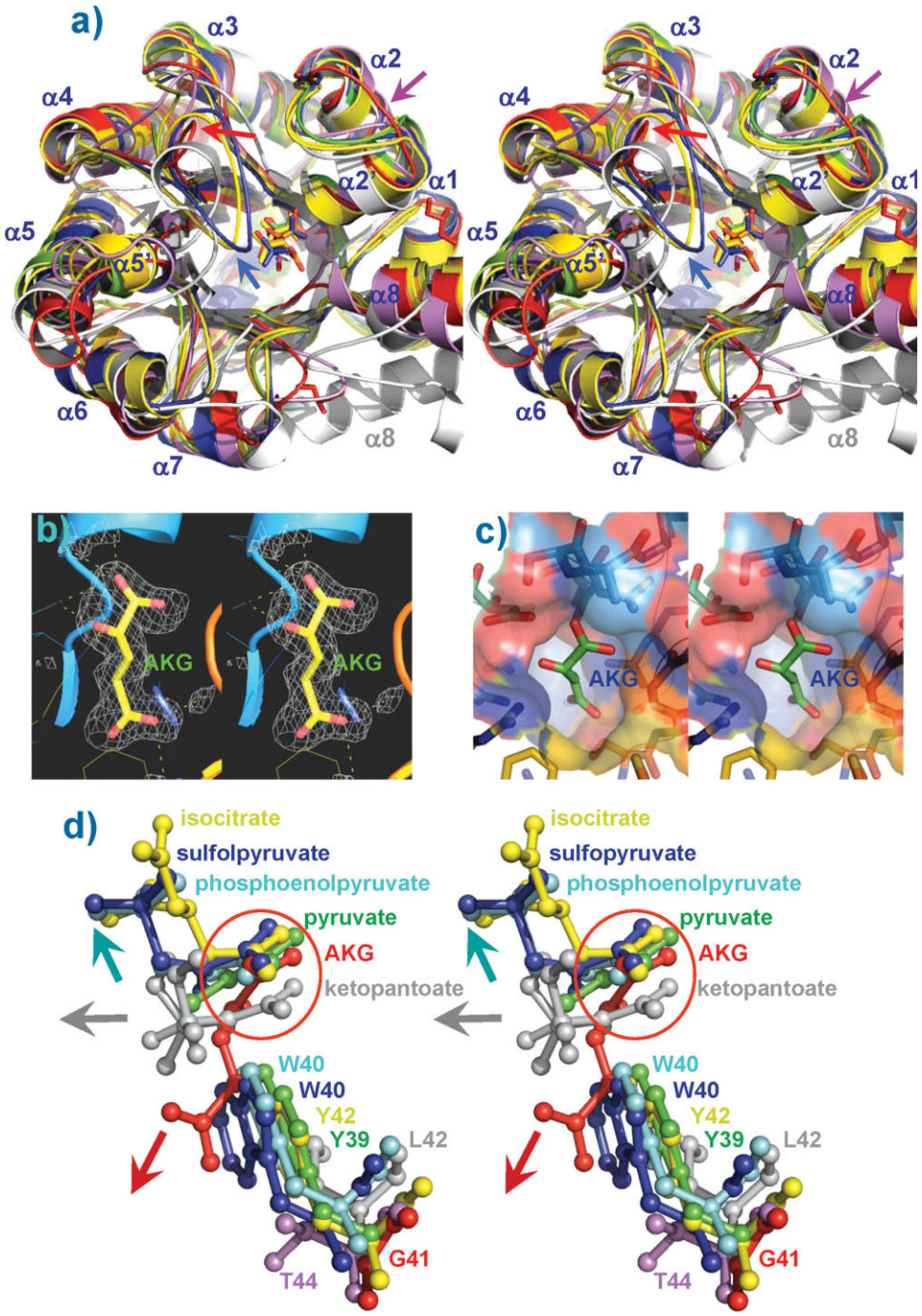

Figure 3

*(**a**) The multiple superimposition of seven representative PDB structures in the ICL/PEPM superfamily generated by the MUSTANG program[31] and plotted by the PyMol program (http://www.delanoscientific.com). DFA0005 exhibits good pairwise rmsd values with the ICL/PEPM superfamily protein, ranging from 1.30 to 1.71 Å for 152–209 Cα atoms, respectively. The DFA0005, 2QIW, 1MIB, 1XG4, 1O5Q, 2HJP, and 1M3U ribbon structures are colored in red, violet, blue, yellow, green, cyan, and gray, respectively. The corresponding ligands located in the central tunnel of the barrel are drawn in sticks in the same colors. The uniquely twisted L2 loop of DFA0005 is indicated by a purple arrow, and the short L4 loop by a red arrow. The corresponding long L4 loops of 1MIB and 1XG4 in their closed forms are marked by a blue arrow. The special L5 loop of 1M3U occupying the same space of the L4 loops of other members is indicated by a gray arrow. (**b**) The stereo picture of the F_o-F_c electron density map (contoured at 1.8 σ) of the AKG ligand for DFA0005. The AKG is drawn in stick, with oxygen atoms colored in red and carbon atoms in yellow. (**c**) The AKG binding pocket of DFA0005 shown in electrostatic plot. AKG is drawn in stick, with oxygen atoms colored in red and carbon atoms in green. Blue or light blue surfaces stand for positive charge, while red or light red for negative charge, and yellow for neutral region. The plots are drawn in a perspective of 30 Å clipping to better show the bottom of binding pocket. (**d**) The superimposition of ligands and critical residues in the ICL/PEPM superfamily. The ligands are drawn in ball-and-stick in the same colors as Figure 3(a) in the top. The bulky residues (Trp, Tyr, or Leu) in other ICL/PEPM superfamily members that sterically hinder the accommodation of their ligands in the DFA0005 AKG binding pocket are shown in the lower right. Three different orientations for the tail moiety of the various ligands are indicated by a blue, gray, and red arrow, respectively, for isocitrate, sulfopyruvate, and phosphonopyruvate, for ketopantoate, and for AKG, respectively.*

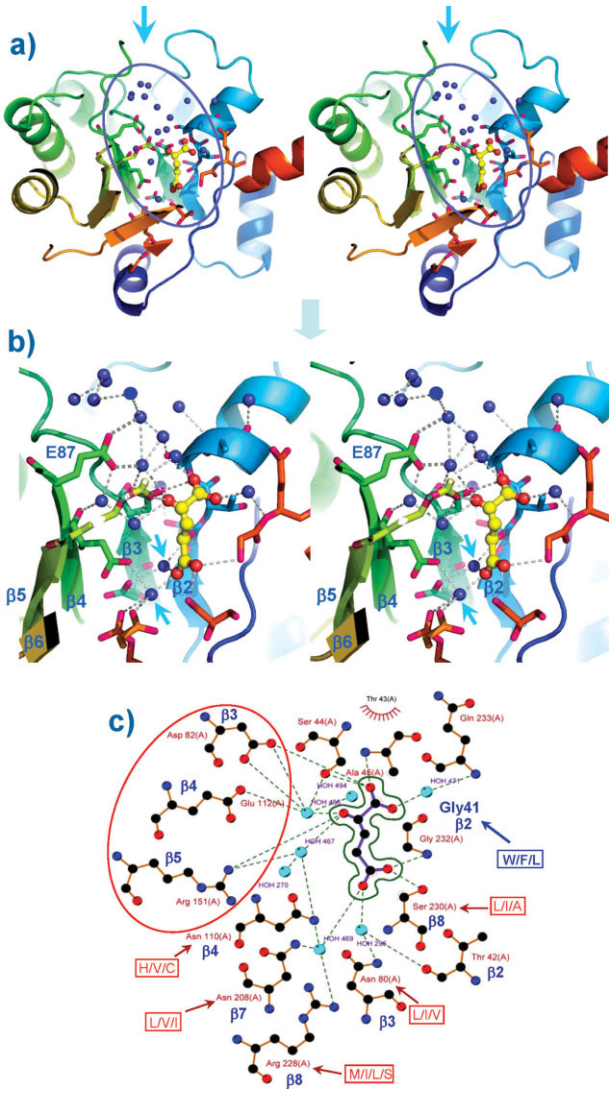

Figure 4

*(**a**) The stereo picture of the unique AKG binding pocket and water channel of DFA0005. The entry of the channel on top of the β-barrel is indicated by a blue arrow, with the fourteen well visible water molecules and AKG ligand circled in blue. The water oxygen atoms are drawn in solid blue circles and the AKG ligand in ball-and-stick. (**b**) The close-up of the circled region in Figure 4(a). The H-bonds between the water molecules and the surrounding amino acid residues are connected by gray dotted arrows. The side chain oxygen atoms of Glu87 are located in the entry of water channel and participate extensively in the H-bonding network. The two deeply buried water molecules are indicated by blue arrows. The surrounding β-strands are also annotated. (**c**) The scheme of the AKG binding pocket drawn by Ligplot.[37] The AKG is traced in a curved green line. The oxygen atoms are colored in red, while those of the seven water oxygen atoms in light blue. H-bonds are connected in green dotted lines. Conserved amino acids contacting the head α-ketocarboxylate moiety among the ICL/PEPM superfamily are circled in red, while those amino acids located in the β3, β4, β7, and β8 strands and contact the tail carboxylate of AKG in DFA0005 are indicated by arrows. The critical residue in position 41 is labeled in blue letters.*

replaced with the slimmest Gly residue to leave room for the AKG penetration [the lower right in Fig. 3(d)] in DFA0005. Obviously the very bulky nature of Trp/Tyr/Leu in this position can exhibit considerable steric hindrance toward AKG penetration for other ICL/PEPM superfamily members.

DISCUSSION

After the crystal structure of the DFA0005 gene product from the novel alkaliphilic *Deinococcus ficus* was determined, it was submitted to the ProFunc server (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/)[30] to help delineate its function. The BLAST search against the UniProt database returns sequences as belonging to the phosphonoenolpyruvate phosphonomutase-like protein (A5G4A7), carboxyphosphonoenolpyruvate phosphonomutase-like proteins (Q8F693) or putative uncharacterized proteins (Q72PY6, A5NWK7, and A1RBE9) with high sequence identities above 40%. Fold matching search by the SSM[4] program returns structures in the ICL/PEPM superfamily (2qiw, 1oqf, 1xg4, 1mum, 1xg3, 2hrw, 2dua, 2hjp, 1zlp) with Z-score above 10 and RMSD values ranging from 1.25 to 1.45 Å. The sequence identities are in the gray area of 32.9%–27.0%. So far all these data seem to indicate that DFA0005 belongs to the ICL/PEPM superfamily.

Nevertheless, the conserved signature sequences of Trp44Gly45Ser46Gly47 for the PEPM/PEPH family and of Tyr43Leu44Ser45Gly46 for the ICL/MICL family (boxed in the β2-α2′ region Fig. 2) are not found in DFA0005, nor is the Lys120Thr121Asn122Ser123Leu124 sequence for the PEPM family, Lys116Asp117Thr1182Ser119Leu120 sequence for the PEPH family, Lys121Arg122Cys123Gly124His125Arg126 sequence for the MICL family or the Lys124Arg125Ser126Gly127His128Arg129 sequence for the ICL family (boxed in the β4-α4 region Fig. 2).[9,13,17] These characteristic features, along with the absence of a long flexible β4-α4 loop for gating the ligand entrance or exit of the ICL/PEPM superfamily proteins [indicated by the blue and red arrows in Fig. 3(a)], indicate that DFA0005 does not belong to the ICL/MICL/PEPM/PEPH family. Compared with MOBL, although both DFA0005/MOBL have a short L4 and a long L5 loops, the L4 loop in MOBL is even shorter, and the L5 loop much longer (see Fig. 2). Interestingly, the long L5 loop of MOBL has occupied the same space originally taken by the long L4 loops in the ICL/PEPM family [indicated by the gray arrow in Fig. 3(a)]. MOBL also has a much longer α8 helix located at a different position with other superfamily members [the gray helix at the lower right of Fig. 3(a)], and forms a pentamer either with[19] or without[18] an α8 helix domain swapping. These, along with the different ligand binding pocket described below [Fig. 3(d)], indicate that DFA0005 is less likely a MOBL (3-methyl-2-oxobutanoate hydroxymethyltransferase)[20] or KPHMT (ketopantoate hydroxymethyltransferase)[18,19] type protein.

The story turns even more bizarre when an unexpected and well-defined electron density region is revealed in the final map. At first, all known substrate/inhibitor/ligand in the ICL/PEPM superfamily such as pyruvate, oxalate, succinate and/or isocitrate[8,9,11,13,17] were tried to fit the map. However, none of them seems to fit. This electron density region seems to be long enough to accommodate five carbon atoms with a branching at the second position. Finally, AKG was tested for fitting since it is a moderate inhibitor (K_i of 0.9 m*M*) for the *T. pyriformis* and *S. hygroscopicus* phosphoenolpyruvate mutase in the ICL/PEPM superfamily.[12] Indeed, it was finally found to fit into the map very well.

Another surprise came when the binding residues of AKG of DFA0005 are compared with other AKG binding residues in this superfamily. As shown in Figure 3(d), the tail moiety (CH_2CH_2COOH) of AKG is found to penetrate toward the bottom of the barrel and locate at a position significantly different from the tail moieties of other ligands in this superfamily, although the head moiety of AKG (HOOC—C=O) can still superimpose. In other words, the ligands of other classical ICL/PEPM superfamily seem to "float" on the top of the barrel, while the AKG ligand in DFA0005 seems to "penetrate" into the bottom of the barrel instead. Structural alignment of DFA0005 with other ICL/PEPM superfamily reveals that residues Asp82, Glu112, and Arg151 (based on the DFA0005 numbering system) interacting with the head HOOC—C=O moiety are conserved, while residues interacting with the tail moiety of the various ligands are very different [Fig. 4(c)]. Unlike the hydrophobic residues employed in the ICL/PEPM superfamily, those present in DFA0005 have been changed to polar residues, that is His/Val/Cys→Asn110, Leu/Val/Ile→Asn208, Met/Ile/Leu/Sser→Arg228, Leu/Ile/Val→Asn80, and Leu/Ile/Ala→Ser230 [Figs. 2 and 4(c)]. Furthermore, one bulky residue (Trp, Tyr or Leu) in the ICL/PEPM superfamily has been mutated to a Gly residue to leave room for the entry of the long tail moiety of the AKG ligand in DFA0005. Electrostatic plot of DFA0005 indicates that the bottom of the AKG binding pocket is highly positive, ready to accept AKG into the binding pocket [Fig. 3(c)].

From the above descriptions, DFA0005 seems to contain many distinguishing features of its own, and can be classified as a novel member in this ICL/PEPM superfamily. From sequence and structural comparison, DFA0005 is more similar to 2QIW recently deposited in PDB. Although 2QIW lacks the C-terminal α9 and α10 helices (see Fig. 2), it has similar lengths in most loops with those of DFA0005. And importantly, they contain similar polar residues around the AKG binding pocket (see Fig. 2). Although 2QIW has a Thr44 in the critical Gly41 position, this amino acid is possibly not bulky enough to block the AKG entrance [colored by purple in Fig. 3(d)]. So 2QIW is still possible to accommodate the AKG ligand in the same binding pocket. However, the details of 2QIW are not published yet, and 2QIW is temporarily annotated as a putative PEPM with no ligand found in the electron density map.

Finally, it is interesting to see if DFA0005 is similar to proteins in other superfamily observed to date that also contain an AKG ligand. PDB search for complexes containing AKG returns four major enzyme families unrelated to the ICL/PEPM superfamily; that is the glutamate dehydrogenase, isocitrate dehydrogenase, aminotransferase, and dioxygenase families. Glutamate dehydrogenase catalyzes the oxidation of glutamate to AKG with the help of a NAD^+ cofactor. It is a regular α/β protein, with AKG product bound in a pocket enclosed by two helices and three β-strands (1HWY).[32] However, no NAD^+ cofactor is found for DFA0005. Isocitrate dehydrogenase catalyzes the oxidation of isocitrate to oxalosuccinate which further undergoes decarboxylation to AKG. It is mainly an open twisted β-sheet protein packed by a number of α-helices at both sides of the β-sheet (1IKA).[33] Aminotransferase catalyzes amino group transfer reactions, and in many such reactions, AKG serves as the amino group acceptor. All aminotransferases incorporate pyridoxal phosphate as the cofactor. However, in the aspartate aminotransferase (1AIB),[34] one of the aminotransferase families, the tail moiety of AKG is found to open to the solvent. Finally, a family of dioxygenases uses AKG as a cofactor to carry out various oxidation and reduction reactions through a ferrous ion. A typical one is the non-haem iron halogenase SyrB2 involved in the syringomycin biosynthesis (2FCU).[35] Although the AKG cofactor of SyrB2 can overlap with that of DFA0005, there is no ferrous ion found in DFA0005. Importantly, in all these four families, amino acid residues surrounding the AKG binding pocket are very different from those of DFA0005, and do not seem to superimpose at all except for the bound AKG ligand (figure not shown). DFA0005 thus contains a novel binding pocket for AKG that is completely different from those published to date in the PDB.

In conclusion, DFA0005 from the alkali-tolerant *Deinococcus ficus* belongs to the ICL/PEPM superfamily, but contains a novel AKG ligand binding pocket unknown until to date. It can be considered a novel member in this superfamily, but needs biochemical data to identify its function. The chemical identity of AKG is also not confirmed yet.

ACKNOWLEDGMENTS

We thank the Core Facilities for Protein X-ray Crystallography in the Academia Sinica, Taiwan, in the National Synchrotron Radiation Research Center, Taiwan, and in the SPring-8 Synchrotron facility in Japan for assistance in X-ray data collection.

REFERENCES

1. Lai W-A, Kampfer P, Arun AB, Shen F-T, Rekha PD, Young C-C. *Deinococcus ficus* sp. nov., isolated from the rhizosphere of *Ficus religiosa* L. Int J Syst Evol Microbiol 2006;56:787–791.
2. Shirai T, Igarashi K, Ozawa T, Hagihara H, Kobayashi T, Ozaki K, Ito S. Ancestral sequence evolutionary trace and crystal structure analyses of alkaline α-amylase from *Bacillus sp.* KSM-1378 to clarify the alkaline adaptation process of proteins. Proteins 2007;66:600–610.
3. Dubnovitsky AP, Kapetaniou EG, Papageorgiou AC. Enzyme adaptation to alkaline pH: atomic resolution (1.08 Å) structure of phosphoserine aminotransferase from *Bacillus alcalophilus.*. Protein Sci 2005;14:97–110.
4. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr 2004;D60:2256–2268.
5. Holm L, Sander C. Dali: a network tool for protein structure comparison. TIBS 1995;20:478–480.
6. Sharma V, Sharma S, zu Bentrup KH, McKinney JD, Russell DG, Jacobs WR, Sacchettini JC. Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. Nat Struct Biol 2000;7: 663–668.
7. Simanshu DK, Satheshkumar PS, Savithri HS, Murthy MRN. Crystal structure of Salmonella typhimurium 2-methylisocitrate lyase (PrpB) and its complex with pyruvate and Mg^{2+}. Biochem Biophys Res Com 2003;311:193–201.
8. Grimm C, Evers A, Brock M, Maerker C, Klebe G, Buckel W, Reuter K. Crystal structure of 2-methylisocitrate lyase (PrpB) from *Escherichia coli* and modelling of its ligand bound active centre. J Mol Biol 2003;328:609–621.
9. Liu S, Lu Z, Han Y, Melamud E, Dunaway-Mariano D, Herzberg O. Crystal structures of 2-methylisocitrate lyase in complex with product and with isocitrate inhibitor provide insight into lyase substrate specificity, catalysis and evolution. Biochemistry 2005;44:2949–2962.
10. Huang K, Li Z, Jia Y, Dunaway-Mariano D, Herzberg O. Helix swapping between two α/β barrels: crystal structure of phosphoenolpyruvate mutase with bound Mg^{2+}-oxalate. Structure 1999;7: 539–548.
11. Liu S, Lu Z, Jia Y, Dunaway-Mariano D, Herzberg O. Dissociative phosphoryl transfer in PEP mutase catalysis: structure of enzyme/sulfopyruvate complex and kinetic properties of mutants. Biochemistry 2002;41:10270–10276.
12. Seidel HM, Knowles JR. Interaction of inhibitors with phosphoenolpyruvate mutase: implications for the reaction mechanism and the nature of the active site. Biochemistry 1994;33:5641–5646.
13. Liu S, Lu Z, Han Y, Jia Y, Howar A, Dunaway-Mariano D, Herzberg O. Conformational flexibility of PEP mutase. Biochemistry 2004;43: 4447–4453.
14. Seidel HM, Pompliano DL, Knowles JR. Phosphonate biosynthesis: molecular cloning of the gene for phosphoenolpyruvate mutase from *Tetrahymena pyriformis* and overexpression of the gene product in *Eshcerichia coli*. Biochemistry 1992;31:2598–2608.
15. Kim A, Kim J, Martin BM, Dunaway-Mariano D. Isolation and characterization of the carbon-phosphorus bond-forming enzyme phosphoenolpyruvate mutase from the Mollusk *Mytilus edulis*. J Biol Chem 1998;273:4443–4448.
16. Hidaka T, Imai S, Hara O, Anzai H, Murakami T, Nagaoka K, Seto H. Carboxyphosphonoenolpyruvate phosphonomutase, a novel enzyme catalyzing C-P bond formation. J Bacteriol 1990;172:3066–3072.
17. Chen CCH, Han Y, Niu W, Kulakova AN, Howard A, Quinn JP, Dunaway-Mariano D, Herzberg O. Structure and kinetics of phosphonopyruvate hydrolase from *Voriovorax sp* Pal2: new insight into the divergence of catalysis within the PEP mutase/Isocitrate lyase superfamily. Biochemistry 2006;45:11491–11504.
18. von Delft F, Inoue T, Saldanha SA, Ottenhof HH, Schmitzberger F, Birch LM, Dhanaraj V, Witty M, Smith AG, Blundell TL, Abell C. Structure of *E. coli* ketopantoate hydroxymethyl transferase complexed with ketopantoate and Mg^{2+}, solved by locating 160 selenomethionine sites. Structure 2003;11:985–996.
19. Chaudhuri BN, Sawaya MR, Kim CY, Waldo GS, Park MS, Terwilliger TC, Yeates TO. The crystal structure of the first enzyme in the pantothenate biosynthetic pathways, ketopantoate hydroxymethyltransferase, from *M. tuberculosis*. Structure 2003;11:753–764.
20. Lu Z, Feng X, Song L, Han Y, Kim A, Herzberg O, Woodson WR, Martin BM, Mariano PS, Dunaway-Mariano D. Diversity of function in the isocitrate lyase enzyme superfamily: The *Dianthus caryophyllus* petal death protein cleaves α-keto and α-hydroxycarboxylic acids. Biochemistry 2005;44:16365–16376.
21. Teplyakov A, Liu S, Lu Z, Howard A, Dunaway-Mariano D, Herzberg O. Crystal structure of the petal death protein from carnation flower. Biochemistry 2005;44:16377–16384.
22. Stols L, Gu M, Dieckman L, Raffen R, Collart FR, Donnelly MI. A new vector for high-throughput, ligation-independent cloning encoding a Tobacco Etch Virus protease cleavage site. Protein Expr Purif 2001;25:8–15.
23. Aslanidis C, de Jong PJ. Ligation-independent cloning of PCR products (LIC-PCR). Nucl Acids Res 1990;18:6069–6074.
24. Otwinowski Z, Minor W. Processing of the X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
25. Terwilliger TC, Berendzen J. Automated MAD and MIR structure solution. Acta Crystallogr 1999;D55:849–861.
26. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr 1998;D54:905–921.
27. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 1997;18:2714–2723.
28. Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
29. Jones S, Thornton JM. Protein-protein interactions: A review of protein dimer structures. Prog Biophy Mol Biol 1995;63:31–65.
30. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. Nucl Acids Res 2005; 33:w89–w93.
31. Knonagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. MUSTANG: A multiple structural alignment algorithm. Proteins: Struct Funct Bioinf 2006;64:559–574.
32. Smith TJ, Peterson PE, Schmidt T, Fang J, Stanley CA. Structures of bovine glutamate dehydrogenase complexes elucidate the mechanism of purine regulation. J Mol Biol 2001;307:707–720.
33. Stoddard BL, Koshland Jr. DE. Structure of isocitrate dehydrogenase with alpha-ketoglutarate at 2.7-A resolution: conformational changes induced by decarboxylation of isocitrate. Biochemistry 1993;32: 9317–9322.
34. Malashkevich VN, Jager J, Ziak M, Sauder U, Gehring H, Christen P, Jansonius JN. Structural basis for the catalytic activity of aspartate aminotransferase K258H lacking the pyridoxal 5′-phosphate-binding lysine residue. Biochemistry 1995;34:405–414.
35. Blasiak LC, Vaillancourt FH, Walsh CT, Drennan CL. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. Nature 2006;440:368–371.
36. Barton GJ. ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng 1993;6:37–40.
37. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 1995;8:127–134.