Crystal Structure and Functional Assignment of YfaU, a Metal Ion Dependent Class II Aldolase from *Escherichia coli* K12[†,‡]

Dean Rea,[§,||] Rebecca Hovington,[||] John F. Rakus,[⊥] John A. Gerlt,[⊥] Vilmos Fülöp,[§] Timothy D. H. Bugg,[||] and David I. Roper*[,§]

Departments of Biological Sciences and Chemistry, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, U.K., and Departments of Biochemistry and Chemistry, University of Illinois, Urbana−Champaign, 600 South Mathews Avenue, Urbana, Illinois 61801

Received May 20, 2008; Revised Manuscript Received July 16, 2008

ABSTRACT: One of the major challenges in the postgenomic era is the functional assignment of proteins using sequence- and structure-based predictive methods coupled with experimental validation. We have used these approaches to investigate the structure and function of the *Escherichia coli* K-12 protein YfaU, annotated as a putative 4-hydroxy-2-ketoheptane-1,7-dioate aldolase (HpcH) in the sequence databases. HpcH is the final enzyme in the degradation pathway of the aromatic compound homoprotocatechuate. We have determined the crystal structure of apo-YfaU and the Mg^{2+}−pyruvate product complex. Despite greater sequence and structural similarity to HpcH, genomic context suggests YfaU is instead a 2-keto-3-deoxy sugar aldolase like the homologous 2-dehydro-3-deoxygalactarate aldolase (DDGA). Enzyme kinetic measurements show activity with the probable physiological substrate 2-keto-3-deoxy-L-rhamnonate, supporting the functional assignment, as well as the structurally similar 2-keto-3-deoxy-L-mannonate and 2-keto-3-deoxy-L-lyxonate (see accompanying paper: Rakus, J. F., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Hubbard, B. K., Delli, J. D., Babbitt, P. C., Almo, S. C., and Gerlt, J. A. (2008) *Biochemistry 47*, 9944−9954). YfaU has similar activity toward the HpcH substrate 4-hydroxy-2-ketoheptane-1,7-dioate and synthetic substrates 4-hydroxy-2-ketopentanoic acid and 4-hydroxy-2-ketohexanoic acid. This indicates a relaxed substrate specificity that complicates the functional assignment of members of this enzyme superfamily. Crystal structures suggest these enzymes use an Asp-His intersubunit dyad to activate a metal-bound water or hydroxide for proton transfer during catalysis.

Functional annotation of genomes presents a major challenge for biology in the 21st century (*1, 2*). Although the function of some gene products can be predicted by sequence comparison with well-studied homologues, the functions of many others remain unknown (*3*). Between these groups lie genes whose functional designation is susceptible to misinterpretation as a result of incorrect or misleading comparison to other genes. It is long established that proteins with highly related sequences can perform different functions *in vivo,* and conversely many proteins have similar functions despite sharing little or no sequence similarity. Structural comparisons of proteins can provide important functional clues since structure is more conserved through evolution than sequence. Structural genomics initiatives are providing an ever-expanding database of protein structural information that provides a valuable resource for functional genome annotation (*4, 5*). However, greater effort is required if incomplete or incorrect functional annotation is to be avoided. Recent reports in the literature highlight the power of combining computational and experimental approaches for functional annotation (*6, 7*). In these studies virtual ligand screening and docking using experimentally determined and modeled structures were used to identify potential substrates. Experimental validation of the substrates and determination of the enzyme−ligand complexes confirmed the correct identification of the substrates and provided justification for this approach to functional annotation.

Aromatic degradative pathways utilize a diverse array of enzymatic activities that may be appropriate for use in synthetic chemistry for the stereoselective synthesis of organic compounds. Research in this area may facilitate the development of biodegradable or biotechnologically useful chemicals and engineered enzymes or organisms useful for bioremediation (*8, 9*). One of the longest characterized

[†] This work was supported by a research grant from the BBSRC (Grant Reference BBC0044501).

[‡] The crystal structure coordinates and structure factors for *Escherichia coli* K12 Apo-YfaU and the YfaU-Mg^{2+}−pyruvate complex have been deposited in the Protein Data Bank (PDB accession codes 2VWS, and 1VWT, respectively).

* To whom correspondence should be addressed: Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, U.K. Phone: 44 (0)24 76528369. Fax: 44 (0)24 76523701. E-mail: david.roper@warwick.ac.uk.

[§] Department of Biological Sciences, University of Warwick.

[||] Department of Chemistry, University of Warwick.

[⊥] University of Illinois at Urbana−Champaign.

[1] Abbreviations: 4-HPA, 4-hydroxyphenylacetate; AkbF, *Rhodococcus sp*. DK17 4-hydroxy-2-oxovalerate aldolase; CL, *Deinococcus radiodurans* citrate lyase; DDGA, 2-dehydro-3-deoxygalactarate aldolase; HPC, homoprotocatechuate; HpcH, 4-hyroxy-2-ketoheptane-1,7-dioate; LB, Luria−Bertani broth; LDH, L-lactate dehydrogenase; MPS, *Macrophoma commelinae* macrophomate synthase; MS, malate synthase; PykF, pyruvate kinase; RhamD, L-rhamnonate dehydratase; YfaU, 2-dehydro-3-deoxy-L-rhamnonate aldolase.

10.1021/bi800943g CCC: $40.75
Published on Web 08/29/2008

```
HpcH            ----------------------MEN-----SFKAALKA-GRPQIGLWLGLSSSYSAELLA 32
AkbF            ----------------------MQSPI--NSFKKALAE-GRTQIGFWLALGDAYSAEVCA 35
YfaU            ----------------------MNALLS-NPFKERLRK-GEVQIGLWLSSTTAYMAEIAA 36
DDGA            ----------------------MNNDVFPNKFKAALAA-KQVQIGCWSALSNPISTEVLG 37
PykF            -------------------MNVFDKAFLHNKFKAAVLE-HKTQIGFGLVSGSAVNAEIVA 40
MPS             MAKSYSEQPELHAKAPYRSAMLTYPGNLRQALKDAMADPSKTLMGVAHGIPSTFVTKVLA 60
                                           :*  :    .  :*       .  ::: .

HpcH            GAGFDWLLIDGEHAPNNVQTVLTQLQAIAPYP---SQPVVRPSWNDPVQIKQLLDVGTQT 89
AkbF            GAGFDWLLIDGEHAPQDLRSVLAQLQVIGAYRD--CHAAVRVPSADTTVIKQYLDLGAQS 93
YfaU            TSGYDWLLIDGEHAPNTIQDLYHQLQAVAPYA---SQPVIRPVEGSKPLIKQVLDIGAQT 93
DDGA            LAGFDWLVLDGEHAPNDISTFIPQLMALKGSA---SAPVVRVPTNEPVIIKRLLDIGFYN 94
PykF            GSGYDFIWIDGEHGPNTVTTIIDQARAIAPYG---SHVIVRPLEADRALIKQLLDAGIQS 97
MPS             ATKPDFVWIDVEHGMFNRLELHDAIHAAQHHSEGRSLVIVRVPKHDEISLSTALDAGAAG 120
                 :  *:: :* **.      .     .        .   :*    .   :.  ** *

HpcH            LLVPMVQNADEAREAVRATRYPPAGIR-----GVGSALARASRWNRIPDYLQKANDQMCV 144
AkbF            LLVPMVDTADEAAAVVRACRYPPGGIR-----GVGG--ARASRWGRYPRYLHEADEQVCV 146
YfaU            LLIPMVDTAEQARQVVSATRYPPYGER-----GVGASVARAARWGRIENYMAQVNDSLCL 148
DDGA            FLIPFVETKEEAELAVASTRYPPEGIR-----GVSVS-HRANMFGTVADYFAQSNKNITI 148
PykF            IIAPMVESGEQAEYIAQSMYYPSRGKR-----GFGAPAVRAGRWGRLPEYIKHAEDELFL 152
MPS             IVIPHVETVEEVREFVKEMYYGPIGRRSFSPWTFSPGIADASLFPNDPYNVATSNNHVCI 180
                :: * *:. ::.   .    * . * *      ..     *  :      .   :. : :

HpcH            LVQIETREAMKNLPQILDVEGVDGVFIGPADLSADMGYAGN-----PQHPEVQAAIEQAI 199
AkbF            VVQAETALALSNLEAIAEVDGIDGVFIGTADLAASLGFPGN-----PAHPEVQDAILDAL 201
YfaU            LVQVESKTALDNLDEILDVEGIDGVFIGPADLSASLGYPDN-----AGHPEVQRIIETSI 203
DDGA            LVQIESQQGVDNVDAIAATEGVDGIFVGPSDLAAALGHLGN-----ASHPDVQKAIQHIF 203
PykF            AVQIESKKGVENLKDIVTTDGVDAVFLGPADLAVDMGYFGD-----FSGEEMQATIEKLI 207
MPS             IPQIESVKGVENVDAIAAMPEIHGLMFGPGDYMIDAGLDLNGALSGVPHPTFVEAMTKFS 240
                  * *:  .:.*:  *     :..::.*..*     *   :          .   :

HpcH            VQIRESGKAPGILIANEQLAKRYLELGALFVAVGVDTTLLARAAEALAARFGAQATAVKP 259
AkbF            QRVRAAGKAPGVLTPVEDLAQKYLAHGAVFVAVGIDTHLLAKQTSALAARFAQVAYS--- 258
YfaU            RRIRAAGKAAGFLAVAPDMAQQCLAWGANFVAVGVDTMLYSDALDQRLAMFKSGKNGPRI 263
DDGA            NRASAHGKPSGILAPVEADARRYLEWGATFVAVGSDLGVFRSATQKLADTFKK------- 256
PykF            KDIRALGKPVGTIAGSPEEAKRYIDWGASFVVVGVDTIFLAHMADSVLGACRSIVK---- 263
MPS             TAAQRNGVPIFGGALSVDMVPSLIEQGYRAIAVQFDVWGLSRLVHGSLAQARASAKQFAG 300
                     * .          .   :  *   :.*  *

HpcH            -GVY---------------------------------- 262
AkbF            ------------------------------------
YfaU            KGSY---------------------------------- 267
DDGA            ------------------------------------
PykF            ------------------------------------
MPS             QGKAATDGTTDETVDGAAEEVANGVSKVKLDEAGDEDKA 339
```

FIGURE 1: Sequence alignment of YfaU with the closest sequence homologues. Identical, highly conserved and conserved residues are indicated by stars, double dots and single dots, respectively. Active site residues are colored red. HpcH, 2-keto-4-hydroxyheptane-1,7-dioate aldolase; AkbF, 4-hydroxy-2-oxovalerate aldolase; DDGA, 2-dehydro-3-deoxygalactarate aldolase; PykF, pyruvate kinase; MPS, macrophomate synthase. Alignment made with ClustalW (*32*).

aromatic degradative pathways involves the catabolism of 4-hydroxyphenylacetate (4-HPA[1]), which is converted to 3,4-dihydroxyphenylacetate or homoprotocatechuate (HPC) and subsequently degraded by a series of enzymes to the central metabolites pyruvate and succinic semialdehyde (*10, 11*). 4-HPA degradation pathway genes are present in a wide range of bacteria, and homologues of these genes are present in many other organisms including higher eukaryotes. HpcH catalyzes the final step in this pathway, the retro-aldol conversion of 4-hydroxy-2-ketoheptane-1,7-dioate to pyruvate and succinic semialdehyde (*12, 13*). HpcH is a divalent metal ion dependent class II aldolase belonging to the pyruvate/enolpyruvate-binding $(\beta/\alpha)_8$ barrel superfamily. The K12 strain of *Escherichia coli* does not contain the HPC pathway genes and is consequently incapable of utilizing 4-HPA as a sole carbon source. It is therefore surprising to find a gene in *E. coli* K12 (gene b2245) encoding YfaU (GI: 18206229), a protein which is 53% identical to HpcH (Figure 1) and is annotated as a putative 4-hydroxy-2-ketoheptane-1,7-dioate aldolase in the sequence databases. The next closest sequence homologue of YfaU is the *E. coli* K12 enzyme 2-dehydro-3-deoxygalactarate aldolase (DDGA, 43% sequence identity). Sequence identity alone therefore suggests YfaU is a putative aldolase involved in aromatic degradation. To investigate the function of YfaU, and to further investigate the structure−function and evolutionary relationships of this biotechnologically important class of enzymes, we have determined the crystal structures of apo-YfaU to 1.39 Å resolution and a Mg^{2+}−pyruvate product complex to 1.93 Å resolution. The structure is typical of the $(\beta/\alpha)_8$ barrel superfamily of divalent metal ion dependent class II aldolases. Of particular note is the Asp-His intersubunit dyad that appears to activate a metal-bound water or hydroxide for proton transfer during catalysis.

MATERIALS AND METHODS

Materials. 4-Hydroxy-2-ketoheptane-1,7-dioate was prepared as previously described (*14*). 4-Hydroxy-2-ketopentanoic acid, 4-hydroxy-2-keto-5-phenyl-pentanoic acid and 4-hydroxy-2-keto-6-phenyl-hexanoic acid were synthesized according to the method of Rossi and Schinz (*15*), via reaction of the appropriate aldehyde with diethyl oxaloacetate to form a 3-carboethoxy-2-keto-4-substituted γ-butyrolactone, followed by decarboxylation to form the 2-keto-4-substituted γ-butyrolactone, and hydrolysis to form the 2-keto-4-hydroxyacid, potassium salt. *4-Hydroxy-2-ketopentanoic acid, K^+ salt:* δ_H (300 MHz, D_2O) 4.30 (1H, m, H-4), 2.93 (2H, d, $J = 6.0$ Hz, H-3), 1.23 (3H, d, $J = 6.6$ Hz, H-5); δ_C (75 MHz, D_2O) 207.4, 172.6, 63.3, 50.7, 24.9

Table 1: Data Collection and Refinement Statistics[a]

	Apo-YfaU	YfaU-Mg^{2+}−pyruvate
	Data Collection	
synchrotron radiation, detector and wavelength (Å)	SRS 14.2, ADSC Q4 CCD, 0.979	SRS 14.2, ADSC Q4 CCD, 0.980
unit cell (Å)	$a = 105.2$, $b = 136.8$, $c = 123.2$	$a = 105.6$, $b = 137.7$, $c = 124.1$
space group	$C222_1$	$C222_1$
resolution (Å)	60−1.39 (1.44−1.39)	53−1.93 (2.0−1.93)
observations	757535	296821
unique reflections	176395	63843
$I/\sigma(I)$	27.2 (2.1)	32.2 (2.7)
R_{sym}[b]	0.058 (0.397)	0.045 (0.342)
completeness (%)	99.5 (99.3)	94.0 (95.5)
	Refinement	
non-hydrogen atoms	6848 (including 2 glycerol, 6 phosphate and 959 waters)	6443 (including 3 Mg^{2+}, 3 pyruvate, 1 glycerol, 3 phosphate and 543 waters)
R_{cryst}[c]	0.172 (0.257)	0.168 (0.242)
reflections used	169354 (12410)	61264 (4467)
R_{free}[d]	0.191 (0.258)	0.206 (0.312)
reflections used	7041 (513)	2579 (200)
R_{cryst} (all data)[c]	0.172	0.169
mean temperature factor (Å^2)	23.7	37.5
rmsds from ideal values		
bonds (Å)	0.013	0.016
angles (deg)	1.6	1.5
DPI[e] coordinate error (Å)	0.05	0.14

[a] Numbers in parentheses refer to values in the highest resolution shell. [b] $R_{sym} = \sum_j\sum_h|I_{h,j} - \langle I_h\rangle|/\sum_j\sum_h\langle I_h\rangle$ where $I_{h,j}$ is the jth observation of reflection h, and $\langle I_h\rangle$ is the mean intensity of that reflection. [c] $R_{cryst} = \sum||F_{obs}| - |F_{calc}||/\sum|F_{obs}|$ where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. [d] R_{free} is equivalent to R_{cryst} for a 4% subset of reflections not used in the refinement (*44*). [e] DPI refers to the diffraction component precision index (*45*).

ppm; MS (ES −ve ion) 131 (M − H). *4-Hydroxy-2-keto-5-phenyl-pentanoic acid, K^+ salt:* δ_H (400 MHz, D_2O) 7.80 (2H, d, $J = 8.0$ Hz), 7.49 (1H, t, $J = 8.0$ Hz), 7.40 (2H, t, $J = 8.0$ Hz), 5.40 (1H, t, $J = 5.6$ Hz, H-4), 3.30 (2H, d, $J = 5.6$ Hz, H-3); δ_C (100 MHz, D_2O) 136.9, 131.2, 128.7, 128.3, 76.3, 51.3 ppm (carbonyls not seen); MS (ES +ve ion) 194 (MH^+). *4-Hydroxy-2-keto-6-phenyl-hexanoic acid, K^+ salt:* δ_H (400 MHz, D_2O) 7.32 (5H, m), 4.03 (1H, m, H-4), 3.26 (1H, dd, $J = 15.0$, 2.5 Hz, H-3a), 2.55 (1H, dd, $J = 15.0$, 9.8 Hz, H-3b), 1.96 (1H, dd, $J = 15.0$, 10.5 Hz, H-5a), 1.80 (1H, dd, $J = 15.0$, 6.0 Hz, H-5b); δ_C (100 MHz, D_2O) 205.8, 163.7, 139.0, 129.6, 127.4, 126.4, 73.9, 49.3, 32.3 ppm; MS (ES −ve ion) 206 (M − H).

Site-Directed Mutagenesis. Site-directed mutagenesis was carried out according to the Quikchange (Stratagene) site-directed mutagenesis protocol with slight modifications. The His49Ala mutation was achieved using forward primer ctgattgacggggagGCcgcgccaaacaccattc and reverse primer gaatggtgtttggcgcgGCctccccgtcaatcag. The Arg74Ala mutation was achieved using forward primer agccaacccgtgatcGCtccggtggaaggcag and reverse primer ctgccttccaccggaGCgatcacgggttggct (both shown 5′ to 3′, and the altered bases are shown in uppercase). Oligonucleotides were custom synthesized and purified by VH Bio Ltd. (Newcastle, U.K.). Sequencing of the resultant plasmids verified the mutations.

Expression, Purification and Crystallization. The cloning, expression, purification and crystallization of YfaU have been described (*16*). However modified expression and purification procedures were used to improve protein yields. The pT7-7-YfaU plasmid was used to transform *E. coli* strain B834 (DE3) pLacIRARE2 (Novagen). Bacteria were grown at 37 °C in 2 L of LB medium containing 0.1 mg/ml ampicillin and 0.35 mg/mL chloramphenicol. When the optical density of the culture at 600 nm reached 0.6−0.8, protein expression was induced by the addition of IPTG to a final concentration of 1 mM and growth was continued at 20 °C for 16 h. The cells were harvested and sonicated, and the YfaU enzyme was purified using metal affinity and size exclusion chromatography as described for the related HpcH enzyme (*13*), except a Superdex 200 column was used instead of a Superdex 75 column.

Enzyme Kinetic Assays. Aldolase activity with 2-keto-3-deoxy-L-rhamnonate, 2-keto-3-deoxy-L-lyxonate and 2-keto-3-deoxy-L-mannonate was measured as described (see accompanying paper (*47*)). Aldolase activity with 4-hydroxy-2-ketopentanoic acid, 4-hydroxy-2-ketohexanoic acid and 4-hydroxy-2-ketoheptane-1,7-dioate was measured using a coupled assay containing lactate dehydrogenase (LDH) by monitoring the decrease in absorbance at 340 nM in a spectrophotometer. The reactions contained varying amounts of substrate, 0.5 mM divalent metal ion chloride, 0.3 mM NADH, 3 μL 1 mg/mL LDH in a total of 1 mL of 100 mM sodium HEPES (pH 8.0). The reaction was initiated by the addition of purified YfaU. Enzyme kinetic parameters were calculated using GraphPad Prism (San Diego, CA).

Structure Determination and Model Refinement. Crystals were picked up from their mother liquor (which contained 30% glycerol) using a cryoloop, placed in a nitrogen stream at 100 K and stored in liquid nitrogen until needed for data collection. To introduce Mg^{2+} and pyruvate, $MgCl_2$ and Na^+ pyruvate were cocrystallized with apo-YfaU at a final concentration of 10 mM. Diffraction data were collected at the Synchrotron Radiation Source, Daresbury. Integration and scaling were performed using DENZO and SCALEPACK (*17*). Subsequent data handling was carried out using the CCP4 software package (*18*). A summary of data collection statistics is shown in Table 1. Molecular replacement was carried out using the coordinates of 2-dehydro-3-deoxygalactarate aldolase (*19*) (DDGA, pdb code 1DXE) as a search model with the AMORE (*20*) program. Automatic chain tracing was performed by ARP-wARP (*21*), resulting in the unambiguous building of almost the entire sequence. Alter-

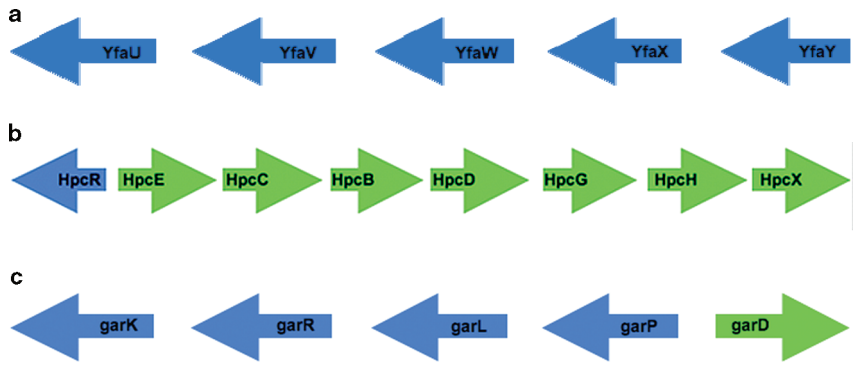

FIGURE 2: (a) *E. coli* K12 operon containing YfaU. YfaU, putative 2-keto-4-hydroxy-1,7-dioate aldolase; YfaV, putative transporter of the major facilitator superfamily; YfaW, putative dehydratase of the enolase superfamily; YfaX, putative transcriptional regulator; YfaY, related to molybdopterin biosynthesis protein MoeA. (b) *E. coli* C 4-hydroxyphenylacetate degradation pathway operon. HpcB, 3,4-dihydroxyphenylacetate dioxygenase; HpcC, 5-carboxymethyl-2-hydroxymuconate semialdehyde reductase; HpcD, 5-carboxymethyl-2-hydroxymuconate isomerase; HpcE, 5-carboxymethyl-2-oxo-hex-3-ene-1,6-dioate decarboxylase/ 2-hydroxyhepta-2,4-diene1,7-dioate isomerase; HpcG, 2-oxohepta-4-ene-1,7-dioate hydratase; HpcH, 2-keto-4-hydroxyheptane-1,7-dioate aldolase; HpcR, transcriptional regulator; HpcX, probable 4-HPA transporter. (c) *E. coli* K12 galactarate degradation pathway operon. garD, probable galactarate dehydratase; garR, glycerate kinase (glycerate- or glucarate-inducible); gar L (DDGA), 2-dehydro-3-deoxygalactarate aldolase; garP, probable D-galactarate transporter; garR, tartronate semialdehyde reductase. Green and blue indicate transcription in a clockwise and anticlockwise direction, respectively.

nate cycles of refinement using Refmac (*22*) and manual rebuilding using O (*23*) were used to complete the structure. The Apo-YfaU coordinates were used in the refinement of the YfaU-Mg^{2+}−pyruvate complex. Mg^{2+} and pyruvate were manually included, and ARP-wARP (*21*) was used to automatically place solvent molecules in electron density peaks where acceptable hydrogen bonds could be formed. Evaluation of the model with PROCHECK (*18*) shows no residues in disallowed regions of the Ramachandran plot. A summary of the refinement statistics is given in Table 1. Figures were drawn using Pymol (*24*) and Molscript (*25*).

RESULTS AND DISCUSSION

Genomic Context and Proposed Function. Prokaryotic genomes are generally organized into operons that contain a number of functionally related genes, and functional information can often be gained from analysis of the genomic context. YfaU is part of a functionally unassigned operon in *E. coli* K12 (Figure 2a) that also includes a putative transcriptional regulator (YfaX), a putative transporter of the major facilitator superfamily (YfaV), a putative dehydratase belonging to the enolase superfamily (YfaW), and a protein related to the molybdopterin biosynthesis protein MoeA (YfaY). This gene organization is very different from the 4-HPA operon that contains HpcH, the closest sequence homologue of YfaU (Figure 2b). Equivalents of the reductase (HpcC), isomerase (HpcD), decarboxylase (HpcE) and hydratase (HpcG) genes of the HPC degradation operon are not present in the operon containing YfaU. Rather, the gene organization more closely resembles the *E. coli* K12 galactarate degradation pathway operon (Figure 2c). This operon comprises the probable D-galactarate transporter garP, the D-galactarate dehydratase garD, and the 2-dehydro-3-deoxygalactarate aldolase garL. DDGA, the protein encoded by garL, produces pyruvate and tartronate semialdehyde, and the operon also contains the tartronate semialdehyde reductase gene garR to convert tartronate semialdehyde to tartronate. While no direct functional information is available for YfaU, YfaW functions as an L-rhamnonate dehydratase, and has been given the name RhamD (see accompanying paper (*47*)). This function was assigned by screening a panel of mono- and diacid sugars as potential substrates as used for the functional assignment of other dehydratse enzymes belonging to the enolase superfamily (*26*−*29*). YfaU does indeed catalyze the retro-aldol reaction on 2-dehydro-3-deoxy-L-rhamnonate, the product of the YfaW/RhamD reaction (discussed below). This illustrates the importance of using operon context as well as homology-based approaches for the functional assignment of prokaryotic genes.

Crystal Structure. The crystal structure of YfaU was determined by molecular replacement using the coordinates of (pdb code 1DXE) as a search model (*19*). Refinement resulted in a final model with a resolution of 1.39 Å and an *R*-factor of 0.17. The Ramachandran plot (*18*) shows the refined structure to have good geometry. The YfaU-Mg^{2+}−pyruvate structure was determined using the coordinates of apo-YfaU. Statistics for the structure determination and refinement are found in Table 1.

YfaU shares the same overall monomer and oligomer structure as the homologous enzymes HpcH (*13*), DDGA (*19*) and macrophomate synthase (MPS) (*30*); a hexameric assembly comprising a trimer of domain-swapped $(\beta/\alpha)_8$ barrel dimers (Figure 3a). Seven α-helices from one subunit and 1 α-helix from the domain-swapped subunit form the central core of the $(\beta/\alpha)_8$ barrel. The asymmetric unit contains the 3-fold related subunits, and the crystallographic 2-fold axis generates the biologically relevant hexamer possessing *32* symmetry. The 2-fold interface is primarily composed of hydrophobic interactions resulting from the deep burial of helix 8 between helices 1 and 8 of the neighboring 2-fold related subunit. The 3-fold interface is much less intimate, but part of it is very close to, and forms part of, the active site. An important salt bridge between His49 $N^{\delta 1}$ and Asp88 $O^{\delta 2}$ of the 3-fold related subunit (Figure 3b) forms a possible catalytic triad with a metal-bound water molecule (catalytic mechanism discussed below). The carbonyl group of Val122 from the 3-fold related subunit also forms a hydrogen bond with this active site water molecule. All 6 subunits of the hexamer adopt very similar conformations; the 2-fold related subunits of a domain-swapped dimer superpose with an rmsd

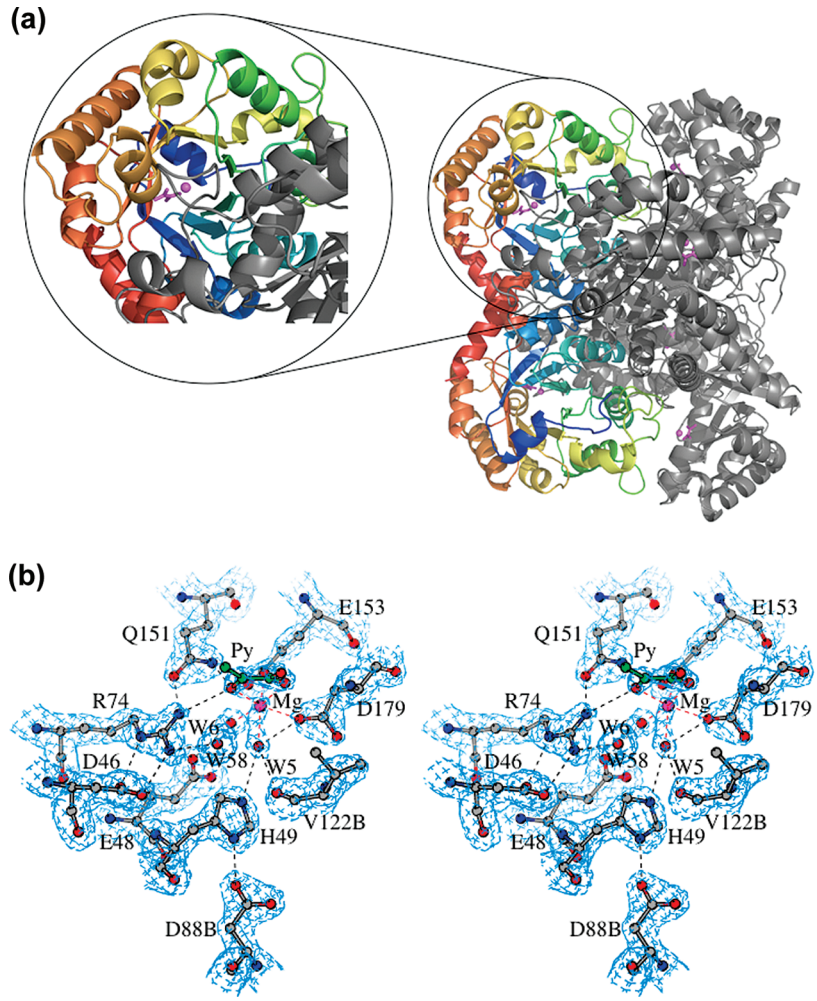

FIGURE 3: (a) Cartoon representation of the YfaU hexameric assembly. The YfaU monomer is a $(\beta/\alpha)_8$ barrel, and 2 barrels form a domain-swapped dimer via swapping of helix 8. Three domain-swapped dimers coalesce to form the biologically relevant hexamer with *32* symmetry. The polypeptide chain is color-ramped blue to red from the N to the C terminus for each of the subunits of one of the domain-swapped dimers. The 6 Mg^{2+} ions (magenta spheres) and 6 pyruvate molecules (magenta stick representation) indicate the position of the active sites. Figure drawn with Pymol (*24*). (b) Stereoview of the active site of the YfaU-Mg^{2+}−pyruvate complex. The SIGMAA weighted $2mF_o - \Delta F_c$ electron density (*46*) is contoured at the 1.5σ level, where σ represents the rms electron density for the unit cell. Contours more than 1.4 Å from any of the displayed atoms have been removed for clarity. The carbon atoms of the enzyme and the pyruvate ligand are colored gray and green, respectively. The Mg^{2+} ion is shown as a magenta sphere. Red and black dashed lines indicate metal coordination bonds and hydrogen bonds, respectively. Figure drawn with MolScript (*25*).

of 0.25 Å for 235 Cα atoms, and a 3-fold related subunit superposes with an rmsd of 0.26 Å.

Active Site Architecture. The active site of YfaU was defined by determination of the structure of the Mg^{2+}−pyruvate product complex (Figure 3b). This structure was solved by molecular replacement using the coordinates of apo-YfaU and refined to an *R*-factor of 0.17. Clear electron density is visible for both the Mg^{2+} ion and the pyruvate molecule. These are located at the C-terminal ends of the beta strands that form the core of the barrel, as are the active sites of HpcH, DDGA, MPS and most other $(\beta/\alpha)_8$ barrel enzymes. Apo-YfaU and the Mg^{2+}−pyruvate complex superpose with an rmsd of 0.16 Å for 238 Cα atoms, and the uncomplexed and liganded domain-swapped dimer superpose with an rmsd of 0.35 Å for 472 Cα atoms, indicating negligible conformational changes upon ligand binding. The Mg^{2+} ion is coordinated by Glu153-$O^{\varepsilon 1}$ (2.1 Å) and Asp179 and $O^{\delta 2}$ (2.2 Å). The carboxyl group and the carbonyl group of pyruvate provide two further Mg^{2+} ligands (2.5 Å and 2.4 Å, respectively), and 2 ordered water molecules (W5, 2.1 Å and W6, 2.3 Å) complete the octahedral coordination shell. This ligand binding mode and coordination sphere are very similar to those observed in the structures of homologous enzymes (Figure 4). Other enzyme−ligand interactions include hydrogen bonds between the pyruvate carboxylate group and the main chain NH groups of Ala178 (3.3 Å) and Asp179 (3.4 Å). The pyruvate carbonyl group also has hydrogen bonds to Arg74-$N^{\delta 1}$ (3.2 Å) and Gln151-$N^{\varepsilon 2}$ (2.9 Å). These hydrogen bonds and the metal ion appear to polarize the carbonyl group, favoring formation of the enolate, thus promoting proton transfer at C3. Arg74 is held in position with salt bridges between Asp46-$O^{\delta 1}$ and Nε (3 Å), and Asp46-$O^{\delta 2}$ and $N^{\eta 2}$. There are a number of ordered water molecules (W) in the active site of the Mg^{2+}−pyruvate complex that may be catalytically important (Figures 3b and 4a). Two of these are the metal-bound W5 and W6. W5 is sandwiched between the metal ion (2.1 Å) and His49-$N^{\delta 1}$ (3.1 Å), and there are further hydrogen bonds to Asp79-$O^{\delta 2}$ (2.8 Å) and the carbonyl O of Val122 from the 3-fold related subunit (2.9 Å). His49 is further hydrogen bonded via $N^{\varepsilon 2}$ to Asp88-$O^{\delta 2}$ (2.4 Å) from the 3-fold related subunit. W6 is sandwiched between the

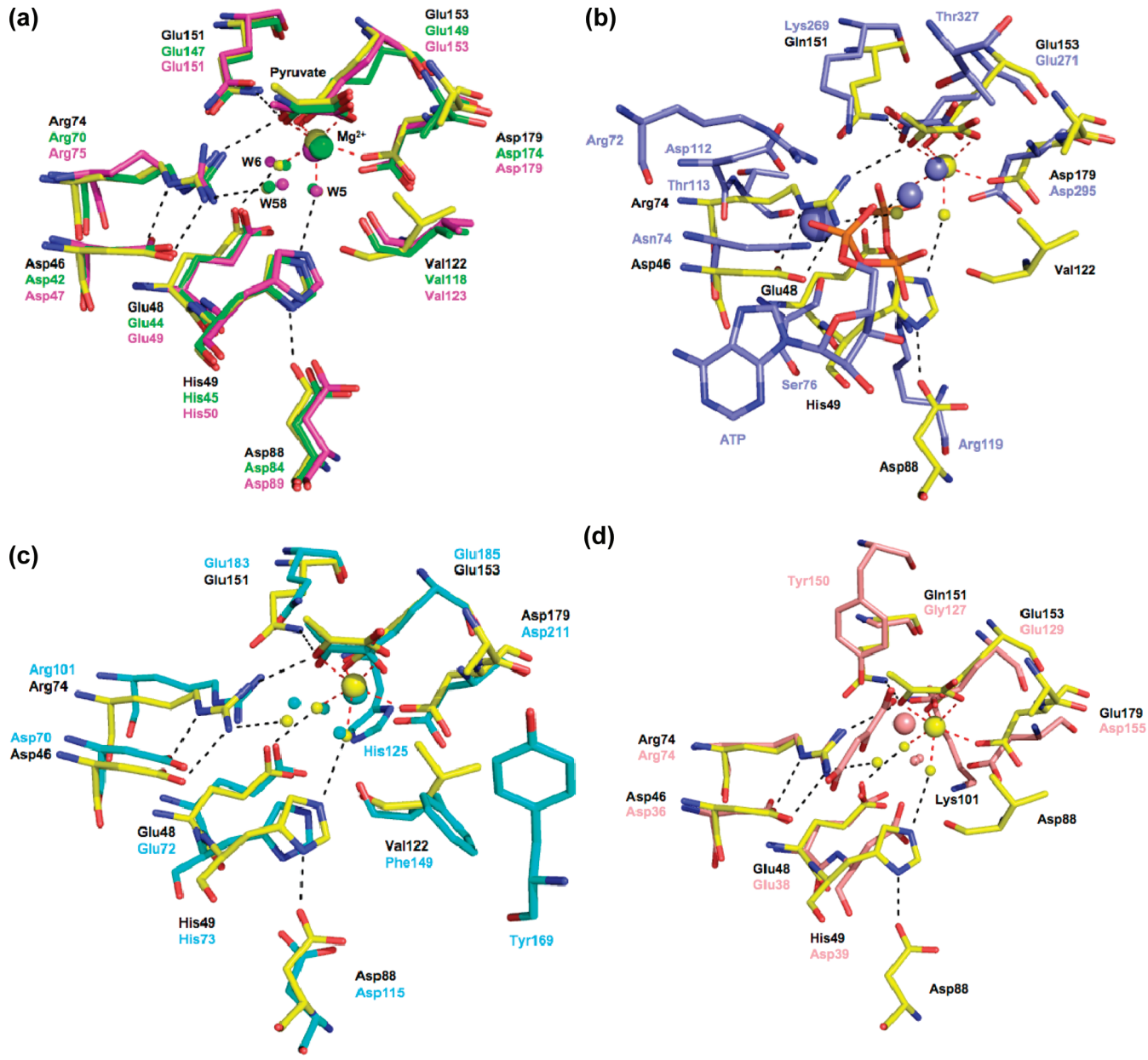

FIGURE 4: Structural alignment of the YfaU-Mg^{2+}−pyruvate complex (yellow) with (a) the HpcH-Mg^{2+}−oxamate (green) and DDGA-Mg^{2+}−pyruvate (magenta) complexes, (b) the PykF-Mg^{2+}−oxalate−ATP (blue) complex, (c) the MPS-Mg^{2+}−pyruvate (cyan) complex, and (d) the CL-Mg^{2+}−oxaloacetate (salmon) complex. Metal ions and ordered water molecules are shown as large and small spheres, respectively. Metal coordination bonds and hydrogen bonds for the YfaU complex are shown as red and black dashed lines, respectively. Figure drawn with Pymol (*24*).

metal ion (2.3 Å) and Glu48-$O^{\varepsilon 1}$ (2.7 Å), with a further hydrogen bond to Glu153-$O^{\varepsilon 2}$ (2.9 Å). W58 is bound to Arg74-$N^{\eta 2}$ (3.2 Å).

Comparison with Related Enzymes. Close homologues were identified using the BLAST algorithm (*31*) at the NCBI server, and sequences were aligned using ClustalW (*32*) (Figure 1). YfaU shares the following sequence identities: 53% with HpcH, 49% with AkbF (4-hydroxy-2-oxovalerate aldolase from *Rhodococcus sp. DK17*), 43% with DDGA, 42% with PykF (pyruvate kinase from *Mannheimia succiniciproducens*) and 21% with MPS (macrophomate synthase from *Macrophoma commelinae*). The closest homologues HpcH, AkbF and DDGA are all metal dependent class II aldolases. The structure of YfaU was compared with structures of these homologous enzymes to identify the structural features responsible for the different catalytic capabilities present in this superfamily. Structural superpositions were performed using Pymol (*24*). HpcH, the closest sequence homologue, is also the closest structural homologue. The HpcH-Mg^{2+}−oxamate complex superposes with the YfaU-Mg^{2+}−pyruvate complex with an rmsd of 0.6 Å for 231 Cα atoms (Figure 4a). The DDGA-Mg^{2+}−pyruvate complex superposes with the YfaU-Mg^{2+}−pyruvate complex with an rmsd of 0.83 Å for 237 Cα atoms (Figure 4a). The lower structural homology with DDGA reflects the lower sequence identity. The close structural homology with HpcH and DDGA is further evidence that YfaU is a class II aldolase catalyzing a retro-aldol reaction similar to those of HpcH and DDGA (Figure 5a).

As with other branches of the $(\beta/\alpha)_8$ barrel enzymes, the pyruvate/enolpyruvate-binding superfamily is a robust framework upon which a number of different catalytic activities

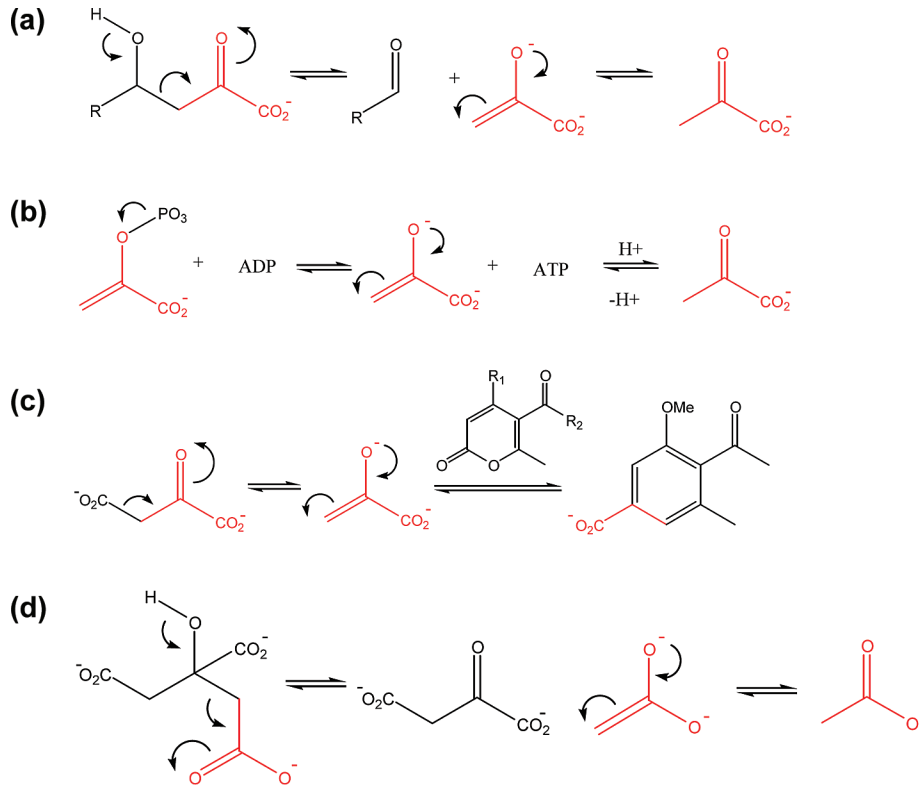

FIGURE 5: Reactions catalyzed by the pyruvate/enolpyruvate-binding $(\beta/\alpha)_8$ barrel superfamily. (a) In the reaction catalyzed by the class II aldolases the condensed substrate is converted by a retro-aldol reaction to an aldehyde moiety and pyruvate enolate that is subsequently ketonized to pyruvate. R = $CH(OH)CH_3$, $CH_2CH_2CO_2$, and $CH(OH)CO_2$ for YfaU, HpcH and DDGA, respectively. (b) In the reaction catalyzed by PykF, phosphoenolpyruvate and ADP (and a proton) are converted to pyruvate (via the enolate) and ATP. (c) In the reaction catalyzed by MPS, oxaloacetate is decarboxylated to pyruvate enolate. A Diels−Alder cycloaddition with a 2-pyrone and subsequent decarboxylation and dehydration results in a six-membered aromatic benzoate. (d) In the reaction catalyzed by CL, citrate is converted to oxaloacetate and acetate. The enolate-forming parts are colored red and curly arrows are included to highlight the mechanistic similarities of the reactions.

can be built. YfaU shares a remarkably high sequence similarity with PykF (42%, Figure 1) despite the differences in the reactions catalyzed by these enzymes (Figures 5a and 5b). PykF catalyzes the final step of glycolysis, the reaction of phosphoenolpyruvate, ADP and a proton to give ATP and pyruvate (*33*). The PykF reaction proceeds in two chemical steps. The first step is phosphoryl transfer from phosphoenolpyruvate to give ATP and pyruvate enolate (*34*). The second step is the addition of a proton to the 2 *si* face of the pyruvate enolate to produce pyruvate (*35*). A close structural alignment is obtained when residues 250−350 of the PykF-Mg^{2+}−oxalate−ATP complex are aligned with residues 60−190 of the YfaU-Mg^{2+}−pyruvate complex (42 Cα atoms are superposed with an rmsd of 2.2 Å). The divalent metal ion of PykF is coordinated in the same fashion as aldolases with Glu271 and Asp295 equivalent to Glu153 and Glu179 of YfaU (Figure 4b). Oxalate, an enolate analogue, is bound in an equivalent bidentate mode, and 2 water molecules complete the coordination shell. Promotion and stabilization of the enolate is even more pronounced in PykF since Arg72 is functionally equivalent to Arg74 of YfaU, but Lys269 is present instead of the equivalent Gln151. This basic pair of residues, along with the divalent metal ion, forms a very strong electropositive region in PykF highly suited for stabilizing the enolate and hence driving phosphotransfer. The main differences result from the structurally distinct substrates bound by the enzymes. The side chains of Asn74, Ser76, Asp112 and Thr113 of PykF are involved in binding the K^+ ion that is in turn involved in binding the terminal phosphate group of the ATP product (Figure 4b). PykF has Arg119 in place of His49, and this is also involved in binding the negative β- and χ-phosphate groups of ATP. Thus the pyruvate-binding half of the active site resembles a class II aldolase and the nucleotide binding half-is quite different. There is an additional divalent metal ion associated with the phosphate groups of the bound ATP.

YfaU shares a much lower sequence identity with MPS (21%, Figure 1). MPS carries out an unusual natural Diels−Alder reaction forming a 6-membered aromatic benzoate from an alkene and a 1,3-diene (Figure 5c). This is a complicated multistep reaction involving the initial decarboxylation of oxaloacetate to form enolpyruvate, followed by the Diels−Alder cycloaddition of the pyruvate enolate and the 2-pyrone, and subsequent decarboxylation and dehydration (*30, 36*). A close structural alignment is obtained when residues 50−220 of the MPS-Mg^{2+}−pyruvate complex are aligned with residues 70−190 of the YfaU-Mg^{2+}−pyruvate complex (100 Cα atoms are superposed with an rmsd of 1.4 Å). This represents a substantially greater degree of structural similarity than with PykF despite the much lower sequence identity. The divalent metal ion of MPS is coordinated by Glu185 and Asp211 that are equivalent to Glu153 and Glu179 of YfaU (Figure 4c). Pyruvate is again bound in a bidentate fashion, and 2 water molecules complete the coordination shell. MPS has a very similar pyruvate enolate binding site, with Arg101 and Gln183 directly equivalent to Arg74 Gln151 of YfaU. Asp70, Glu72 and His73 all have identical counterparts in YfaU

Table 2: Kinetic Data for Mg^{2+}-Activated YfaU

substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} M^{-1})
2-keto-3-deoxy- L-lyxonate	0.3	0.8	3.8×10^2
2-keto-3-deoxy- L-mannonate	0.3	0.14	2.1×10^3
2-keto-3-deoxy- L-rhamnonate	0.4	0.078	5.1×10^3
4-hydroxy-2-ketoheptane-1,7-dioate	0.54	0.15	3.6×10^3

(Asp46, Glu48 and His49, respectively). The MPS active site has Asp115 and Phe149 from the 3-fold related subunit that are equivalent to Asp88 and Val122 from the 3-fold related subunit in YfaU. The main difference in the MPS active site is the presence of His125 that sits directly beneath the divalent metal ion, and Tyr169 from the 3-fold related subunit. His125 is not believed to be involved in the catalytic mechanism. Instead, Tyr169 from the 3-fold related subunit is involved in the cycloaddition step, as shown by site-directed mutagenesis (*36*). The binding site for the 2-pyrone substrate is unknown at present, preventing a full mechanistic understanding.

The β-subunit of citrate lyase (CL) is an even more distant homologue of YfaU, sharing a sequence identity of only 13% with the *Deinococcus radiodurans* enzyme. CL catalyzes a reversible retro-aldol-like reaction in which citrate is converted to acetate and oxaloacetate (*37*) (Figure 5d). Despite the low level of sequence identity, an appreciable degree of structural similarity is observed (Figure 4d). A close structural alignment is obtained when residues 50−220 of the CL-Mg^{2+}−oxalate complex are aligned with residues 70−190 of the YfaU-Mg^{2+}−pyruvate complex (101 Cα atoms are superposed with an rmsd of 3 Å). The divalent metal ion is coordinated by Glu129 and Asp155 that are equivalent to Glu153 and Glu179 of YfaU. The product oxaloacetate is again bound in a bidentate manner, but in a significantly different orientation to pyruvate in the class II aldolase structures. The enolate stabilization motif differs slightly; Arg74 is equivalent to Arg74 of YfaU, but CL has Tyr150 in place of Gln151. Asp36 and Glu38 have direct counterparts in YfaU (Asp46 and Glu48, respectively). However CL has Asp39 in place of His49 in YfaU. The reaction catalyzed by CL involves the initial proton abstraction from the C3 hydroxyl of citrate, and Asp39 is ideally placed for this general base role, 3.6 Å from the C3 hydroxyl. Given the similarity of the reactions catalyzed by YfaU (Figure 5a) and CL (Figure 5d) and the similar locations of these residues in their respective active sites, His49 is a good candidate for involvement in the initial proton abstraction at the C4 hydroxyl in the YfaU reaction.

Although the PykF, MPS and CL reactions are different from those of the class II aldolases and different from each other, they all have features in common. They share a partial reaction that generates an enolate intermediate that is stabilized by interaction with a divalent metal ion and an active site Arg residue (catalytic mechanism discussed below).

Substrate Specificity. L-Rhamnonate, L-lyxonate, and L-mannonate are all dehydrated by RhamD, and the resulting 2-keto-3-deoxy acid sugars are all substrates for the similarly promiscuous YfaU (Table 2, and Table 2 of ref *47*). 2-Keto-3-deoxy L-rhamnonate is the best substrate ($k_{cat} = 0.4$ s^{-1}, $K_m = 0.078$ mM, $k_{cat}/K_m = 5.1 \times 10^3$ s^{-1} M^{-1}), but 2-keto-3-deoxy-L-mannonate ($k_{cat} = 0.3$ s^{-1}; $K_m = 0.8$ mM, $k_{cat}/K_m = 3.8 \times 10^2$ s^{-1} M^{-1}) and 2-keto-3-deoxy-L-lyxonate ($k_{cat} = 0.3$ s^{-1}; $K_m = 0.14$ mM, $k_{cat}/K_m = 2.1 \times 10^3$ s^{-1} M^{-1}) are also reasonably good substrates. We tested YfaU with HKHD, the substrate of the closely related HpcH aldolase (Table 2). The kinetic constants with HKHD are similar to those obtained for the aforementioned substrates. 2-Keto-3-deoxy-L-rhamnonate has a slighter higher affinity and substrate specificity constant, further supporting this as the physiological substrate, but it is apparent that the rather relaxed substrate specificity of the class II aldolases can complicate the functional assignment of these enzymes. Significantly, these data were obtained with Mg^{2+}-activated enzyme, but we observed that YfaU is more catalytically proficient when Ni^{2+} is reconstituted in the active site (Table 3). While Ni^{2+} is not likely to be the physiologically relevant active site metal, this enhanced activity could be useful for future biocatalytic applications. Ni^{2+}-activated and Mg^{2+}-activated enzymes have similar affinity for HKHD, but the Ni^{2+}-activated enzyme has a higher turnover ($k_{cat} = 299$ s^{-1}), resulting in an elevated substrate specificity constant. Co^{2+} was previously determined to be the optimal metal for HpcH (*38*), and Co^{2+}-activated HpcH is even more active toward its own substrate, as expected ($k_{cat} = 332$ s^{-1}; $K_m = 0.03$ mM, $k_{cat}/K_m = 1.1 \times 10^7$ s^{-1} M^{-1}, unpublished data). Similarly to Mg^{2+}-activated YfaU, Mg^{2+}-activated HpcH was considerably less active when assayed with HKHD ($k_{cat} = 9.4$ s^{-1}; $K_m = 0.05$ mM, $k_{cat}/K_m = 1.9 \times 10^5$ s^{-1} M^{-1}, unpublished data). The synthetic substrates 4-hydroxy-2-ketopentanoic acid ($k_{cat} = 396$ s^{-1}; $K_m = 0.1$ mM, $k_{cat}/K_m = 3.9 \times 10^6$ s^{-1} M^{-1}) and 4-hydroxy-2-ketohexanoic acid ($k_{cat} = 447$ s^{-1}; $K_m = 0.05$ mM, $k_{cat}/K_m = 9.9 \times 10^6$ s^{-1} M^{-1}) also proved very good substrates with Ni^{2+}-activated YfaU (Table 3). These kinetic constants are comparable to Co^{2+}-activated HpcH with HKP and HKH (*38*). 4-Hydroxy-2-keto-5-phenyl-pentanoic acid and 4-hydroxy-2-keto-6-phenyl-hexanoic acid were also synthesized and tested, but these were not turned over by either YfaU or HpcH, suggesting a requirement for an aliphatic or less bulky distal end of the substrate. Therefore HpcH and YfaU both show rather broad specificity toward the keto-acid substrate, but show differences in their metal ion specificity.

Table 3: Kinetic Data for Ni^{2+}-Activated YfaU

substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} M^{-1})
4-hydroxy-2-ketoheptane-1,7-dioate	299	0.1	2.7×10^6
4-hydroxy-2-ketopentanoic acid	396	0.1	3.9×10^6
4-hydroxy-2-ketohexanoic acid	447	0.05	9.9×10^6

At present only crystal structures containing the 3-carbon pyruvate product and not the full-length substrates have been determined for these enzymes. The exact binding positions and hence substrate specificity determinants cannot therefore be fully understood at present. There do not appear to be any obvious substrate-binding residues that can explain the preferences of YfaU, HpcH or DDGA for their particular substrates. The enzyme−product complexes obtained to date suggest that the substrate is tethered in the active site by the carboxylate and ketone group that are bound to the divalent metal ion, while the distal end remains solvent-exposed and flexible, allowing substrates of differing lengths to be cleaved. This explanation is consistent with the substrate specificity data presented here and elsewhere (*38*).

Catalytic Mechanism. The retro-aldol reaction carried out by YfaU, HpcH, DDGA and related enzymes involves (1)

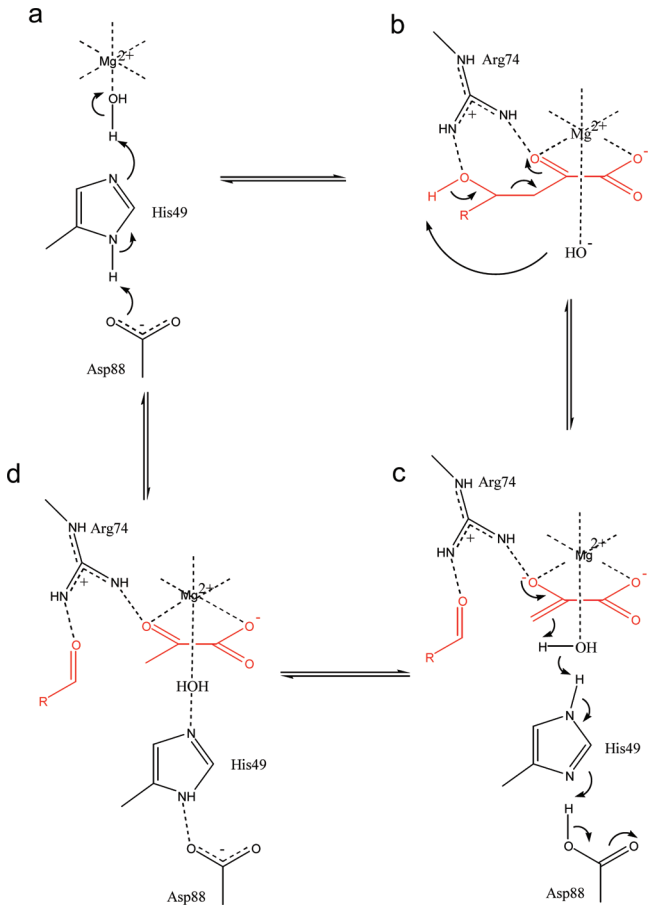

FIGURE 6: Proposed catalytic mechanism of YfaU. (a) The Asp88-His49 intersubunit dyad deprotonates the metal-bound water molecule.(b) The condensed substrate binds to the metal ion in a bidentate fashion with the C3 carbonyl and C4 hydroxyl interacting with Arg74. The metal-bound hydroxide ion, activated by the intersubunit Asp88-His49 dyad, abstracts the C4 hydroxyl proton, coupled with breakage of the C3−C4 carbon−carbon bond. (c) The resulting aldehyde product and pyruvate enolate are stabilized by the metal ion and Arg74. Protonation of pyruvate enolate at C3 by the Asp88-His49-metal-bound water results in the other product pyruvate (d), returning the enzyme to the ground state and completing the catalytic cycle.

abstraction of the proton from the C4 hydroxyl group, (2) formation of the C4-keto group coupled with breakage of the C3−C4 carbon−carbon bond to form an aldehyde moiety and pyruvate enolate, and (3) ketonization of the enolate and protonation at C3 to generate pyruvate (Figure 6). One or 2 general acid/base groups are required, depending on whether the same or different catalytic groups carry out the proton transfers at the C3 and C4 carbons. It is not immediately obvious from the crystal structures which residue(s) fulfill this requirement. The YfaU-Mg^{2+}−pyruvate and equivalent HpcH and DDGA complexes do not reveal the position of the condensed substrate, only the 3-carbon pyruvate, and only represent 1 snapshot of the catalytic cycle. The structures clearly show the importance of Arg74 that binds and activates the C2 keto group and stabilizes the enolate. The catalytic importance of Arg74 was confirmed by site-directed mutagenesis; the Arg74Ala mutant was totally devoid of activity. This was also the case for the equivalent Arg70Ala mutant of HpcH (*13, 38*). Although no structures are available of a complex containing the condensed substrate or aldehyde moiety for the class II aldolases, we can obtain clues from the structure of the related malate synthase (MS). Sequence identity with the class II aldolases is very low (8% with YfaU), but MS is also a $(\beta/\alpha)_8$ barrel enzyme (*39*). MS catalyzes the Claisen condensation of acetyl-coA and glyoxylate to form malate (*40, 41*), which has features in common with the retro-aldol reaction catalyzed by the class II aldolases (in the condensation direction). MS first abstracts a proton from acetyl-CoA to generate an enolate that is stabilized by a divalent metal ion and the active site Arg338. MS then condenses the acetyl-CoA enolate with the aldehyde group of glyoxylate that is also bound to and activated by Arg388 (*40, 41*). Given the similarities between the enzymes, Arg74 of YfaU is probably functionally equivalent to Arg338 of MS, binding to and activating the keto group of the aldehyde acceptor as well as the pyruvate moiety. If this is the case, W58 bound to Arg74 $N^{\eta 2}$ (3.2 Å) occupies the expected position of the C4 hydroxyl of the condensed substrate. W58 and, hence, the C4 hydroxyl of the condensed substrate are 4 Å from the C3 atom of pyruvate, and this distance is only 3.5 Å and 3.1 Å in the equivalent HpcH and DDGA complexes, respectively (Figure 4a). The only potential acid/base in the vicinity of W58 is W5 (3.8 Å), and again this distance is even shorter in the HpcH and DDGA complexes (3 Å and 2.7 Å, respectively). W5 is hydrogen bonded to His49 $N^{\varepsilon 2}$ (3.1 Å) that is in turn hydrogen bonded via $N^{\delta 1}$ to Asp88 $O^{\delta 2}$ (2.4 Å). This very short hydrogen bond is reminiscent of the low energy hydrogen bonds found between the Asp and His residues of Ser-Asp-His catalytic triads found in peptidases and related hydrolytic enzymes (*42*). Thus W5, activated by the intersubunit Asp88-His49 dyad, is well poised to deprotonate the C4 hydroxyl. Additionally, given that W5 is also coordinated to the divalent metal ion (2.1 Å), it is probable that W5 is stabilized as a metal-bound hydroxide ion that would be an even stronger general base. The Asp88-His49-W5 motif occupies the same location as Asp39 of CL (Figure 4d), the general base that abstracts the proton from the hydroxyl group in this analogous reaction, further supporting the role of this motif as the C4 general base. Site-directed mutagenesis confirmed the catalytic importance of His49; the His49Ala mutant was completely inactive. This was also the case with the equivalent His45Ala mutant of HpcH (*13*). Deprotonation of the C4 hydroxyl group is coupled with keto formation and C−C bond breakage, resulting in the aldehyde moiety and pyruvate enolate (Figure 6). The aldehyde moiety is released and the enolate is protonated at C3 to yield pyruvate. The metal-bound W5 that participates as a general base in the first part of the reaction may act as the general acid in this step. The pyruvate C3−W5 distance is 5.4 Å (4.4 Å and 4.3 Å for the HpcH and DDGA complexes, respectively). This distance could be considerably shorter during the actual catalytic cycle given modest plasticity. No other obvious candidates for general acid/base groups catalyzing proton transfer at C3 are apparent from any of the existing structures.

CONCLUSIONS

We used structural and biochemical studies to confirm that YfaU, an enzyme in a previously uncharacterized operon in *E. coli* K12, is a metal-dependent class II aldolase belonging to the pyruvate/enolpyruvate binding $(\beta/\alpha)_8$ barrel superfamily. Genomic context suggested a 2-dehydro-3-deoxy

sugar aldolase function rather than the 4-hydroxy-2-keto-heptane-1,7-dioate aldolase function as annotated in the databases based on homology with the related HpcH. Functional assignment of YfaW as a L-rhamnonate dehydratase suggested 2-dehydro-3-deoxy-L-rhamnonate is the physiological substrate of YfaU, and enzyme kinetic measurements support this assignment. A combination of approaches therefore appears to be the best strategy for functional assignment of metal-dependent class II aldolases, as has been shown for other enzyme superfamilies (*26*−*29*). Crystal structures of the closely related YfaU, HpcH and DDGA and structures of more distantly related enzymes allowed us to make mechanistic proposals. An intersubunit Asp-His dyad appears to activate a metal-bound water or hydroxide that acts as the general acid/base for this class of enzymes.

YfaU and HpcH are able to catalyze the retro-aldol reaction on structural analogues of the physiological substrates including 4-hydroxy-2-ketopentanoate and 4-hydroxy-2-ketohexanoate (*38*). A recent study has also shown that pyruvate is not absolutely required as the nucleophile in the condensation direction for this class of enzymes (*43*). Specific aldolases including YfaU were used to catalyze the aldol condensation of acetaldehyde and α-ketobutyrate, resulting in the formation of 4-hydroxy-3-methyl-2-keto-pentanoate. Amination by a branched-chain amino acid aminotransferase resulted in 4-hydroxy-L-isoleucine in a 2-step synthesis (*43*). This demonstrates one way of overcoming the unfavorable equilibrium position to utilize these enzymes in the condensation direction, and highlights the huge potential of this class of enzymes for stereochemically controlled carbon−carbon bond formations. Structure−function studies such as this are important if this potential is to be realized.

ACKNOWLEDGMENT

We wish to thank the Synchrotron Radiation Source, Daresbury, for assistance with X-ray diffraction data collection.

REFERENCES

1. Saghatelian, A., and Cravatt, B. F. (2005) Assignment of protein function in the postgenomic era. *Nat. Chem. Biol. 1*, 130–142.
2. Watson, J. D., Sanderson, S., Ezersky, A., Savchenko, A., Edwards, A., Orengo, C., Joachimiak, A., Laskowski, R. A., and Thornton, J. M. (2007) Towards fully automated structure-based function prediction in structural genomics: a case study. *J. Mol. Biol. 367*, 1511–1522.
3. Lee, D., Redfern, O., and Orengo, C. (2007) Predicting protein function from sequence and structure. *Nat. Rev. Mol. Cell Biol. 8*, 995–1005.
4. Janin, J. (2007) Structural genomics: winning the second half of the game. *Structure 15*, 1347–1349.
5. Gileadi, O., Knapp, S., Lee, W. H., Marsden, B. D., Muller, S., Niesen, F. H., Kavanagh, K. L., Ball, L. J., von Delft, F., Doyle, D. A., Oppermann, U. C., and Sundstrom, M. (2007) The scientific impact of the Structural Genomics Consortium: a protein family and ligand-centered approach to medically-relevant human proteins. *J. Struct. Funct. Genomics 8*, 107–119.
6. Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., and Raushel, F. M. (2007) Structure-based activity prediction for an enzyme of unknown function. *Nature 448*, 775–779.
7. Song, L., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Brown, S., Imker, H. J., Babbitt, P. C., Almo, S. C., Jacobson, M. P., and Gerlt, J. A. (2007) Prediction and assignment of function for a divergent N-succinyl amino acid racemase. *Nat. Chem. Biol. 3*, 486–491.
8. Diaz, E., Ferrandez, A., Prieto, M. A., and Garcia, J. L. (2001) Biodegradation of aromatic compounds by *Escherichia coli*. *Microbiol. Mol. Biol. Rev. 65*, 523–569.
9. Diaz, E. (2004) Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. *Int. Microbiol. 7*, 173–180.
10. Roper, D. I., Fawcett, T., and Cooper, R. A. (1993) The *Escherichia coli* C homoprotocatechuate degradative operon: hpc gene order, direction of transcription and control of expression. *Mol. Gen. Genet. 237*, 241–250.
11. Prieto, M. A., Diaz, E., and Garcia, J. L. (1996) Molecular characterization of the 4-hydroxyphenylacetate catabolic pathway of *Escherichia coli* W: engineering a mobile aromatic degradative cluster. *J. Bacteriol. 178*, 111–120.
12. Rea, D., Fulop, V., Bugg, T. D., and Roper, D. I. (2005) Expression, purification and preliminary crystallographic analysis of 2,4-dihydroxy-hepta-2-ene-1,7-dioate aldolase (HpcH) from *Escherichia coli* C. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. 61*, 821–824.
13. Rea, D., Fulop, V., Bugg, T. D., and Roper, D. I. (2007) Structure and mechanism of HpcH: a metal ion dependent class II aldolase from the homoprotocatechuate degradation pathway of *Escherichia coli*. *J. Mol. Biol. 373*, 866–876.
14. Jenkins, J. R., and Cooper, R. A. (1988) Molecular cloning, expression, and analysis of the genes of the homoprotocatechuate catabolic pathway of *Escherichia coli* C. *J. Bacteriol. 170*, 5317–5324.
15. Rossi, A., and Schinz, H. (1948) Alcuni α-cheto-γ-lattoni con sostituenti alchilici in posizione γ. *Helv. Chim. Acta 31*, 377–385.
16. Wright, A., Blewett, A., Fulop, V., Cooper, R., Burrows, S., Jones, C., and Roper, D. (2002) Expression, purification, crystallization and preliminary characterization of an HHED aldolase homologue from *Escherichia coli* K12. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 58*, 2191–2193.
17. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. 276*, 307–326.
18. CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 760−763.
19. Izard, T., and Blackwell, N. C. (2000) Crystal structures of the metal-dependent 2-dehydro-3-deoxy-galactarate aldolase suggest a novel reaction mechanism. *EMBO J. 19*, 3849–3856.
20. Navaza, J. (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 57*, 1367–1372.
21. Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 53*, 448–455.
22. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 53*, 240–255.
23. Jones, T. A., Zou, J. Y., Cowan, S. W., Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A 47* (Part 2), 110−119.
24. DeLano, W. L. *The PyMOL User's Manual* (2002) DeLano Scientific, Palo Alto,CA.
25. Esnouf, R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics Modell. 15*, 133–138.
26. Yew, W. S., Fedorov, A. A., Fedorov, E. V., Almo, S. C., and Gerlt, J. A. (2007) Evolution of enzymatic activities in the enolase superfamily: L-talarate/galactarate dehydratase from *Salmonella typhimurium* LT2. *Biochemistry 46*, 9564–9577.
27. Yew, W. S., Fedorov, A. A., Fedorov, E. V., Rakus, J. F., Pierce, R. W., Almo, S. C., and Gerlt, J. A. (2006) Evolution of enzymatic activities in the enolase superfamily: L-fuconate dehydratase from *Xanthomonas campestris*. *Biochemistry 45*, 14582–14597.
28. Yew, W. S., Fedorov, A. A., Fedorov, E. V., Wood, B. M., Almo, S. C., and Gerlt, J. A. (2006) Evolution of enzymatic activities in the enolase superfamily: D-tartrate dehydratase from *Bradyrhizobium japonicum*. *Biochemistry 45*, 14598–14608.
29. Rakus, J. F., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Vick, J. E., Babbitt, P. C., Almo, S. C., and Gerlt, J. A. (2007) Evolution of enzymatic activities in the enolase superfamily: D-Mannonate dehydratase from *Novosphingobium aromaticivorans*. *Biochemistry 46*, 12896–12908.

30. Ose, T., Watanabe, K., Yao, M., Honma, M., Oikawa, H., and Tanaka, I. (2004) Structure of macrophomate synthase. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 1187–1197.
31. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389–3402.
32. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res. 22*, 4673–4680.
33. Larsen, T. M., Laughlin, L. T., Holden, H. M., Rayment, I., and Reed, G. H. (1994) Structure of rabbit muscle pyruvate kinase complexed with Mn2+, K+, and pyruvate. *Biochemistry 33*, 6301–6309.
34. Seeholzer, S. H., Jaworowski, A., and Rose, I. A. (1991) Enolpyruvate: chemical determination as a pyruvate kinase intermediate. *Biochemistry 30*, 727–732.
35. Rose, I. A. (1970) Stereochemistry of pyruvate kinase, pyruvate carboxylase, and malate enzyme reactions. *J. Biol. Chem. 245*, 6052–6056.
36. Ose, T., Watanabe, K., Mie, T., Honma, M., Watanabe, H., Yao, M., Oikawa, H., and Tanaka, I. (2003) Insight into a natural Diels-Alder reaction from the structure of macrophomate synthase. *Nature 422*, 185–189.
37. Dimroth, P., and Eggerer, H. (1975) Isolation of subunits of citrate lyase and characterization of their function in the enzyme complex. *Proc. Natl. Acad. Sci. U.S.A. 72*, 3458–3462.
38. Wang, W., and Seah, S. Y. (2005) Purification and biochemical characterization of a pyruvate-specific class II aldolase, HpaI. *Biochemistry 44*, 9447–9455.
39. Howard, B. R., Endrizzi, J. A., and Remington, S. J. (2000) Crystal structure of Escherichia coli malate synthase G complexed with magnesium and glyoxylate at 2.0 A resolution: mechanistic implications. *Biochemistry 39*, 3156–3168.
40. Anstrom, D. M., Kallio, K., and Remington, S. J. (2003) Structure of the *Escherichia coli* malate synthase G:pyruvate:acetyl-coenzyme A abortive ternary complex at 1.95 Å resolution. *Protein Sci. 12*, 1822–1832.
41. Anstrom, D. M., and Remington, S. J. (2006) The product complex of *M. tuberculosis* malate synthase revisited. *Protein Sci. 15*, 2002–2007.
42. Schutz, C. N., and Warshel, A. (2004) The low barrier hydrogen bond (LBHB) proposal revisited: the case of the Asp His pair in serine proteases. *Proteins 55*, 711–723.
43. Smirnov, S. V., Samsonova, N. N., Novikova, A. E., Matrosov, N. G., Rushkevich, N. Y., Kodera, T., Ogawa, J., Yamanaka, H., and Shimizu, S. (2007) A novel strategy for enzymatic synthesis of 4-hydroxyisoleucine: identification of an enzyme possessing HMKP (4-hydroxy-3-methyl-2-keto-pentanoate) aldolase activity. *FEMS Microbiol. Lett. 273*, 70–77.
44. Brunger, A. T. (1992) Free R-Value - a Novel Statistical Quantity for Assessing the Accuracy of Crystal-Structures. *Nature 355*, 472–475.
45. Cruickshank, D. W. (1999) Remarks about protein structure precision. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 55*, 583–601.
46. Read, R. J. (1986) Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. *Acta Crystallogr., Sect. A 42*, 140–149.
47. Rakus, J. F., Fedorov, A. A., Fedorov, E. V., Glasner, M. E., Hubbard, B. K., Delli, J. D., Babbitt, P. C., Almo, S. C., and Gerlt, J. A. (2008) Evolution of Enzymatic Activities in the Enolase Superfamily: L-Rhamnonate Dehydratase. *Biochemistry 47*, 9944–9954.

BI800943G