

# Structural and Functional Characterization of MppR, an Enduracididine Biosynthetic Enzyme from *Streptomyces hygroscopicus*: Functional Diversity in the Acetoacetate Decarboxylase-like Superfamily

A. Maxwell Burroughs,<sup>†</sup> Robert W. Hoppe,<sup>‡</sup> Neal C. Goebel,<sup>§</sup> Bilal H. Sayyed,<sup>‡</sup> Tyler J. Voegtline,<sup>‡</sup> Alan W. Schwabacher,<sup>‡</sup> T. Mark Zabriskie,<sup>§</sup> and Nicholas R. Silvaggi<sup>\*‡</sup>

<sup>†</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, United States

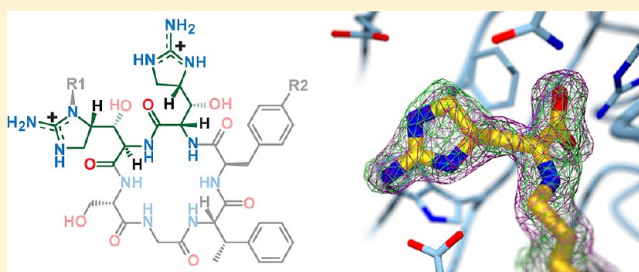
<sup>‡</sup>Department of Chemistry and Biochemistry, University of Wisconsin—Milwaukee, 3210 North Cramer Street, Milwaukee, Wisconsin 53211, United States

<sup>§</sup>Department of Pharmaceutical Sciences, Oregon State University, Corvallis, Oregon 97331, United States

## S Supporting Information

**ABSTRACT:** The nonproteinogenic amino acid enduracididine is a critical component of the mannopeptimycins, cyclic glycopeptide antibiotics with activity against drug-resistant pathogens, including methicillin-resistant *Staphylococcus aureus*. Enduracididine is produced in *Streptomyces hygroscopicus* by three enzymes, MppP, MppQ, and MppR. On the basis of primary sequence analysis, MppP and MppQ are pyridoxal 5'-phosphate-dependent aminotransferases; MppR shares a low, but significant, level of sequence identity with acetoacetate decarboxylase. The exact reactions catalyzed by each enzyme

and the intermediates involved in the route to enduracididine are currently unknown. Herein, we present biochemical and structural characterization of MppR that demonstrates a catalytic activity for this enzyme and provides clues about its role in enduracididine biosynthesis. Bioinformatic analysis shows that MppR belongs to a previously uncharacterized family within the acetoacetate decarboxylase-like superfamily (ADCSF) and suggests that MppR-like enzymes may catalyze reactions diverging from the well-characterized, prototypical ADCSF decarboxylase activity. MppR shares a high degree of structural similarity with acetoacetate decarboxylase, though the respective quaternary structures differ markedly and structural differences in the active site explain the observed loss of decarboxylase activity. The crystal structure of MppR in the presence of a mixture of pyruvate and 4-imidazolecarboxaldehyde shows that MppR catalyzes the aldol condensation of these compounds and subsequent dehydration. Surprisingly, the structure of MppR in the presence of "4-hydroxy-2-ketoarginine" shows the correct 4R enantiomer of "2-ketoenduracididine" bound to the enzyme. These data, together with bioinformatic analysis of MppR homologues, identify a novel family within the acetoacetate decarboxylase-like superfamily with divergent active site structure and, consequently, biochemical function.



Bacterial strains resistant to multiple classes of antibiotics, including last-line drugs like vancomycin, are a serious threat to public health.<sup>1–3</sup> Meeting this challenge requires the development of new antibiotics, preferably with modes of action that are distinct from those of current antimicrobial compounds. The mannopeptimycins make up a group of five related cyclic glycopeptide antibiotics produced by *Streptomyces hygroscopicus* NRRL 30439 that are active against an array of Gram-positive pathogens.<sup>4–6</sup> Of particular importance is the activity of the mannopeptimycins against methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant enterococci (VRE). Mannopeptimycins exert their antimicrobial effects by interfering with cell wall biosynthesis, though in a manner distinct from that of glycopeptide antibiotics like vancomycin.<sup>7</sup> The activity

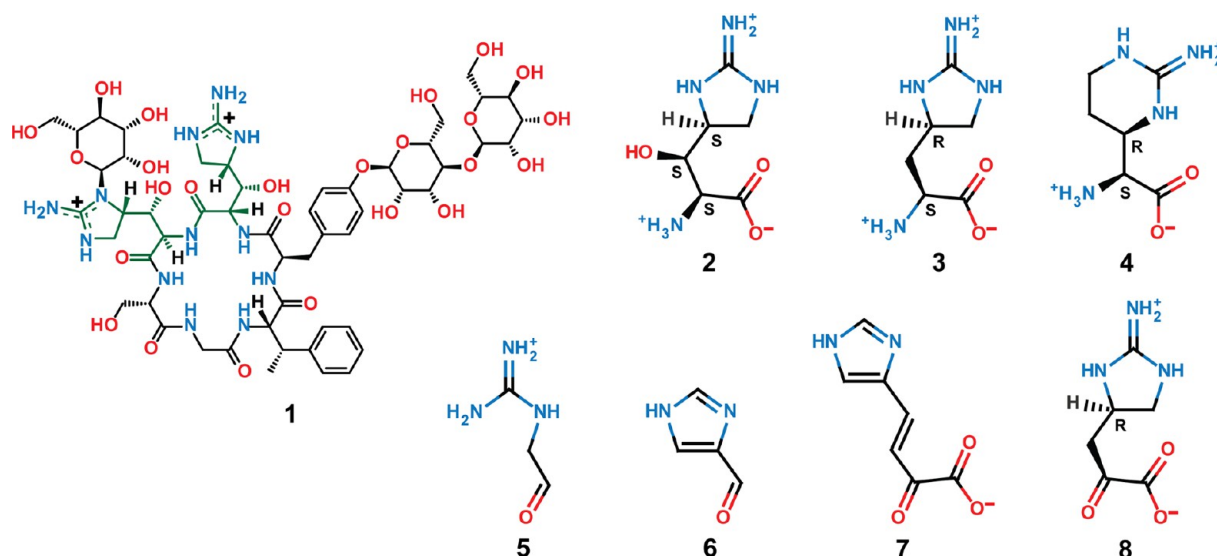
and novel mode of action of these compounds have led to significant interest in developing this scaffold into a clinically viable antibiotic.<sup>8–11</sup>

Drug development efforts based on the mannopeptimycin scaffold will benefit from a ready supply of enduracididine (End). Furthermore, this unusual building block may be incorporated into semisynthetic nonribosomal peptide derivatives to give products with altered biological activity. While a synthetic route to  $\beta$ -hydroxyenduracididine ( $\beta$ hEnd) has been developed,<sup>12</sup> an enzymatic or chemoenzymatic approach to End production may

Received: March 27, 2013

Revised: June 10, 2013

Published: June 11, 2013



**Figure 1.** Chemical structures of mannopeptimycin  $\beta$  (1),  $\beta$ -hydroxyenduracididine (2), L-enduracididine (3), L-capreomycin (4), guanidinoacetaldehyde (5), imidazole 4-carboxaldehyde (6), 2-oxo-3-(imidazol-5-yl)pyruvate (7), and 3-[(4R)-2-iminoimidazolidin-4-yl]-2-oxopropanoic acid (8).

ultimately be more facile, economical, and environmentally friendly.

The mannopeptimycins [e.g., mannopeptimycin  $\alpha$  (1, Figure 1)] consist of a cyclic hexapeptide core that is glycosylated with mannose residues at two positions.<sup>13</sup> A distinguishing feature of these compounds is the presence of  $\beta$ hEnd (2, Figure 1), which is derived from hydroxylation of End (3, Figure 1).<sup>14</sup> However, the biosynthetic origin of End (and thus  $\beta$ hEnd) is currently unknown. The presence of a guanidinium group, together with evidence from feeding experiments using radiolabeled amino acids,<sup>15</sup> suggests that End is derived from arginine. Furthermore, comparison of the mannopeptimycin biosynthetic cluster with that of the unrelated lipodepsipeptide enduracidin, which also contains multiple End residues, shows that there are only three pairs of proteins with a high level of sequence identity: MppP and EndP (80%), MppQ and EndQ (68%), and MppR and EndR (75%).<sup>5,16</sup> Finally, the enzyme MppO, which has no homologue in the enduracidin biosynthetic cluster, has been demonstrated to be a non-heme iron,  $\alpha$ -ketoglutarate-dependent oxygenase that selectively hydroxylates the  $\beta$ -carbon of free End.<sup>14</sup> Thus, the gene products MppP, MppQ, and MppR are thought to act through an unknown series of steps to transform arginine to enduracididine.

Analysis of the MppP, MppQ, and MppR amino acid sequences reveals that MppP and MppQ are likely both pyridoxal 5'-phosphate (PLP)-dependent aminotransferases and that MppR is somewhat similar to members of the acetoacetate decarboxylase-like superfamily (ADCSF; 17–25% identical in sequence). The nonproteinogenic amino acid capreomycin (4, Figure 1), which is analogous to End in that it is a cyclized form of arginine, is produced by two enzymes: a non-heme iron oxygenase (e.g., VioC) and a PLP-dependent aminotransferase (e.g., VioD).<sup>17</sup> VioC hydroxylates L-Arg at the  $\beta$ -carbon to give (2S,3S)- $\beta$ -hydroxyarginine.<sup>18</sup> VioD uses a PLP cofactor to catalyze an intramolecular  $\beta$ -elimination–replacement reaction to give (2S,3R)-capreomycin.<sup>19</sup> Given that the MppPQR system lacks an oxygenase and includes two apparent aminotransferases, it is clear that End and  $\beta$ hEnd must be produced by a chemical logic different from that of capreomycin. If arginine is the biosynthetic precursor of

End, how the oxidation takes place is an open question. As MppO does not oxidize arginine,<sup>14</sup> this question remains open for both the mannopeptimycin and enduracidin biosynthetic clusters.

The exact nature of this chemical logic, however, is not immediately obvious. The most glaring question concerns the role of an enzyme belonging to the ADCSF. The only characterized members of the ADCSF are confirmed acetoacetate decarboxylases (ADCs), what we term the “classical” ADCs.<sup>20</sup> The unique feature of these enzymes is the presence in the active site of a lysine side chain with a  $pK_a$  of 6.5, a shift of  $\sim 4$  pH units from the solution value.<sup>20,21</sup> The structure of *Chromobacterium violaceum* ADC (CvADC) indicates that the markedly hydrophobic environment of the active site is responsible for this dramatic  $pK_a$  perturbation.<sup>22</sup> The catalytic mechanism involves the formation of a Schiff base between the nucleophilic lysine amino group and the  $\beta$ -carbon of acetoacetate.<sup>20</sup> Decarboxylation is encouraged by the juxtaposition of two negatively charged glutamate side chains and facilitated by the Schiff base.<sup>22</sup> The imine is then hydrolyzed to release acetone and regenerate the active form of the enzyme. With no clear rationale for the involvement of a decarboxylase in End biosynthesis, it appears likely that MppR has some other catalytic function. Consistent with this, a preliminary comparative genomics analysis of the ADC fold specifically supports a functional shift from the decarboxylase activity for MppR and its closest homologues while suggesting that the ADC fold represents a scaffold with diverse substrate specificities and biochemical functions. To gain a better understanding of the role of MppR in End biosynthesis and its relationship to acetoacetate decarboxylase enzymes, the structures of MppR with the buffer HEPES bound, as well as the covalent complexes with pyruvate, 2-oxo-3-(imidazol-5-yl)pyruvate (7, Figure 1), and 3-[(4R)-2-iminoimidazolidin-4-yl]-2-oxopropanoic acid (8, i.e., “2-keto-enduracididine”, Figure 1), have been determined. The structures and preliminary biochemical characterization suggest a possible function of MppR in End biosynthesis.

## MATERIALS AND METHODS

**Cloning, Expression, and Purification of MppR.** The coding sequence of MppR was optimized for expression in *Escherichia coli* and synthesized by GenScript Inc. (Piscataway, NJ). This synthetic gene was subcloned into the pE-SUMO<sub>kan</sub> expression vector (LifeSensors Inc., Malvern, PA) using primers containing BsaI and XbaI restriction sites (forward, 5'-GGTC-TCAAGGTATGGAAAACCTGTACTTTCAGG-3'; reverse, 5'-GCTCTAGATCATTAGCTACGCACACCGATTT-3'). The His<sub>6</sub>-tagged SUMO-MppR fusion protein was expressed from *E. coli* BL21 Star (DE3) cells (Invitrogen Inc., Carlsbad, CA) carrying the pE-SUMO-MppR plasmid. Cultures were grown in 2 L of Luria-Bertani medium (50 μg/mL kanamycin) at 37 °C to an OD<sub>600</sub> of 0.8–1.0, at which point protein expression was induced with 0.4 mM IPTG. The temperature was reduced to 25 °C, and the cultures were grown overnight. Cells were harvested by centrifugation, resuspended in 5 mL/g of buffer A [25 mM Tris (pH 8.0), 300 mM NaCl, and 10 mM imidazole] supplemented with 1 mg/mL hen egg lysozyme, and stored overnight at –20 °C. The cell suspension was thawed at room temperature for 2 h and then treated with 0.1 mg/mL DNase I (Worthington Biochemical Corp., Lakewood, NJ). The lysate was clarified by centrifugation at 39000g for 45 min and then applied to a 5 mL HisTrap column (GE Lifesciences, Piscataway, NJ) at a flow rate of 5 mL/min to isolate the His<sub>6</sub>-SUMO-MppR fusion protein. The protein was eluted by a four-step gradient of buffer B [25 mM Tris (pH 8.0), 300 mM NaCl, and 250 mM imidazole]. The His<sub>6</sub>-SUMO-MppR fusion protein eluted in the third and fourth steps and was ~90% pure. Peak fractions were pooled and dialyzed overnight against 3.5 L of 25 mM Tris (pH 8.0) and 150 mM NaCl in the presence of ~2 μM SUMO protease (LifeSensors Inc.). The dialysate was passed through the HisTrap column a second time to remove the cleaved His<sub>6</sub>-SUMO tag as well as the protease. The resulting MppR preparation was >95% pure, as judged on Coomassie-stained sodium dodecyl sulfate–polyacrylamide gel electrophoresis gels. Selenomethionine-labeled MppR was purified using the same protocol, except that SelenoMethionine Medium Complete (Molecular Dimensions, Newmarket, Suffolk, U.K.) was used as the growth medium rather than LB and T7 Express Crystal cells (New England Biolabs, Ipswich, MA) were used in place of the BL21 Star (DE3) cells.

**Activity Tests with Arginine and Hydroxylated Derivatives.** MppR (10 μM) was incubated with 1 mM L-Arg, (3S)-hydroxy-L-Arg, (3R)-hydroxy-L-Arg, (4R,S)-hydroxy-L-Arg, or 2-oxo-5-guanidinovaleric acid for 6 h at 25 °C. The protein was removed by ultrafiltration using Amicon centrifugal filter devices. Prior to high-performance liquid chromatography (HPLC) analysis, 50 μL aliquots of the reaction mixtures containing amino acid substrates were mixed with 50 μL of 80 mM LiCO<sub>3</sub><sup>–</sup>, 70 μL of acetonitrile (MeCN), and 30 μL of 5 mM dansyl chloride (DNS-Cl) in MeCN. After incubation for 30 min at 50 °C, 40 μL of 2% ethylamine was added to consume unreacted DNS-Cl. The samples were incubated for 10 min at 25 °C. The reaction mixture containing 2-oxo-5-guanidinovaleric acid was derivatized with OPD by mixing 100 μL of the reaction mixture with 100 μL of water and 100 μL of 100 mM OPD in 2 N HCl. The samples were incubated at 80 °C for 20 min and then cooled on ice for 5 min. All derivatized samples were centrifuged at 30000 rcf for 10 min prior to HPLC analysis. Aliquots of the supernatants (5–10 μL) were analyzed on an Agilent 1220 Infinity HPLC system with a 4.6 mm × 150 mm Aquasil C18

column (Thermo Scientific). The separation was conducted at 50 °C using a linear gradient from 2 to 50% MeCN in a 0.1% TFA/water mixture over 14 min at a flow rate of 0.5 mL/min.

**Size Exclusion Chromatography.** The solution molecular mass of MppR was estimated by gel filtration chromatography using a Superdex S-200 10/300 column (GE Lifesciences) and a mixture of molecular mass standards, including rabbit muscle aldolase (156.8 kDa), *Saccharomyces cerevisiae* alkaline phosphatase (126.0 kDa), and *Streptomyces rubiginosus* xylose isomerase (86.4 kDa). To define the standard curve, each protein was dissolved at a concentration of 10 mg/mL in 50 mM NaPO<sub>4</sub> and 150 mM NaCl (pH 7.0). The mobile phase consisted of 50 mM NaPO<sub>4</sub> and 150 mM NaCl (pH 7.0), and the separation was performed with 250 μL of the standard mixture at a flow rate of 0.5 mL/min and 4 °C. A sample of MppR (12 mg/mL) was run separately over the same column and under identical conditions.

**Labeling with 5-Nitrosalicylaldehyde (2-hydroxy-5-nitrobenzaldehyde).** MppR (380 μM) was incubated with 5 mM 5-nitrosalicylaldehyde (5-NSA) for 15 min at 25 °C. NaBH<sub>4</sub> was added to a final concentration of 10 mM to reduce the aldimine and trap the *p*-nitrophenol reporter group on the enzyme. The presence of the reporter was verified by UV spectroscopy on a sample of labeled protein after thorough buffer exchange to remove unreacted 5-NSA.

**Preparation of 2-Oxo Acids.** The 2-oxo acids used in this study were prepared by a modification of the procedure first described by Meister.<sup>23</sup> For example, 2-oxo-5-guanidinovaleric acid was prepared by dissolving 2 g of L-Arg in 50 mL of water and the pH was adjusted to 7.2 with concentrated HCl. Next, 100 mg of *Crotalus atrox* L-amino acid oxidase (LAAO) and 1.5 mg of catalase were dissolved in 10 mL of water and added to the amino acid solution. The volume was adjusted to 100 mL with water, and the reaction mixture was incubated in the dark at room temperature while being vigorously stirred. After 18 h, the 100 mL reaction mixture was concentrated to 7 mL at 50 °C under vacuum. The 2-oxo acid crystallized from the concentrated reaction mixture at 4 °C after approximately 1 week. The crystalline material was harvested, dried, and stored at –20 °C. The sample of 2-oxo-4(*R/S*)-hydroxy-5-guanidinovaleric acid (i.e., “4-hydroxy-2-ketoarginine”) was prepared similarly, starting from 4(*R/S*)-hydroxy-L-Arg. Both compounds were characterized by nuclear magnetic resonance (NMR) and found to be predominantly the desired product (see the Supporting Information).

**Crystallization, Structure Determination, and Model Refinement.** Initial crystallization conditions were identified by screening 14 mg/mL MppR against the Index HT screen (Hampton Research). After optimization of the most promising hits, diffraction-quality crystals were obtained by the hanging drop vapor diffusion method from 25 to 30% polyethylene glycol (PEG) 3350, 0.2 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 1–10 mM HEPES (pH 7.5). Drops contained 2 μL of protein solution at 12–18 mg/mL (~380–750 μM) and 1 μL of crystallization solution. Crystals appeared after 3–4 days and grew to maximal dimensions of ~300 μm × 300 μm × 100 μm. Crystals of selenomethionine (SeMet)-substituted MppR were grown using the same conditions but were significantly smaller (~100 μm × 100 μm × 20 μm). Crystals were harvested from the hanging drops, coated with LV Cryo Oil (MiTeGen, LLC), and flashed-cooled by being plunged into liquid nitrogen. Structures of MppR with compounds 7 and 8 (Figure 1) bound were obtained by transferring crystals of native MppR into 30 μL drops of a soaking solution containing 36% PEG 3350, 0.18 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and

Table 1. Crystallographic Data Collection and Refinement Statistics

	SeMet MppR	MppR-HEPES	MppR-pyruvate	MppR-7	MppR-8
Data Collection					
resolution (Å) (last shell) <sup>a</sup>	34.33–2.20 (2.28–2.20)	34.26–1.85 (1.95–1.85)	34.30–1.72 (1.81–1.72)	41.87–1.67 (1.73–1.67)	33.28–1.70 (1.73–1.70)
wavelength (Å)	0.97625	0.97950	0.97856	0.97856	0.97856
no. of reflections					
observed	321309 (21370)	323463 (46672)	606665 (87494)	645358 (43289)	373589 (17118)
unique	31470 (3019)	52362 (7559)	65173 (9382)	71344 (6515)	67228 (3321)
completeness (%) <sup>a</sup>	99.7 (97.4)	100.0 (100.0)	100.0 (100.0)	99.1 (91.8)	99.9 (100.0)
R <sub>merge</sub> (%) <sup>a,b</sup>	0.092 (0.401)	0.109 (0.320)	0.063 (0.468)	0.062 (0.475)	0.048 (0.420)
multiplicity	10.2 (7.1)	6.2 (6.2)	9.3 (9.3)	9.0 (6.6)	5.6 (5.2)
⟨I/σ(I)⟩ <sup>a</sup>	24.2 (4.7)	9.4 (4.2)	16.7 (4.1)	26.3 (3.5)	39.2 (3.8)
figure of merit	0.348				
phasing power	1.111				
Refinement					
no. of reflections in the working set	31436	51432	65133	67732	63304
no. of reflections in the test set	1559	1958	3298	3573	3318
R <sub>cryst</sub> (R <sub>free</sub> )	0.191 (0.212)	0.145 (0.176)	0.157 (0.179)	0.156 (0.177)	0.150 (0.173)
no. of non-H atoms	4493	4565	4514	4768	4544
no. of solvent atoms	460	555	504	629	597
no. of TLS groups	NA	24	13	17	18
average B factor (Å <sup>2</sup> ) <sup>c</sup>					
protein atoms	26.7	22.2	26.0	25.3	22.9
ligand atoms	28.3	16.4	26.9	29.5	23.9
solvent	35.9	29.3	31.3	36.8	33.8
rmsd					
bond lengths (Å)	0.022	0.010	0.009	0.010	0.010
bond angles (deg)	1.886	1.267	1.258	1.307	1.297
coordinate error (Å)	0.30	0.15	0.16	0.22	0.14

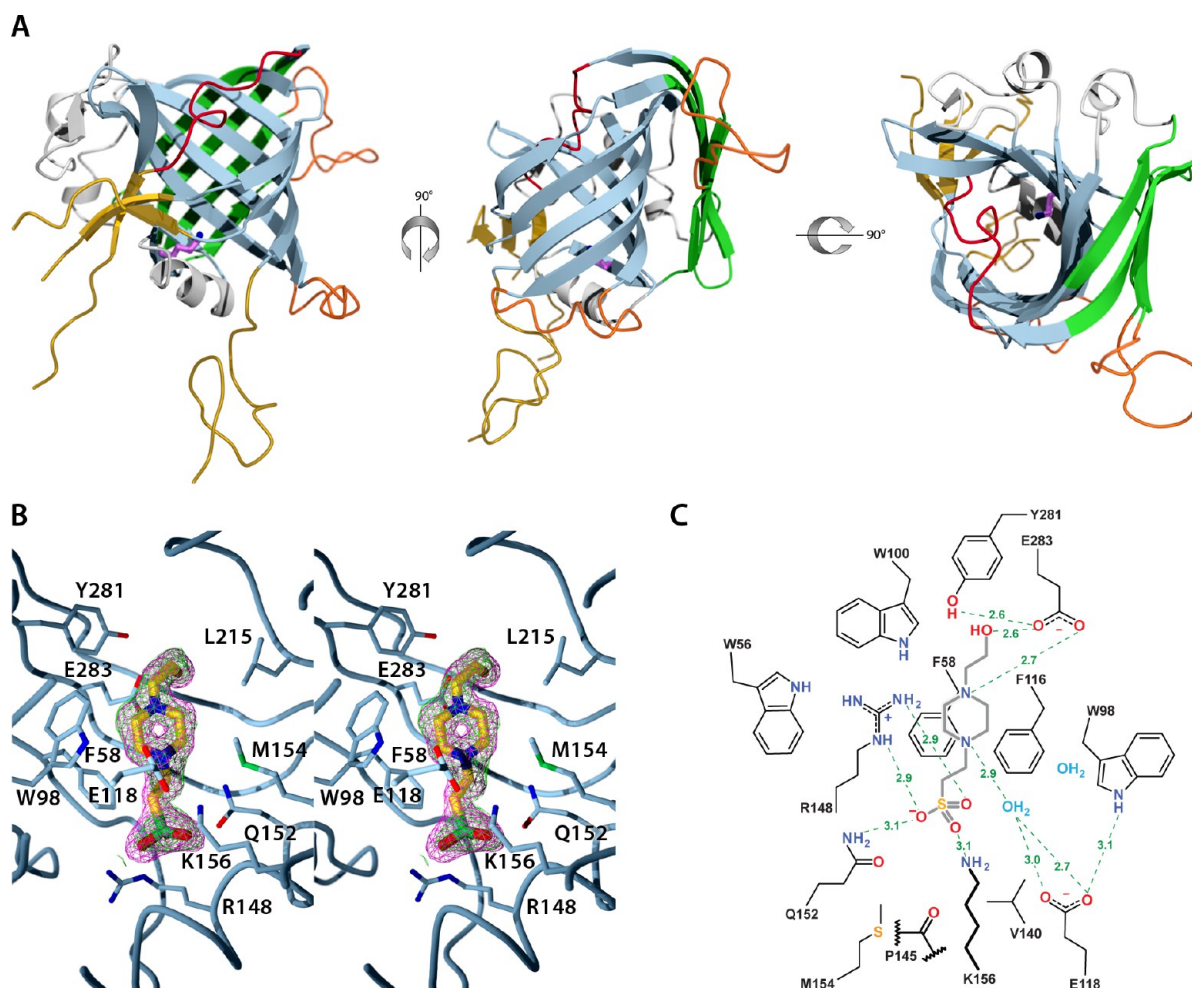
<sup>a</sup>Values in parentheses apply to the high-resolution shell indicated in the resolution row. <sup>b</sup> $R = \sum(|F_{\text{obs}}| - \text{scale} \times |F_{\text{calc}}|) / \sum |F_{\text{obs}}|$ . <sup>c</sup>Isotropic equivalent B factors, including contribution from TLS refinement.

either 30 mM pyruvate with 50 mM **6** or <300 mM **8**. After being soaked for 2–24 h, crystals were coated with LV Cryo Oil and flash-cooled. X-ray diffraction data for SeMet MppR and MppR-HEPES were collected at beamline 21-ID-D of the Life Science Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS). The MppR-pyruvate, MppR-7, and MppR-8 data sets were collected at LS-CAT beamline 21-ID-G. All crystals were screened for diffraction quality using the rotating anode X-ray source at Marquette University (Milwaukee, WI). Data were processed with HKL2000<sup>24</sup> or MOSFLM<sup>25,26</sup> and SCALA<sup>27</sup> of the CCP4 Program Suite.<sup>28</sup>

The structure of MppR was determined by the single-wavelength anomalous diffraction (SAD) method using 2.2 Å resolution data collected from a crystal of SeMet-substituted MppR at 0.97625 Å, 41 eV above the tabulated K-edge wavelength for Se (0.97950 Å). autoSHARP<sup>29</sup> was used to determine the Se substructure, which contained 12 of the 14 Se atoms in the asymmetric unit, and calculate high-quality, density-modified phases. An initial model comprising ~85% of the asymmetric unit contents was built automatically using the PHENIX package (phenix.autobuild<sup>30,31</sup>). After iterative cycles of manual model building in COOT<sup>32</sup> and maximum likelihood-based refinement using the PHENIX package (phenix.refine<sup>33</sup>), ordered solvent molecules were added automatically in

phenix.refine and culled manually in COOT. Hydrogen atoms were added to the model using phenix.reduce<sup>34</sup> and were included in the later stages of refinement to improve the stereochemistry of the model. Positions of H atoms were refined using the riding model with a global B factor. For all models, except the original SeMet MppR model, regions for translation-libration-screw (TLS) refinement were identified using the TLSMD server<sup>35</sup> or phenix.find\_tls\_groups and the TLS parameters were refined in phenix.refine. Once the refinement converged (e.g., SeMet MppR; R = 0.191, and R<sub>free</sub> = 0.212), the model was validated using the tools implemented in COOT and PHENIX.<sup>36,37</sup> Sections of the backbone with missing or uninterpretable electron density were not included in the final model. Side chains with poor or missing electron density were modeled in favored rotameric conformations; the B factors were allowed to refine without additional restraints, and the occupancies were held to 1.0.

The final, refined model of SeMet MppR was stripped of water molecules, H atoms, and B factor information (all B factors set to 20.00) and used to determine the structures of MppR with HEPES, pyruvate, **7**, or **8** bound by the difference Fourier method or by molecular replacement in PHASER.<sup>38</sup> A similar refinement protocol was used for all of the models presented here. Restraints and coordinates for the modified Lys156 residue



**Figure 2.** (A) Ribbon representation of the MppR protomer, shown in three orthogonal orientations to highlight the dimerization and tetramerization loops (yellow and orange, respectively) and the double-barrel fold. The entrance loop (red) may play a role in gating access to the active site, which is identified by the position of Lys156 (shown as sticks with magenta carbon atoms). In the absence of  $\alpha$ -keto acids, the enzyme crystallizes with a molecule of HEPES buffer bound. (B) Stereoview showing the  $2F_o - |F_c|$  (magenta) and  $2F_o - |F_c|$  simulated annealing composite omit (green) electron density maps, both contoured at  $1.0\sigma$ . (C) Schematic showing the interactions between HEPES and MppR, and the associated distances. This figure and subsequent figures showing crystallographic structures were rendered with the POVSCRIPT<sup>54</sup> modification of MOLSCRIPT<sup>55</sup> and POVRAY.

in the liganded structures were generated with phenix.elbow<sup>39</sup> and added to the model in COOT. Data collection and model refinement statistics for all structures are listed in Table 1. Coordinates and structure factors have been deposited in the Protein Data Bank as entries 4JM3, 4JMC, 4JMD, and 4JME.

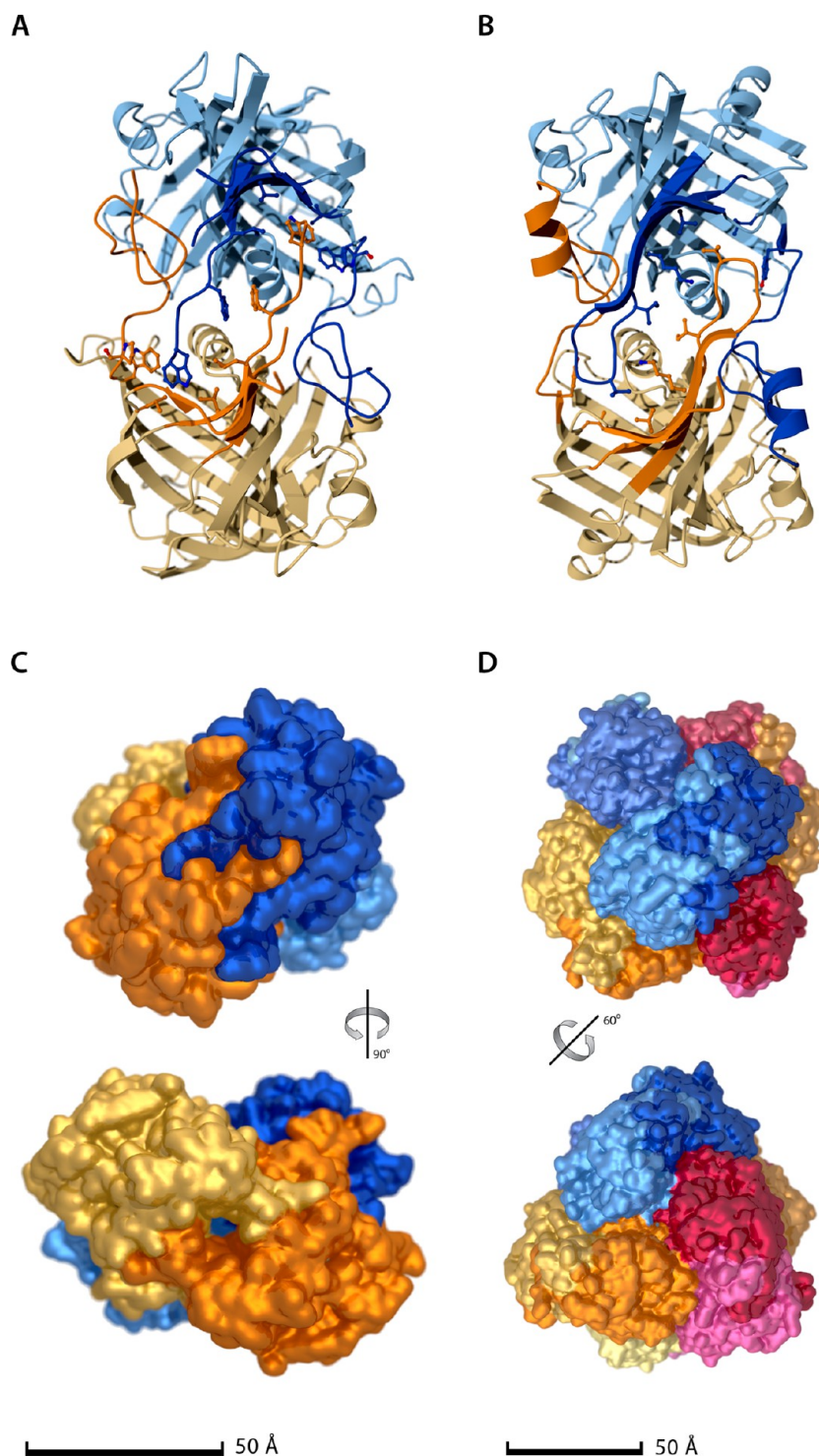
## RESULTS AND DISCUSSION

**Biochemical Characterization.** Recombinant *S. hygroscopicus* MppR was expressed in *E. coli* as a His<sub>6</sub>-SUMO fusion protein and purified using conventional chromatographic techniques. MppR was expressed at a high level in *E. coli*, yielding 20–30 mg of >95% pure protein per liter of culture. The molecular mass was estimated to be 131.5 kDa by size exclusion chromatography, consistent with a tetramer of 32.2 kDa protomers (128.8 kDa). ADC, in contrast, is dodecameric with a solution molecular mass of ~330 kDa.<sup>22,40</sup>

Given the sequence identity between ADC and MppR, the first experiments undertaken were aimed at determining whether these two proteins share the same catalytic activity. The catalytic Lys115 of ADC is known to react with 5-nitrosalicylaldehyde (5-NSA) to give a Schiff base that can be trapped by reduction of the imine with NaBH<sub>4</sub> to give 2-hydroxy-5-nitrobenzyl-ADC.<sup>41,42</sup>

Treatment of MppR with 5-NSA and NaBH<sub>4</sub> at pH 6.5 results in stable labeling of the enzyme with the *p*-nitrophenol reporter group as verified by UV spectroscopy (Figure S1 of the Supporting Information). After unreacted 5-NSA had been removed by buffer exchange, the sample retained an intense yellow color and a strong UV absorbance peak at 405 nm. This result indicates that MppR, like ADC, contains a reactive lysine side chain in the active site that is deprotonated to a significant extent at pH 6.5.

MppR was assayed for acetoacetate decarboxylase activity by monitoring the disappearance of the enolate form of acetoacetate as described previously.<sup>20</sup> There was no change in absorbance at 270 nm after 60 min, indicating that acetoacetate is not a substrate for MppR. Because enduracididine is hypothesized to originate from *L*-Arg, MppR was tested for activity against *L*-Arg, (3*S*)-hydroxy-*L*-Arg, (3*R*)-hydroxy-*L*-Arg, (4*R,S*)-hydroxy-*L*-Arg, and 2-oxo-5-guanidinovaleric acid (i.e.,  $\alpha$ -keto-Arg). After 6 h at 25 °C, reaction mixtures were derivatized with dansyl chloride (DNS) or *o*-phenylenediamine (OPD) and analyzed by HPLC. Comparison of the reaction mixtures with authentic standards and controls lacking MppR indicates that none of the compounds tested are substrates for MppR.



**Figure 3.** Ribbon representation of the MppR “dimer” (A) that is very similar to the corresponding unit of *C. violaceum* acetoacetate decarboxylase (B; PDB entry 3BGT). The two protomers of each protein shown are colored light blue and light yellow, while the respective interface loops are colored orange and dark blue. Both interfaces are comprised almost entirely of hydrophobic interactions. The full biological units of MppR (C) and CvADC (D), on the other hand, are quite different. The functional significance, if any, of the tetrameric vs dodecameric quaternary structure is currently unclear.

**Overall Structure of MppR.** MppR crystallized from a solution containing 25–30% polyethylene glycol 3350, 0.2–0.3 M  $(\text{NH}_4)_2\text{SO}_4$ , and 1–100 mM HEPES (pH 7.5) in space group  $P3_121$  (unit cell dimensions of  $a = b = 109.8 \text{ \AA}$  and  $c = 87.8 \text{ \AA}$ ) with two molecules in the asymmetric unit. Experimental phases were obtained to 2.2 Å using single-wavelength anomalous diffraction from crystals of the selenomethionine-substituted

enzyme (Table 1). The resulting model was used to phase a higher-resolution (1.9 Å) data set by molecular replacement. The final model of MppR·HEPES contains 257 (monomer A) or 259 (monomer B) of 302 total amino acids, 589 water molecules, and two HEPES molecules. At the N-terminus, 31–33 residues are disordered, as well as seven at the C-terminus and an additional six from a surface loop (residues 226–231). There are no

Ramachandran outliers in the model, and 99.2% of the residues are in the favored regions of the Ramachandran plot. The protein chains from the SeMet MppR model were used to obtain initial phases for the structures of MppR described below.

The overall fold of MppR is similar to the  $\beta$ -cone<sup>22</sup> or double-barrel<sup>43</sup> fold observed in acetoacetate decarboxylases. The structure contains a high proportion of  $\beta$ -strand (45%) and relatively little  $\alpha$ -helix (11%). The 13  $\beta$ -strands are organized into one discontinuous, antiparallel  $\beta$ -sheet with a compound fold that forms a large barrel (light blue in Figure 2A) and a second, smaller barrel-like feature orthogonal to the large barrel (green in Figure 2A). The large barrel is closed at one end by the longest  $\alpha$ -helix in the structure ( $\alpha 3$ , residues 142–152). At the opposite end, residues 205–214 form a loop with a short, four-residue stretch of  $\alpha$ -helix ( $\alpha 4$ ). We dub this part of the structure the “entrance loop”. Given its position at the open end of the large barrel, and the predominance of coil structure in this region, it is plausible that this part of the structure may control access to the active site. While the *B* factors of residues 205–214 are comparable to the average *B* factor of the model, the observations that (1) the opening to the barrel is too small ( $\sim 6.5 \times 4.5$  Å) to allow HEPES ( $\sim 6.5 \times 5.0$  Å) or a similarly sized molecule into the barrel and (2) the residues in the entrance loop make indirect hydrogen bonding interactions and/or direct packing interactions with bound ligands suggest that this loop might be mobile in the absence of ligand. This may explain, at least in part, why MppR failed to crystallize in the absence of a bound ligand (e.g., HEPES). Finally, MppR possesses two pairs of protuberances (yellow and orange in Figure 2A) that are involved in oligomerization (see below).

**MppR Active Site.** The interior of the large barrel defines a flattened, elliptical pocket with a narrow opening exposed to solvent and, at the opposite end of the barrel, an anion binding site and the lysine residue (Lys156) that aligns with the catalytic lysine side chain of ADC (Figure 2B). The environment of the presumed catalytic Lys156 is overwhelmingly hydrophobic. The protein atoms closest to the lysine amino group are Met154 *S* $\delta$  (4.1 Å), Val140 *C* $\gamma$  (4.2 Å), Phe116 *C* $\epsilon$  (4.5 Å), Gly149 *C* $\alpha$  (4.6 Å), Phe145 *C* $\beta$  (5.1 Å), Ala138 *C* $\beta$  (5.7 Å), Trp98 *C* $\delta$  (6.5 Å), Leu120 *C* $\delta$  (6.5 Å), Phe58 *C* $\epsilon$  (6.8 Å), and Arg148 *N* $\epsilon$ /*N* $\eta$  (6.9/7.0 Å). In this environment, the only potential hydrogen-bonding partners are Glu118 *O* $\epsilon$  (2.7 Å), Pro145 *O* (4.1 Å), and an ordered solvent molecule (3.0 Å). The relative orientations of Glu118 and the solvent molecule suggest that only one would be able to form a hydrogen bond with Lys156 at any given time. Also, formation of the hydrogen bond with the carbonyl oxygen of Pro145 would require a reorientation of the lysine side chain that would likely preclude hydrogen bonding interactions with Glu118 or the solvent molecule. As in acetoacetate decarboxylase, the hydrophobic active site environment and dearth of potential hydrogen bonding interactions are likely responsible for the apparent *pK<sub>a</sub>* perturbation observed for Lys156. Given its propensity for binding negatively charged species (see below), Arg148 is likely to be charged. A charge on this residue could be stabilized by hydrogen bonds with multiple solvent molecules. Also, the indole ring of Trp100 is  $\sim 3.5$  Å from the guanidinium group and is suitably oriented to stabilize a positive charge via a  $\pi$ -cation interaction.<sup>44</sup>

Each molecule of MppR contains one molecule of the HEPES buffer from the crystallization solution bound in the active site (Figure 2B,C). The buffer molecule appears to be tightly held, with an average *B* factor of 26.9 Å<sup>2</sup>, agreeing well with the average *B* factor of all protein atoms (26.0 Å<sup>2</sup>). The sulfonate group of the

HEPES molecule makes a bidentate interaction with the guanidinium group of Arg148 (both 2.9 Å) and hydrogen bonding interactions with the side chain amide of Glu152 (3.1 Å) and the  $\epsilon$ -amino group of Lys156 (3.1 Å). The piperazine nitrogens of HEPES make additional hydrogen bonding interactions with Glu283 (2.7 Å) and, through a bridging water, with Glu118 (Glu–H<sub>2</sub>O, 2.7 Å; H<sub>2</sub>O–HEPES, 2.9 Å). One side of the piperazine ring is in a hydrophobic pocket defined by the side chains of Phe58 and Trp98 (4.0–4.2 Å), while the other side is exposed to the side chains of Met154 (4.0 Å) and Leu215 (3.9 Å) and a pair of water molecules (3.2–3.5 Å). Finally, the hydroxyethyl moiety of the buffer forms a hydrogen bond to the carboxylate of Glu283 (2.6 Å). Several attempts to determine the structure of MppR with no buffer bound by substituting different buffers into the crystallization solution resulted in a failure to obtain crystals or in structures with alternative buffers bound [e.g., Tris (data not shown)].

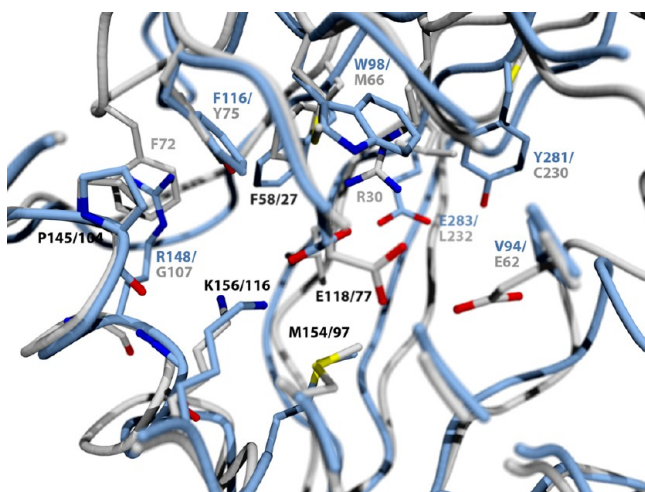
**Quaternary Structure.** The two protomers in the crystallographic asymmetric unit are arranged as a dimer (Figure 3A) mediated by the “small interface loops” [residues 100–114 and 166–176 (orange in Figure 2A)]. The resulting interface buries  $\sim 1300$  Å<sup>2</sup> of surface area (determined using PISA<sup>45</sup>) and consists of predominantly van der Waals contacts with only six hydrogen bonds and no ionic interactions. This dimer is related to a second dimer by the crystallographic 2-fold rotation axis to form a tetramer. This interface is mediated by the “large interface loops” [residues 34–55 and 218–235 (yellow in Figure 2A)], which form extensive domain-swapping contacts that bury an additional  $\sim 2400$  Å<sup>2</sup> of surface area. This interface is also dominated by van der Waals contacts, having only four hydrogen bonds and no ionic interactions. The extensive buried surface area and specific intermolecular interactions strongly suggest that the tetramer observed in the crystal represents the true oligomeric state of MppR and is not a consequence of crystallization. A tetramer would also be consistent with the SEC results described above. Further evidence of the physiological significance of the observed MppR tetramer comes from the structure of the putative acetoacetate decarboxylase from *Methanococcus marisnigri* [Protein Data Bank (PDB) entry 3CMB (Joint Center for Structural Genomics, unpublished observations)], which exhibits an almost identical quaternary structure.

**Comparison to Acetoacetate Decarboxylase.** In spite of the modest sequence identity with CaADC (PDB entry 3BH2) and CvADC (PDB entry 3BGT), the tertiary structure of MppR is extremely similar to those of the classical ADCs (Figure 3A,B). Least-squares fitting with SSM<sup>46</sup> results in root-mean-square deviations (rmsd) for *C* $\alpha$  atoms of 2.3 and 1.9 Å for CaADC and CvADC, respectively. The interface mediated by the large interface loops of MppR is also strongly conserved (Figure 3A,B). This is not unexpected, because in ADC this interface was shown to be important for maintaining the catalytic lysine in the hydrophobic active site. The lysine side chain adjacent to the catalytic lysine (Lys116, CaADC numbering) projects into the interface, acting as a physical wedge to prevent the catalytic lysine from backing out of the active site.<sup>22</sup> Mutating Lys116 of CaADC to cysteine abolished ADC activity,<sup>20</sup> presumably because the catalytic lysine was out of position. Chemical rescue of the Lys116Cys mutant with 2-bromoethylamine restored ADC activity. It was thought, in the absence of structural data, that the positive charge adjacent to the catalytic lysine was responsible for the *pK<sub>a</sub>* perturbation of that residue. However, the structural data of Ho et al. indicate that the recovery was actually due to the increased length of the resulting S-(2-aminoethyl)cysteine side

chain forcing the catalytic lysine back into position. Interestingly, the residue adjacent to the catalytic Lys156 of MppR is Gln157, and contrary to the case for ADC, the environment around this side chain is not crowded by the adjacent protomer. In fact, Gln157 interacts primarily with solvent and so does not seem to be acting as a wedge analogous to Lys116 of CaADC. It may be that the role of the residue adjacent to the catalytic lysine in this fold is more complicated or that the wedge mechanism at work in the classical ADCs is not universal.

While one interface is conserved between MppR and the classical ADCs, the higher-order quaternary structures are very different. Whereas six copies of the ADC dimer trimerize to give a homododecamer, two copies of the MppR dimer associate back to back to form a homotetramer (Figure 3C,D). The structural basis for this difference is clear: one of the small interface loops of MppR disrupts the packing of the “trimerization platform” observed in the classical ADCs.<sup>22</sup> The functional significance of the higher-order quaternary structures (tetrameric vs dodecameric) is less clear, and its elucidation will require additional structural and biochemical data from MppR-like and classical ADCSF enzymes.

As expected from the high degree of structural similarity between MppR and ADC, the active sites of both enzymes are strikingly similar (Figure 4). Both active sites, as discussed earlier,



**Figure 4.** Superposition of the active sites of MppR and CvADC (PDB entry 3BGT). MppR is colored light blue, while CvADC is colored gray. Conserved residues are denoted by black labels. Unconserved positions are labeled with both the MppR (blue) and CvADC (gray) residues.

are highly hydrophobic; both contain a lysine side chain with an apparently perturbed  $pK_a$ , and both contain a glutamate side chain near the catalytic lysine (MppR, Glu118; CvADC, Glu77). Perhaps more informative than the similarities are the subtle, but important, differences between the two active sites. First, Glu62 of CvADC, which is adjacent to Glu77 in that active site and is important for the decarboxylase activity,<sup>22</sup> has been replaced with valine (Val94) in MppR. The position occupied by the guanidinium group of CvADC Arg30, which coordinates the carboxylate group of acetoacetate, is occupied by the indole ring of Trp98 in MppR. The aliphatic Leu232 of CvADC is replaced with Glu283 in MppR (the carboxylate is 9.0 Å from the catalytic Lys156). Finally, the conserved Gly107 and Trp112 of CvADC are replaced with Arg148 and Gln152, respectively, in MppR. These two hydrophobic-to-polar substitutions create a polar subsite in the active site that we propose as a carboxylate-binding

site (see below). These amino acid substitutions provide a structural rationale for the lack of decarboxylase activity in MppR. The positive charge that positions the carboxylate of acetoacetate in ADC (Arg30) is missing in MppR, which has the (presumably) positively charged Arg148 closer to, and on the opposite side of, the catalytic lysine residue. The Glu62 side chain of CvADC that helps create an electrostatic environment that favors decarboxylation is lost in MppR, where it is replaced by the much smaller, hydrophobic valine side chain. Clearly, these few changes result in significant changes in the electrostatic environment of the active site.

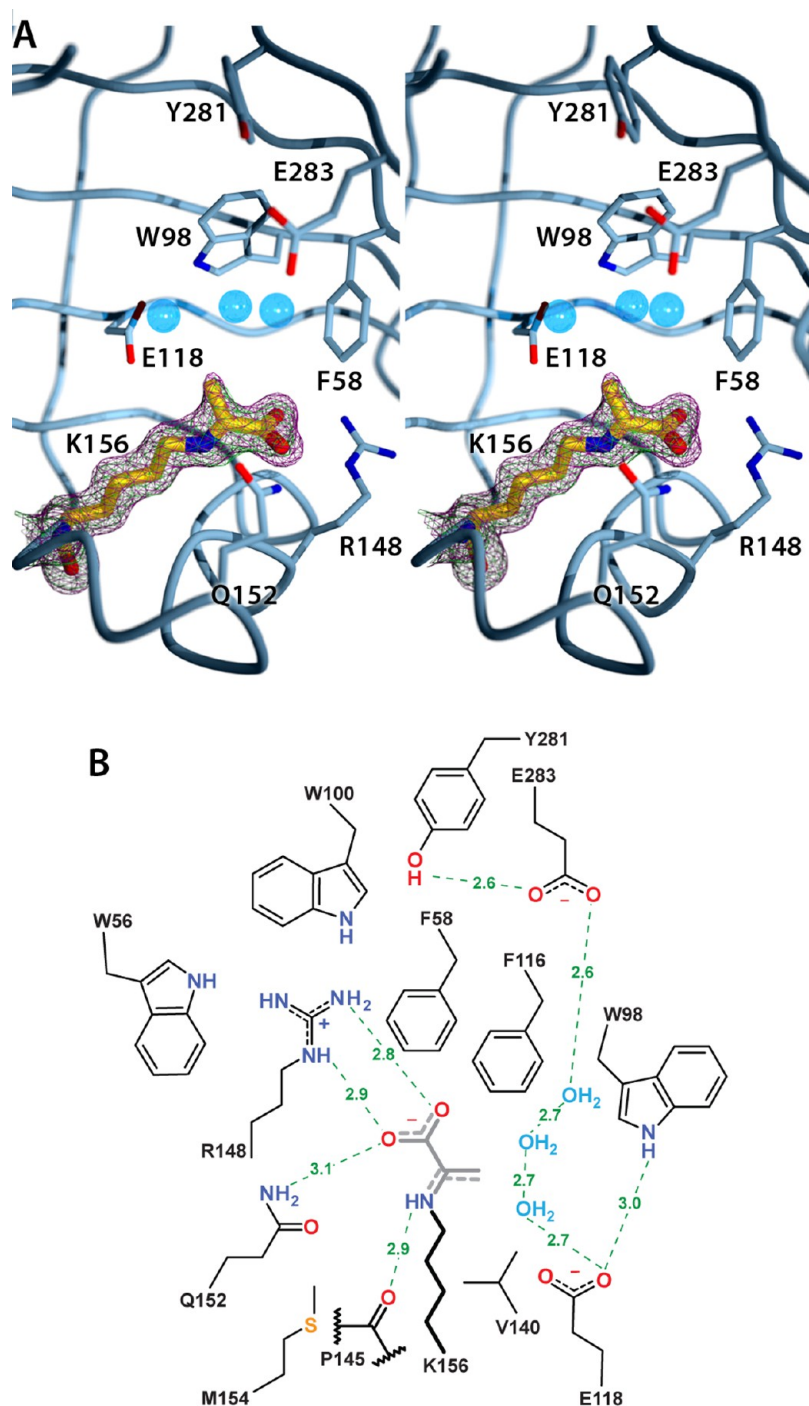
#### Structures of MppR Complexes with $\alpha$ -Keto Acids.

Because acetoacetate was not a substrate for MppR and the side chain of Arg148 is quite close to the putative nucleophilic  $\epsilon$ -amino group of Lys156, we hypothesized that MppR might react with  $\alpha$ -keto acids. This is based primarily on the observation that the carboxylate, which we envisioned interacting with the guanidinium of Arg148 analogous to the sulfonate group of HEPES, is too far from the ketone in  $\beta$ -keto acids to permit favorable alignment of the reacting groups. The crystal structure with HEPES bound suggested that a shorter  $\alpha$ -keto acid might fit well. To test this possibility, crystals of MppR were soaked in a solution containing PEG 3350,  $(\text{NH}_4)_2\text{SO}_4$ , and 30 mM sodium pyruvate for 24 h. Data were collected from a pyruvate-treated crystal, and the resulting structure showed pyruvate covalently bound to Lys156 as the Schiff base (Figure 5). Unlike the MppR-HEPES structure, the main chain carbonyl of Pro145 is close enough (2.9 Å) and in a suitable orientation to form a hydrogen bond to the  $\zeta$ -nitrogen of Lys156, which could stabilize the complex in the iminium form. The  $\alpha$ -carboxylate of pyruvate makes a bidentate interaction with Arg148 (2.8 and 2.9 Å) and a hydrogen bonding interaction to the amide nitrogen of Gln152 (3.1 Å). These two residues constitute a carboxylate-binding subsite in the MppR active site. The  $\beta$ -carbon of pyruvate is oriented toward Glu118 and the opening of the active site. Structures were determined using data collected from crystals that had soaked for 2 weeks in solutions containing pyruvate, and from 3-week-old crystals grown in the presence of pyruvate (data not shown). In both cases, the enzyme was in the Schiff base form, suggesting that the imine is long-lived. However, because hydrolysis of the Schiff base would simply regenerate the  $\alpha$ -keto acid, we cannot distinguish between the kinetic and thermodynamic stability of this complex. Still, the environment around the imine is hydrophobic and quite crowded, which would make it difficult for a water molecule to gain access to the  $\alpha$ -carbon. Indeed, in almost every Schiff base complex structure we examined, there are no water molecules proximal to the imine. Thus, it would not be surprising to learn that the rate of hydrolysis of the Schiff base is very slow.

The structure of the MppR-pyruvate complex led to a series of structures of MppR in complexes with other  $\alpha$ -keto acids for the probing of substrate specificity and the possible roles of key residues near the catalytic Lys156. Eight of the nine  $\alpha$ -keto acids tested formed a stable Schiff base complex with the enzyme (data not shown). In the interest of clarity and brevity, only the complexes of MppR with 7 and 8 will be discussed in detail.

If an MppR Schiff base with an  $\alpha$ -keto acid is involved in the formation of End, compounds 10 and/or 11 are plausible participants in the reaction (Figure 6). These compounds both resemble aldol condensation products between pyruvate and guanidinoacetaldehyde (5, Figure 1). Such an aldol process could, in principle, lead to End; however, a source of guanidinoacetaldehyde (5) is unknown. The reverse reaction,

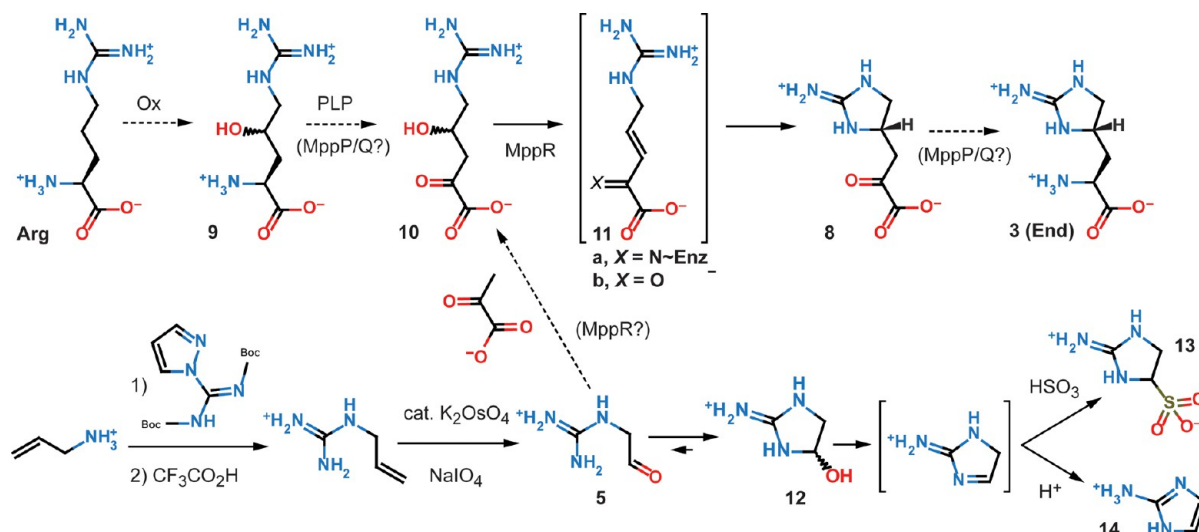




**Figure 5.** Stereoview of the MppR-pyruvate Schiff base complex (A) showing the experimental  $2|F_o| - |F_c|$  electron density (magenta) and the simulated annealing composite omit electron density (green), both contoured at  $1.0\sigma$ . The schematic of the complex (B) shows potential interactions and their distances (in angstroms). Pyruvate is covalently bound to the enzyme as indicated by continuous electron density between Lys156 and the  $\alpha$ -carbon of pyruvate. The side chain of pyruvate is too short to unequivocally differentiate between the imine and enamine forms of the Schiff base. This result suggested that MppR takes an  $\alpha$ -keto acid as its substrate.

cleavage of 2-oxo-4-hydroxyarginine (**10**) to pyruvate and **5**, has been proposed in the synthesis of the antibiotic 2-aminoimidazole by *Streptomyces eurocidicus*.<sup>47</sup> On the basis of these proposed intermediates, 4-imidazolecarboxaldehyde (**6**, Figure 1) was chosen as a stable analogue of **5**. To test the hypothesis that MppR might be capable of catalyzing an aldol condensation, we determined the structure of MppR in the presence of both pyruvate and **6**.

Crystals of MppR were soaked for 2.5 h in a solution containing 30% PEG 3350, 0.3 M  $(\text{NH}_4)_2\text{SO}_4$ , and 30 mM pyruvate, after which an equal volume of 100 mM **6** in PEG and  $(\text{NH}_4)_2\text{SO}_4$  was added. After an additional 4.0 h, crystals were flash-cooled and diffraction data were collected. The resulting structure showed the condensation product 2-oxo-3-(imidazol-5-yl)pyruvate (**7**, Figure 1) bound covalently to Lys156 (Figure 7). The absence of the aldehyde oxygen and the C2–C3–C4–C5 torsion angle of  $-178.0^\circ$  (Figure 7B) is consistent with the



**Figure 6.** Hypothetical enduracididine biosynthetic scheme starting from L-Arg (top). The structures reported here suggest that the biosynthesis could also originate with a metabolite of Arg, like guanidinoacetaldehyde (5) and pyruvate. The bottom row shows a possible synthetic route to 5 and potential byproducts.

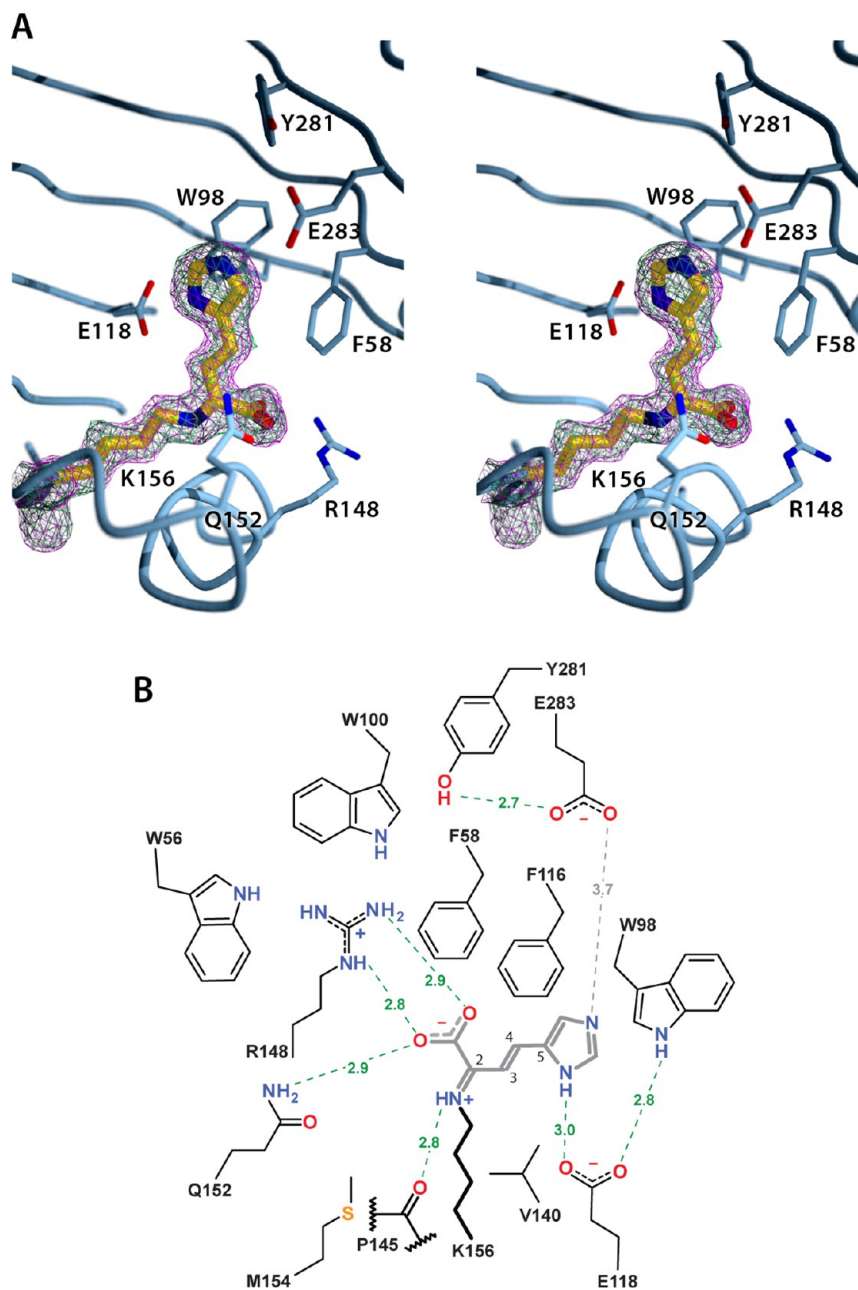
observed condensation product being unsaturated. The pyruvate moiety is identical to the MppR-pyruvate complex. When the MppR-7 and MppR-pyruvate complexes are superimposed using all C $\alpha$  atoms, the five pyruvate-derived atoms (which do not influence the superposition) overlay with an rmsd of 0.20 Å. The imidazole moiety makes a single hydrogen bond to Glu118 [3.0 Å (Figure 6B)], one of the presumed key Schiff base-forming residues based on analogy to ADC. The other imidazole nitrogen atom is close to Glu283 but is too far away (3.7 Å) and likely in the wrong protonation state to make a hydrogen bonding interaction with that residue. The observed aldol condensation and subsequent dehydration appear to be catalyzed by MppR, because no 7 could be detected by HPLC analysis of the crystal soaking solution that had not been treated with enzyme after standing overnight at room temperature. Thus, we have demonstrated the first observed enzymatic activity for MppR (Scheme 1). Preliminary kinetic analysis (see the Supporting Information) suggests that the reaction obeys Michaelis-Menten kinetics, though the  $K_M$  value for 6 is very high (>50 mM). We did not observe saturation of the reaction rate below the solubility limit of 6.

However, we were also not able to identify any 7 in overnight reaction mixtures containing 10  $\mu$ M MppR, 30 mM pyruvate, and 100 mM 4-imidazolecarboxaldehyde, a surprising observation in light of our crystallographic data. Further experiments with pyruvate and the alternative aldehyde substrate 3-(2-furyl)acrolein, the condensation product of which is bright yellow and absorbs UV light at 360 nm, cleared up this apparent paradox. In these tests, high concentrations of MppR (>100  $\mu$ M) were required for the accumulation of sufficient condensation product to give a measurable absorbance signal, suggesting that the condensation product remained covalently linked to the enzyme. This suspicion was confirmed after filtering the reaction mixture through a centrifugal concentrator to remove the enzyme. The yellow color remained trapped in the concentrator, while the filtrate lacked any absorbance at 360 nm, indicating that the product remained bound to the enzyme. It is currently unclear why the rate of hydrolysis of this Schiff base complex is so slow (it is stable for at least 1 day), but it may be that substance 7 mimics an intermediate, 11a (Figure 6), that must be converted

to 8, with concomitant conformational changes of individual active site residues and/or larger structural elements, before it can be released from the enzyme. It is also possible that the observed aldol condensation is a side reaction resulting from the presence of the Schiff base with pyruvate and is not relevant to End biosynthesis.

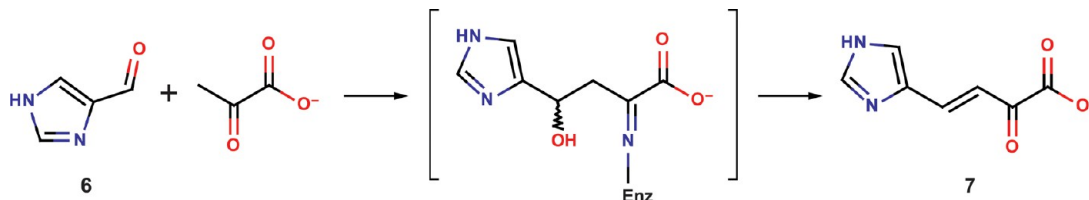
On the basis of analogy to capreomycin biosynthesis, it is possible that 4-hydroxyarginine is an intermediate in End biosynthesis and thus a potential substrate for MppR. To determine if MppR can react with 4-hydroxyarginine, a diastereomeric mixture of 4(*R/S*)-hydroxy-2(*S*)-arginine was treated with *C. atrox* L-amino acid oxidase (LAAO)<sup>48</sup> to convert the 2*S* diastereomers to the corresponding  $\alpha$ -keto acids (10, Figure 6). The LAAO was removed from the solution using a centrifugal concentrator, and the filtrate was evaporated to an amber residue in a rotary evaporator at 50 °C. When crystals of MppR were soaked in a solution containing this material at a nominal concentration of 300 mM (assuming all of the material was 10), the resulting crystal structure clearly showed 3-[(4*R*)-2-iminoimidazolidin-4-yl]-2-oxopropanoic acid (8, i.e., “2-keto-enduracididine”, Figure 1) bound in the active site (Figure 8) as the Schiff base. The pyruvate moiety of 8 makes the same interactions with the enzyme as pyruvate (Figure 5B) and 7 (Figure 7B). The two nitrogen atoms in the iminoimidazolidine ring of 8 make hydrogen bonding interactions with the carboxylate oxygen atoms of Glu283 (2.8 and 3.0 Å). Significantly, the chirality at C4 of 8 is that observed in the mannopeptimycins.<sup>6</sup>

Comparison of the NMR spectrum of the LAAO reaction product used to prepare the crystal soaking solution (Figure S2A of the Supporting Information) with that of the model compound 2,3-dihydroxypropylguanidinium (Figure S2B of the Supporting Information) shows that 10 is the major product from the LAAO reaction. Specifically, CH<sub>2</sub>N signals at  $\delta$  3.35 and 3.22 show geminal coupling constants of 14.4 Hz and vicinal coupling of 3.6 or 7.2 Hz, matching closely those of the linear model compound. Because the sample used to soak crystals contained predominantly the linear molecule 10, the observation of the cyclized 8 bound to the enzyme suggests that MppR catalyzed the cyclization of 10 to produce the iminoimidazolidine



**Figure 7.** Stereoview of the MppR-7 Schiff base complex (A) showing the experimental  $2|F_o| - |F_c|$  electron density (magenta) and the simulated annealing composite omit electron density (green), both contoured at  $1.2\sigma$ . The schematic of the complex (B) shows potential interactions and their distances (in angstroms). Note the continuous electron density between 7 and Lys156 N $\epsilon$ , indicating that 7 is covalently bound to the enzyme. The C2–C3–C4–C5 torsion angle is  $-178.0^\circ$ , which puts all four atoms in the same plane. Thus, this structure is most likely in the imine form.

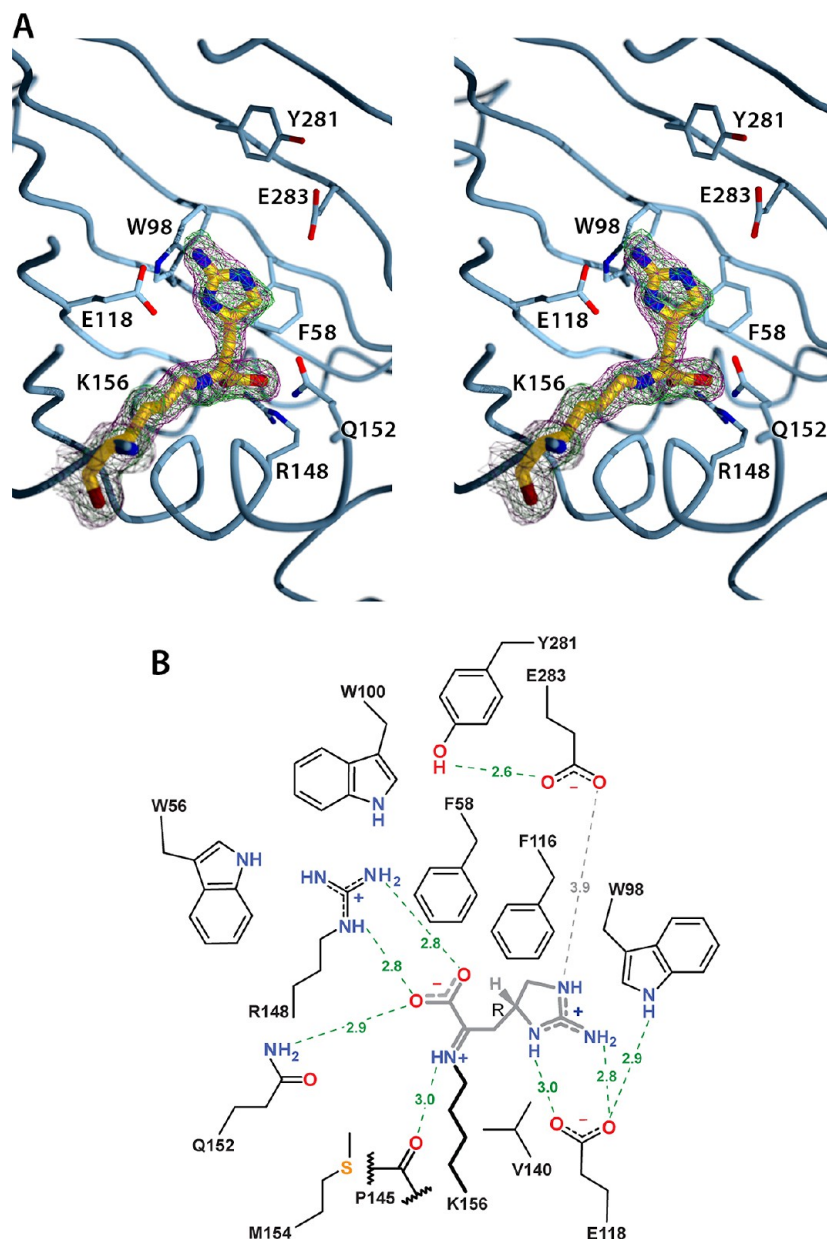
**Scheme 1**



ring of enduracididine. It should be noted that the sample of 10 does contain some uncharacterized impurities (<15%), so we cannot rule out the presence of a small amount of 8 in the crystal soaking solution. Consequently, it is possible, though perhaps

less likely, that MppR has simply selected keto acid 8 from the mixture rather than catalyzing its formation.

**Catalytic Implications and Proposed Function of MppR.** The three structures of MppR in complexes with  $\alpha$ -



**Figure 8.** Stereoview of the MppR-8 Schiff base complex (A) showing the experimental  $2|F_o| - |F_c|$  electron density (magenta) and the simulated annealing composite omit electron density (green), both contoured at  $1.2\sigma$ . The schematic of the complex (B) shows potential interactions and their distances (in angstroms). As expected, **8** is bound covalently to Lys156  $N\zeta$  as the Schiff base imine. The fact that the soak solution contained predominantly **10** suggests that MppR catalyzed the cyclization to give **8**.

keto acids described here show that the Schiff base intermediate is shared between both the classical ADC enzymes and MppR. More importantly, the MppR-7 structure demonstrates that the enzyme can catalyze the aldol condensation and subsequent dehydration of pyruvate and 4-imidazolecarboxaldehyde (**6**) to give the corresponding enone **7** (if it were released from the enzyme). While it is not clear at this point that the condensation activity is relevant to the biosynthesis of End, it does raise the possibility that MppR may synthesize an oxidized arginine derivative suitable for cyclization to form the iminoimidazoline ring of End directly from pyruvate and guanidinoacetaldehyde (**5**). Finally, the fact that the *R* enantiomer of **8** was trapped in the MppR-8 structure suggests that, whatever the origin of the presumed oxidized arginine intermediate, MppR may catalyze its cyclization to give the iminoimidazoline ring of **8**.

**Natural History and Genome Context Analysis of MppR-like Domains.** Comparative genome analysis of MppR and its closely related orthologs defined a monophyletic family of MppR-like ADCSF enzymes clearly distinct from the classical ADCs<sup>22</sup> (see Supporting Information and Materials and Methods). MppR family members are found in both bacteria and archaea. The distribution of MppR domains across archaeal lineages, including its presence in the deep-branching korarchaeota lineage,<sup>49</sup> relative to their sporadic distribution in bacteria strongly suggests an archaeal origin for MppR-like enzymes followed by horizontal transfer into bacteria. Phylogenetic trees constructed for the family support this scenario (Figure S3 of the Supporting Information), with strong grouping of archaeal sequences to the exclusion of bacterial sequences. Given that bacterial membership is concentrated in

actinobacteria and only sporadically found in  $\alpha$ -,  $\beta$ -, and  $\delta$ -proteobacteria and synergistetes, it is likely the transfer occurred initially into actinobacteria followed by subsequent dispersion to additional lineages through horizontal gene transfer events (Figure S3 and text file of the Supporting Information).

Analysis of the genome context of the MppR family reveals a near-constant association with a gene closely related to the classical aldehyde dehydrogenases comprised of a NAD-binding domain fused to a Rossmannoid catalytic domain.<sup>50</sup> Several neighborhoods spanning both archaeal and bacterial lineages additionally associate with a distinct enzyme containing an oxidoreductase-like Rossmannoid domain. Further associations observed almost exclusively in bacteria occur with a variety of genes encoding enzymes likely acting on amino acid or amino acid-derived substrates, including PLP-dependent aminotransferases, GATase-like glutamine amidotransferases, and GlnA-like glutamine synthetases. These associations suggest that archaeal members of the family could act as decarboxylases as part of a pathway involved in the general processing of highly oxidized substrates, while bacterial members were recruited to more specialized functional niches likely involving the processing of amino acids or associated metabolites. The MppR protein itself clusters with several actinobacterial proteins all likely to play a role in End biosynthesis; other bacterial MppR-like domains could act on similar, End-like substrates (text file of the Supporting Information).

## CONCLUSIONS

In spite of significant sequence and structural identity with the classical acetoacetate decarboxylases, as suggested by a comparative genome analysis and shown through targeted biochemical assays, MppR is not a decarboxylase. Rather, it has type I pyruvate aldolase activity. The demonstrated aldolase-dehydratase activity corroborates the idea that MppR catalyzes the dehydration of an oxidized arginine derivative (e.g., **10**) to give the 3–4 unsaturated derivative **11b**, which is subsequently cyclized to give **8**. It is also possible that MppR may catalyze the aldol condensation of pyruvate and guanidinoacetaldehyde (**5**) to produce compound **10**. From a broader perspective, the contextual analysis of this and additional ADCSF families suggests that the ADC fold displays considerable adaptability for the acquisition of novel biochemical functions. Rapid exploration of substrate space in ADC-like enzymes appears to proceed through retention of core catalytic residues with punctuating loss and gain of ancillary structural and sequence features concomitant with extensive horizontal gene flow. Similar patterns have been observed previously in other rapidly evolving enzyme families.<sup>51–53</sup> The overriding implication of these findings is that the ADC domains represent possible fruitful avenues of biophysical and biochemical investigations, potentially leading to the discovery of useful biocatalysts like the classical ADCs.

## ASSOCIATED CONTENT

### Supporting Information

Methods used in the bioinformatic analysis of the ADCSF described in detail, additional images illustrating the 5-NSA labeling of MppR and the distribution of the gene context of MppR-like sequences, and a text file with sequence alignments. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [silvaggi@uw.edu](mailto:silvaggi@uw.edu). Phone: (414) 229-2647.

### Funding

Work at Oregon State University was supported by National Institute of Allergy and Infectious Diseases Grant AI073784. This work is presently supported by National Science Foundation MRI Grant 0959442 and by National Institute of Allergy and Infectious Diseases Grant 1R03AI090339-01A1.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357. Use of LS-CAT Sector 21 was supported by the Michigan Economic Development Corp. and the Michigan Technology Tri-Corridor (Grant 085P1000817).

## ABBREVIATIONS

ADC, acetoacetate decarboxylase; ADCSF, acetoacetate decarboxylase-like superfamily; End, enduracididine;  $\beta$ hEnd,  $\beta$ -hydroxyenduracididine; rmsd, root-mean-square deviation.

## REFERENCES

- (1) Foglia, E. E., Fraser, V. J., and Elward, A. M. (2007) Effect of nosocomial infections due to antibiotic-resistant organisms on length of stay and mortality in the pediatric intensive care unit. *Infection Control and Hospital Epidemiology* 28, 299–306.
- (2) Mauldin, P. D., Salgado, C. D., Durkalski, V. L., and Bosso, J. A. (2008) Nosocomial infections due to methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococcus: Relationships with antibiotic use and cost drivers. *Ann. Pharmacother.* 42, 317–326.
- (3) Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H. W., Scheld, W. M., Bartlett, J. G., and Edwards, J., Jr. (2008) The epidemic of antibiotic-resistant infections: A call to action for the medical community from the Infectious Diseases Society of America. *Clin. Infect. Dis.* 46, 155–164.
- (4) He, H. (2005) Mannopeptimycins, a novel class of glycopeptide antibiotics active against Gram-positive bacteria. *Appl. Microbiol. Biotechnol.* 67, 444–452.
- (5) Magarvey, N. A., Haltli, B., He, M., Greenstein, M., and Hucul, J. A. (2006) Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant Gram-positive pathogens. *Antimicrob. Agents Chemother.* 50, 2167–2177.
- (6) Singh, M. P., Petersen, P. J., Weiss, W. J., Janso, J. E., Luckman, S. W., Lenoy, E. B., Bradford, P. A., Testa, R. T., and Greenstein, M. (2003) Mannopeptimycins, new cyclic glycopeptide antibiotics produced by *Streptomyces hygroscopicus* LL-AC98: Antibacterial and mechanistic activities. *Antimicrob. Agents Chemother.* 47, 62–69.
- (7) Ruzin, A., Singh, G., Severin, A., Yang, Y., Dushin, R. G., Sutherland, A. G., Minnick, A., Greenstein, M., May, M. K., Shlaes, D. M., and Bradford, P. A. (2004) Mechanism of action of the mannopeptimycins, a novel class of glycopeptide antibiotics active against vancomycin-resistant Gram-positive bacteria. *Antimicrob. Agents Chemother.* 48, 728–738.
- (8) Guppi, S. R., and O'Doherty, G. A. (2007) Synthesis of azanalogues of the glycosylated tyrosine portion of mannopeptimycin-E. *J. Org. Chem.* 72, 4966–4969.
- (9) Petersen, P. J., Wang, T. Z., Dushin, R. G., and Bradford, P. A. (2004) Comparative in vitro activities of AC98-6446, a novel semisynthetic glycopeptide derivative of the natural product mannopeptimycin  $\alpha$ , and other antimicrobial agents against Gram-positive clinical isolates. *Antimicrob. Agents Chemother.* 48, 739–746.

- (10) He, H., Shen, B., Petersen, P. J., Weiss, W. J., Yang, H. Y., Wang, T. Z., Dushin, R. G., Koehn, F. E., and Carter, G. T. (2004) Mannopeptimycin esters and carbonates, potent antibiotic agents against drug-resistant bacteria. *Bioorg. Med. Chem. Lett.* 14, 279–282.
- (11) Sum, P. E., How, D., Torres, N., Petersen, P. J., Ashcroft, J., Graziani, E. I., Koehn, F. E., and Mansour, T. S. (2003) Synthesis and evaluation of ether and halogenated derivatives of mannopeptimycin glycopeptide antibiotics. *Bioorg. Med. Chem. Lett.* 13, 2805–2808.
- (12) Olivier, K. S., and Van Nieuwenhze, M. S. (2010) Synthetic studies toward the mannopeptimycins: Synthesis of orthogonally protected  $\beta$ -hydroxyenduracidinidines. *Org. Lett.* 12, 1680–1683.
- (13) He, H., Williamson, R. T., Shen, B., Graziani, E. I., Yang, H. Y., Saky, S. M., Petersen, P. J., and Carter, G. T. (2002) Mannopeptimycins, novel antibacterial glycopeptides from *Streptomyces hygroscopicus*, LL-AC98. *J. Am. Chem. Soc.* 124, 9729–9736.
- (14) Haltli, B., Tan, Y., Magarvey, N. A., Wagenaar, M., Yin, X., Greenstein, M., Hucul, J. A., and Zabriskie, T. M. (2005) Investigating  $\beta$ -hydroxyenduracidinidone formation in the biosynthesis of the mannopeptimycins. *Chem. Biol.* 12, 1163–1168.
- (15) Hatano, K., Nogami, I., Higashide, E., and Kishi, T. (1984) Biosynthesis of Enduracidin: Origin of Enduracidinidone and Other Amino Acids. *Agric. Biol. Chem.* 48, 1503–1508.
- (16) Yin, X., and Zabriskie, T. M. (2006) The enduracidin biosynthetic gene cluster from *Streptomyces fungicidicus*. *Microbiology* 152, 2969–2983.
- (17) Ju, J., Ozanick, S. G., Shen, B., and Thomas, M. G. (2004) Conversion of (2S)-arginine to (2S,3R)-capreomycinidine by VioC and VioD from the viomycin biosynthetic pathway of *Streptomyces* sp. strain ATCC11861. *ChemBioChem* 5, 1281–1285.
- (18) Yin, X., and Zabriskie, T. M. (2004) VioC is a non-heme iron,  $\alpha$ -ketoglutarate-dependent oxygenase that catalyzes the formation of 3S-hydroxy-L-arginine during viomycin biosynthesis. *ChemBioChem* 5, 1274–1277.
- (19) Yin, X., McPhail, K. L., Kim, K. J., and Zabriskie, T. M. (2004) Formation of the nonproteinogenic amino acid 2S,3R-capreomycinidine by VioD from the viomycin biosynthesis pathway. *ChemBioChem* 5, 1278–1281.
- (20) Highbarger, L. A., Gerlt, J. A., and Kenyon, G. L. (1996) Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the pKa of active-site lysine 115. *Biochemistry* 35, 41–46.
- (21) Warren, S., Zerner, B., and Westheimer, F. H. (1966) Acetoacetate decarboxylase. Identification of lysine at the active site. *Biochemistry* 5, 817–823.
- (22) Ho, M. C., Menetret, J. F., Tsuruta, H., and Allen, K. N. (2009) The origin of the electrostatic perturbation in acetoacetate decarboxylase. *Nature* 459, 393–397.
- (23) Meister, A. (1952) Enzymatic preparation of  $\alpha$ -keto acids. *J. Biol. Chem.* 197, 309–317.
- (24) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods Enzymol.* 276, 307–326.
- (25) Batty, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* 67, 271–281.
- (26) Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. In *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, Vol. 26, Daresbury Laboratory, Warrington, U.K.
- (27) Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D* 62, 72–82.
- (28) Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* 50, 760–763.
- (29) Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.* 364, 215–230.
- (30) Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., and Adams, P. D. (2008) Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* 426, 419–435.
- (31) Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* 64, 61–69.
- (32) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D* 66, 486–501.
- (33) Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P., and Adams, P. D. (2010) Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr. D* 66, 1153–1163.
- (34) Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* 285, 1735–1747.
- (35) Painter, J., and Merritt, E. A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D* 62, 439–450.
- (36) Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* 66, 12–21.
- (37) Urzhumtseva, L., Afonine, P. V., Adams, P. D., and Urzhumtsev, A. (2009) Crystallographic model quality at a glance. *Acta Crystallogr. D* 65, 297–300.
- (38) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.
- (39) Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* 65, 1074–1080.
- (40) Tagaki, W., and Westheimer, F. H. (1968) Acetoacetate decarboxylase. The molecular weight of the enzyme and subunits. *Biochemistry* 7, 895–900.
- (41) Frey, P. A., Kokesh, F. C., and Westheimer, F. H. (1971) A reporter group at the active site of acetoacetate decarboxylase. I. Ionization constant of the nitrophenol. *J. Am. Chem. Soc.* 93, 7266–7269.
- (42) Kokesh, F. C., and Westheimer, F. H. (1971) A reporter group at the active site of acetoacetate decarboxylase. II. Ionization constant of the amino group. *J. Am. Chem. Soc.* 93, 7270–7274.
- (43) Andreeva, A., and Murzin, A. G. (2010) Structural classification of proteins and structural genomics: New insights into protein folding and evolution. *Acta Crystallogr. F* 66, 1190–1197.
- (44) Gallivan, J. P., and Dougherty, D. A. (1999) Cation- $\pi$  interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9459–9464.
- (45) Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.
- (46) Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* 60, 2256–2268.
- (47) Nakane, A., Nakamura, T., and Eguchi, Y. (1977) A novel metabolic fate of arginine in *Streptomyces eurocidicus*. Partial resolution of the pathway and identification of an intermediate. *J. Biol. Chem.* 252, 5267–5273.
- (48) Meister, A. (1954) The  $\alpha$ -keto analogues of arginine, ornithine, and lysine. *J. Biol. Chem.* 206, 577–585.
- (49) Elkins, J. G., Podar, M., Graham, D. E., Makarova, K. S., Wolf, Y., Randau, L., Hedlund, B. P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltzman, E., Barry, K., Koonin, E. V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M., and Stetter, K. O. (2008) A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8102–8107.

(50) Liu, Z. J., Sun, Y. J., Rose, J., Chung, Y. J., Hsiao, C. D., Chang, W. R., Kuo, I., Perozich, J., Lindahl, R., Hempel, J., and Wang, B. C. (1997) The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nat. Struct. Biol.* 4, 317–326.

(51) Burroughs, A. M., Iyer, L. M., and Aravind, L. (2009) Natural history of the E1-like superfamily: Implication for adenylation, sulfur transfer, and ubiquitin conjugation. *Proteins* 75, 895–910.

(52) Aravind, L., Anantharaman, V., Zhang, D., de Souza, R. F., and Iyer, L. M. (2012) Gene flow and biological conflict systems in the origin and evolution of eukaryotes. *Front. Cell. Infect. Microbiol.* 2, 89.

(53) Iyer, L. M., Abhiman, S., Maxwell Burroughs, A., and Aravind, L. (2009) Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: Synthesis of novel metabolites and peptide modifications of proteins. *Mol. BioSyst.* 5, 1636–1660.

(54) Fenn, T. D., Ringe, D., and Petsko, G. A. (2003) POVScript+: A program for model and data visualization using persistence of vision ray-tracing. *J. Appl. Crystallogr.* 36, 944–947.

(55) Kraulis, P. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.