Structure and Function of the Pimeloyl-CoA Synthetase BioW Defines a New Fold for Adenylate-Forming Enzymes

Paola Estrada[1], Miglena Manandhar[2], Shi-Hui Dong[1,3], Jaigeeth Deveryshetty[3], Vinayak Agarwal[3,4], John E. Cronan[1,2,3], and Satish K. Nair[1,3,4*]

[1]Department of Biochemistry, [2]Department of Microbiology, [3]Institute for Genomic Biology and [4]Center for Biophysics and Computational Biology, University of Illinois at Urbana Champaign, 600 S. Mathews Ave, Urbana IL USA

ABSTRACT

Reactions that activate carboxylate groups through acyl-adenylate intermediates are found throughout biology. Enzymes catalyzing such reactions include acyl/aryl-CoA synthetases that conjugate activated acids to the thiol of CoA and tRNA synthetases that attach the α-carboxylate of amino acids to the 2'- or 3'-hydroxyl of the terminal nucleotide in tRNA. Here, we describe the structure and biochemical characterization of BioW, which represents a new protein fold within the large superfamily of adenylating enzyme. Nucleotide, CoA, and substrate bound structures identify the enzyme active site, and elucidate the mechanistic strategy used to conjugate CoA to the seven carbon α,ω-dicarboxylic pimelate, a biotin precursor. Proper position of reactive groups for the two half reactions is achieved solely through minor movements of active site residues, as confirmed by site-directed mutational analysis. The ability of BioW to hydrolyze adenylates of non-cognate substrates is reminiscent of pre-transfer proofreading observed in some tRNA synthetases, and we show that this activity can be abolished by a single residue active site mutation. These studies illustrate how BioW can carry out three different, biologically prevalent chemical reactions (adenylation, thioesterification, and proofreading) in the context of a new protein fold.

INTRODUCTION

Biotin (vitamin B7 or coenzyme R) is a water soluble, essential cofactor in all domains of life, where it serves as the prosthetic group for numerous metabolic enzymes that catalyze carboxyl-transfer reactions[1]. The chemical structure of biotin consists of a tetrahydroimidizalone ring fused with organosulfur-containing tetrahydrothiopene ring that bears a valeric acid substituent[2]. Attachment of biotin to constituent enzymes occurs via an amide linkage between the carboxylate of the valeric acid moiety and the ε-amine of a specific Lys in biotin carrier protein, a small domain of roughly 80 residues[3]. The bicyclic rings of the resultant biotinylated protein extends outwards, where the ureido ring N8 nitrogen is poised to carry equivalents of CO_2 (carboxybiotin) between the carboxylation and carboxyltransfer domains of biotin-dependent enzymes[4].

Biosynthesis of biotin only occurs in bacteria, fungi, and plants. Although mammals do not synthesize biotin *de novo*, intestinal bacteria produce biotin well in excess of the necessary daily requirements[5]. The unusual structure of biotin is derived from two precursors, alanine and a thioester of the C7 α,ω-dicarboxylic acid pimelate[6] (**Figure 1**). Genetic studies of *E. coli* biotin auxotrophic mutants helped elucidate the structures of the intermediates in the late stages of the biosynthetic pathway[7,8], and subsequent studies established that four enzymes catalyze the assembly of the bicyclic ring in an ATP- and *S*-adenosylmethionine (SAM)-dependent manner[9]. However, the pathway for incorporation of pimelate into biotin has only recently been determined, and only for the Gram-negative proteobacterium *E. coli* and the Gram-positive Firmicute *B. subtilis* and closely related species[10]. Notably, the biosynthetic routes to the common pimeloyl-CoA intermediate are divergent between *E. coli* and *B. subtilis*.

Isotopic labeling studies using *E. coli* cultures grown in differentially labeled ^{13}C acetate suggested that three acetate units are incorporated into the pimelate group of biotin[11,12]. The biotin carbon atoms derived from the α and ω carboxylates of biotin demonstrated different labeling patterns, ruling out symmetric pimelate, and consistent with a pimeloyl thioester, as the likely intermediate. Genetic and biochemical studies show that two genes essential for *E. coli* biotin biosynthesis (BioC and BioH) provide a route for the assembly of pimeloyl-ACP thioester by hijacking the fatty acid biosynthetic machinery[13]. Specifically, BioC is a SAM-dependent

methyltransferase that converts the ω-carboxylate of malonyl-ACP into the corresponding methyl ester, which condenses with malonyl-ACP. Esterification both protects the ω-carboxylate from spontaneous decarboxylation, and provides a neutral charge mimic of fatty acyl chains that can be accommodated in the hydrophobic active sites of fatty acid synthetic enzymes. Following each round of condensation, the inceptive β-keto group is reduced to the methylene through the fatty acid synthesis machinery, and after two condensation/reduction cycles, BioH hydrolyzes the pimeloyl-ACP methyl ester and prevents further elongation[14].

In contrast, *B. subtilis* and related strains (but not all bacilli), utilize two different orthogonal strategies for the production of pimeloyl thioesters (**Figure 1A**). In the first scheme, the cytochrome P450 enzyme BioI catalyzes the *de novo* synthesis of pimeloyl thioester from long-chain acyl-ACPs[15-17]. Orientation of the ACP-linked fatty acyl thioester on the surface of BioI locks the long-chain acyl group into the active site along a bent-tunnel, thus positioning the C7-C8 bond directly above the heme iron, where consecutive hydroxylations at C7 and C8, and subsequent oxidation of the vicinal diol produces the pimeloyl-ACP product[18]. The physiological role of BioI is unclear as many *bio* operons lack this gene. A second route is the ATP-dependent synthesis of pimeloyl-CoA from the free α,ω-dicarboxylic acid pimelate by the acyl-CoA synthetase BioW[19]. The activity of BioW is strictly dependent on ATP, and the enzyme demonstrates specificity for pimelate[13,15]. This substrate specificity is strict, as the enzyme cannot utilize longer or shorter dicarboxylic acids and monocarboxylates of equivalent length (C6 or C7) are also not substrates[20].

The BioW catalyzed reaction is shown to proceed through formation of stoichiometric quantities of an acyl-adenylate intermediate[20], a common feature amongst acyl-CoA ligases. However, the primary sequence of the enzyme does not reveal any of the sequence motifs that are common amongst the different clades of acyl-CoA ligases[21], and the polypeptide is about half the size of the canonical acyl-CoA ligases[22]. In general, enzymes that carry out adenylation reactions can activate an otherwise inert substrate carboxylate through the attachment of adenosine monophosphate, thereby replacing a poor leaving group (hydroxide) with a much better one (AMP). Adenylation chemistry is prevalent throughout biology, and a recent classification of adenylating enzymes superfamily defines three classes; class I that includes adenylation domains

from non-ribosomal peptide synthetases (NRPS), and acyl/aryl-CoA ligases, class II that encompass the amino acyl-tRNA syntheases (aaRSs), and class III that include the siderophore synthetases[21]. Biochemical analysis of reaction intermediates demonstrated that recombinant *B. subtilis* BioW can hydrolyze non-cognate substrates[20], an activity that is similar to pretransfer proofreading observed in aaRSs that define class II members of the superfamily. The primary sequence of BioW also lacks any notable motifs common amongst aaRSs.

Here we report several crystal structures of a functionally active BioW from *Aquifex aeolicus*. These structures reveal a previously uncharacterized fold for an acyl-CoA ligase. The enzyme demonstrates a strict specificity for pimelic acid and shows the ability to proofread and hydrolyze non-cognate acyl-adenylates. Biochemical characterization of structure-based variants identifies the residues that establish both the strict specificity for the correct length dicarboxylic acid, as well as dictate the proofreading of non-cognate acyl-adenylates.

RESULTS

Identification and characterization of functional BioW homologs

As primary sequence analysis failed to identify obvious structural homologs, we utilized the Enzyme Similarity Tool from the EFI[23] in order to analyze the relationship among different sequences in the BioW superfamily. Using a default E-value cutoff of 10^{-10}, a total of 293 sequences were culled from UniProt. At an alignment score corresponding to a cutoff of at least 60% identity in sequence produced a total of 2857 edges in the similarity network. Each of the protein sequences/nodes are colored by class, but most of the sequences clustered according to genus. We cloned representative BioW homologs from different genera, and screened each for crystallization. Crystals could be readily obtained for *A. aeolicus* (hereafter AaBioW; related nodes colored in yellow in **Figure 1**), and for *B. amyloliquefaciens* (hereafter BaBioW; related nodes colored in cyan). We determined the kinetic parameters for AaBioW by measuring the rate of pimeloyl-CoA production using analytical HPLC (**Figure 2A,D**). As *A. aeolicus* is a thermophilic organism, all of the kinetic experiments were carried out at 323K, a suitable temperature below the natural growth environment of the organism (358K) but one where formation of non-enzymatic side products was negligible. The K_M of AaBioW for pimelate is 10.7 +/- 0.96 μM with a k_{cat} value of 7.45 +/- 0.023 x 10^{-1} s^{-1}. The K_M value is similar to that

obtained for the well-characterized BioW from *Bacillus subtilis*, the kinetic parameters for which were determined using TLC analyses[20]. In contrast, the k_{cat} value for AaBioW is nearly four orders of magnitude greater than that of the *Bacillus* enzyme (k_{cat} = 8.4 +/- 1.3 x 10^{-5} s^{-1}). Hence, the enzyme from Aquificae is a much more robust catalyst than that from Firmicutes. Attempts to determine the kinetic parameters for BaBioW were hampered by a propensity for the enzyme to precipitate and loss of activity upon prolonged incubation. AaBioW proved to be more amenable for biochemical studies and thus structural efforts focused solely on this enzyme.

The overall ligase reaction catalyzed by BioW consists of two half reactions, namely ATP-dependent adenylation of pimelate, followed by acyl transfer to CoA. We sought to characterize the interaction between AaBioW and each of the two substrates (ATP and CoA) using isothermal titration calorimetry (**Figure 2B,C**). Experiments were conducted in duplicate, and the results are the means of both experiments. Each of the ligands bound to the enzyme with a stoichiometry of one, and with respective dissociation constant (K_d) values of 13 +/- 3 μM for ATP and 14 +/- 2.5 μM for CoA. The calorimetric data support the biochemical results and demonstrate direct binding of each ligand by the enzyme.

All known adenylate forming enzymes that have been characterized are metal dependent, with a preference for Mg^{2+} [21]. Typically, the metal ion coordinates to the phosphate groups of ATP and acts to stabilize charges that develop in the transition state and facilitates release of the pyrophosphate product. We tested the Mg^{2+} requirement for AaBioW by measuring total acyl-CoA production in end-point experiments at various concentrations of the metal ion (**SI Figure S1**). In reactions carried out with 1 mM EDTA and in the absence of Mg^{2+}, no consumption of ATP can be observed. The production of both AMP and the cognate pimeloyl-CoA varied in a linear fashion with increasing concentration of exogenous Mg^{2+} and saturated above stoichiometric concentrations of the metal. These data demonstrate that, like *B. subtilis* BioW[20], the adenyl transfer reaction catalyzed by AaBioW is Mg^{2+}-dependent.

In addition to catalyzing the synthesis of pimeloyl-CoA, *B. subtilis* BioW is demonstrated to catalyze the hydrolysis of non-cognate substrates (proofreading)[20]. We tested the proofreading activity of AaBioW against a panel of α,ω-dicarboxylate substrates including glutaric (C5),

adipic (C6), and suberic (C8) acids using thin-layer chromatography (**Figure 2E**). In the absence of CoA, adenylated intermediates can be observed for the cognate C7, as well as the non-cognate C6, and C8 substrates. However, while addition of CoA to the pimeloyl (C7)-adenylate produced the corresponding pimeloyl-CoA, CoA addition to the non-cognate C6 and C8 adenylates simply resulted in hydrolysis to yield AMP and the free acids **(Figure 2E and SI Figure S2**). In contrast, reactions carried out using glutarate (C5) in the absence of CoA resulted only in the accumulation of AMP, suggesting that AaBioW can catalyze the cleavage of glutaryl-adenylate, as was previously established for *B. subtilis* BioW. Likewise, increase in the pH of the reaction buffer resulted in the expected hydrolysis of the cognate pimeloyl-adenylate, but had minimal effect on the hydrolysis of the non-cognate glutaryl-adenylate (**SI Figure S2**). These data are consistent with the prior observations with *B. subtilis* BioW[20] that AMP production from the corresponding non-cognate adenylates is a function of proofreading by AaBioW, and not the result of non-enzymatic solvent hydrolysis.

Crystal Structures of BioW

We determined the crystal structures of BaBioW to 3.1 Å resolution and of AaBioW to 2.5 Å resolution. Crystallographic phases for each structure were determined independently using single wavelength anomalous diffraction data collected on crystals of selenomethionine labeled proteins. Relevant data collection and refinement statistics are reported in **SI Table S2.** The structures of the two homologs are architecturally similar, and the overall fold consists of a N-terminal domain of roughly 55 residues that contain three anti-parallel β-strands with a single intervening α-helix and a C-terminal domain of roughly 200 residues that constitutes a α/β domain (**Figure 3A**). A DALI search[26] against the Protein Data Bank[27] fails to identify any structural homologs of the entire polypeptide. Consequently, the overall topology and structure of BioW represents a new protein fold with no resemblance to those of known adenylating enzymes **Figure 3B**).

Structure-based similarity searches using the isolated domains of the protein revealed that the N-terminus resembles the P_{II} proteins that modulate bacterial signaling pathways via interactions with a wide-range of protein targets[28] (50-60 Cα aligned with a Z-scores around 5.0) (**SI Figure S3A,B**). The physiological role of P_{II} proteins involves regulation of cellular nitrogen levels. The

homotrimeric P_{II} proteins are modulated by binding to ligand effectors, most notably nucleotides such as ATP and ADP[29]. More distant topological relationship can also be discerned with the B-domains of ATP-grasp enzymes, such as L-amino acid ligases (43 Cα aligned with a Z-score of 3.2). The B-domain is positioned away from the main body of ATP-grasp enzymes, and undergoes conformational movements to encapsulate the active site upon binding of nucleotide[30].

The BioW C-terminal domain resembles a modified Rossmann fold, a structural motif that is commonly found in nucleotide binding proteins[31]. However, the connectivity of the secondary structural elements of this domain in BioW is markedly different from the alternating β-α-β motif found in canonical Rossmann fold enzymes (**Figure 3B**), such as protein kinases, and class II aaRSs[32,33]. The closest structural homolog of this domain is the amino-terminal nucleotide-binding domain of the di-adenylate cyclase DisA (PDB Code 3C23; Z-score of 4.1; RMSD of 3.7 Å over 102 aligned Cα atoms) (**SI Figure S3C**)[34]. This enzyme is a homotetramer that orients bound ATP from each protomer in an antiparallel fashion to facilitate the synthesis of bis-(3',5')-cyclic dimeric adenosine monophosphate (c-di-AMP), and this activity is coordinated with the binding of a branched DNA substrate by the carboxy-terminal domain. The location of the bound nucleotide in AaBioW is considerably different from that found in DisA (**SI Figure S3A, C**), as would be expected given their different topologies and overall reactions catalyzed by each enzyme. A more distant structural conservation is observed with (3',5')-cyclic phosphodiesterases (PDB Code 3DBA; Z-score of 3.0; RMSD of 4.0 Å over 88 aligned Cα atoms)[35], for which the nucleotide-binding site is entirely different from that in BioW. There are also some structural similarities with the "little finger"/polymerase associated domain (LF/PAD) unique to the Y-family of translesion DNA polymerases. The LF/PAD domain is flexible and interacts with the DNA duplex upstream from a mismatch, to help guide the template strand into the polymerase active site[36].

Characterization of the substrate binding sites in AaBioW

Individual cocrystal structures of AaBioW with substrate pimelate (2.45 Å), the nonhydrolyzable analog AMP-CPP+Mg^{2+}+pimelate (2.55 Å), and with AMP+CoA (2.25 Å) allow elucidation of the active sites of the enzyme (**Figure 3C-E**). The AaBioW-AMP+CoA, and AMP-PCP+Mg^{2+}+pimelate structures show unambiguous density corresponding to the nucleotide in

the C-terminal Rossmann fold domain, where it is flanked on the opposite side by the N-terminal P_{II}-like domain. In both structures, the adenine moiety is bracketed between two Arg residues (Arg135 and Arg215) mainly through van der Waals contacts rather than the more common cation-π stacking interaction (**Figure 3D,E; Figure 4A**). These residues, along with a third Arg113, and two aliphatic residues (Gly114 and Leu166) delineate the binding pocket for the nucleobase. There are minimal hydrogen-bonding interactions with the enzyme, and such interactions are only through main chain atoms, namely between N6 and N1 of adenine and the backbone amide and carbonyl oxygen, respectively, of Thr136. Similarly, the O2' atom of the ribose is within hydrogen-bonding distance to the backbone carbonyls of Met112 and Gly114, and the O3' atom is engaged by the side chain of Asp183. The α phosphate oxygen is engaged via interactions with Ser182, Asp183, and Arg215, with the side chain Arg159 located ~3.4 Å away. In the AMP-PCP+Mg^{2+}+pimelate structure, one of the oxygen atoms on the β phosphate is within hydrogen-bonding distance to Asp183, and the oxygen atoms of the γ phosphate are proximal to, but beyond hydrogen bonding distance from, His16, Arg39, and Arg135. A single Mg^{2+} ion is positioned at the β phosphate of the nucleotide, is additionally coordinated by Asp183 and is located ~3.3 Å away from the α phosphate. Notably, the nucleotide-binding site lacks all of the canonical features of representative adenyltransferases, such as stacking interactions that stabilize the nucleobase and carboxylate residues that are poised to facilitate metal-assisted pyrophosphate hydrolysis. Lastly, although some ATP-grasp enzymes contain a multi-domain architecture consisting of an A-domain that bears a Rossmann fold and a B-domain that superficially resembles a P_{II}-like fold[30], the architecture of BioW is distinct, as is the nature of the nucleotide binding site.

The binding site for pimelate could be readily discerned in electron density maps derived from the AaBioW-pimelate and AMP-PCP+Mg^{2+}+pimelate cocrystal structures. Bifurcated density, characteristic of the α and ω carboxylates, situates the bound pimelate at a location adjacent to the nucleotide-binding site, where we designate the α carboxylate as that closest to the phosphate of AMP (**Figure 3C,D**). Both of the substrate carboxylates are engaged through numerous interactions with polar residues, and a hydrophobic environment that is established by Ala163, Tyr187, and Tyr191 encapsulate the methylene carbons of the substrate. The α carboxylate is poised to interact with the side chains of Arg159, Ser182, and Arg215, and the side chains of

Tyr187, Tyr199, and Arg201 engage the ω carboxylate (**Figure 4B**). Most notably, the orientation of the side chain of Arg159 precisely matches the length of the pimelate, and may serve as a "ruler" to set the length of the correct substrate. The orientation of Arg159 also sequesters the ligand away from bulk solvent.

The structure of AaBioW in complex with AMP+CoA establishes the location of the CoA binding site (**Figure 3E**). The ligand is positioned almost exclusively within the N-terminal P_{II}-like domain, an unexpected finding given that CoA binding is not a function that has been previously attributed to P_{II}-like domains. The adenosine 3'-phosphate is accommodated by minor rearrangements of side chains. The CoA adenine is engaged via π-stacking interactions with the side chain of His16, and the 3'-phosphate is positioned to interact with the side chains of Arg113, Arg132, and Arg135 (**Figure 4B**). Side chain rotations of all of these residues are necessary in order to accommodate the binding of CoA (**SI Figure S4**). Numerous basic residues including Arg7, Arg9, and Lys49, engage the CoA diphosphate. The phosphopantetheine moiety runs along a groove between the N- and C-terminal domains and the thiol is positioned in proximity to the nucleotide and pimelate binding regions of the active site. All of the residues involved in interactions with nucleotide, CoA and pimelate are conserved across BioW sequences (**SI Figure S5**).

Mutational analysis of the function of active site residues

A superposition of all of the ligand bound structures yields plausible models for Michaelis complexes for the two half reactions (adenylation and thioester formation) (**Figure 4A, B**). These models suggest that each of the substrates is bound in the enzyme active site with the optimal orientation to facilitate catalysis. In the adenyation complex model, the α phosphate of ATP is oriented roughly 3.6 Å away from the α carboxylate of pimelate, and minor movement of either substrate would be necessary to facilitate the attack of the weakly nucleophilic carboxylate onto the weakly electrophilic phosphate (**Figure 4A**). The β and γ phosphates of ATP are positioned in the region between the P_{II} and Rossmann fold domains, to allow displacement to bulk solvent upon adenyl transfer. The function of residues that interact with ATP and pimelate were probed through site-specific variants (**Figure 4C, SI Figure S6**). End-point analyses demonstrate that the Tyr199→Ala and Arg201→Ala mutations have little effect on product

formation, consistent with the role for these residues in simply anchoring the ω carboxylate of the substrate pimelate. In contrast, the Arg159→Ala and Tyr187→Ala variants demonstrate a notable reduction in turnover, which is in line with the function of these residues in forming the exterior wall of the pimelate-binding cavity. Lastly, both Ser182 and Arg215 are situated immediately adjacent to the location of the incipient phosphoester bond, where they are poised to stabilize the intermediate formed during the in-line attack of the pimelate oxygen onto the α phosphorous atom of ATP in the adenylation reaction. Consequently, the Ser182→Ala and Arg215→Ala each demonstrate a significant reduction in product formation.

The hypothetical model for the pre-catalytic thioester forming complex shows a paucity of residues that could participate in chemistry, consistent with the energetic favorability of this half reaction (**Figure 4B**). In the model, the thiol sulfur atom is positioned at a distance suitable to favor nucleophilic attack onto the carbon atom of the adenylate anhydride. Although His16 forms a π-stacking interaction with the adenine, the His16→Ala variant displays only a modest 20% loss in activity relative to the wild type, reflecting the importance of these other interacting residues in stabilizing CoA binding. A slight 2.5 Å movement of Arg159 in the thioester forming model positions this residue to stabilize the negative charge that would develop on the oxygen atom of the α carboxylate upon attack by the thiol. Similar movements of amino acid side chains occur throughout the CoA binding site, notably at Arg7, Arg113, Arg132, and Arg135 (**SI Figure S4**). These changes serve to both avoid steric clashes upon ligand binding and to facilitate favorable interactions with polar groups in CoA. In the adenylation conformation, formation of the adenylate would be presumably protected from solvent hydrolysis by these Arg residues, which would then be displaced upon CoA binding. An interesting discovery from the mutational analysis is the identification of an Arg132→Ala variant that demonstrates far more robust activity than the wild-type enzyme. Inspection of the product distribution of this variant demonstrates a notable increase in production of both pimeloyl-CoA and AMP relative to the wild-type AaBioW (**SI Figure S7**). In the AMP+CoA crystal structure, Arg132 engages the adenosine 3'-phosphate of CoA.

A gatekeeper residue for proofreading of non-cognate adenylates

An unusual property of *B. subtilis* BioW is the ability of this enzyme to catalyze hydrolytic cleavage of acyl-adenylates other than the cognate pimeloyl (C7)-adenylate[20], as we have also demonstrated for AaBioW for the non-cognate glutaryl (C5)-adenylate. A comparison of the AaBioW cocrystal structures provides some hints regarding residues that may facilitate this proofreading activity (**SI Figure S4**). The movement of Arg159 that serves to position this residue to assist in thioester formation is reminiscent of the domain movement in acetyl-CoA synthetase that positions a catalytic lysine important for adenylation away from the active site to facilitate thioester formation[37]. A surface rendering of the thioester formation state illustrates the position of Arg159 directly above the α-phosphate consistent with a role for this residue in protecting the cognate pimeloyl-adenylate (**Figure 5A**). Conceivably, misalignment of the Arg159 side chain upon formation of complexes with non-cognate adenylates could facilitate hydrolytic cleavage, perhaps through electronic effects. In order to decipher any roles that this residue may play in proofreading, we carried out HPLC analysis of Arg159→Ala AaBioW using cognate pimelate, as well as non-cognate di- and mono-acids as substrates. The Arg159→Ala variants demonstrates the ability to generate the CoA thioester using pimelate as a substrate, albeit at lower concentrations than that of the wild-type enzyme (**Figure 5B**). However, while the variant enzyme can generate the corresponding adenylates using any of the non-cognate mono- or di-acid substrates, the acyl-adenylates are no longer hydrolyzed or converted into the corresponding thioesters in the presence of CoA (**Figures 5C-F)**. Hence, Arg159→Ala AaBioW lacks the ability to proofread, which strongly suggests a role for Arg159 in this activity.

DISCUSSION

Our biochemical and structural analyses of BioW elucidate a new architectural fold for an acyl-CoA ligase, and expands upon the growing superfamily of adenylating enzymes. A recent re-classification for adenylate-forming enzymes subdivides family members based on structural and sequence conservation, and consists of class I (adenylation domains of NRPSs, and acyl/aryl-CoA synthetases), class II (aminoacyl-tRNA synthetases) and the class III (NRPS-independent siderophore synthetases)[21]. Although structurally distinct from any of these other class members, BioW employs catalytic strategies that are congruous with those employed by other adenylating enzymes, including the use of a two-step reaction scheme to activate substrate by coupling it to AMP, followed by the attack of the adenylate by a nucleophile. All of these enzymes are metal dependent, typically using one or more Mg^{2+} ions, which coordinate to the ATP phosphates to facilitate pyrophosphate release.

Members of all three classes of adenylation enzymes contain one or more basic residues located adjacent to the α phosphate of ATP: class I enzymes contain an invariant Lys (Lys517 in gramicidin synthetase S[38]; Lys529 in firefly luciferase[39]) located in motif A10 within this subfamily. A highly conserved Thr (Thr190 in gramicidin synthetase S) is also located near the α phosphate in class I members. Similarly, both class II and III enzymes employ a conserved Arg. In class II members, this Arg is found in motif 2 of tRNA synthetases (Arg262 in lysine tRNA synthetase)[40], and or the conserved Arg305 in the class III NRPS-independent siderophore synthetase AcsD[41]. Along with the catalytically requisite divalent metal ion, these positively charged residues both enhance the electrophilicity of the α-phosphorus, and track the accumulation of negative charge through the pentavalent transition state to the pyrophosphate product. In AaBioW, Arg215 is poised to play an equivalent role in catalysis, and Ser182 likely serves a role equivalent to the conserved Thr in class I members. We show that single residue mutations of these residues compromises, but does not eliminate, catalytic activity and this may reflect the overlapping roles that these two residues, in conjunction with the requisite Mg^{2+} ion, play in catalysis.

A superposition of the various co-crystal structures illustrates that many of the active site Arg residues move in the thioester forming state (**SI Figure S4**), and this movement is necessary for

binding of CoA. Side chain movement may explain, in part, the observed greater activity of the Arg132→Ala variant, as Arg132 buttresses the position of Arg113 and Arg135. Movement of the latter two residues could only occur after Arg132 is shifted out of the way, and consequently, the Arg132→Ala variant may provide a more expedient means of accommodating the necessary shifts in Arg113, and Arg135. Although Arg132 engages the adenine 3'-phosphate of CoA, multiple other residues are also involved in interactions with the ligand. Biotin synthetic enzymes are generally poor catalysts[9], given the low physiological demand from biotin, and consequently, there would be no evolutionary pressure to produce a more efficient BioW. Hence, given the role in ligand binding, Arg132 is preserved amongst BioWs as mutations that may confer an increase in catalytic efficiency do not yield a biosynthetic competitive advantage.

Crystal structures of AaBioW with bound substrate reveal that the enzyme engages both carboxylates of the substrate through numerous interactions, using the side chain of Arg159 to establish the distance between the two carboxylates in the cognate pimelate (C7) substrate. Hence, it seemed plausible that single residues variants of AaBioW at residues that engage either of the two carboxylates in the substrate might utilize either monocarboxylate substrates or dicarboxylate substrates of varying length. To test this theory, we generate Ala mutants at Arg159, and Arg215 (which engage the α carboxylate) at Arg201 (which engages the ω carboxylate), and tested these variants against a panel of mono- and di-carboxylate substrates (**SI Figure S9, S10**). However, none of these variants demonstrated any activity against any substrate other than pimelate. Given that adipic acid (C6) does not occur naturally, it is not surprising that BioW does not show ligase activity for the molecule. However, both glutaric acid (C5), and suberic acid (C8) occur naturally and are byproducts of amino acid metabolism, and oxidation of castor oil, respectively. The lack of activity of either wild-type or variant AaBioWs against these other dicarboxylates suggest that the enzyme has been adapted to utilize strictly only the C7 dicarboxylate for CoA-ligation. This is an important property because biotin homologues having shortened or extended valerate moieties cannot replace biotin[42].

These above data provide some insights into a possible evolutionary role of this catalyst, with respect to class I enzymes that carry out the adenylation of fatty acid substrates. Massive domain movements occur during the transition from the adenylation to the transesterification half-

reactions in the class I catalysts[37]. In contrast, the various cocrystal structures of AaBioW show no change in the overall structure of the enzyme throughout the presumed catalytic cycle, other than movements of the side chains of various Arg residues for optimal contact with the respective substrate. Gulick has proposed that the domain alteration strategy utilized by class I adenylating enzymes was necessary as simple fatty acid substrates lack any additional functional groups that could be utilized to properly position the substrate during the adenylation reaction[22]. In such enzymes, substrate positioning during adenylation occurs through the carboxylate, and, domain movement would be necessary to allow access of the CoA thiol to the adenylate during the thioesterification step. In contrast, as the substrates for class II and class III adenylating enzyme contain additional functional groups (i.e. polar side chains and/or the α-amine of amino acids for class II members such as tRNA synthetases, additional carboxylates for class III NRPS-independent siderophore synthetases), positioning of the reactive carboxylate for adenylation can be achieved through interactions with these other groups, eliminating the need for a domain alteration strategy. The cocrystal structures of AaBioW presented here reinforce this hypothesis as the ω carboxylate provides an additional chemical handle for substrate positioning.

Biochemical studies of AaBioW recapitulate the unexpected observation that the enzyme can hydrolyze adenylates of non-cognate dicarboxylates, and this proof reading activity was previously observed for *B. subtilis* BioW[20]. As proofreading of the misactivated dicarboxylate occurs after formation of the adenylate, the activity is analogous to the pretransfer proofreading activities observed in aminoacyl tRNA synthetases. Our structural and biochemical data demonstrate that this activity does not require a second active site, as is the case with several tRNA synthetases, where proofreading activity occurs in a region of the polypeptide distinct from where adenylation is catalyzed[43]. However, proofreading activities are also observed in tRNA synthetases that lack obvious proofreading domains, including MetRS[44], LysRS-II[45], SerRS[46], and yeast mitochondrial ThrRS[47], each of which use unique strategies to avoid misactivation. For example, production of the non-cognate homocysteinyl-adenylate in the MetRS active site results in spontaneous intramolecular cyclization to produce the nonproductive thiolactone[48], and likewise misadenylation of homoserine or ornithine in the LysRS-II active site results in cyclization into the corresponding lactone or lactam, respectively[45]. The proofreading activity of BioW is most comparable to that observed in SerRS[46] and mitochondrial ThrRS

(mThrRS)[49], wherein the enzyme active site can actively, and directly catalyze the hydrolysis of misactivated aminoacyl-adenylate. Biochemical studies of mThrRS with the misactivated Ser adenylate show that the enzyme catalyzes the selective and preferential hydrolysis of Ser-AMP, and the cocrystal structure of mThrRS with an inert analog of Ser-AMP reveal conformational changes at the active site, relative to the Thr-AMP structure, that repositions a solvent molecule for attack on the misactivated Ser-adenylate[47]. In the case of BioW, pH stability experiments demonstrate that the non-cognate adenylates are not as easily hydrolyzed as cognate adenylates (**SI Figure S2**), and such differences may facilitate the proofreading activity of AaBioW. Analysis of the Arg159→Ala variant is consistent with at least a partial role for this residue in proofreading, perhaps through electronic or substrate orientation effects.

Prior sequence-based searches against various databases failed to identify any recognizable motif in *B. subtilis* BioW[20], which is surprising in light of the fact that the enzyme uses two cofactors (CoA and ATP) for which there is extensive literature on structure-based classifications of binding motifs[50]. The structural and biochemical data presented here demonstrate that the structure of BioW consists of a Rossmann fold observed in the amino-terminal domain (3',5')-di-cAMP cyclase DisA appended to a P_{II} fold to yield a new class of adenylating enzyme. The diadenylate cyclase active site of DisA is comprised of two different subunits, which results in proper positioning of two molecules of ATP (i.e. one from each monomer) to facilitate catalysis. The orientation of the AaBioW P_{II} domain superimposes with the location of the second monomer of the DisA heterodimer. Consequently, the CoA substrate bound at the AaBioW P_{II} domain is positioned for the nucleophilic attack of its thiol onto the pimeloyl-adenylate. In light of the small size of the protein, it is surprising that multiple activities, including the capability of proofreading misadenylated substrates, can be encoded within the BioW fold. Nature has achieved this feat through judicious use of the flexibility of multiple Arg residues to provide the necessary interactions to facilitate productive catalysis. The strategy is successful only because the chemical makeup of the substrate, which provides unconstrained flexibility, along with dual carboxylate functional groups as chemical handles to provide a straightforward means to achieve an otherwise elaborate catalytic scheme.

Acknowledgements

We thank the Mining Microbial Genomes theme at the Carl R. Woese Institute for Genomic Biology for access to the LC-MS equipment. We also thank Keith Brister and the staff at the Life Sciences Collaborative Access Team (Sector 21) at the Argonne National Labs for facilitating data collection. This work was supported by NIH grant AI15650 (to J.C.).

REFERENCES

1. McMahon, R.J. Biotin in metabolism and molecular biology. *Annu Rev Nutr* **22**, 221-39 (2002).

2. DeTitta, G.T., Edmonds, J.W., Stallings, W. & Donohue, J. Molecular structure of biotin. Results of two independent crystal structure investigations. *J Am Chem Soc* **98**, 1920-6 (1976).

3. Chapman-Smith, A. & Cronan, J.E., Jr. Molecular biology of biotin attachment to proteins. *J Nutr* **129**, 477S-484S (1999).

4. Tong, L. Structure and function of biotin-dependent carboxylases. *Cell Mol Life Sci* **70**, 863-91 (2013).

5. Streit, W.R. & Entcheva, P. Biotin in microbes, the genes involved in its biosynthesis, its biochemical role and perspectives for biotechnological production. *Appl Microbiol Biotechnol* **61**, 21-31 (2003).

6. Eisenberg, M.A. & Star, C. Synthesis of 7-oxo-8-aminopelargonic acid, a biotin vitamer, in cell-free extracts of Escherichia coli biotin auxotrophs. *J Bacteriol* **96**, 1291-7 (1968).

7. Eisenberg, M.A. & Krell, K. Dethiobiotin synthesis from 7,8-diaminolargonic acid in cell-free extracts of a biotin auxotroph of Escherichia coli K-12. *J Biol Chem* **244**, 5503-9 (1969).

8. Eisenberg, M.A. & Krell, K. Synthesis of desthiobiotin from 7,8-diaminopelargonic acid in biotin auxotrophs of Escherichia coli K-12. *J Bacteriol* **98**, 1227-31 (1969).

9. Lin, S. & Cronan, J.E. Closing in on complete pathways of biotin biosynthesis. *Mol Biosyst* **7**, 1811-21 (2011).

10. Cronan, J.E. & Lin, S. Synthesis of the alpha,omega-dicarboxylic acid precursor of biotin by the canonical fatty acid biosynthetic pathway. *Curr Opin Chem Biol* **15**, 407-13 (2011).

11. Ifuku, O. et al. Origin of carbon atoms of biotin. 13C-NMR studies on biotin biosynthesis in Escherichia coli. *Eur J Biochem* **220**, 585-91 (1994).

12. Sanyal, I., Lee, S.-L. & Flint, D.H. Biosynthesis of pimeloyl-CoA, a biotin precursor in Escherichia coli, follows a modified fatty acid synthesis pathway: 13C-labeling studies. *J. Am. Chem. Soc.*, 2637-2638 (1994).

13. Lin, S., Hanson, R.E. & Cronan, J.E. Biotin synthesis begins by hijacking the fatty acid synthetic pathway. *Nat Chem Biol* **6**, 682-8 (2010).

14. Agarwal, V., Lin, S., Lukk, T., Nair, S.K. & Cronan, J.E. Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proc Natl Acad Sci U S A* **109**, 17406-11 (2012).

15. Bower, S. et al. Cloning, sequencing, and characterization of the Bacillus subtilis biotin biosynthetic operon. *J Bacteriol* **178**, 4122-30 (1996).

16. Cryle, M.J. & De Voss, J.J. Carbon-carbon bond cleavage by cytochrome p450(BioI)(CYP107H1). *Chem Commun (Camb)*, 86-7 (2004).

17. Stok, J.E. & De Voss, J. Expression, purification, and characterization of BioI: a carbon-carbon bond cleaving cytochrome P450 involved in biotin biosynthesis in Bacillus subtilis. *Arch Biochem Biophys* **384**, 351-60 (2000).

18. Cryle, M.J. & Schlichting, I. Structural insights from a P450 Carrier Protein complex reveal how specificity is achieved in the P450(BioI) ACP complex. *Proc Natl Acad Sci U S A* **105**, 15696-701 (2008).

19. Ploux, O., Soularue, P., Marquet, A., Gloeckler, R. & Lemoine, Y. Investigation of the first step of biotin biosynthesis in Bacillus sphaericus. Purification and characterization of the pimeloyl-CoA synthase, and uptake of pimelate. *Biochem J* **287 (Pt 3)**, 685-90 (1992).

20. Manandhar, M. & Cronan, J.E. Proofreading of noncognate acyl adenylates by an acyl-coenzyme a ligase. *Chem Biol* **20**, 1441-6 (2013).

21. Schmelz, S. & Naismith, J.H. Adenylate-forming enzymes. *Curr Opin Struct Biol* **19**, 666-71 (2009).

22. Gulick, A.M. Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem Biol* **4**, 811-27 (2009).

23. Gerlt, J.A. et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* **1854**, 1019-37 (2015).

24. Ploux, O. & Marquet, A. Mechanistic studies on the 8-amino-7-oxopelargonate synthase, a pyridoxal-5'-phosphate-dependent enzyme involved in biotin biosynthesis. *Eur J Biochem* **236**, 301-8 (1996).

25. Webster, S.P. et al. Characterisation of 8-amino-7-oxononanoate synthase: a bacterial PLP-dependent, acyl CoA condensing enzyme. *Biochem Soc Trans* **26**, S268 (1998).

26. Holm, L. & Park, J. DaliLite workbench for protein structure comparison. *Bioinformatics* **16**, 566-7 (2000).

27. Berman, H.M. et al. The Protein Data Bank. *Nucleic Acids Res* **28**, 235-42 (2000).

28. Osanai, T. & Tanaka, K. Keeping in touch with PII: PII-interacting proteins in unicellular cyanobacteria. *Plant Cell Physiol* **48**, 908-14 (2007).

29. Forchhammer, K. & Luddecke, J. Sensory properties of the PII signalling protein family. *FEBS J* **283**, 425-37 (2016).

30. Fawaz, M.V., Topper, M.E. & Firestine, S.M. The ATP-grasp enzymes. *Bioorg Chem* **39**, 185-91 (2011).

31. Rossmann, M.G. & Argos, P. The taxonomy of binding sites in proteins. *Mol Cell Biochem* **21**, 161-82 (1978).

32. Cheek, S., Zhang, H. & Grishin, N.V. Sequence and structure classification of kinases. *J Mol Biol* **320**, 855-81 (2002).

33. Moras, D. Structural and functional relationships between aminoacyl-tRNA synthetases. *Trends Biochem Sci* **17**, 159-64 (1992).

34. Witte, G., Hartung, S., Buttner, K. & Hopfner, K.P. Structural biochemistry of a bacterial checkpoint protein reveals diadenylate cyclase activity regulated by DNA recombination intermediates. *Mol Cell* **30**, 167-78 (2008).

35. Martinez, S.E., Heikaus, C.C., Klevit, R.E. & Beavo, J.A. The structure of the GAF A domain from phosphodiesterase 6C reveals determinants of cGMP binding, a conserved binding surface, and a large cGMP-dependent conformational change. *J Biol Chem* **283**, 25913-9 (2008).

36. Prakash, S., Johnson, R.E. & Prakash, L. Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem* **74**, 317-53 (2005).

37. Gulick, A.M., Starai, V.J., Horswill, A.R., Homick, K.M. & Escalante-Semerena, J.C. The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry* **42**, 2866-73 (2003).

38. Conti, E., Stachelhaus, T., Marahiel, M.A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* **16**, 4174-83 (1997).

39. Branchini, B.R., Murtiashaw, M.H., Magyar, R.A. & Anderson, S.M. The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase. *Biochemistry* **39**, 5433-40 (2000).

40. Onesti, S., Miller, A.D. & Brick, P. The crystal structure of the lysyl-tRNA synthetase (LysU) from Escherichia coli. *Structure* **3**, 163-76 (1995).

41. Schmelz, S. et al. AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis. *Nat Chem Biol* **5**, 174-82 (2009).

42. Balcher, M.R. & Lichstein, H.C. Growth promotion and antibiotin effect of homobiotin and norbiotin. *J Bacteriol* **58**, 579-83 (1949).

43. Yadavalli, S.S. & Ibba, M. Quality control in aminoacyl-tRNA synthesis its role in translational fidelity. *Adv Protein Chem Struct Biol* **86**, 1-43 (2012).

44. Fersht, A.R. & Dingwall, C. An editing mechanism for the methionyl-tRNA synthetase in the selection of amino acids in protein synthesis. *Biochemistry* **18**, 1250-6 (1979).

45. Jakubowski, H. Misacylation of tRNALys with noncognate amino acids by lysyl-tRNA synthetase. *Biochemistry* **38**, 8088-93 (1999).

46. Gruic-Sovulj, I., Rokov-Plavec, J. & Weygand-Durasevic, I. Hydrolysis of non-cognate aminoacyl-adenylates by a class II aminoacyl-tRNA synthetase lacking an editing domain. *FEBS Lett* **581**, 5110-4 (2007).

47. Ling, J., Peterson, K.M., Simonovic, I., Soll, D. & Simonovic, M. The mechanism of pre-transfer editing in yeast mitochondrial threonyl-tRNA synthetase. *J Biol Chem* **287**, 28518-25 (2012).

48. Jakubowski, H. & Fersht, A.R. Alternative pathways for editing non-cognate amino acids by aminoacyl-tRNA synthetases. *Nucleic Acids Res* **9**, 3105-17 (1981).

49. Minajigi, A. & Francklyn, C.S. Aminoacyl transfer rate dictates choice of editing pathway in threonyl-tRNA synthetase. *J Biol Chem* **285**, 23810-7 (2010).

50. Kinjo, A.R. & Nakamura, H. Comprehensive structural classification of ligand-binding motifs in proteins. *Structure* **17**, 234-46 (2009).

FIGURE LEGENDS

Figure 1. Reaction catalyzed by BioW and representative homologs. Convergent biosynthetic pathways for the incorporation of the pimelate dicarboxylate into biotin. The pathway highlighted in the rectangle illustrates the direct ligation of pimelic acid to CoA via a two-step reaction catalyzed by BioW.

Figure 2. Biochemical activity of AaBioW. (A) Formation of pimeloyl-CoA by AaBioW analyzed by HPLC analysis (black trace). The elution profiles for the isolated standards are shown in the panels above. (B, C) Binding isotherms from ITC measurement of AaBioW binding to ATP and CoA, respectively. (D) Michaelis-Menten curve obtained by measuring the AMP production over varying concentrations of pimelate with fixed concentrations of ATP (0.4 mM) and CoA (0.3 mM). Measurements were conducted in triplicate. (E) The proofreading activity of AaBioW was monitored using a range of dicarboxylic acids, including the cognate (C7) pimelate. The lanes marked -E are reactions that lacked enzyme.

Figure 3. Crystal structures of AaBioW and ligand complexes. (A) Ribbon diagram of AaBioW illustrating the orientation of the P_{II} domain (in pink) and the Rossmann fold domain (in light blue). Secondary structural elements are demarcated. (B) Topology diagram showing the overall fold of BioW. (C-E) Simulated annealing difference Fourier maps (Fo-Fc) of AaBioW complexes contoured to 2.5σ (blue) showing the bound (C) pimelate, (D) AMP-PCP+Mg^{2+}, and (E) AMP+CoA. The coordinates for ligand were omitted prior map calculations. The coordinates of the complexes are superimposed and the ligands are shown in ball-and-stick representation.

Figure 4. Structure-based mutational analysis of the AaBioW active site. (A, B) Model of the AaBioW active site structure during the (A) adenylation, and (B) thioester formation steps, generated by superimposing the crystal structures of relevant ligand-bond complexes. The nucleotide is shown as green sticks, with the pimelate shown in purple, and CoA in yellow. Active site residues that may play a role in catalysis are shown as tan sticks. (C) Biochemical activities of site-specific variants of active site residues in AaBioW identified in panels A and B. The efficiency of each variant is measured as amount of pimeloyl-CoA formed, relative to the wild-type enzyme.

Figure 5. Proofreading activity of the wild type and Arg159→Ala AaBioWs. (A) Surface representation of the AaBioW active site based on the model from Figure5B. The P_{II} and the Rossmann fold domains are shown as pink and blue surfaces, respectively, and CoA (yellow) and pimeloyl-AMP (green) are shown as sticks. (B) Although the Arg159→Ala variant can no longer proofread, the enzyme still retains ligase activity and can catalyze the formation of pimeloyl-CoA. (C-F) HPLC traces (absorbance at 254 nM) traces for AaBioW Arg159→Ala mutant with different mono- and di-acid substrates.

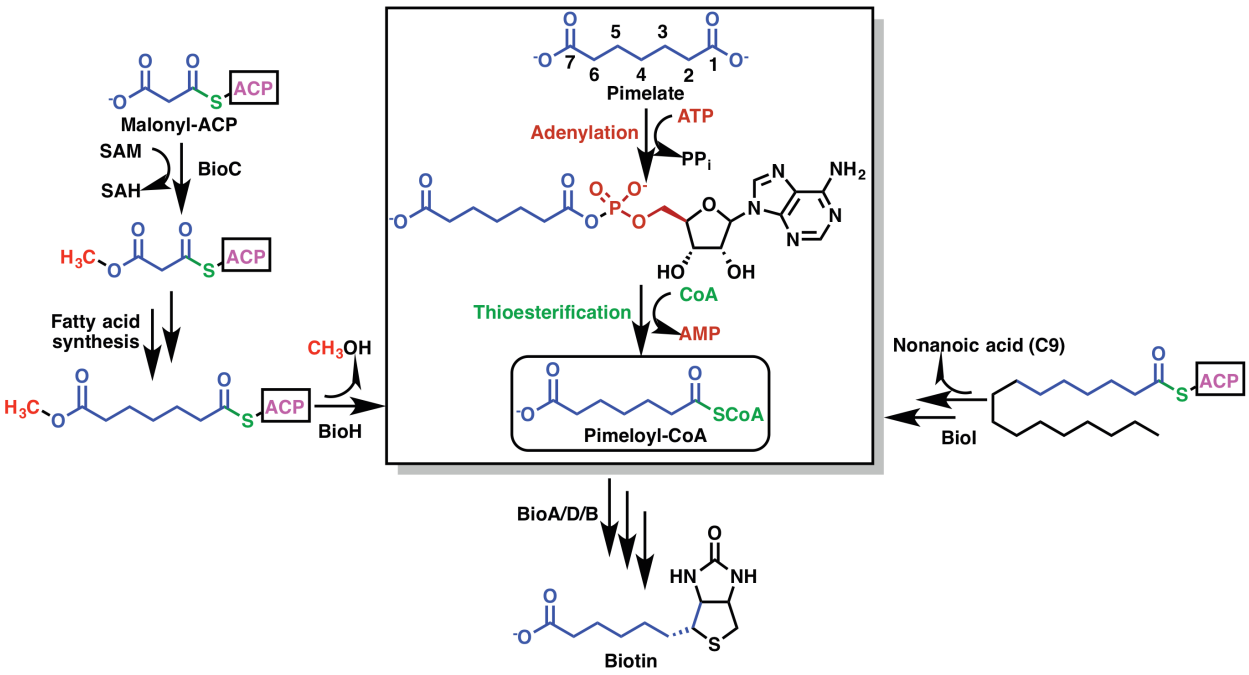

Figure 1. Estrada et al., 2016.

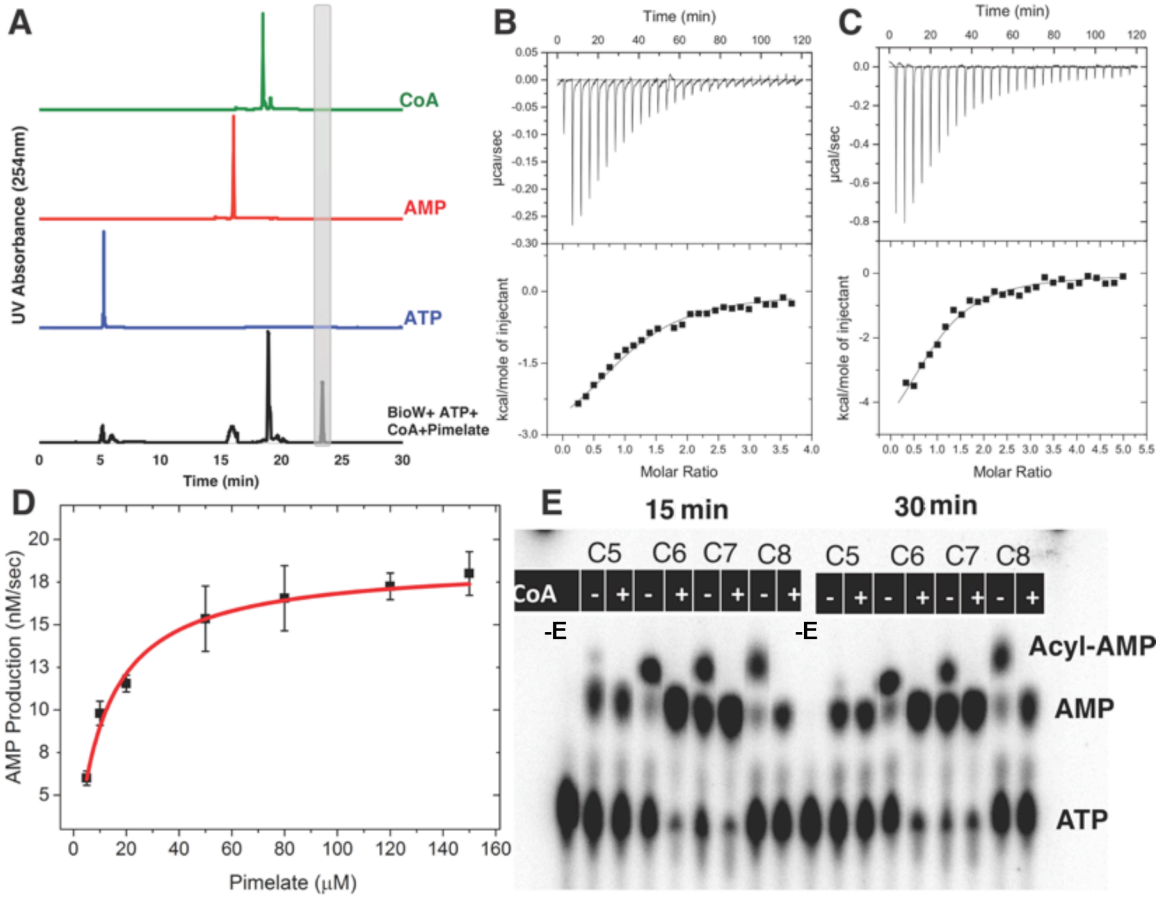

Figure 2. Estrada et al., 2016.

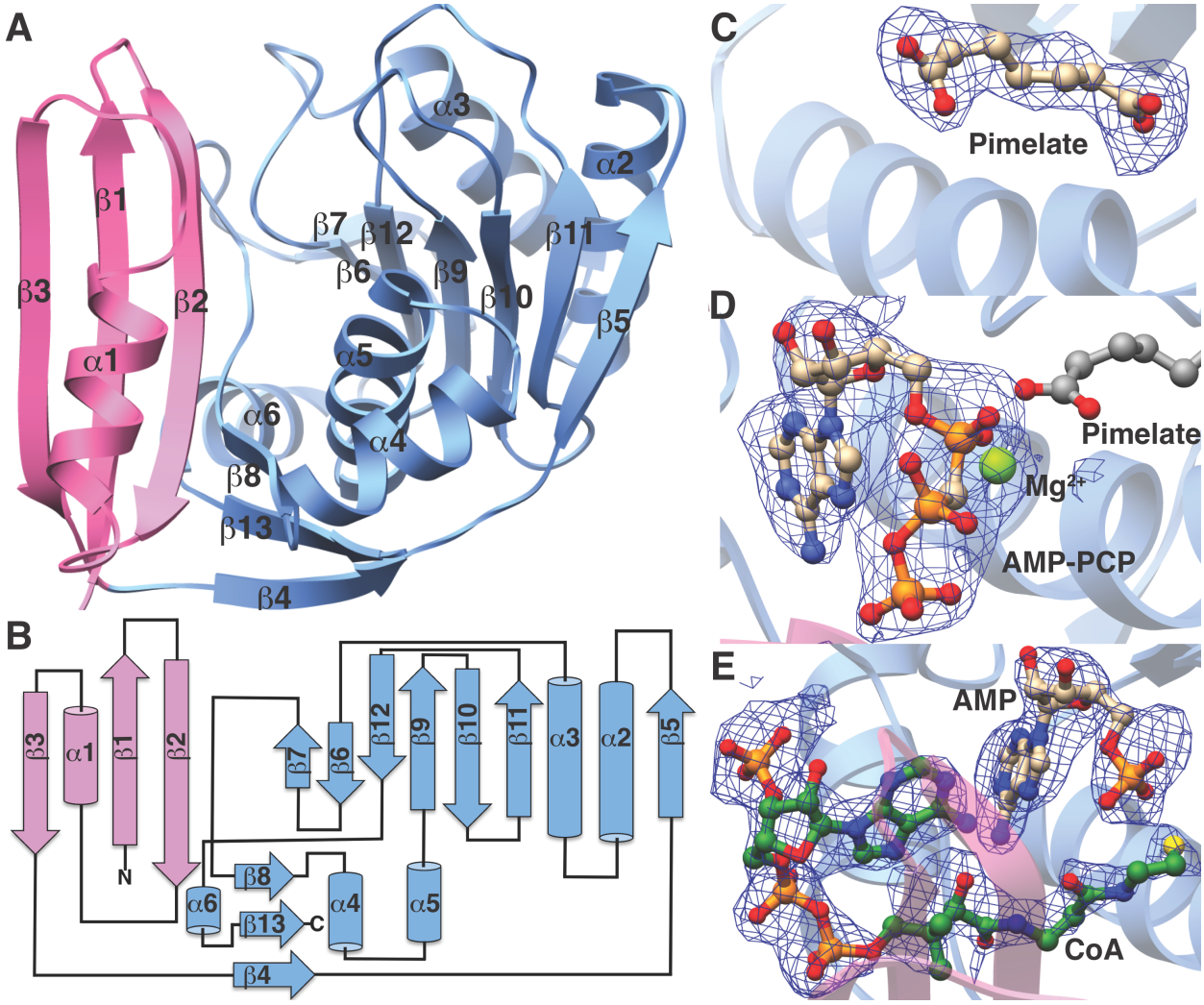

Figure 3. Estrada et al., 2016.

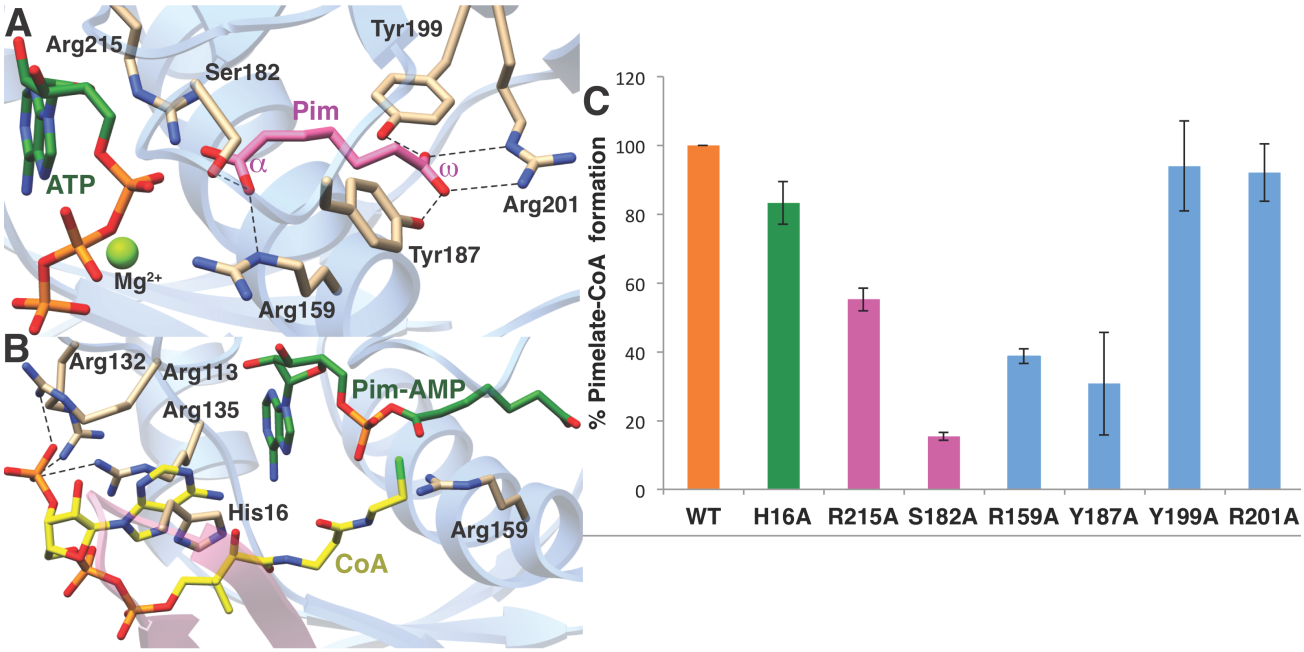

Figure 4. Estrada et al., 2016.

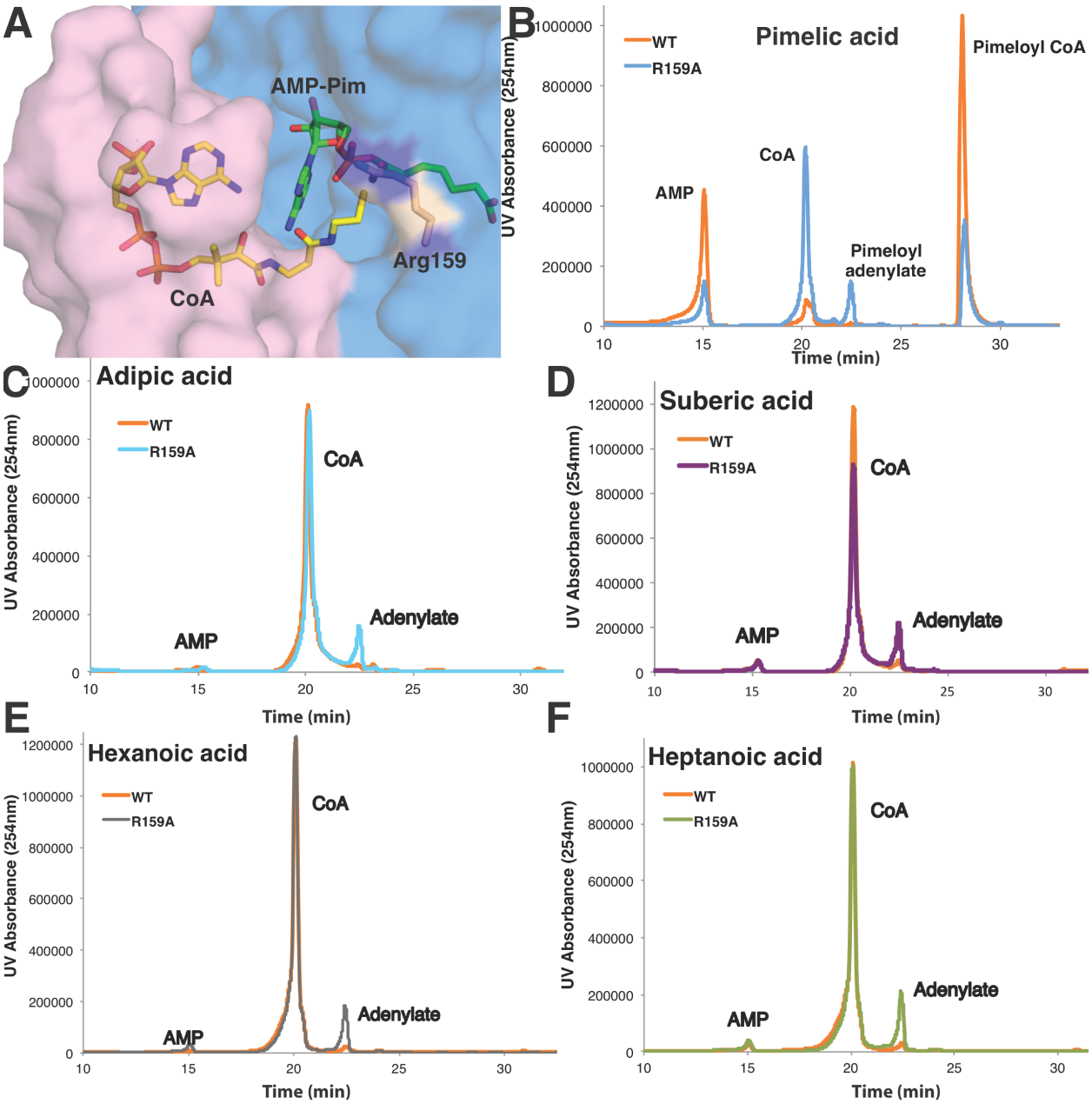

Figure 5. Estrada et al., 2016.