Using the pimeloyl-CoA synthetase adenylation fold to synthesise fatty acid thioesters

Menglu Wang[1a]; Lucile Moynié[2a], Peter J. Harrison[1], Van Kelly[1], Andrew Piper[1], James H. Naismith[2, 3*] and Dominic J. Campopiano[1*]

1 EastChem School of Chemistry, David Brewster Road, University of Edinburgh, Edinburgh, EH9 3FJ.

2 Biomedical Sciences Research complex, University of St. Andrews, St Andrews, KY16 9ST

3 State Key Laboratory of Biotherapy and Cancer Center, West China Hospital of Sichuan University and Collaborative Innovation Center of Biotherapy and Cancer, Chengdu 610041, China

a equal first authors

* joint corresponding authors; naismith@st-andrews.ac.uk, Dominic.Campopiano@ed.ac.uk.

Abstract

Biotin is an essential vitamin in plants and mammals functioning as the carbon dioxide carrier within central lipid metabolism. Bacterial pimeloyl-CoA synthetase (BioW) acts as a highly specific substrate selection gate ensuring the integrity of the carbon chain in biotin synthesis. BioW catalyses the condensation of pimelic acid (C7 dicarboxylic acid) with CoASH in an ATP dependent manner to form pimeloyl-CoA, the first dedicated biotin building block. Multiple structures of *Bacillus subtilis* BioW together capture all three substrates as well as the intermediate pimeloyl-adenylate and product pyrophosphate (PP_i), and indicates that the enzyme uses an internal ruler to select the correct dicarboxylic acid substrate. Both the catalytic mechanism and the surprising stability of the adenylate intermediate were rationalized through site-directed mutagenesis. Building on this understanding, BioW was engineered to synthesise high value heptanoyl (C7) and octanoyl (C8) mono carboxylic acid-CoA and C8 dicarboxylic-CoA products, highlighting the enzyme's synthetic potential.

INTRODUCTION

In fatty acid biosynthesis biotin is an essential carbon dioxide carrier cofactor within large enzyme complexes, such as acetyl-CoA carboxylase (ACC), that catalyse carboxylation of acetyl-CoA to form malonyl-CoA [1]. Biotin (**Fig. 1a**) possesses two fused rings (ureido and thiophane) whose carbon skeleton derives from the unusual C7 α,ω-dicarboxylic acid, pimelic acid (**1**). Pioneering studies identified the intermediates involved in forming the fused bicyclic structure that have been shown to be conserved in all bacterial species [2-5]. *E. coli* has proved a useful model organism to study biotin biosynthesis and sequencing of the genomic locus, the "bio operon", revealed five genes (bioA, bioB, bioF, bioC and bioD) involved in biotin formation [6]. Four of the five genes encode highly conserved enzymes that catalyse the conversion of pimeloyl-CoA to biotin (8-amino-7-oxononanoate synthase (AONS, bioF gene product), diaminononanoate synthase (DANS, bioA gene product), dethiobiotin synthetase (DTBS, bioD gene product) and biotin synthase (BS, bioB gene product, **Fig. 1a**) [7]. The x-ray structures and catalytic mechanisms of the *E. coli* AONS, DANS, DTBS and BS enzymes have been reported [8-13].

In contrast to the early biosynthetic steps, the origin of pimelic acid varies across bacterial species, and has been found in various forms: as a free di-acid, a mono-methyl ester, and transformed into a pimeloyl thioester by attachment to either coenzyme A (CoASH) or an acyl carrier protein (ACP) [14]. The thioester is condensed with L-alanine in an AONS-catalysed reaction to begin the late stages of biotin biosynthesis [8, 13]. Labelling studies using ^{13}C acetate provided insight into the origin of the carbon atoms of pimeloyl-CoA and biotin in *E. coli* [15, 16]. Two enzymes from the biotin pathway, BioC (a SAM-dependent methyltransferase) and BioH (an esterase), hijack fatty acid biosynthesis by methylating malonyl-ACP, then extending it through two rounds of the fatty acid synthase cycle [17]. The

crystal structure of an inactive BioH S82A mutant:pimeloyl-ACP methyl ester complex revealed the origin of the acyl-chain length specificity [18].

In *Bacillus subtilis*, the presence of genes encoding BioI and BioW enzymes has suggested that this microbe makes and processes the pimelic acid precursor differently from *E. coli* [19]. BioI shows homology to cytochrome P450 enzymes but is able to complement biotin production in bioC or bioH mutant *E. coli* strains. Subsequently, crystal structures of BioI captured in complex with acyl-ACPs of varying chain lengths have provided a mechanistic understanding of how heme chemistry is used to act upon the acyl chain and generate pimelic acid [20, 21]. BioW from *Bacillus sphaericus* has been characterised to have pimeloyl-CoA synthetase activity (EC number 6.2.1.14) [22, 23]. This homolog displays high sequence identity to the predicted enzyme in *B. subtilis* so it is somewhat surprising that this enzyme was not characterised until recently [24]. BioW uses pimelic acid and MgATP to generate a pimeloyl-adenylate intermediate which subsequently reacts with CoASH to give the acyl-CoA thioester product (**Fig. 1a**). The BioW reaction is highly specific to pimelic acid; mono- and di-acids of different chain lengths are either not processed at all or are hydrolysed at the acyl-adenylate stage in a proofreading manner [24]. Whilst this ensures that only the correct vitamin is made *in vivo* rather than potentially toxic analogues with shorter and/or longer chains, it represents an unfortunate limitation for biotechnological applications. It also suggests that BioW has been under great selective pressure to accept only pimelic acid, since biotin vitamers with different chain lengths have never been isolated.

The activation of carboxylic acids as adenylates is a common strategy in nature and exemplified in the ANL superfamily of adenylating enzymes that includes the acyl-CoA synthetases [25, 26]. The carboxylate of biotin is also adenylated by a biotin protein ligase (BPL) for final transfer to the lysine side chain of carboxylases [27]. The mechanism involves the attack of the carboxylate at the α-phosphate of ATP to form the adenylate, driven by

pyrophosphate (PP_i) elimination (**Fig. 1a**). The proposed transition state is the negatively charged, penta-coordinate trigonal bipyramidal phosphorous atom, which is common to other triphosphate nucleotide utilising enzymes including kinases and DNA polymerases [26]. Interestingly a structural link between kinases and one type of adenylase has been identified [28]. Sequence analysis of the *B. subtilis* BioW showed no homology with other known adenylases or ATP-dependent synthetases, suggesting it may represent another member of the adenylase enzyme superfamily.

Here we present structural and mechanistic studies of *B. subtilis* BioW and reveal that it extends the number of folds capable of catalysing adenylation. It is the C-terminal domain that recognises pimelic acid with exquisite specificity, thus controlling entry into the biotin biosynthetic pathway. Structure-guided mutagenesis was used to generate an enzyme that can processes a range of fatty acid substrates, unlocking the synthetic potential of this enzyme for the production of acyl-CoAs. The ability to generate acyl-CoA thioesters through biocatalytic methods would be extremely valuable since they are used as substrates in studies of important metabolic enzymes such as those involved in fatty acid biosynthesis [29].

RESULTS

B. subtilis BioW isolation and assay

In contrast to previous studies [24] we found that a 6xHis-tag positioned at the N-terminus of *B. subtilis* BioW (**Supplementary Results, Supplementary Fig. 1a, b and c**), allowed the isolation of milligram amounts of active enzyme. By HPLC and mass spectrometry (**Supplementary Fig. 2**) we confirmed that BioW converted pimelic acid (C7 dicarboxylic acid, **1**), MgATP and CoASH to pimeloyl-CoA (**2**), eluting at 16.2 minutes, with a mass of 910.1865 Da ($M+H^+$, expected 910.1855 for $C_{28}H_{46}N_7O_{19}P_3S$) and AMP. For all crystallisation, native mass spectrometry and substrate specificity assays the His-tag was removed and the resultant enzyme was a homodimer (**Supplementary Fig. 1c**).

For enzyme kinetics we modified a commercially-available (ENZCHECK), coupled assay (pyrophosphatase (PPase), purine nucleoside phosphorylase (PNP)) which links PP_i formation in the first half of the reaction to the formation of a colorimetric product detectable at 360 nm (**Supplementary Fig. 3** [30, 31]). We also monitored acyl-CoA thioester bond formation at 230 nm [23] and observed pimeloyl-CoA (**2**) product formation by HPLC and mass spectrometry. The coupled assay proved to be more useful than the 230 nm assay due to lower background absorbance. We used the coupled assay to determine the kinetic parameters for the BioW substrates pimelic acid (K_m = 70.5 ± 6.8 μM; k_{cat} = 0.48 ± 0.02 s^{-1}) and ATP (299.6 ± 37.6 μM, k_{cat} = 0.44 ± 0.03 s^{-1}) and the 230 nm assay used to measure the value for CoASH (229.3 ± 25.4 μM, k_{cat} = 0.87 ± 0.05 s^{-1}) (**Supplementary Fig. 4**, **Supplementary Table 1**). We observed substantially higher enzyme activity than previously reported for pimelic acid (~0.480 s^{-1} compared with 8.4 × 10^{-5} s^{-1}) [24] and our results are comparable with that observed for the *B. sphaericus* BioW; K_m pimelate (145 μM), ATP (170 μM), CoASH (33 μM) and a k_{cat} of 0.4 s^{-1} [23]. We also used the coupled assay to

determine relative activities (based on the k_{cat}) of dicarboxylic acid substrates of various carbon chain lengths (C5, glutaric acid (**3**); C6, adipic acid (**4**); C8, suberic acid (**5**); C9, azelaic acid (**6**)). This confirmed that the wild-type enzyme is highly specific for the native substrate since the k_{cat} values (0.0051 s^{-1} for **3**, 0.0066 s^{-1} for **4**, 0.0038 s^{-1} for **5** and 0.0001 s^{-1} for **6**) were all ≤1% compared with pimelic acid (0.480 s^{-1}) (**Fig. 1b and Supplementary Table 2**). Importantly, very little activity (with large errors) was detected with both the monocarboxylate substrates heptanoic acid (**7**, 0.0016 s^{-1}) and octanoic acid (**8**, 0.0020 s^{-1}) with wild-type BioW **(Fig. 1b and Supplementary Table 2**).

Structural analysis of BioW

The asymmetric unit of BioW is a homodimer (**Fig. 2a**), and analysis with the PISA server [32] identified that the observed dimer is stable, burying approximately 900 $Å^2$ of surface area per monomer. The crystal structure was consistent with native mass spectrometry and gel filtration experiments. Each monomer consists of two domains, a small 70 residue N-terminal domain and a larger C-terminal domain (**Fig. 2b**). The N-terminal domain has a three-stranded anti-parallel β-sheet (β1-3) with an α-helix lying on one face. The dimer interface is confined to the N-terminal domain that brings together the β3 strands of each monomer such that the dimer has a 6 stranded antiparallel β-sheet. In addition to these main chain hydrogen bonds, there are a few additional polar and hydrophobic interactions that involve the α-helix (α1). Superimposing monomers reveals that the relative orientation of domains differs within each monomer, suggesting the linkage between the domains is flexible. The C-terminal domain has a central, predominantly anti-parallel seven-stranded β-sheet; each face of the sheet has two α-helices and one of the helices (α5) stacks against a small two stranded parallel β-sheet (β4 and 13) (**Fig. 2b**). The domain has a fifth helix (α4) that packs against

the loop of the C-terminus. Analysis with PDB fold does not identify other enzymes with this fold, suggesting the structure represents an extension to the ANL family of enzymes known to catalyse the adenylation reaction [26]. This family currently contains three classes (I-III) and so we propose BioW as the first representative of class IV.

Two co-complexes, pimeloyl-adenylate plus PPi (**Fig. 2c and Supplementary Fig. 5a**) and AMP-PNP plus pimelic acid (**Supplementary Fig. 5b and 5c**) localize the active site almost entirely within the C-terminal domain and identify the residues in contact with ATP and pimelic acid. Although the AMP-PNP could be expected to be active (since the enzyme breaks the α-β not the β-γ bond) the inactivity of this derivative has been observed before [33]. In both complexes the adenosine and ribose rings, the β and γ phosphates and the two Mg^{2+} ions are located in essentially identical positions making the same contacts with the protein (**Fig. 2d** and **2e** and **Supplementary Fig. 5a, 5b, 5d and 5e**). The adenosine ring sits in a pocket with the side chain of Arg145 stacking on its face; the ring itself makes two hydrogen bonds to the main chain at Val146, a hydrogen bond to Glu31 and to a network of water molecules. The ribose ring hydrogen bonds to Arg227, the main chain of Val122 and to the network of water molecules. Comparing the atoms at the α phosphate in both complexes reveals that the phosphorous atom has shifted less than 1.0 Å and the three oxygen atoms that remain are unaffected by bond breaking/making and have shifted less than 0.2 Å. Two of these oxygen atoms are each bound to a different Mg^{2+} ion, the third oxygen atom which connects to the ribose ring is salt bridged to Arg227.

Combining the complexes reveals that the α oxygen-phosphorous bond which connects to the β-phosphate makes an angle of 180° with the oxygen atom that links the α phosphorous to the carboxylate of pimelic acid (**Fig. 2d**). From these structures, a model of the penta-covalent bipyramidal transition state can be constructed. The carboxyl-adenylate bond sits

between Ser194 on one side and on the other side two water molecules (denoted W1 and W2) which are hydrogen bonded to each other (**Fig. 2e**). The Ser194 side chain is hydrogen bonded to the protein main chain. The hydrogen bond network of these water molecules is extensive but neither is ideally positioned to attack the carbonyl. W1 is bound to the guanidyl group of Arg227, which would be predicted to lower the nucleophilicity of the water, whilst W2 is hydrogen-bonded to W1 and would require some re-arrangement to be correctly positioned for attack. In the adenylate complex Arg170 is hydrogen bonded to the pimelic acid carboxylate oxygen atom that is not linked to the phosphorous. In contrast, in the enzyme complex with AMP-PNP and pimelic acid, the carboxylate group is in a different orientation and not aligned for attack at the α phosphate (**Supplementary Fig. 5b**). The carboxylate makes a different salt bridge to Arg170 and hydrogen bonds to Ser194 as well as water molecules ligated to the Mg^{2+} ions. The β and γ phosphate groups bind two Mg^{2+} ions, the β phosphate makes interactions with the N-terminal domain side chains of His26 and Arg145 and the γ phosphate makes contacts with the side chains of Lys49 and His53. The Mg^{2+} ions both have octahedral coordination and are ligated by three phosphate oxygen atoms. One Mg^{2+} ion has three bound water molecules and the other Mg^{2+} ion binds to the side chains of Asp195, Asp196 and one water to complete the coordination shell.

Pimelic acid binds in an elongated tunnel (**Supplementary Fig. 5d and 5e**) and the carboxylate group at the C7 (ω) position, that does not react with ATP is anchored in both structures by a salt bridge to Arg213, as well as forming hydrogen bonds to two tyrosine side-chains (Tyr199 and Tyr211, **Figs. 2e** and **Supplementary Fig. 5d**). There are some small differences in the precise conformation of the alkyl chains between the two structures that arise from the different location of the reactive carboxylate (C1, α) in the two subunits. Despite this, in both structures the alkyl chain makes contacts (< 4.0 Å) along its length with

the enzyme such that any substitution of hydrogens with larger heteroatoms would result in clashes.

A crystal of the enzyme with CoASH and pimelic acid has density for two pimelic acid molecules but only one CoASH (**Fig. 2f and Supplementary Fig. 5f and 5g**). There is however no reliable density for the pantetheine arm (only the adenine, ribose and three phosphates clearly defined). CoASH is bound close to the dimer interface where it interacts with the strands of the N-terminal domain. The adenine ring makes a hydrogen bond to Glu66 from the other subunit, whilst the 3-phosphate is bound by a cluster of three Arg residues (Arg13, Arg123, Arg145). A surface plot (**Supplementary Fig. 5h**) of the complex reveals two possible tunnels that reach from the CoASH binding sites to the pimelic acid. One tunnel is formed by residues Arg11, Arg13, Glu31, Arg123 and Arg145, whilst the other is formed by His26, Arg55, Asp195. Although the ordered pyrophosphate group of CoASH points towards the first of these tunnels, the disorder of the pantothenate arm and the fact that the complex was obtained by soaking (thus restricting any conformational change) precludes a definitive assignment of the correct tunnel. The 5-pyrophosphate group does not make any hydrogen bonds. We suggest that due to the two fold non-crystallographic axis, were a second molecule of CoASH to be bound, the terminal 5′ phosphates of the CoASH would be closer than 4.0 Å.

Native mass spectrometry analysis of BioW

We used native ESI-MS analysis to detect BioW-bound intermediates [34] and observed charge state distributions consistent with monomeric and dimeric protein (**Fig. 3a**). Deconvolution of the +14 charge state is consistent with the mass of the ligand-free BioW dimer (**Fig 3a**,**b**, observed *m/z* = 4233.9, theoretical *m/z* = 4233.6).

Upon addition of pimelic acid and MgATP, we observed two species, a dominant mass of *m/z* = 4304.0 consistent with the formation of a BioW dimer: pimeloyl-adenylate complex with 2:2 stoichiometry (mass difference 980.7 Da, **Fig. 3c**) and a smaller peak corresponding to a single pimeloyl-adenylate bound to the dimer (*m/z* = 4268.9). When BioW was mixed with pimelic acid, MgATP and CoASH we observed an increase in the relative abundance of the BioW dimer with only one pimeloyl-adenylate bound (*m/z* = 4269.0) and also of the ligand-free BioW (*m/z* = 4234.4) (**Fig. 3d**). We only observed ions consistent with the binding of the pimeloyl-adenylate and failed to detect BioW complexes with either pimelic acid, MgATP, PP_i or the product pimeloy-CoA bound. This suggested that the BioW: pimeloyl-adenylate intermediate complex is highly stable and is broken down in the presence of CoASH.

Residues controlling substrate binding and catalysis

Structural data and multiple sequence alignment of BioW isoforms (**Supplementary Fig. 6**) guided the selection of four residues for mutagenesis (Tyr199, Tyr211, Arg213, Arg227). Each BioW mutant was expressed and purified in a similar manner to the wild type enzyme. We measured enzyme activity (turnover) using the PP_i release-based coupled assay with saturating concentrations of the acid, MgATP and CoASH and set the wild-type BioW activity ($k_{cat,}$ s^{-1}) as 100%. Mutants R227E and R227K showed reduced BioW turnover with the natural substrate pimelic acid by around 20 fold (~4% activity remaining) (**Fig. 4a**). The R213A mutation reduced BioW activity >99% compared with wild type and was effectively inactive. In contrast, mutant Y199F retained 55% activity and similarly, mutant Y211F retained 36% activity (**Fig. 4a**).

The wild type BioW displays high specificity for pimelic acid with much reduced turnover observed (k_{cat} value ≤1%) with both shorter and longer dicarboxylic acids substrates (**3**, **4**, **5**

and **6**, **Fig. 1b**, **Fig. 4b**, **Supplementary Table 2**). Both Y199F and Y211F mutants displayed reduced or barely detectable turnover compared to the WT BioW with shorter chain length substrates glutaric acid and adipic acid (**Fig. 4b**). In contrast, these mutants displayed increased activity relative to wild type BioW with the longer dicarboxylic acid substrates suberic acid and azelaic acid (**Fig. 4b**). Using an increased enzyme concentration of 3.0 μM we observed that the Y199F mutant displayed a two-fold improved activity with but no improvement with azelaic acid over wild type (**Supplementary Fig. 7**). In contrast, the Y211F mutant displayed ~4 fold increased activity with the suberic acid substrate and ~3 fold increased activity with the azelaic acid substrate relative to the wild-type BioW (**Supplementary Fig. 7**).

The BioW crystal structure with CoASH bound (albeit without clear density for the phosphopantetheine arm) identified the side-chains of Arg11 and Arg13 as potentially involved in binding the phosphate groups of this substrate and we generated mutants at these two sites (R11A and R13A). Both mutants were essentially inactive with pimelic acid (**Supplementary Fig. 8**, <1.0 % activity compared with WT).

An engineered BioW displays novel activity

The wild-type BioW enzyme displayed little activity when incubated with heptanoic acid and octanoic acid at both low (0.1 μM) and high (3.0 μM) enzyme concentrations respectively (**Fig. 4b** and **Supplementary Fig. 7**). Similarly, the Y199F mutant was also inactive with heptanoic acid and octanoic acid (**Fig. 4b** and **Supplementary Fig. 7**). In contrast, the Y211F displayed activity with both monocarboxylic acid substrates at 0.1 μM concentration (**Fig. 4b**) and at the higher concentration (3.0 μM) the Y211F mutant displayed a k_{cat} of 0.012 ± 0.001 s^{-1} and 0.008 s^{-1} with heptanoic acid and octanoic acid respectively (**Supplementary**

Fig. 7) Indeed, turnover was such that we could determine the K_m for heptanoic (529.0 ± 79.7 μM) and ATP (479.4 ± 81.5 μM) (**Supplementary Fig. 9**). We set up "synthetic reactions" by incubation of the wild type and the Y211F mutant enzyme (5.0 μM) with saturating concentrations of heptanoic acid, MgATP and CoASH for 5 hours at 30 ^{o}C. HPLC analysis showed that the wild type BioW produced no detectable product (**Fig. 5a**). In contrast, the BioW Y211F mutant generated a peak eluting at 20.3 minutes (**Fig. 5a**) which was confirmed as the heptanoyl-CoA (**9**) thioester product using mass spectrometry (**Fig. 5b**, predicted mass for $C_{28}H_{48}N_7O_{17}P_3S$ $[M+H]^+$ = 880.2113 Da, observed mass = 880.2124 Da**)**. It is worth noting that despite producing no detectable product with pimelic acid, we found that the BioW R213A mutant did give rise to detectable amounts of the heptanoyl-CoA and octanoyl-CoA (**10**) (**Supplementary Fig. 7**).

Since the BioW Y211F mutant displayed novel heptanoyl-CoA synthetase activity we tested whether it could also accept fatty acid substrates substituted at the reduced end of the acyl chain. Five commercially available derivatives (7-bromoheptanoic acid, **11**; 7-aminoheptanoic acid, **12**; 6-methylheptanoic acid, **13**; 7-phenylheptanoic acid, **14** and 7-octenoic acid, **15**) were screened using a convenient LC-MS method to detect the formation of the corresponding acyl-CoA products (**Supplementary Figs. 10 a-f)**. We observed product formation for all of the substrates apart from the 7-aminoheptanoic acid derivative confirming the broadened synthetic utility of the BioW Y211F mutant.

DISCUSSION

An early step in biotin biosynthesis requires the AONS-catalysed condensation of L-alanine and a thioester activated form of pimelic acid (either pimeloyl-CoA or pimeloyl-acyl carrier

protein (ACP), **Fig. 1a**). We found that BioW is highly specific for pimelic acid, turning over this substrate at least 100 times faster than either adipic acid or suberic acid.

The fatty acid specificity of BioW is controlled by a cluster of residues that bind to and fix the position of the ω- carboxylate of pimelic acid while the carbon chain sits in a hydrophobic tunnel (**Fig. 2e** and **Supplementary Fig. 5d and 5e**). Only when the intervening seven carbon atoms adopt an extended conformation will the other (α) carboxylate group be appropriately positioned to attack the α-phosphate of ATP required to generate the pimeloyl-adenylate. Since the tunnel is restricted in volume and the ω position fixed, additional carbons will result in clashes with ATP and shorter chains will not reach it. We identified this internal ruler as the origin of the chain length selectivity of BioW.

Superposition of the ligand-bound BioW structures revealed that the geometry of the incoming nucleophilic oxygen atom of the carboxylate and the phosphate is ideal for addition to form the pentavalent phosphorus transition state (**Fig. 2d**). We suggest Arg227 and Arg170 stabilise the negative charge on the oxygen atom of the carboxylate enhancing its nucleophilicity. In the AMP-PNP analog structure, we identified two Mg^{2+} ions that will bridge the α and γ phosphates of ATP and in doing so each Mg^{2+} binds to and stabilises the terminal oxygen atom of the α phosphate in the transition state. In the BioW:pimeloyl-adenylate:PP_i complex the two Mg^{2+} ions bind to and, we presume, stabilise the negatively charged PP_i leaving group.

We found that PPi release (and thus activity in our coupled assay) only occurs in a substantial amount when CoASH is bound to the enzyme (**Supplementary Fig. 9**). Mutation of predicted CoASH binding residues (Arg11, Arg13) stopped turnover even in the presence of CoASH (**Supplementary Fig. 9**). Such a tight coupling would be desirable *in vivo* since futile consumption of ATP would be deleterious to the cell. Since ATP and pimelic acid bind

to the protein in the absence of CoASH, these data suggested either the adenylate does not form in the absence of CoASH or that the adenylate is stable until CoASH binds. We favour the latter explanation as, in the absence of CoASH, we observed the adenylate intermediate in crystals (**Fig. 2c**) and by native mass spectrometry (**Fig. 3**). Adenylates are usually highly reactive, as exemplified in studies of biotin protein ligase (BPL), which generates the biotin-adenylate intermediate that is used to transfer biotin onto the biotin carrier domain of ACC. Mutation of a key arginine in BPL releases the adenylate intermediate, allowing it to biotinylate in a promiscuous manner [35, 36], and this has been used in biotin proximity assays (named BioID) [37]. The BioW assay had a detectable background rate without CoASH suggesting water can hydrolyse the adenylate (**Supplementary Fig. 9**). The adenylate intermediate, although bound in the enzyme, is not shielded from solvent with water molecules positioned close to the carboxylate group (**Fig. 2e**). We suggest that the positioning and hydrogen bonding of the water molecules means they cannot rapidly hydrolyse the adenylate and this accounted for the very unusual stability of the adenylate intermediate.

In a crystal soaked with CoASH, all but the pantothenate arm was located experimentally and mutagenesis at the ribose phosphate binding site supported our assignment that this is where CoASH binds (**Fig. 2f**). BioW possesses two possible tunnels from the surface to the catalytic site that could be consistent with the pantothenate arm reaching in to attack the adenylate (**Fig. 6** and **Supplementary Fig. 5h)**. To form the thioester bond, the thiol group must displace at least one water molecule at the adenylate and since thiols are generally regarded as stronger nucleophiles than water, we suggest the thiol is sufficiently reactive to decompose the adenylate in the presence of Arg227. It has been proposed that class I-III ANL enzymes undergo large conformational changes during their catalytic cycle, the C-terminal domain rotates in relation to the N-terminus in the presence of CoASH [25, 38]. We did not observe any

such large conformational changes in the class IV BioW, but we cannot exclude such a mechanism.

Thioesters are chemically synthesised in an analogous way to the enzymatic process; typically by acid activation (e.g. as acid anhydrides or acid chlorides), then reaction with an appropriate thiol. Older methods are low-yielding and the di-acids have to be protected to direct synthesis of mono-thioester targets. Alternative methods such as iron-catalysed coupling of aldehydes with sulfur surrogates [39] and methyl acyl phosphates have been developed [40], but a "green" biocatalytic synthesis of simple fatty acid thioesters would be highly desirable. The relatively simple catalytic domain of BioW provided an ideal template to engineer an enzyme with expanded substrate range. Having defined the chain length specificity and the "internal ruler" (**Fig. 2e**), we made mutants at Tyr199 and Tyr211 that trimmed back H-bond donors and side chains that mediate electrostatic interactions. We found that the BioW Y211F mutant retained activity with pimelic acid and gained the ability to synthesise heptanoyl-CoA from heptanoic acid (**Fig. 5**). We saw no evidence that the adenylate intermediate is released into solution suggesting that the mutant enzyme has retained the ability to stabilise the non-cognate adenylate intermediate.

Encouraged by the gain of function of BioW Y211F, we tested a variety of fatty acid derivatives modified at the terminus of the acyl chain and found that the enzyme was able to convert four of the five substrate analogues into the corresponding acyl-CoA thioester product (**Supplementary Fig. 10**). This result exemplified the synthetic utility of a BioW mutant to generate useful acyl-CoA products. A promiscuous acyl-ACP synthetase (AasS) from *Vibrio harveyi* has also been used to attach various fatty acid substrates to ACPs [41]. With its expanded substrate range, our BioW Y211F mutant is a useful addition to the toolkit that enables enzyme-catalysed preparation of these useful chemical building blocks. BioW is a relatively small enzyme, with few residues defining the active site, so we propose that this

new member of the ANL superfamily (the first class IV member) offers great scope to expand the range of acyl-thioesters that can be synthesised. Exemplifying this idea, the BioW R213A mutant accepted octanoic acid, suggesting that the active site can be further exploited and engineered to accept longer acyl chains. Applying the tools of directed evolution and combining these with the application of a high-throughput assay for acyl-CoA thioester formation activity would allow the full synthetic potential of BioW to be explored [42].

Acknowledgments

This work was supported by grants from the Biotechnology and Biological Sciences Research Council (BBSRC, BB/M003493/1) (DJC & JHN) and Wellcome Trust (WT100209MA (J.H.N.). The native mass spectrometry data was acquired on an instrument funded by an Engineering and Physical Sciences Research Council (EPSRC) grant to the University of Edinburgh (EP/K039717/1). We acknowledge the use of the Diamond synchrotron. J.H.N. is a Royal Society Wolfson Merit Award Holder and 1000 talent scholar at Sichuan University. M.W. was funded by a University of Edinburgh PhD scholarship. We thank Bohdan Mykhaylyk for help with bioinformatics analysis and Dr. Logan Mackay for help with mass spectrometry analysis.

Author Contributions

A.P. carried out the initial cloning of the bioW gene and characterisation of the recombinant BioW. M.W. and P.J.H. carried out the enzyme isolation, characterisation and assay. M.W. generated all the BioW mutants and determined the kinetic parameters of wild type and mutant enzymes. V.K. performed all protein and acyl-CoA mass spectrometry analysis. M.W. and L.M. prepared the enzyme for crystal trials and optimised crystallisation conditions. L.M. carried out the crystallography experiments and with J.H.N. acquired and interpreted the data. L.K., M.W., P.J.H, V.K., J.H.N. and D.J.C. interpreted the data and wrote the paper.

Competing Financial Interests

The authors declare no competing financial interests.

Additional Information

Supplementary information is available in the online version of the paper.

References

1. Tong, L. Structure and function of biotin-dependent carboxylases. *Cell Mol Life Sci* **70**, 863-91 (2013).
2. Eisenberg, M.A. The incorporation of 1,7 C14 pimelic acid into biotin vitamers. *Biochem Biophys Res Commun* **8**, 437-41 (1962).
3. 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).
4. 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).
5. Eisenberg, M.A. & Stoner, G.L. Biosynthesis of 7,8-diaminopelargonic acid, a biotin intermediate, from 7-keto-8-aminopelargonic acid and S-adenosyl-L-methionine. *J Bacteriol* **108**, 1135-40 (1971).
6. Otsuka, A.J. et al. The Escherichia coli biotin biosynthetic enzyme sequences predicted from the nucleotide sequence of the bio operon. *J Biol Chem* **263**, 19577-85 (1988).
7. Lin, S. & Cronan, J.E. Closing in on complete pathways of biotin biosynthesis. *Mol Biosyst* **7**, 1811-21 (2011).
8. Alexeev, D. et al. The crystal structure of 8-amino-7-oxononanoate synthase: a bacterial PLP-dependent, acyl-CoA-condensing enzyme. *J Mol Biol* **284**, 401-19 (1998).
9. Alexeev, D., Baxter, R.L. & Sawyer, L. Mechanistic implications and family relationships from the structure of dethiobiotin synthetase. *Structure* **2**, 1061-72 (1994).
10. Berkovitch, F., Nicolet, Y., Wan, J.T., Jarrett, J.T. & Drennan, C.L. Crystal structure of biotin synthase, an S-adenosylmethionine-dependent radical enzyme. *Science* **303**, 76-9 (2004).
11. Huang, W. et al. Crystal structure of an ATP-dependent carboxylase, dethiobiotin synthetase, at 1.65 A resolution. *Structure* **2**, 407-14 (1994).
12. Kack, H., Sandmark, J., Gibson, K., Schneider, G. & Lindqvist, Y. Crystal structure of diaminopelargonic acid synthase: evolutionary relationships between pyridoxal-5'-phosphate-dependent enzymes. *J Mol Biol* **291**, 857-76 (1999).
13. Webster, S.P. et al. Mechanism of 8-amino-7-oxononanoate synthase: spectroscopic, kinetic, and crystallographic studies. *Biochemistry* **39**, 516-28 (2000).
14. 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).
15. 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).
16. 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-labelling studies. *J Am Chem Soc* **116**, 2637-38 (1994).
17. 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).
18. 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).
19. Bower, S. et al. Cloning, sequencing, and characterization of the Bacillus subtilis biotin biosynthetic operon. *J Bacteriol* **178**, 4122-30 (1996).
20. Cryle, M.J. & De Voss, J.J. Carbon-carbon bond cleavage by cytochrome p450(BioI)(CYP107H1). *Chem Commun (Camb)*, 86-7 (2004).
21. 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).

22. Gloeckler, R. et al. Cloning and characterization of the Bacillus sphaericus genes controlling the bioconversion of pimelate into dethiobiotin. *Gene* **87**, 63-70 (1990).
23. 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).
24. Manandhar, M. & Cronan, J.E. Proofreading of noncognate acyl adenylates by an acyl-coenzyme a ligase. *Chem Biol* **20**, 1441-6 (2013).
25. 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).
26. Schmelz, S. & Naismith, J.H. Adenylate-forming enzymes. *Curr Opin Struct Biol* **19**, 666-71 (2009).
27. Pendini, N.R. et al. Microbial biotin protein ligases aid in understanding holocarboxylase synthetase deficiency. *Biochim Biophys Acta* **1784**, 973-82 (2008).
28. Schmelz, S. et al. AcsD catalyzes enantioselective citrate desymmetrization in siderophore biosynthesis. *Nat Chem Biol* **5**, 174-82 (2009).
29. Finzel, K., Lee, D.J. & Burkart, M.D. Using modern tools to probe the structure-function relationship of fatty acid synthases. *Chembiochem* **16**, 528-47 (2015).
30. Upson, R.H., Haugland, R.P., Malekzadeh, M.N. & Haugland, R.P. A spectrophotometric method to measure enzymatic activity in reactions that generate inorganic pyrophosphate. *Anal Biochem* **243**, 41-5 (1996).
31. Webb, M.R. A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proc Natl Acad Sci U S A* **89**, 4884-7 (1992).
32. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-97 (2007).
33. Yount, R.G., Babcock, D., Ballantyne, W. & Ojala, D. Adenylyl imidodiphosphate, an adenosine triphosphate analog containing a P--N--P linkage. *Biochemistry* **10**, 2484-9 (1971).
34. Walzthoeni, T., Leitner, A., Stengel, F. & Aebersold, R. Mass spectrometry supported determination of protein complex structure. *Curr Opin Struct Biol* **23**, 252-60 (2013).
35. Choi-Rhee, E., Schulman, H. & Cronan, J.E. Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein Sci* **13**, 3043-50 (2004).
36. Tron, C.M. et al. Structural and functional studies of the biotin protein ligase from Aquifex aeolicus reveal a critical role for a conserved residue in target specificity. *J Mol Biol* **387**, 129-46 (2009).
37. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **196**, 801-10 (2012).
38. McElroy, W.D., DeLuca, M. & Travis, J. Molecular uniformity in biological catalyses. The enzymes concerned with firefly luciferin, amino acid, and fatty acid utilization are compared. *Science* **157**, 150-60 (1967).
39. Huang, Y.T., Lu, S.Y., Yi, C.L. & Lee, C.F. Iron-catalyzed synthesis of thioesters from thiols and aldehydes in water. *J Org Chem* **79**, 4561-8 (2014).
40. Pal, M. & Bearne, S.L. Synthesis of coenzyme A thioesters using methyl acyl phosphates in an aqueous medium. *Org Biomol Chem* **12**, 9760-3 (2014).
41. Beld, J., Finzel, K. & Burkart, M.D. Versatility of acyl-acyl carrier protein synthetases. *Chem Biol* **21**, 1293-9 (2014).
42. Renata, H., Wang, Z.J. & Arnold, F.H. Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution. *Angew Chem Int Ed Engl* **54**, 3351-67 (2015).

Figure Legends

Figure 1. The BioW reaction highlighting its role in biotin biosynthesis and the activity of wild type BioW with various fatty acid substrates. (A) The enzyme pimeloyl-CoA synthetase (BioW) catalyses the conversion of pimelic acid to the pimeloyl-CoA thioester via the pimeloyl-adenylate intermediate. In the next step the condensation of pimeloyl-CoA with L-alanine is catalysed by the pyridoxal 5'-phosphate (PLP)-dependent enzyme 8-amino-7-oxononanoate synthase (AONS). The product of this reaction, 8-amino-7-oxononanoate, is then converted to 7,8-diaminononanoate (DAN) by the PLP-dependent transaminase diaminononanoate synthase (DANS). The ureido ring of dethiobiotin is formed by the enzyme dethiobiotin synthease (DTBS). The final step involves the insertion of sulfur and is catalysed by the enzyme biotin synthase (BS). AMP is adenosine monophosphate. **(B)** The PPi release assay was used to determine enzyme activity and is based on turnover (k_{cat}, s^{-1}) in the presence of BioW (0.1 μM), ATP (1.0 mM) and CoASH (1.0 mM) and carboxylic acids (1.5 mM) of various carbon chain lengths. The best BioW substrate is pimelic acid (C7; **1**) and the worst is azelaic acid (C9; **6**). Data represent mean values ± s.d. from multiple experiments.

Figure 2. Structural biology of BioW. (A) Overall structure of the BioW dimer. The AMP-PNP and pimelic acid molecules are shown as sticks (carbon in yellow, oxygen in red, nitrogen in blue and phosphate in orange). The protein A and B subunits are colored green and blue, respectively. The N-terminal domain is highlighted by a dashed box.

(B) Overall structure of BioW monomer. Secondary structure elements are labelled and coloured red (α-helices), yellow (β-sheets) and green (loops). AMP-PNP and pimelic acid are

shown as sticks (carbon atoms in white, other atoms as in **(a)**), and Mg^{2+} ions are shown as green spheres.

(C) F_o-F_c electron density omit map at 3 σ around the pimeloyl-adenylate. Compounds are shown as sticks with carbon atoms colored cyan, other atoms as **(a)**.

(D) Superposition of the pimeloyl-adenylate / PPi and AMP-PNP / pimelic acid complexes. For clarity, the protein has been removed. Carbon atoms are coloured in cyan (pimeloyl-adenylate) and yellow (AMP-PNP/pimelic acid), other atoms as **(a)**. The superposition allows a model for the pentavalent phosphorous transition state to be deduced.

(E) The pimeloyl-adenylate is stable, despite two water molecules (W1 and W2) being positioned close to the adenylate ester carbonyl group. The hydrogen bonding network and positioning of waters inhibits their ability to attack the ester.

(F) Binding site of the adenosine 3', 5'-diphosphate group of CoA. Residues within 4.0 Å of the CoA are displayed. Carbon atoms of the CoA are pink and residues of A and B subunits are light cyan and white.

Figure 3. Native mass spectrometry analysis of BioW. **(A)** BioW is observed as a monomer (charge states 12+ to 9+) and a dimer (charge states 16+ to 13+) by native mass spectrometry when ionised from 200 mM ammonium bicarbonate buffer. The 14+ charge state dimer species [boxed in **(A)** and expanded in **(B)**], has observed *m/z* = 4233.9 and is in good agreement with the theoretical *m/z* = 4233.6. **(C)** Incubation of BioW with 1.0 mM pimelic acid and 1.0 mM MgATP led to the formation of a 14+ ion with *m/z* = 4304.0. This mass difference (980.7 Da) is consistent with non-covalent binding of two molecules of pimeloyl-adenylate (theoretical mass 489.4 Da). **(D)** Addition of 0.5 mM CoASH to the

incubation described in **(C)** led to an increase in the intensity of the ion with *m/z* = 4269.0 which is consistent with one molecule of pimeloyl-adenylate.

Figure 4. Activity of wild type BioW and active site mutants towards pimelic acid and assays to determine the chain length specificity of BioW and designed mutants. (A) The pyrophosphate (PPi) release assay was used to determine enzyme activity and is based on turnover (k_{cat}, s^{-1}) in the presence of BioW enzyme (0.1 μM, WT and mutant), pimelic acid (1.5 mM), ATP (1.0 mM), CoASH (1.0 mM). The value for WT BioW (0.48 s^{-1}) was set at 100% and the BioW mutant values are shown relative to WT. **(B)** The turnover (k_{cat}, s^{-1}) of the BioW WT enzyme and each of the mutants Y199F, Y211F and R213A (0.1 μM) in the presence of different mono- and dicarboxylic acids (compounds **3**, **4**, **5**, **6**, **7**, **8**, 1.5 mM), ATP (1.0 mM) and CoASH (1.0 mM). The WT BioW data with pimelic acid (0.48 s^{-1}) is included for comparison. Data represent mean values ± s.d. from multiple experiments.

Figure 5. Synthesis of hepatonyl-CoA by the engineered BioW Y211F mutant. The WT BioW enzyme and Y211F mutant were incubated with the heptanoic acid, ATP and CoASH substrates for 5 hours at 30 ^{o}C before analysis by HPLC. The heptanoyl-CoA product elutes at 20.3 minutes. No detectable product is observed using the WT BioW. **(B)** The peak eluting at 20.3 minutes was collected and confirmed as C7-CoA by FT-ICR Mass spectrometry (predicted mass for $C_{28}H_{48}N_7O_{17}P_3S$ $[M+H]^+$ = 880.2113 Da, observed 880.2124 Da). The expertimental mass spectrum shown (*m/z* range = 878-886) matches that predicted (the red dots are the isotope distribution based on the chemical formula) and the inset is the full mass spectrum (*m/z* range = 0-2500).

Figure 6. Schematic of the BioW active site. The key residues in the C-terminal domain involved in pimelic acid recognition (Tyr199, Tyr211 and Arg213) are highlighted, as are those involved in pimeloyl-adenylate binding (Arg170, Arg227). The N-terminal residues involved in coenzyme A (CoASH) binding (Arg11 and Arg13) are included, as is the PP_i derived from ATP and the coordinating Mg^{2+} ions. We propose formation of the pimeloyl-CoA thioester product by attack of the adenylate intermediate by the thiol of the incoming CoASH.

ONLINE METHODS

Reagents

Standard chemical reagents (pimelic acid or other dicarboxylic acids, mono- fatty acids, ATP, buffers) and enzymes were from Sigma, Alfa Aesar and Acros Organic. CoASH was from Chem-Impex International. UV-transparent plates were from Costar.

Cloning, expression and purification

The full-length, codon-optimised *bioW* gene (*Bacillus subtilis (strain 168)*, UNIPROT code: P53559*)* was purchased from GenScript and cloned into two expression plasmids to give recombinant BioW with a non-cleavable N-terminal 6His tag (pET-28a) and a pEHISTEV plasmid which gives an N-terminal tobacco etch virus (TEV) protease-cleavable His_6 tag [43]. The *B. subtilis* bioW gene was expressed in *E. coli* BL21 (DE3) transformed with the appropriate plasmid and 2 L LB media were grown in 500 ml shake flasks until the OD_{600} reached 0.6-0.8. Expression was induced with 0.1 mM IPTG and the cells were grown at 30 °C for 4 h. All purification steps were performed at 4 °C or on ice. Cells were resuspended in lysis buffer (20 mM sodium phosphate pH 7.5, 500 mM NaCl, 10 mM imidazole, 2 mM TCEP, 10% glycerol with the addition of complete EDTA-free protease inhibitor tablets (Roche), DNase (0.2 mg per 10 mL buffer) and lysozyme (1 mg per 10 mL buffer,) and lysed with a cell disruptor at 30 kPSI (Constant Systems Ltd). The crude lysate was cleared by centrifugation (45,000 *g*, 4 °C, 30 min), filtered through a 0.22 μm filter and loaded onto a

5ml His-trap (GE Healthcare) column attached to an AKTA system. The column was washed with the same lysis buffer without the protease inhibitor tablet and DNase. The BioW was eluted with an increasing imidazole gradient (25-300 mM) over 20 column volumes (100 ml). Eluted BioW was dialysed against 1L of 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM β-mercaptoethanol (BME), 10% glycerol, 0.25 mM PMSF for 1h. Where appropriate (e.g. for crystallisation), the His_6 tag was cleaved by incubation with TEV protease (mass ratio 1:15) overnight. The uncleaved protein and protease were then removed by passage through a second HisTrap column in lysis buffer containing 20 mM imidazole. Finally, the BioW (6His-tagged or TEV-cleaved) was loaded onto a gel filtration Superdex 200 column (GE Healthcare) in 10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.25 mM PMSF, 1 mM dithiothreitol (DTT). After His_6 tag removal, the BioW N-terminus contains the sequence GAMEEE. Each step of the purification was monitored by SDS-PAGE and the masses of both His_6 BioW and TEV-cleaved BioW were confirmed by LC ESI-MS analysis.

Site-directed and site-saturation mutagenesis.

The BioW mutants Y199F, Y211F and R213A were constructed according to the overlapping primer site-directed mutagenesis method [44] with the following primers pairs.

Y199F

Y199F forward (5′-ATGCTGGTCGGACGATCCGGAT**TTC**ATAACAGGCTATGTTGCGGGTAAGAAAATGGGCTAT-3′)

Y199F reverse (5′-CTTACCCGCAACATAGCCTGTTAT**GAA**ATCCGGATCGTCCGACCAGCATAATTCTGCAACG-3′)

Y211F

Y211F forward (5′-ATGTTGCGGGTAAGAAAATGGGC**TTC**CAGCGTATTACAGCAATGAAAGAATACGGGACTG-3′)

Y211F reverse (5′-
CTTTCATTGCTGTAATACGCTG**GAA**GCCCATTTTCTTACCCGCAACATAGCCTGTT
ATG-3′)

R213A

R213A forward (5′-
GCGGGTAAGAAAATGGGCTATCAG**GCA**ATTACAGCAATGAAAGAATACGGGAC
TGAAG-3′)

R213A reverse (5′-
CCCGTATTCTTTCATTGCTGTAAT**TGC**CTGATAGCCCATTTTCTTACCCGCAACAT
AG-3′).

R11A

R11A forward (5′-
GAAGAAACTTTTTATAGTGTC**GCT**ATGAGGGCTTCAATGAATGGATCTCATGAAG
ACGGCGG-3′)

R11A reverse (5′-
GAGATCCATTCATTGAAGCCCTCAT**AGC**GACACTATAAAAAGTTTCTTCTTCCAT
GG-3′)

R13A

R13A forward (5′-
GAAGAAACTTTTTATAGTGTCAGAATG**GCT**GCTTCAATGAATGGATCTCATGAAG
ACGGCGG-3′)

R13A reverse (5′-
GAGATCCATTCATTGAAGC**AGC**CATTCTGACACTATAAAAAGTTTCTTCTTCCAT
GG-3′)

R227E and R227K

The BioW mutants R227E, R227K were generated by saturation mutagenesis with the following primers:

R227X forward (5′-ATACGGGACTGAAGAGGGCTGC**RMA**GTCTTTTTTATTGATGGCTCCAATGATGTAAACA-3′)

R227X reverse (5′-GGAGCCATCAATAAAAAAGAC**TKY**GCAGCCCTCTTCAGTCCCGTATTCTTTCATTGCTG-3′)

Where R = A/G, M = A/C, K = G/T, Y = C/T.

Colonies were picked, sequenced and based on the results BioW mutants R227E and R227K were taken forward for analysis. All BioW point mutant proteins were overexpressed with 0.5 mM IPTG at 20 °C overnight. The purification protocol is the same as the His_6 tagged wild type protein with TEV protease removal of the His tag. The mass of each BioW mutant was confirmed by LC electrospray ionization mass spectrometry (ESI-MS).

Enzyme assays

Monitoring of acyl-CoA thioester bond formation.

Measurements were made by a previously described continuous UV spectrophotometric assay [23]. This assay is based on monitoring the formation of the acyl-CoA thioester bond at 230 nm. The BioW enzyme (0.2 μM) in 100 mM Tris-HCl pH 8.5, 200 mM NaCl, 10 mM $MgCl_2$ was incubated at 30 °C with 0.4 mM CoASH, 0.4 mM ATP in a reaction volume of 200 μL The reactions were started by the addition of 1.5 mM pimelic acid and the 230 nm absorbance was recorded for 30 minutes on a BioTek Synergy HT plate reader with Costar 96-well UV-transparent plates. The rate over the first 10 minutes was analysed using the Michaelis-Menten model and a nonlinear regression fit on GraphPad to give values of K_m and k_{cat}.

Pyrophosphate (PPi) production assay.

The BioW activity was also determined by monitoring the release of PP_i via a two enzyme coupled assay. The final volume of the reactions was 200 μL and contained 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$, 1.0 mM triscarboxyethylphosphine (TCEP), 0 to 1.0 mM CoASH, 0 to 1.0 mM ATP, 0.5 mM 7-methyl-6-thioguanosine (MESG, Berry &

Associates), 0.03 U of inorganic pyrophosphatase (PPase from baker's yeast, Sigma), 1 U of bacterial purine nucleoside phosphorylase (PNP, Sigma) and 0.1 μM BioW enzyme (or up to 3.0 μM for several mutants). These reagents were pre-incubated at 30 °C for 15 min and the reaction was initiated by adding 0 to 1.5 mM pimelic acid (or dicarboxylic acids or fatty acid substrates of varying carbon chain length). The increase in absorbance at 360 nm resulting from the enzymatic conversion of MESG to 7-methyl-6-thioguanine (ε_{360}= 11,000 $M^{-1}cm^{-1}$) was monitored over 30 min on a BioTek Synergy HT plate reader with Costar 96-well UV-transparent plates. The data from the first 10 minutes was analysed using the Michaelis-Menten model and a nonlinear regression fit on GraphPad gave values of K_m and k_{cat}.

High Performance Liquid Chromatography (HPLC) analysis of acyl-CoA formation.

The formation of acyl-CoA esters was detected by an HPLC method. The reactions contained 25 mM Tris-HCl pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 0.2 mM TCEP, 1.5 mM CoASH, 1.5 mM ATP, 5.0 μM BioW and 1.5 mM mono- or dicarboxylic acid substrate (and substituted derivatives) were incubated at 30 °C in a final volume of 500 μL. The reactions were terminated at varying time points by incubation with acetonitrile (0.1% TFA, (trifluoroacetic acid)) (volume ratio 1:1). The samples were clarified by centrifugation at 17, 000 x g for 10 minutes and filtered through a 0.45 μm filter. 10 to 50 μL of the sample was then injected onto Luna 5μ C18(2) RP-HPLC column (100A, 250x4.60mm, Phenomenex),and eluted with water (0.1% TFA v/v) for 5 min followed by a 15 min gradient from 5% to 55% acetonitrile (0.1% TFA, v/v)/water (0.1% TFA, v/v) which was maintained for 5 min, returning to 100% water (0.1% TFA) for the final five minutes. The eluant was monitored at 260 nm and where necessary the peak containing the acyl-CoA product was collected and concentrated under reduced pressure (to remove TFA and acetonitrile) and heating at 35 °C. Samples were either resuspended or diluted into further into water and analysed by FT-ICR Mass Spectrometry on a 12 Tesla SolariX FT-ICR MS (Bruker Daltronics) via positive electrospray ionisation. For MS, samples were co-injected with ESI-L Low Concentration Tuning Mix (Agilent Technologies, G1969-85000). Synthetic pimeloyl-CoA standard was prepared using a published method [45].

Mass spectrometry analysis of acyl-CoA products

2.5 μL of the BioW incubated with various fatty acids (see HPLC analysis above) was injected onto a Kinetex 5μm C18 reverse phase column (100Å, 50x2.1mm, Phenomenex) and

eluted at 0.5ml/min with buffers A (0.1% v/v formic acid) and B (acetonitrile, 0.1% v/v formic acid). The samples were loaded and washed with 0.5% buffer A for 5 min before switching from waste to MS and running a gradient from 0.5% to 55% buffer B, followed by column washing and equilibration. Electrospray ionisation was coupled to a Synapt G2 Q-TOF (Waters) with a 120 °C source temperature, 2kV capillary voltage, and sampling cone voltage of 35. MS spectra were acquired over a 150-2000 mass range in positive polarity and in sensitivity mode. A lockmass correction was applied once at the beginning of each analysis against Leucine Enkephalin peptide.

Crystallization and Structure Determination

Crystals of BioW appeared at 4°C after a week from a hanging drop of 1 µl of protein solution (19 $mg.ml^{-1}$, 5.0 mM pimelic acid, 1.0 mM $MgCl_2$ and 5.0 mM ATP, AMP-PNP or BrATP) with 1 µl of reservoir solution containing 26 % PEG 400, 0.1 M HEPES pH 7.5 in vapor diffusion with reservoir. For phasing, crystals were soaked in 50 mM K_2PtCl_4 for approximately 15 minutes and back-soaked in mother liquor containing pimelic acid, ATP and $MgCl_2$. Crystals crystalized in the presence of BrATP were soaked overnight in a crystallisation solution containing 20 mM CoASH and 5 mM pimelic acid before been flash frozen. BrATP was not present in the crystal structure but a residual density in the chain A active site has been modelled as a PEG400 molecule. Data were collected at 100 °K at the Diamond light source Oxfordshire, beamlines IO3 (λ=1.07206), IO4 (λ=0.97949) and IO2 (λ=0.9795) for K_2PtCl_4, AMP-PNP/pimelic acid and pimelic acid/CoASH complexes respectively. For the Pimeloyl-adenylate/PPi complex, data were collected in house using a Rigaku Micromax^{TM}-007HF Cu anode with VariMax optics and Rigaku Saturn 944+ CCD detector. Data were processed with XIA2 [46-50] or HKL2000 [51]. Phases were determined by SAD and an initial model built using the Crank pipeline [52] as implemented in CCP4i2 [53] and the structure refined with REFMAC5 and adjusted with COOT [54] and Autobuild program of PHENIX [55].

Subsequent structures were solved by molecular replacement with the program PHASER [56] using the coordinates from the experimental structure. Final models were adjusted with COOT [54] and refinement was carried out using REFMAC in the CCP4 program suite with NCS restraints and TLS parameters [57]. Final refinement statistics are given in Supplementary Table 3. Atomic coordinates and structure factors have been deposited in the Protein Data Bank (5fm0, 5flg, 5fll, 5g1f). Coordinates and topologies of ligands were generated by PRODRG [58]. The quality of all structures were checked with MOLPROBITY [59]. The

Ramachandran statistics are as followed: 98.6% favored and 1.4% allowed for K_2PtCl_4; 100% favored for the AMP-PNP/pimelic acid complex; 99.2% favored and 0.8% allowed for the pimeloyl-adenylate/PPi complex; 98.3% favored and 1.7% allowed for the pimelic acid/CoASH complex. Figures were drawn using PYMOL [60].

Native Mass Spectrometry

The BioW sample was desalted twice over P6 Micro Bio-Spin columns (Bio-Rad) into 200 mM ammonium bicarbonate and further diluted to approximately 20 μM (10 μM dimer). Native MS spectra were acquired before and after addition of 1.0 mM MgATP, 1.5 mM pimelic acid and 0.5 mM CoASH. Samples were infused by nano-electrospray ionisation (nano-ESI) using a TriVersa NanoMate (Advion) set at 1.6 kV and 0.8 psi pressure. Electrospray was coupled to a Synapt G2 Q-TOF (Waters) with the source set at 80 ^oC, backing pressure increased to 5 mbar, and sampling cone voltage of 150 V. Spectra were averaged over 4 minutes of acquisition data and presented without further data processing. Peak *m/z* annotations were extracted from smoothed and centroided using Waters MassLynx V4 software.

Data Availability

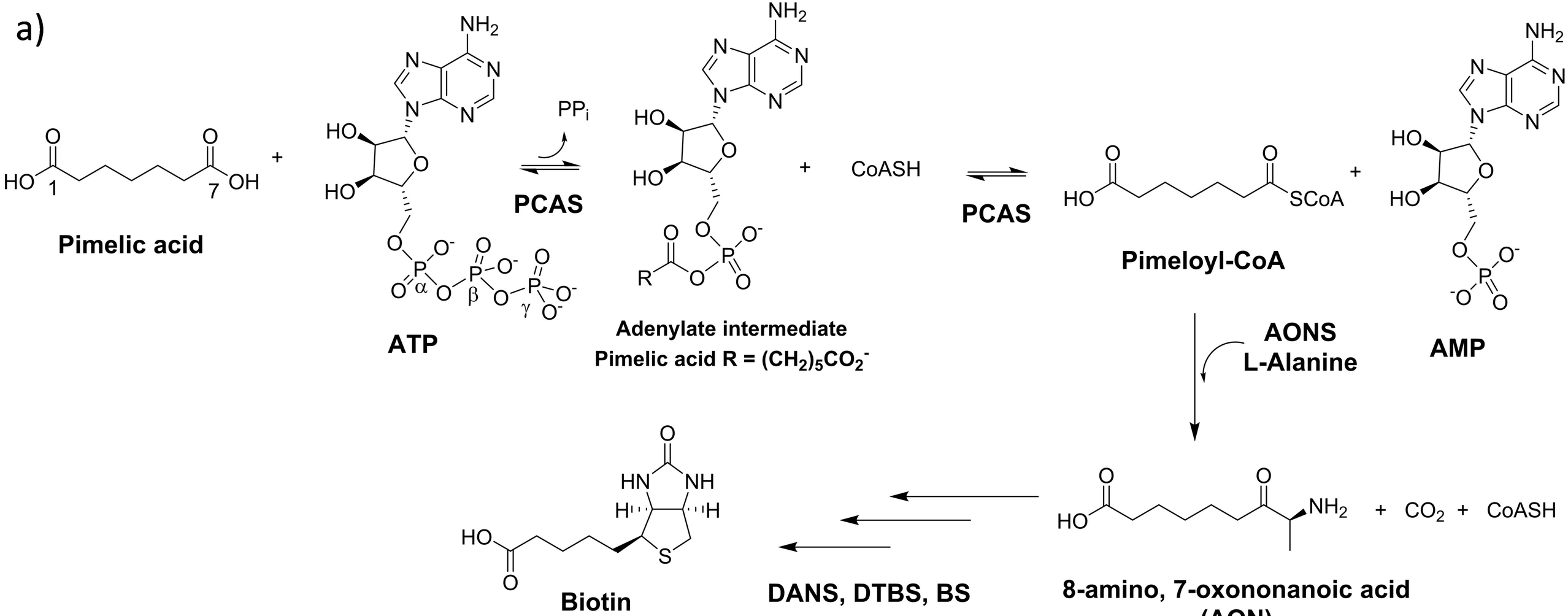
Crystallographic X-Ray crystal structures of all BioW structures reported have been deposited in the Protein Data Bank under accession codes: 5fm0 (BioW:K_2PtCl_4 soak), 5flg (BioW-AMP/PNP/pimelic acid), 5fl1 (BioW: pimeloyl-adenylate/PP_i), 5g1f (BioW: pimelic acid/CoASH soak). The research data supporting this publication can be accessed at Datashare [http://dx.doi.org/10.7488/ds/1560].

Online Methods References

43. Liu, H. & Naismith, J.H. A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif* **63**, 102-11 (2009).
44. Liu, H. & Naismith, J.H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91 (2008).
45. Ploux, O. & Marquet, A. The 8-amino-7-oxopelargonate synthase from Bacillus sphaericus. Purification and preliminary characterization of the cloned enzyme overproduced in Escherichia coli. *Biochem J* **283 (Pt 2)**, 327-31 (1992).
46. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190 (2010).
47. Zhang, Z., Sauter, N.K., van den Bedem, H., Snell, G. & Deacon, A.M. Automated diffraction image analysis and spot searching for high-throughput crystal screening. *Journal of Applied Crystallography* **39**, 112-119 (2006).

48. Sauter, N.K., Grosse-Kunstleve, R.W. & Adams, P.D. Robust indexing for automatic data collection. *Journal of Applied Crystallography* **37**, 399-409 (2004).
49. Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *Journal of Applied Crystallography* **26**, 795-800 (1993).
50. Evans, P. Scaling and assessment of data quality. *Acta Crystallographica Section D-Biological Crystallography* **62**, 72-82 (2006).
51. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymol., Macromolecular Crystallography, Pt A* **276**, 307-326 (1997).
52. Skubak, P. & Pannu, N.S. Automatic protein structure solution from weak X-ray data. *Nat Commun* **4**, 2777 (2013).
53. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D-Biological Crystallography* **67**, 235-242 (2011).
54. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
55. Terwilliger, T.C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* **64**, 61-9 (2008).
56. McCoy, A.J. et al. Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674 (2007).
57. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-55 (1997).
58. Schuttelkopf, A.W. & van Aalten, D.M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **60**, 1355-63 (2004).
59. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21 (2010).
60. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

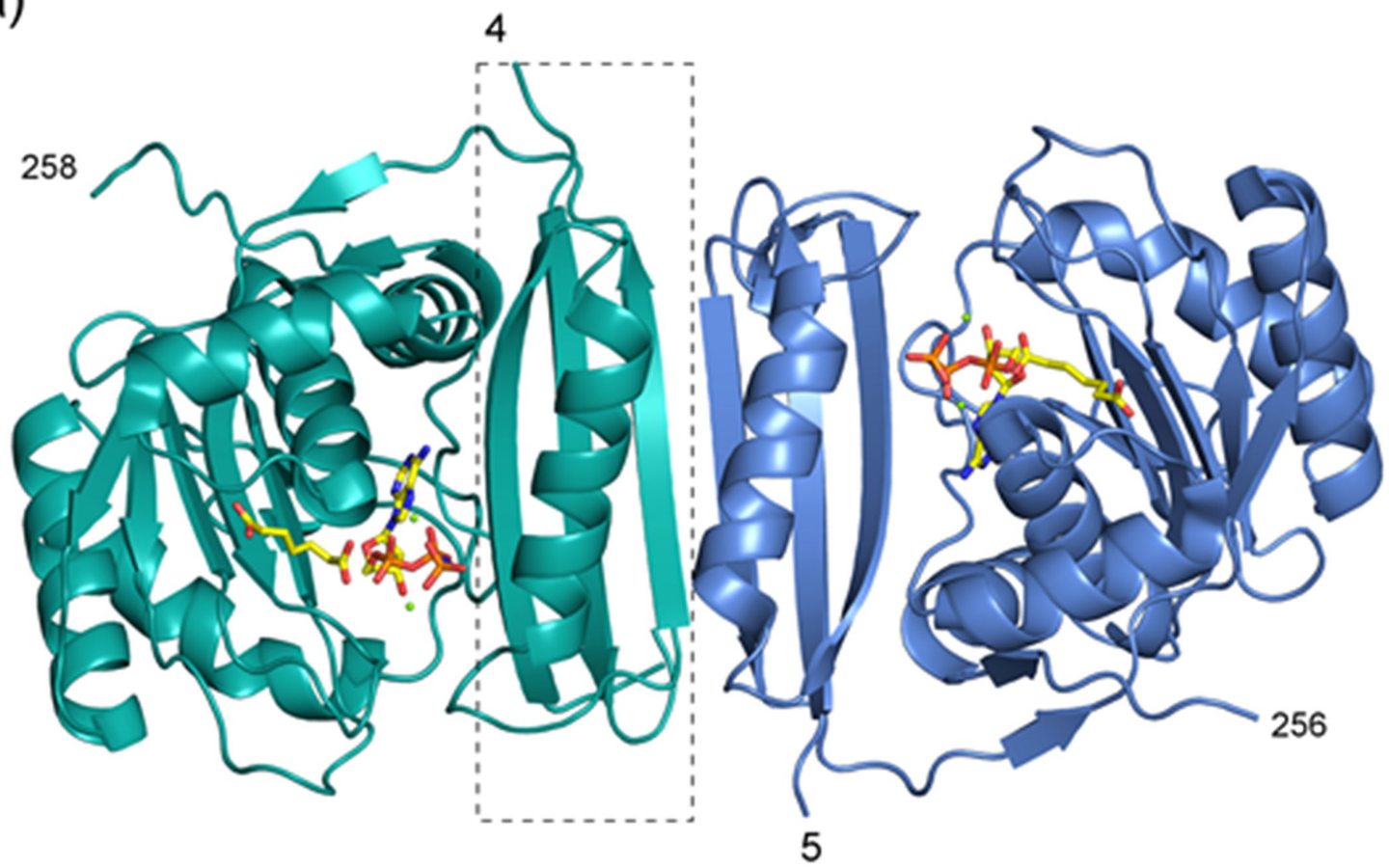
a)


Pimelic acid
ATP
PPi
PCAS
Adenylate intermediate
Pimelic acid R = (CH2)5CO2-
CoASH
PCAS
Pimeloyl-CoA
AMP
AONS
L-Alanine
8-amino, 7-oxononanoic acid
(AON)
+ CO2 + CoASH
DANS, DTBS, BS
Biotin

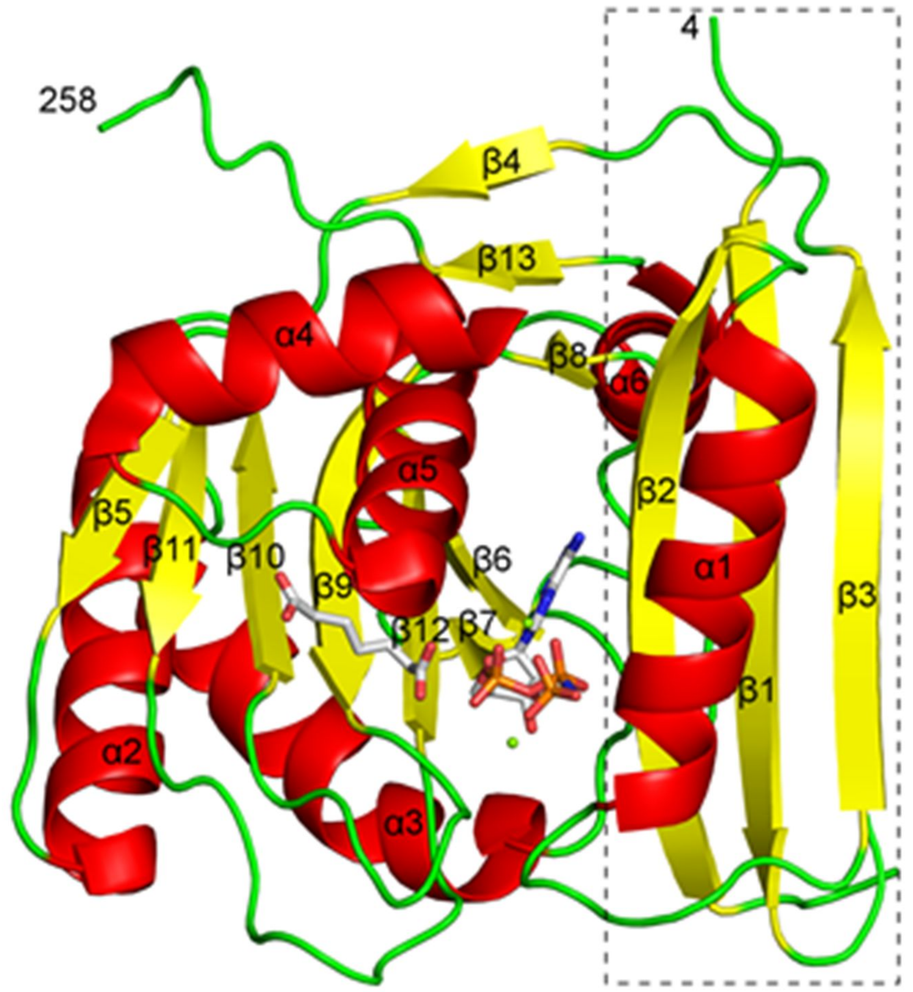
b)

Turnover (s-1)
5.50x10-1
4.40x10-1
3.30x10-1
2.20x10-1
6.0x10-3
4.0x10-3
2.0x10-3
0.0
Glutaric Acid (3)
Adipic Acid (4)
Pimelic Acid (1)
Suberic Acid (5)
Azelaic Acid (6)
Heptanoic Acid (7)
Octanoic Acid (8)

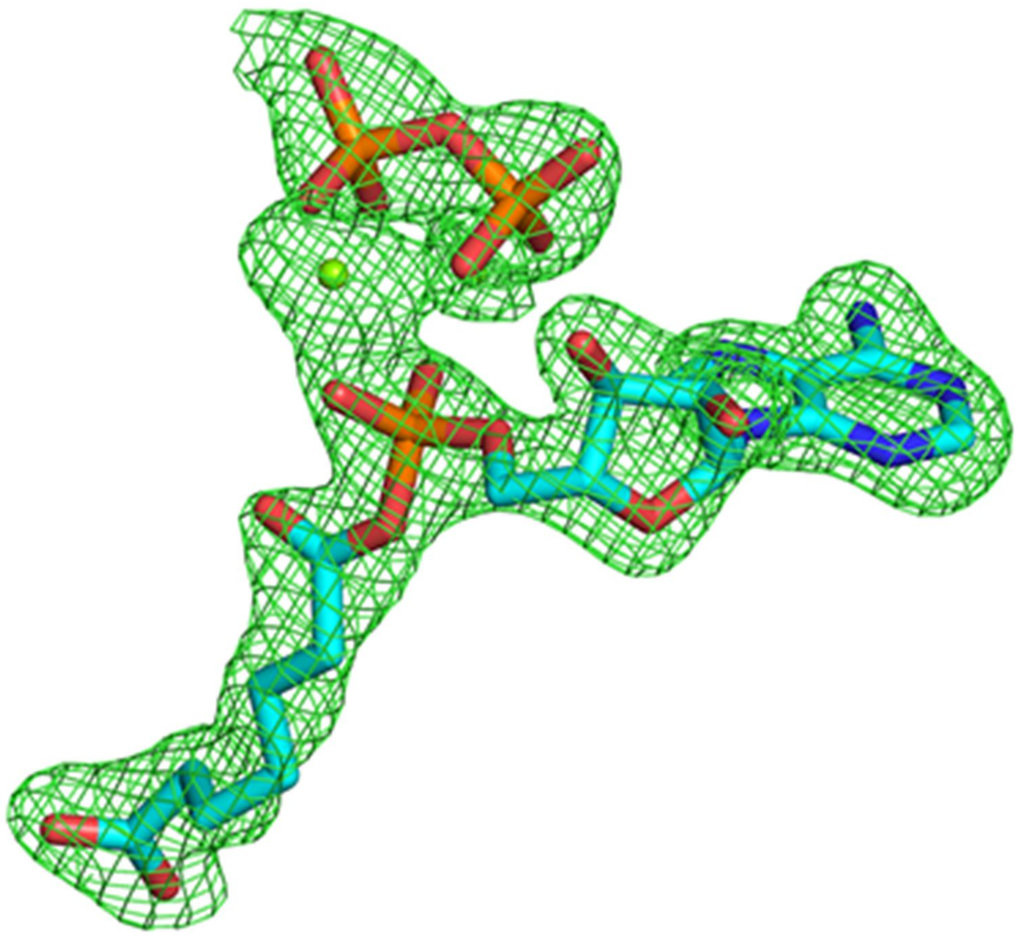
a)


4
258
256
5

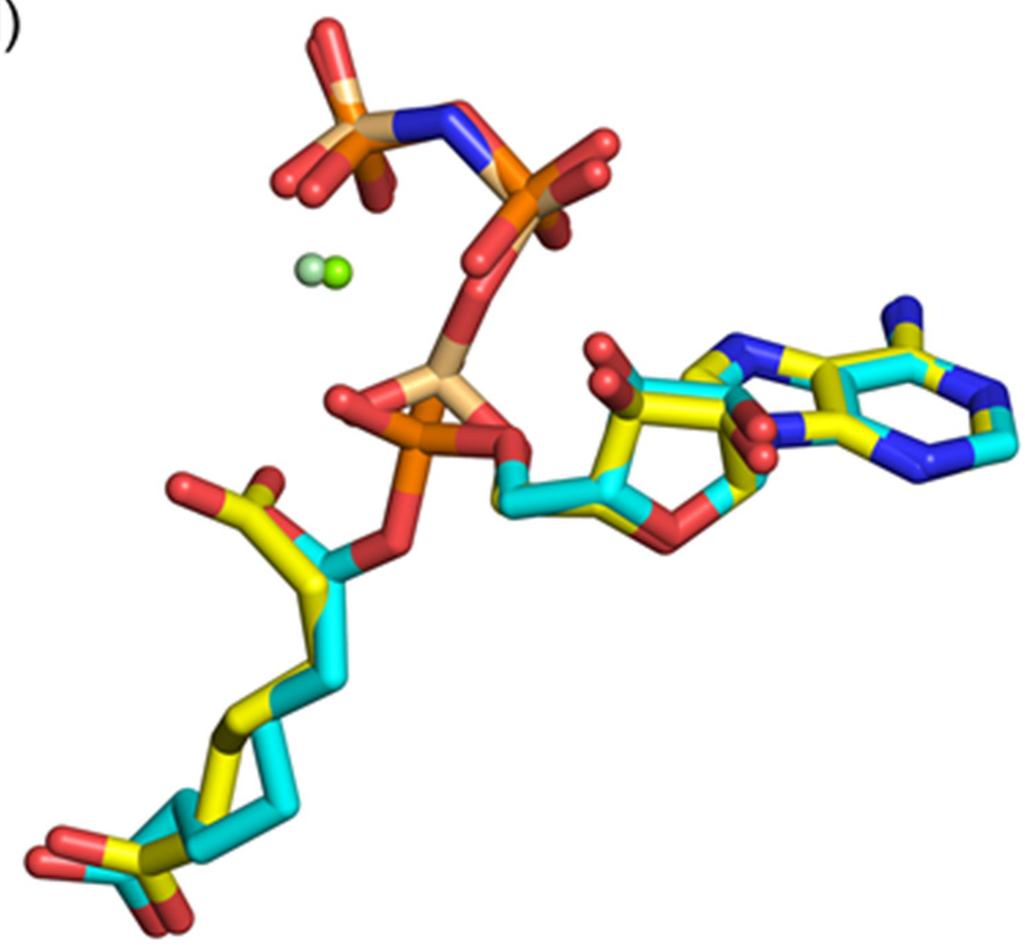
b)


4
258
β4
β13
α4
β8
α6
α5
β2
β5
β11
β10
β9
β6
α1
β3
β12
β7
β1
α2
α3

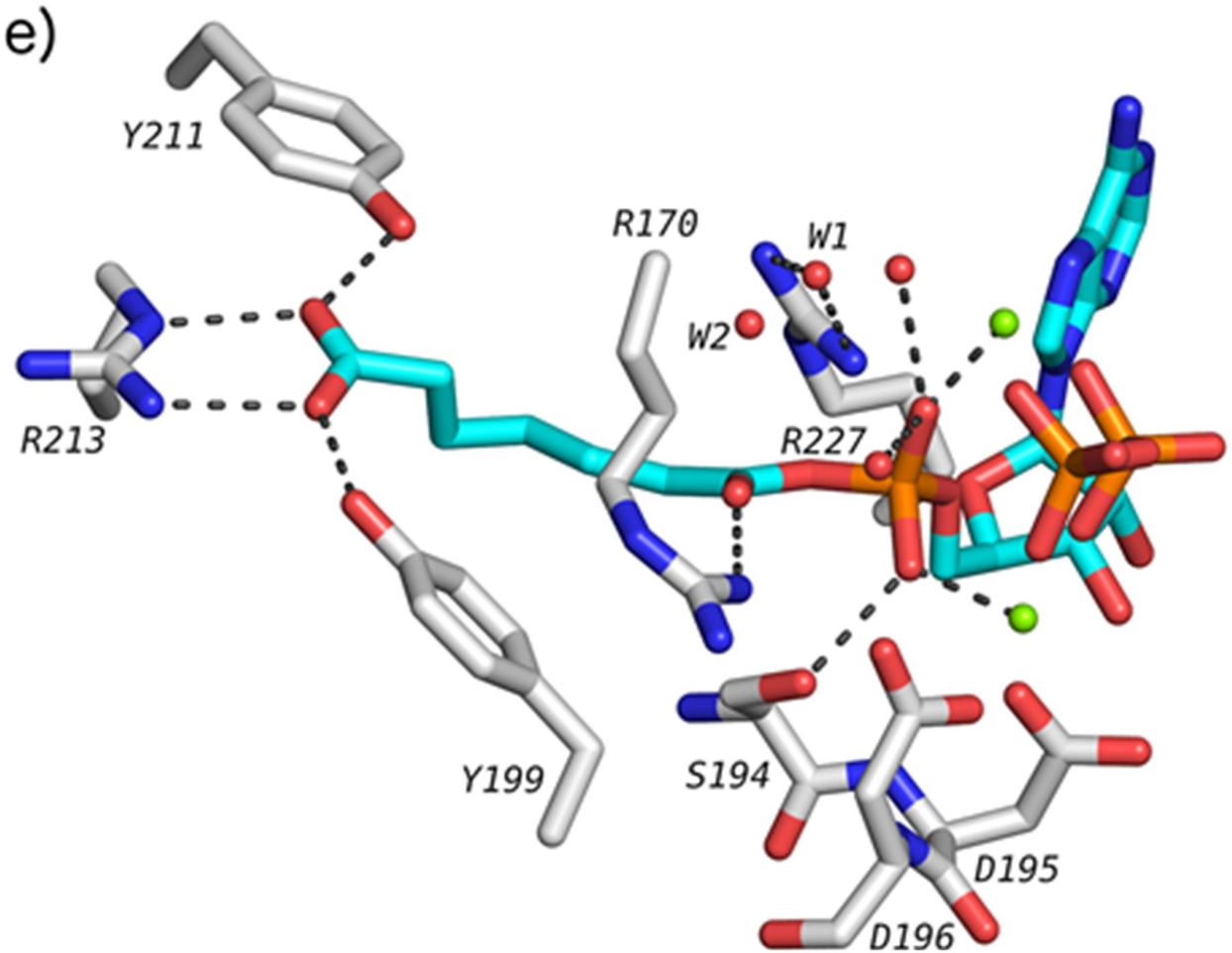
c)



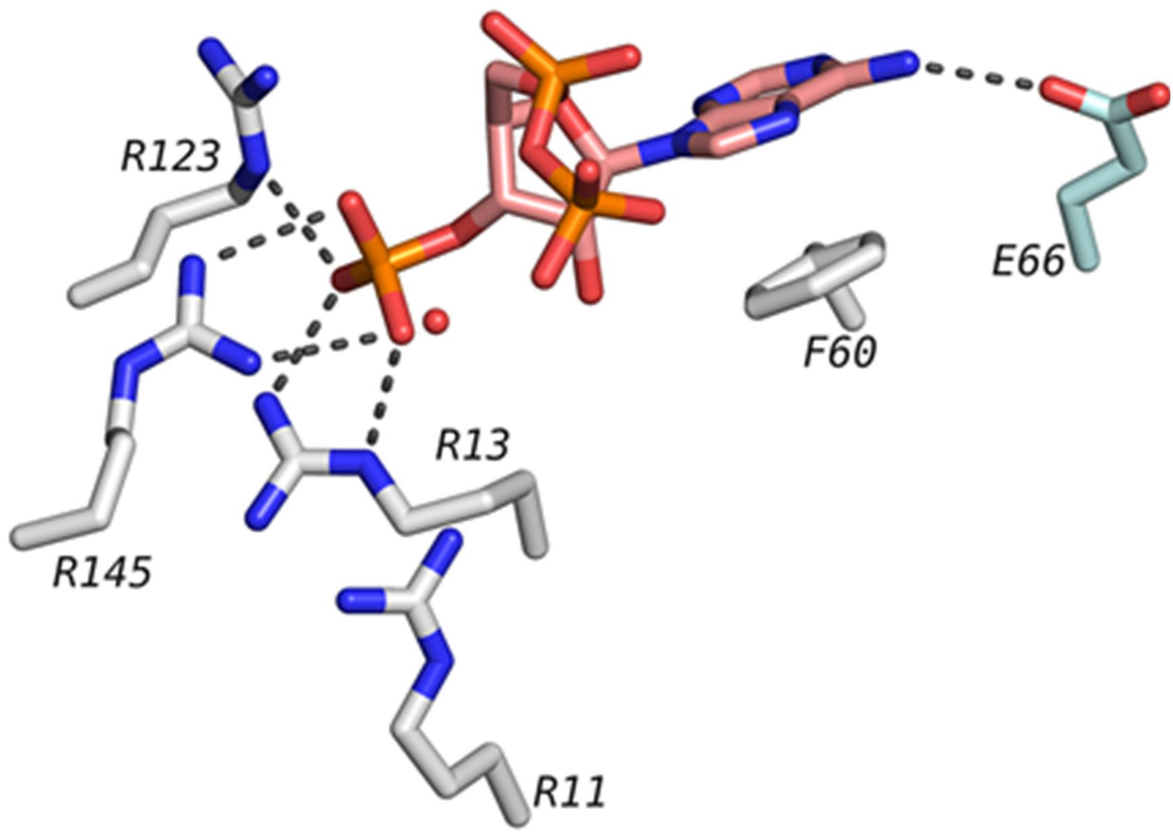
d)



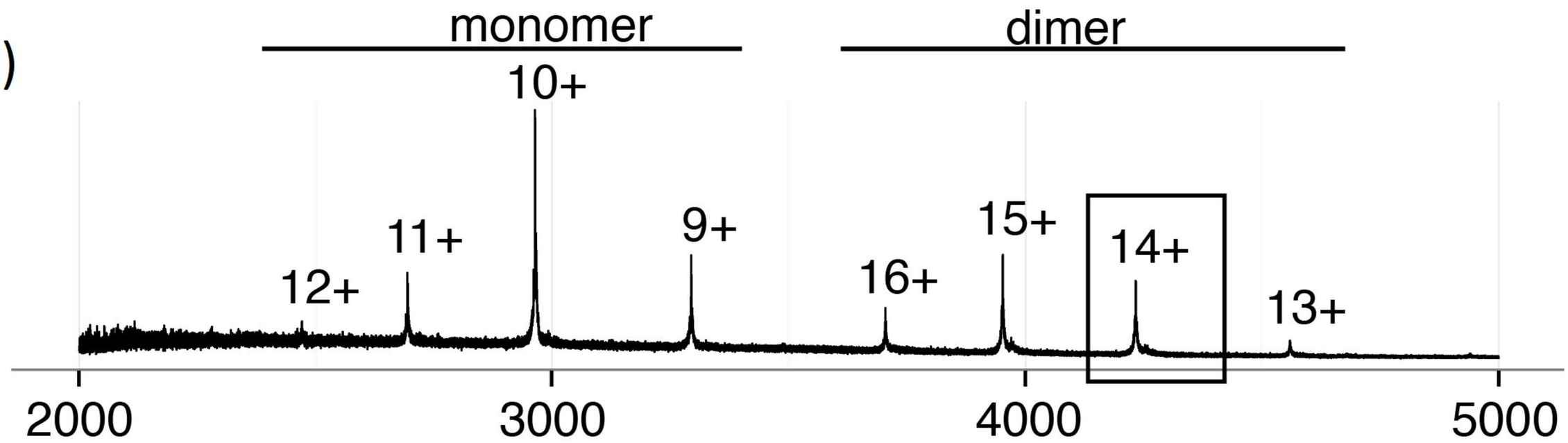
e)


Y211
R170
W1
W2
R213
R227
Y199
S194
D195
D196

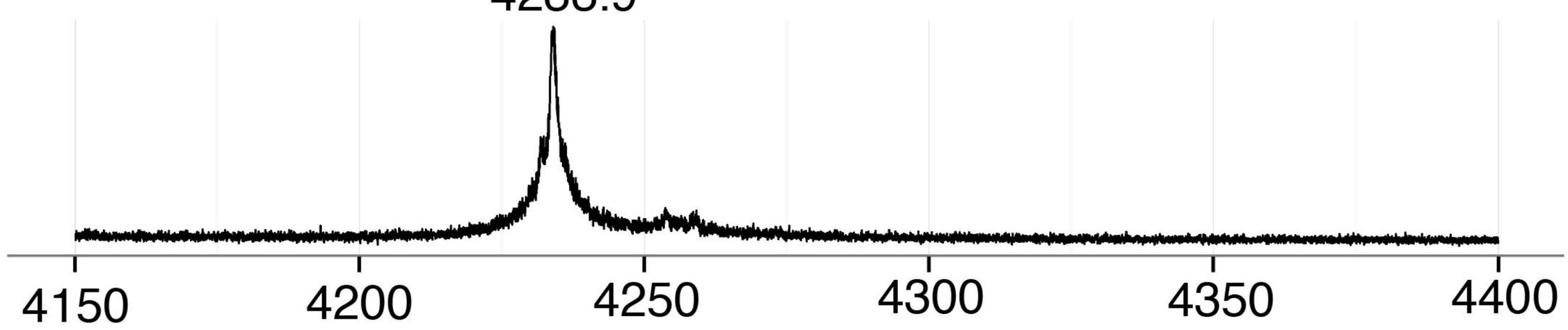
f)


R123
E66
F60
R13
R145
R11

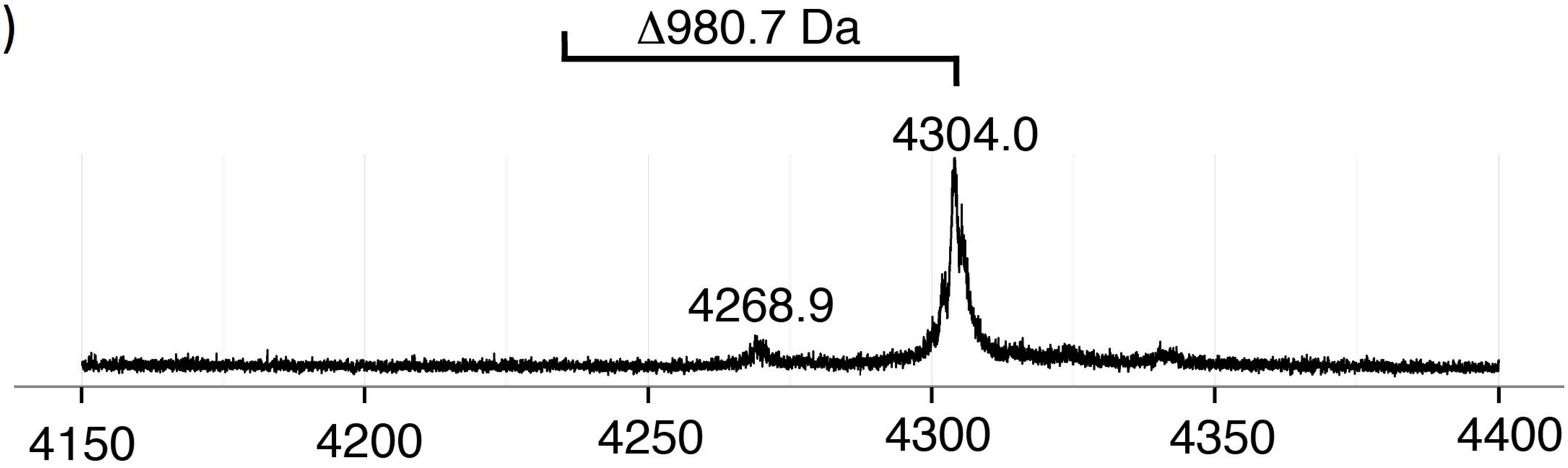
a)



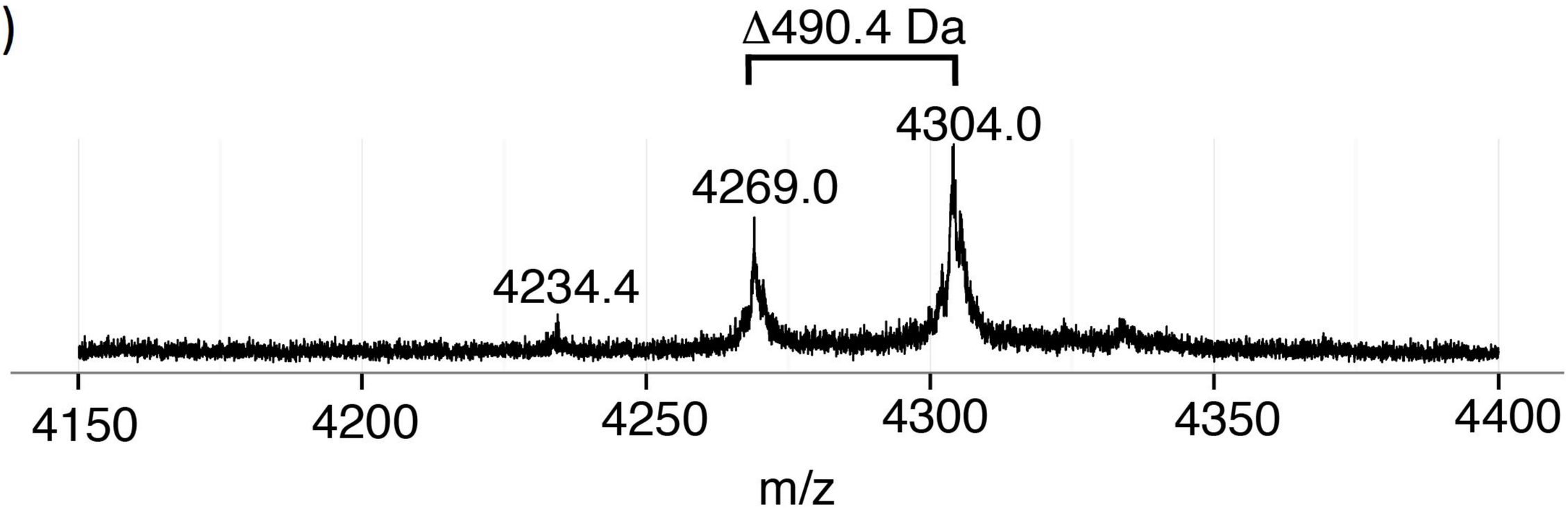
b)



c)

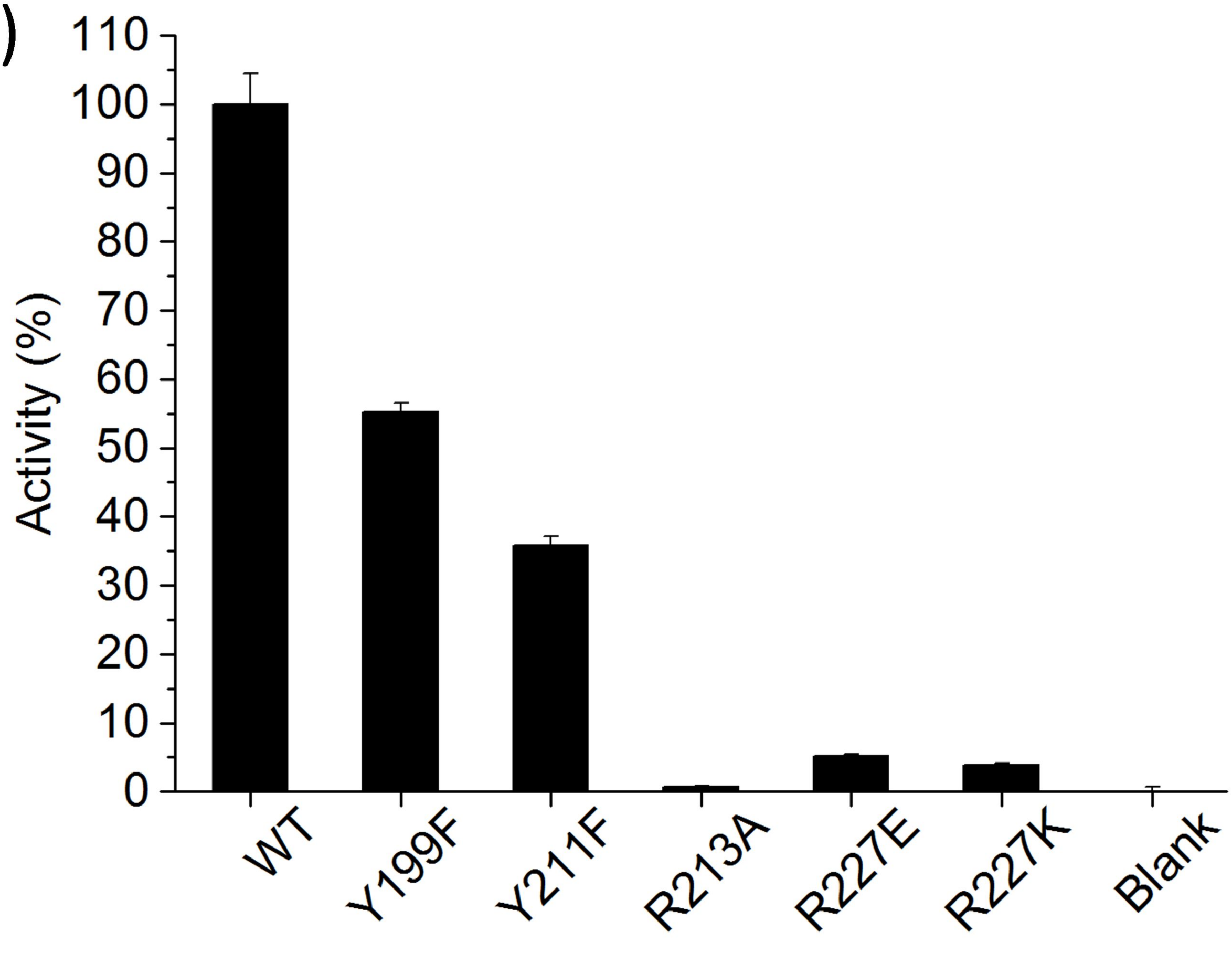


d)

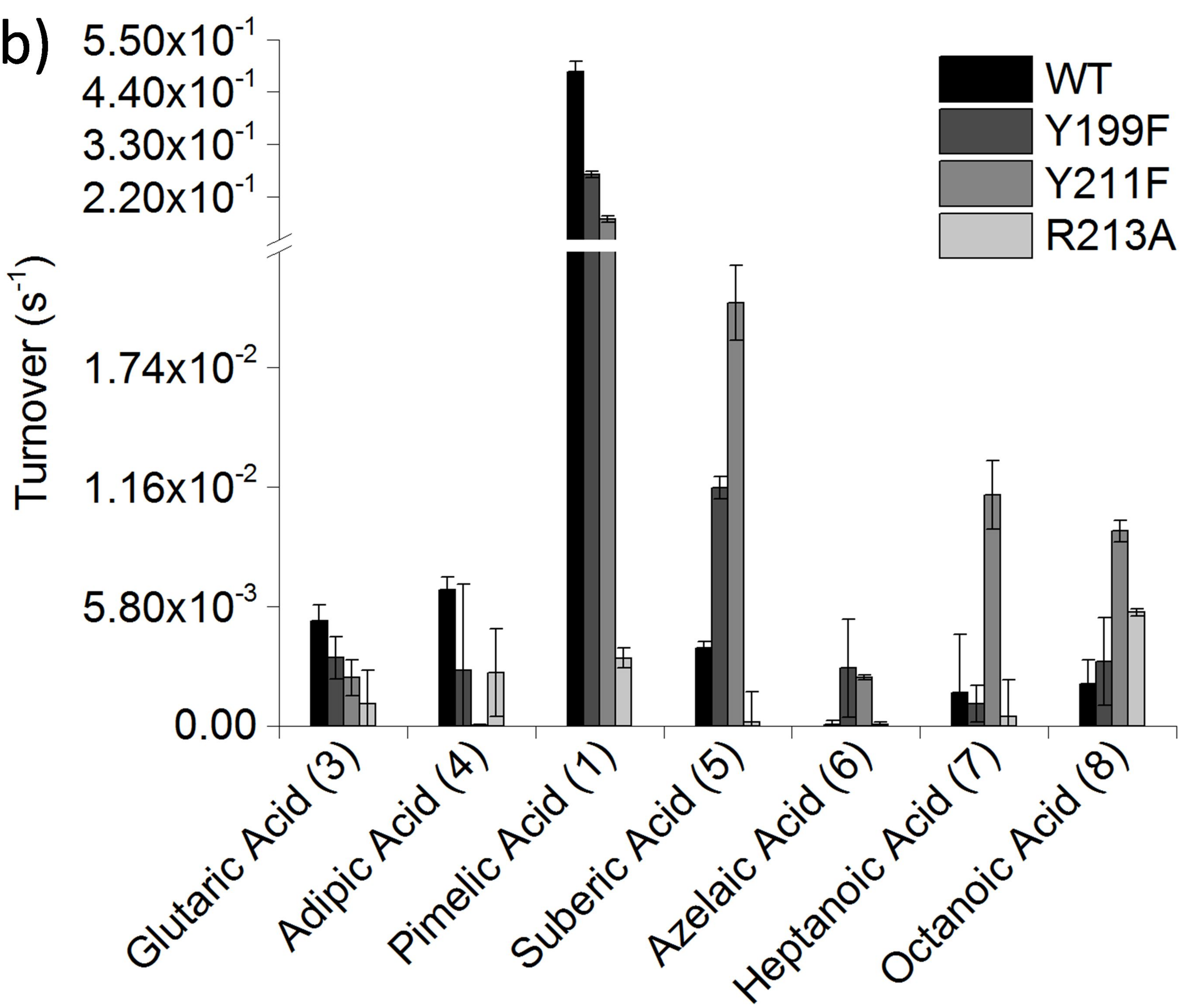


Relative intensity

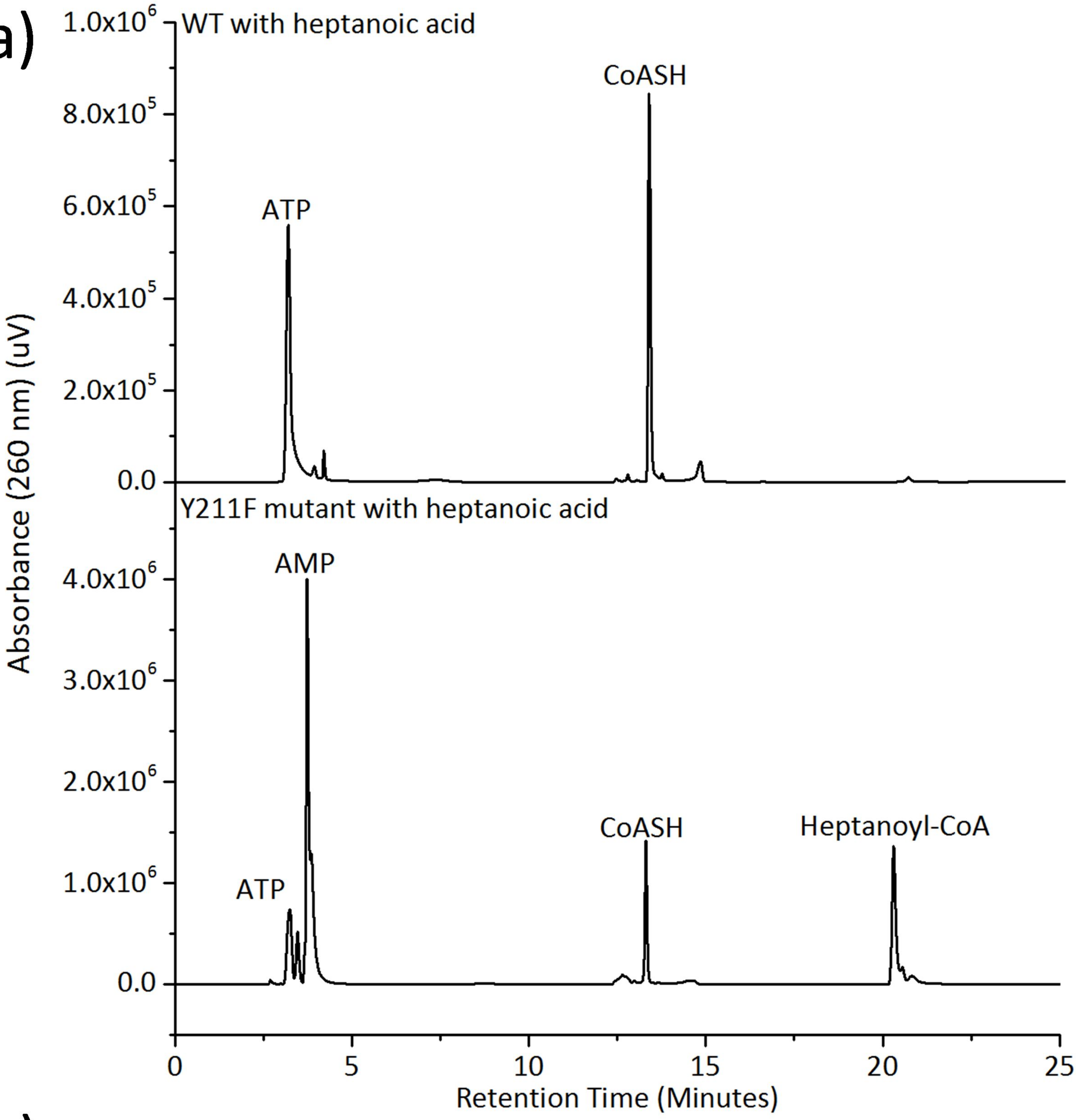
a)


Activity (%)
110
100
90
80
70
60
50
40
30
20
10
0
WT
Y199F
Y211F
R213A
R227E
R227K
Blank

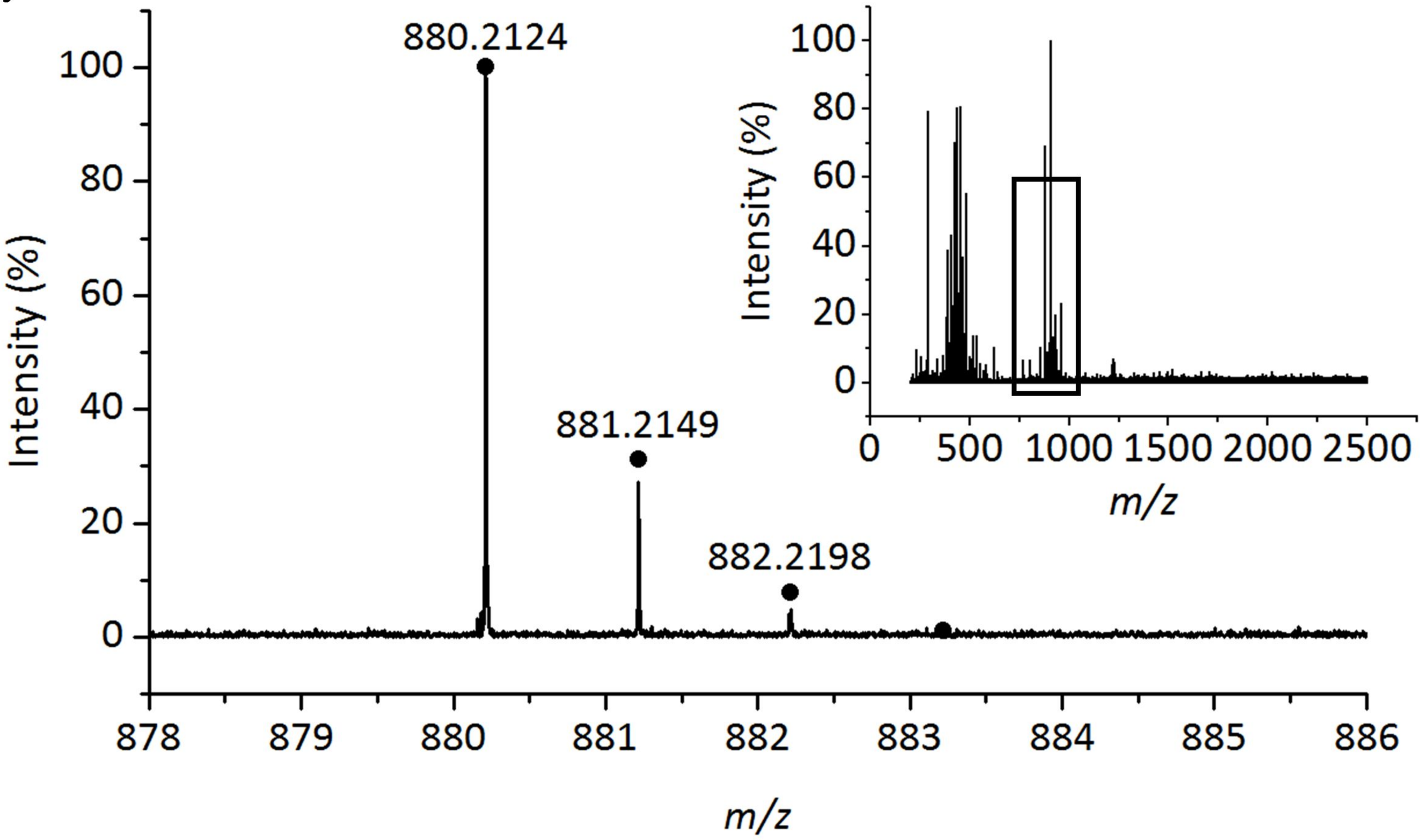
b)

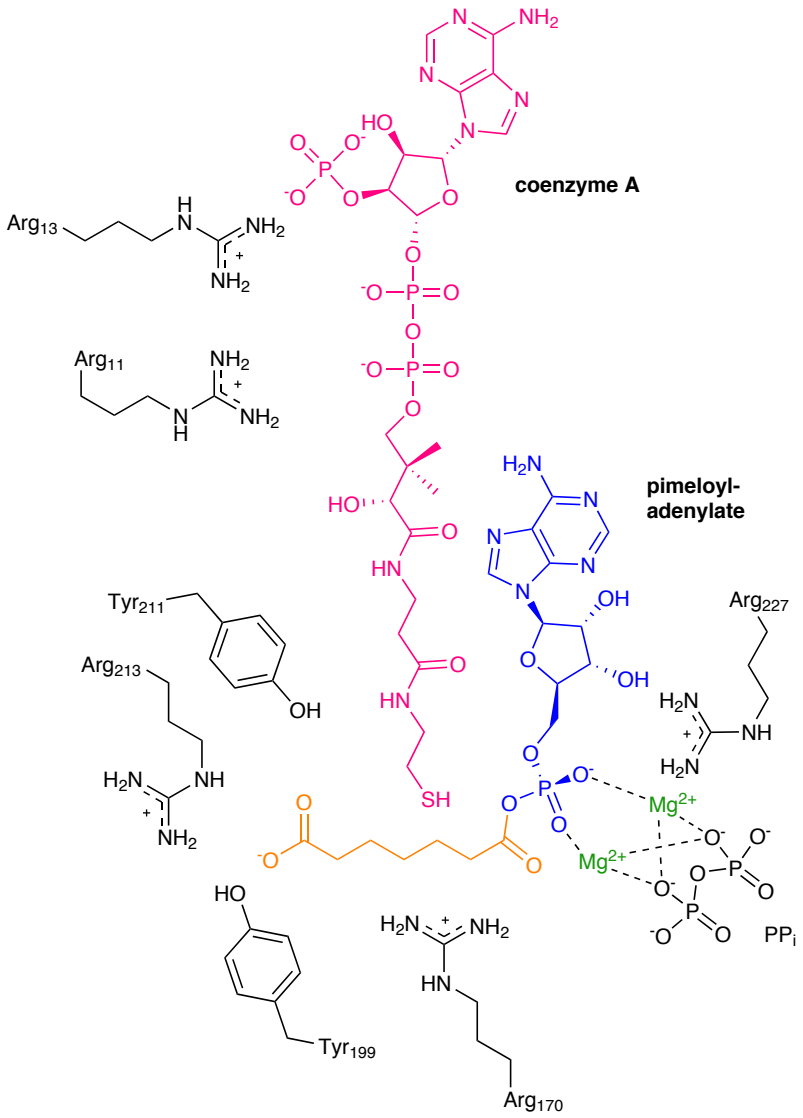

Turnover (s-1)
5.50x10-1
4.40x10-1
3.30x10-1
2.20x10-1
1.74x10-2
1.16x10-2
5.80x10-3
0.00
WT
Y199F
Y211F
R213A
Glutaric Acid (3)
Adipic Acid (4)
Pimelic Acid (1)
Suberic Acid (5)
Azelaic Acid (6)
Heptanoic Acid (7)
Octanoic Acid (8)

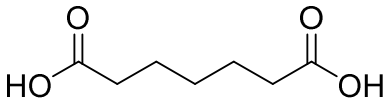
a)


WT with heptanoic acid
ATP
CoASH
Y211F mutant with heptanoic acid
AMP
ATP
CoASH
Heptanoyl-CoA
Absorbance (260 nm) (uV)
Retention Time (Minutes)

b)


880.2124
881.2149
882.2198
Intensity (%)
m/z


coenzyme A
pimeloyl-
adenylate
Arg13
Arg11
Tyr211
Arg213
Arg227
Tyr199
Arg170
Mg2+
Mg2+
PPi

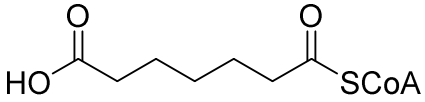


Pimelic Acid

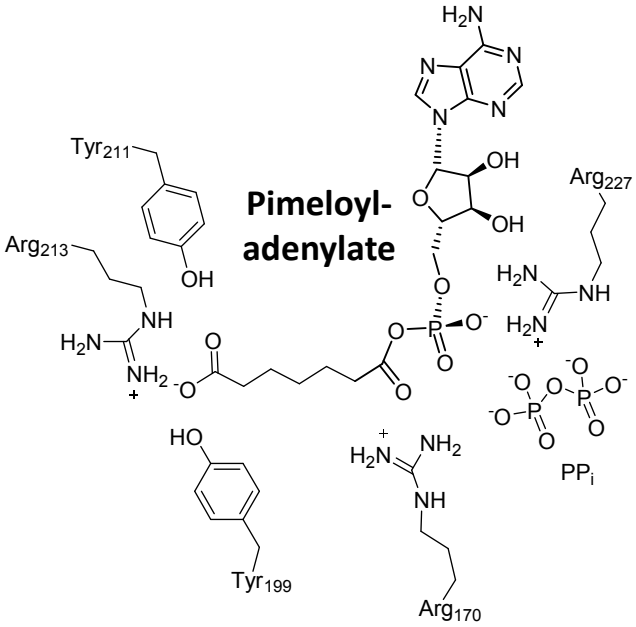
ATP + CoASH

type BioW

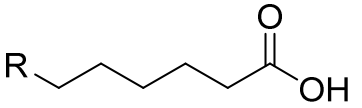
AMP + PP_i



Pimeloyl-CoA



Pimeloyl-adenylate

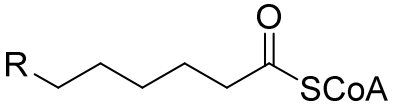


Substituted fatty acids

ATP + CoASH

Engineered BioW Y211F

AMP + PP_i



Substituted fatty acids