STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF ACETOACETATE DECARBOXYLASE-LIKE ENZYMES

by

Lisa S. Mueller

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Chemistry

at

The University of Wisconsin-Milwaukee

August 2016

ABSTRACT

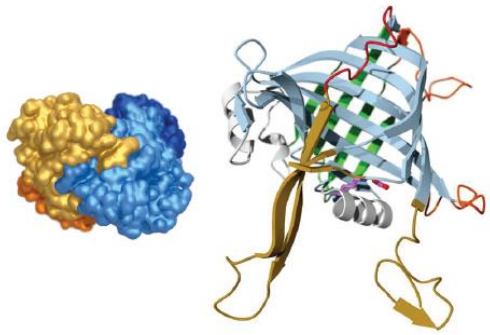
STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF ACETOACETATE DECARBOXYLASE-LIKE ENZYMES

by

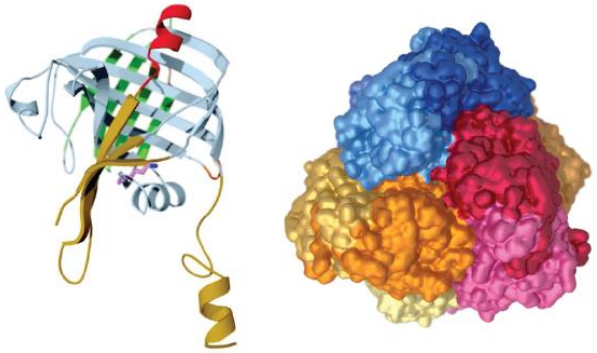
Lisa S. Mueller

The University of Wisconsin-Milwaukee, 2016
Under the Supervision of Professor Nicholas R. Silvaggi

***S. bingchenggensis* Aldolase-Dehydratase**



***C. acetobutylicum* Acetoacetate Decarboxylase**



The acetatoacetate decarboxylase-like superfamily (ADCSF) is a largely unexplored group of enzymes that may be a potential source of new biocatalysts. Bioinformatic analysis has grouped these approximately 2000 enzymes into seven different families based on comparison of predicted active site residues. To date, only the prototypical ADCs (Family I) that catalyze the decarboxylation of acetoacetate have been studied. Analysis of gene context suggests that Family V contains predominantly enzymes predicted to be involved in secondary metabolism. On average, these share about 20% sequence identity to the true ADCs. To learn more about the diversity of chemistries performed by members of Family V, we have been studying two enzymes annotated as "acetoacetate decarboxylase" in the GenBank database. These are Sbi_00515 from *Streptomyces bingchenggensis* and Swit_4259 from *Sphingomonas wittichii*. Steady state

analyses of these enzymes demonstrate that both lack decarboxylase activity with any of the substrates tested. This was surprising given that the crystal structures of both enzymes show that their overall folds are almost indistinguishable from that of the prototypical ADCs, though the quaternary structures are different. An important observation from the bioinformatic and crystallographic analyses is that the catalytic lysine and putative acid/base catalyst residues of the true ADCs are retained in both groups of enzymes, but the active site architectures are different. Specifically, two residues shown to be important for the acetoacetate decarboxylase reaction, Arg29 and Glu61 in *Clostridium acetobutylicum* ADC (CaADC), are not retained in the Family V enzymes. Site-directed mutagenesis, steady state and transient kinetics, and mass spectroscopy data suggest evidence for reversible aldolase-dehydratase and retro-aldolase activities mediated through a Schiff-base mechanism. These are the first Schiff-base-forming aldolases that do not use the TIM barrel fold. Although the physiologically relevant reactions of these Family V enzymes are unknown, these studies illustrate that the ADC fold is a versatile platform that can be adapted to perform different chemistries.

I dedicate my thesis towards advancement in the basic sciences, and myself. For most of my life, I never thought I'd make it this far, let alone achieve anything of significance. Eat it, depression and anxiety.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

LIST OF ABBREVIATIONS

ADC	Acetotacetate decarboxylase
ADCSF	Acetoacetate decarboxylase-like superfamily
BLAST	Basic Local Alignment Sequence Tool
CaADC	*C. acetobutylicum* ADC
CMP	Cinnamylidenepyruvate
CvADC	*C. violaceum* ADC
DSC	Differential scanning calorimetery
KIE	Kinetic isotope effect
4NC	4-nitrocinnamaldehyde
4NCP	4-nitrocinnamylidenepyruvate
PCR	Polymerase chain reaction
PDB	Protein Data Bank
ShMppR	*S. hygroscopicus* MppR
SSM	Site-directed saturation mutagenesis
TIM barrel	Triose phosphate isomerase barrel
UV	Ultraviolet

ACKNOWLEDGEMENTS

These six years have gone by fast. I'd be remiss if I didn't mention those who helped me succeed. First, I am incredibly grateful to my advisor, Nicholas R. Silvaggi, for taking me on as his first graduate student. He saw my potential, brought it to the surface, and shaped me into a biochemist and a better person. The Silvaggi research group was wonderful to be a part of, and I enjoyed working with individuals dedicated to their work and collaboration in learning new techniques.

I would not have understood the nuances of my projects without the aid of my advisor and thesis committee. I am thankful for their interdisciplinary perspectives and patience for my numerous questions, concerns, and overthinking. Their courses and office hours were an invaluable resource to my overall comprehension of the biochemical field. In addition, their research groups welcomed team work when I had questions about their specialties.

I am forever glad to have received the AOP Fellowship from UW-Milwaukee Graduate School. Coming from my background, I had a lot of doubts in my ability to do well in post-baccalaureate work. That award changed my perspective on myself, and allowed me to focus on my research without reserve.

My sanity outside of lab was retained by my husband and close friends, who are my family. The "Ladies of Chemistry," my husband, and close friends reminded me about the big picture of work-life balance, cheered me up after any tough day, and encouraged me to keep striving for improvement. I appreciate and love them in more ways than I can ever express

CHAPTER I: INTRODUCTION

1.1 Biocatalysis

Biocatalysts are attractive tools for green chemistry because they are naturally regio-, chemo-, and stereospecific, and have evolved to catalyze reactions under mild conditions in aqueous solvent [7]. Green chemistry aims to reduce or eliminate the creation or usage of hazardous materials at all steps of chemical production, including reducing the environmental impact from of chemical production processes [8]. It also aims to reduce resource usage by eliminating the need for high or low temperatures or pressures to facilitate catalysis. To date, more than 100 biocatalysts have been used in industrial applications [9], including the pharmaceutical manufacture of Sitagliptin [10], a drug used in the treatment of Type II diabetes, the production of high-fructose corn syrup [11], and environmental remediation, such as aromatic hydrocarbon degradation [12]. In addition, biocatalysts can be used in research or diagnostic roles [13]. Another advantage for biocatalysts is that for a good candidate, they have about 100 % enantiomeric excess (ee) for chiral molecules. The choice of biocatalyst depends on a number of factors, including whole cells or isolated enzymes, the chemistry, substrate specificity, catalytic efficiency, and/or stability at various temperatures or in organic solvents [14]. The paradox for enzymes in organic synthesis is that their specificity, their major strength, also limits their usefulness, because it is difficult to find an enzyme to do the required chemistry with the desired starting materials. This limitation can be overcome by engineering an existing enzyme to fit a specific reaction, or by finding new enzymes with different activity or substrate specificity profiles.

Finding the enzyme that catalyzes the desired reaction could obviate the need for enzyme engineering. In many cases, it is possible to find homologs of known enzymes with improved properties, such as altered substrate specificity, different stereochemistry of the product, altered pH or temperature optima, increased activity, and/or increased temperature or solvent stability. Genome mining has become an effective tool for identifying enzymes with potentially useful catalytic activities. The amino acid sequence of an enzyme of interest can be used to find homologous sequences in the database (*e.g.* the Basic Local Alignment Search Tool [BLAST]). Detailed analysis of the resulting multiple sequence alignments can lead to identification of differences in active site residues that may impact catalytic activity [15, 16]. Nonetheless, functional annotation in sequence databases is often incomplete or wrong [17]. The challenge remains to make an educated guess on protein function from sequence information alone. More information on potentially novel enzymes can be gleaned from sequence similarity networks (SSNs) and the genome neighborhood network (GNN) [18, 19]. Other large scale initiatives to mine data range from protein-protein interactions [20], metabolic pathways [21], to databases specifically for bacteria, archaea, and fungi [22]. Once a new enzyme has been selected for study, it needs to be expressed, purified, and tested for activity, substrate specificity, and stability. During or after that process, the enzyme may need further engineering.

1.2 Protein Engineering

Protein engineering can proceed by three different methods, each with sub-techniques for altering protein properties to meet process conditions: (1) rational design, (2) directed evolution, and (3) computational design [13, 23]. First, rational design uses structural and

mechanistic information, as well as molecular modeling for prediction of changes in the protein structure in order to change its properties [13]. For instance, changing an active site Tyr to Phe by studying an X-ray crystallographic structure by site directed mutagenesis uses rational design. Next, directed evolution comprises of two major steps: random mutagenesis of the gene encoding the enzyme, and identification of desired variants within the mutant library by screening or selection [13]. The most common means of generating the necessary genetic variation are recombination-based methods (*e.g.* gene shuffling) and error-prone PCR. In gene shuffling, related genes are digested with Deoxyribonuclease I (DNase I) and reassembled by polymerase chain reaction (PCR) [24]. This technique may take place *in vitro* or *in vivo* and relies on annealing of homologous sequences on separate fragments of the gene [25]. Error-prone PCR is another way to achieve random mutagenesis in the hope of obtaining a protein variant with the desired properties. This *in vitro* technique takes advantage of a DNA polymerase without proof-reading ability, and/or manipulating the reaction conditions by adding divalent manganese or magnesium during PCR amplification of a gene, or unbalancing nucleotide concentrations [26, 27]. All of these manipulations cause the polymerase to make far more mistakes than it would under optimal conditions. Site-directed saturation mutagenesis (SSM) combines elements of directed evolution with rational design. Rational design targets specific residues with potential roles in catalysis, substrate binding, or release. SSM samples all 20 of the proteinogenic amino acids at a limited number of positions that are identified in structural and/or enzyme kinetic studies of the parent enzyme. Lastly, computational tools like the HotSpot Wizard [13] also aid in the semi-rational design process. This computational algorithm takes into account the protein structure, hypothetically mutating residues in the active site or those lining access tunnels, hence “hot spots”, that the

user may choose to alter by site-directed mutagenesis [23, 28]. In the case where an experimental structure of a protein is not available in the Protein Data Bank (PDB), a homology model can be made. This takes the sequence of the protein of interest, combined with a known structure of a homologous protein with significant sequence identity [29]. The backbone template with desired sequence can be used as a guide for making amino acid mutations. A recent publication has made the push for homology modeling and molecular docking together, which could greatly impact the residues to be mutated [30].

Directed evolution and SSM can generate huge libraries, on the order of $10^6 – 10^9$ variants, all of which must be assayed for the property being optimized [31]. Thus, the primary limitation in directed evolution experiments is the identification of mutants with desirable properties [32]. A 200 amino acid protein with two amino acid substitutions anywhere in the sequence generates 7,183,900 possible variants, three mutations in the same protein produces 9,008,610,600 possibilities [33]. All variants can be screened, or a partial library can be analyzed, with further evolution done on the best mutants from the first library. High-throughput methods are necessary in order to resolve the mutants in a timely manner. Fluorescence-activated cell sorting (FACS) can monitor changes in substrate specificity and stereochemistry in a few hours, with up to 10^8 variants [34]. This method can be coupled to cell surface display, *in vitro* compartmentalization (IVC), or single cell compartmentalization [34]. An *in vivo* selection method with FACS has also been described, by coupling an enantiomer to both a specific carbon source needed for growth and a toxic compound as a growth inhibitor to eliminate null mutation [35]. A limitation of these approaches is that they are not run under realistic conditions; when subject to physiological conditions or substrates, the hits from FACS may not perform as expected [31]. Two

recent papers have found a way around this, yet still incorporating fluorescence. Chen *et al*. developed microcapillary single-cell analysis and laser extraction (µSCALE), which begins with variant libraries expressed in bacteria (*Escherichia coli*) or yeast (*Saccharomyces cerevisiae*) [36]. The method sorted single cells into a microcapillary, and fluorescence microscopy is used on the microcapillary array. The microscope can detect emitting antibodies bound on the surface of a yeast cell, fluorescent proteins in *E. coli* cells, or real-time kinetic measurements of enzyme activity when the enzyme is bound to the surface of *S. cerevisiae* (provided the reaction produces a fluorescent product [36]). The hits can be retrieved by a precise laser extraction system [31, 36]. Colin *et al*. used a metagenomic library of 1.25 x 10^6 variants, pooled from different sources, and expressed in *E. coli* with hopes of identifying promiscuous enzyme activity [37]. Single bacterial cells were encapsulated into picoliter water-in-oil droplets, each presumed to contain one bacterial cell with a different library member, and incubated with substrate and lysis agents. After single cell lysis, the emulsion droplet was re-inserted into a microfluidic sorting chip, and diverted based on fluorescence intensity of a leaving group [31, 37].

Semi-rational design principles discussed were applied by Pavlova *et al.* to develop an enzyme to break down the recalcitrant toxin 1,2,3-trichloropropane (TCP) [38]. TCP is a byproduct in the manufacture of chlorinated compounds, it is used as a solvent, a component of paint and varnish remover, and is proposed to be a human carcinogen [39]. TCP can be converted to 2,3-dichloropropane-1-ol (DCL) by the haloalkane dehalogenase DhaA from *Rhodococcus rhodochrous*, though the rate of this reaction is too modest to support a viable remediation process. Random acceleration molecular dynamics (RAMD) was applied to the X-ray crystallographic structure of WT DhaA from *R. rhodochrous* to simulate the release of product,

the rate-limiting step, from the enzyme active site [38, 40]. This simulation highlighted residues that might affect the movement of substrate and product through the access tunnels that connect the buried active site to bulk solvent to enhance conversion of TCP. Previously, two residues were found to increase activity 3-4 times by DNA shuffling and error-prone PCR [41]. Pavlova *et al*., however, focused on further enhancement of TCP to DCL conversion with updated methods, including SSM to vary the nucleotide codon by more than one base pair at a time to create larger screening libraries. The two previous mutants were starting points for directed evolution. More than 5,000 clones were picked, grown, and assayed for TCP conversion using a high throughput colorimetric assay based on phenol red, a pH-sensitive dye. The vast majority of DhaA variants, 96%, showed little to no improvement over WT. However, 25 clones with greater activity than WT were ultimately sequenced and found to have unique mutations. The steady state kinetics were monitored at 460 nm by mercuric thiocyanate and ferric ammonium sulfate, a more sensitive assay than phenol red. In the end, five mutants were generated by site-directed mutagenesis in the access tunnel to the active site. These mutations hindered water from getting to the active site as suggested by RAMD, yet increased affinity for TCP, and increased k_{cat} 32-fold and k_{cat}/K_M 26-fold over the WT values. This study illustrates the utility of generating large mutant libraries combined with high-throughput screening for protein engineering.

Semi-rational protein engineering and design has gained traction over the large combinatorial libraries of directed evolution [42]. Semi-rational approaches can harbor smaller, functionally rich libraries of variants, and reduce the reliance on high-throughput screening.

De novo design is the last method currently used to engineer enzymes, but it may be the most difficult method to execute. First, a chemically plausible transition state for the desired

reaction is designed, then an active site that stabilizes that transition state is erected around it by quantum mechanics/molecular mechanics (QM/MM) calculations [23, 43]. That transition state and active site are then put into an existing protein scaffold and designed using a protein modelling software package such as ROSETTA [44, 45]. This *in silico*/*in vitro* technique requires

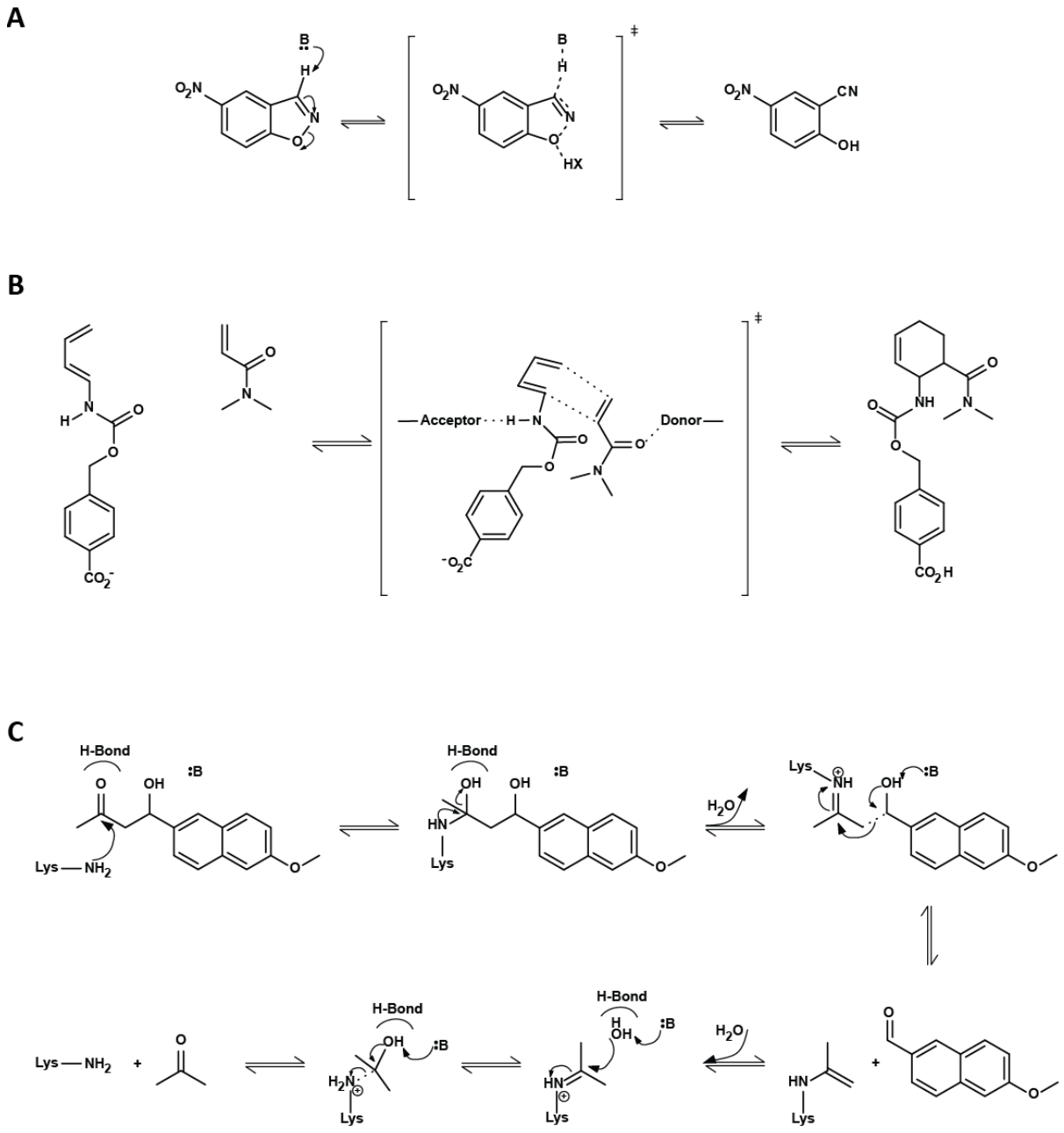


Figure 1. Enzymes have been engineered to perform the Kemp elimination (**A**) [1], Diels-Alder reaction (**B**) [3], retro-aldol cleavage (**C**) [4]. Both the Kemp and Diels-Alder reactions are not found in nature.

structural information and is most useful for designing enzymes specifically for reactions found not to occur in nature to aid in synthetic organic chemistry. Examples of *in silico*-designed proteins have been more encouraging in recent years, though kinetic rates are still much slower than natural enzymes [46]. Directed evolution performed after the *de novo* design has been found to drastically improve activity [44]. Notably, the Baker laboratory has engineered enzymes in this way to perform the non-natural (i.e. not found in biological systems) reactions shown in Figure 1. These include the Kemp elimination (**A**), the Diels-Alder reaction (**B**), and retro-aldol cleavage (**C**) [1, 3, 4]. They were first to tackle the challenge a multistep reaction with a non-natural substrate [4]. Retro-aldol cleavage has been described as the most difficult designed enzyme to date due to its multistep reaction. Multistep reactions are more difficult to engineer, since multiple transition states and intermediates must be considered simultaneously in the design of a composite transition state in order for the reaction to occur [47-49].

Three considerations were given priority for the *in silico* design of retro-aldol cleavage, borrowing themes from naturally occurring Type I aldolases: a hydrophobic pocket to lower the pK_a of the catalytic lysine, proton transfer from an ordered water molecule, and substrate binding interactions [47]. The three major protein scaffolds used were the jelly roll (PDB ID 1f5j), nuclear transport factor 2 (NTF2) (PDB ID 1oho), and triosephosphate isomerase (TIM) barrel (PDB ID 1lbl) folds. Various positions within the active site were tested for the catalytic lysine to optimize the Schiff base chemistry. Rational design was first done to residues in the active site to study the relationship between sequence and activity. Surface residues were next altered to improve solubility. Once an apparent optimum had been reached as determined by steady state kinetics, directed evolution was done. Error prone PCR and DNA shuffling resulted in variants with

improved activity. In the improved variants, the mutations were found typically throughout the protein. The most successful mutant was a TIM barrel that acquired 11 overall amino acid changes concentrated largely in the active site, with a few occurring on the surface. Second shell and buried residues were also changed. It is proposed that these mutations led to a different conformation of an active site loop, as supported by lower *B*-factors, in the crystallographic structure compared to earlier models [48]. A *B*-factor is a measurement related to the atomic displacement of an atom from its average position [50]. Overall, the information gleaned from this study was the need for a flexible protein backbone rather than a rigid structure, and supporting functional groups throughout the protein to increase activity.

The information gained from protein engineering can be applied to interpretation of other work and facilitate understanding enzymes as a whole. Subtle changes of a significant residue in a protein sequence can change protein activity, binding, and/or stability. Instances throughout history can tell the story of biocatalysts before directed evolution or *de novo* work were considered.

1.3 Acetoacetate Decarboxylase

Clostridium acetobutylicum is one of the first examples of an industrial, whole cell biocatalyst, that later became well-studied for its specific contributions to acetone, butanol, and ethanol production. It was in 1905 that acetone was discovered as a product of microbial fermentation by Schardinger, and in 1913 Chaim Weizmann discovered the anaerobic bacterium that would be named *C. acetobutylicum* [51]. Weizmann began producing acetone and butanol from potatoes on an industrial scale in Britain, though the microbe could accept other starchy

substances, as well as glucose. Despite the small initial industrial process for acetone, it was not enough to supply the British war effort at the advent of World War I in 1914. The British Navy, of which Winston Churchill was a high ranking officer, heard about Weizmann and encouraged him to develop large-scale industrial acetone production. Acetone was needed to produce cordite, or smokeless gunpowder, for British munitions manufacture. Throughout World War I, the Weizmann process, the production of acetone, butanol, and ethanol (ABE fermenation), used *C. acetobutylicum* fermentation with maize as the original industrial starting material [52]. The ABE process was the industrial norm until the late 1940s, when more efficient and cheaper petroleum-based methods were developed [51]. Since green chemistry has been a larger concern in recent years, there is renewed interest in *C. acetobutylicum* for its use in production of biofuels [53].

There continued to be an interest in one specific enzyme that aided in the transformation of starch to acetone, acetoacetate decarboxylase (ADC). ADC is involved in solventogenesis, which begins as cell growth slows down and transitions from classical sugar fermentation to solvent production (acetone, butanol, and ethanol) [54]. In the 1960s-70s, Frank Westheimer pioneered the application of physical organic chemistry principles to the study of enzymes, using ADC as a model. He found that the reaction took place via a Schiff base mechanism [55]. Westheimer used a nitrophenol reporter group to propose that a perturbed pK_a of an active site lysine was responsible for the cleavage of acetoacetate to acetone and carbon dioxide. Without a structure of the enzyme, he suggested the perturbed pK_a was due to the electrostatic interaction of an adjacent base, lysine side chain (K116), shifted the pK_a of the active site lysine (K115), from about 10 to about 6 [56-58]. Another study, from Highbarger *et al*., supported that K115

was the catalytic residue by chemical rescue of the active site K115 and adjacent K116 [59]. Residue 115 was changed to a cysteine while K116 remained, and at pH 6.0, the optimum of WT ADC, no cleavage occurred for acetoacetate. However, activity was restored by aminoethylation of the of the mutated C115 residue, which is very nearly the same size as the original lysine residue. In addition, K116 was mutated to arginine, while K115 remained; the activity for K116R was lower than WT, however the pK_a for K115 was similar to WT, presumably due to the proximity of the protonated arginine. Likewise, when K116C was alkylated with 2-bromoethylammonium bromide, activity was restored and the pK_a of K115 matched that of WT [59, 60]. Westheimer passed away in 2007, two years before the structures of acetoacetate decarboxylase from *C. acetobutylicum* (CaADC) and *Chromobacterium violaceum* (CvADC) were solved [5]. These crystallographic models showed that the perturbed pKa of K115 was due to the hydrophobic microenvironment, and not due to the electrostatics of the active site lysine and another protonated base, K116, next to it, as illustrated in Figure 2 [6]. The largely hydrophobic active site

of acetoacetate decarboxylase destabilizes the protonated amine, leading to the effect observed by the nitrophenol reporter group.

1.4 Acetoacetate Decarboxylase-Like Superfamily

Nature is a master of doing more, with less. Enzymes are no exception, as a relatively small number of protein folds have been adapted to perform the myriad chemical reactions occurring in living organisms [61]. It is possible to find numerous examples of putative or hypothetical enzymes in the genome sequence databases, and it is often assumed that the functions of so-called hypothetical proteins will match those of the closest homologs of known

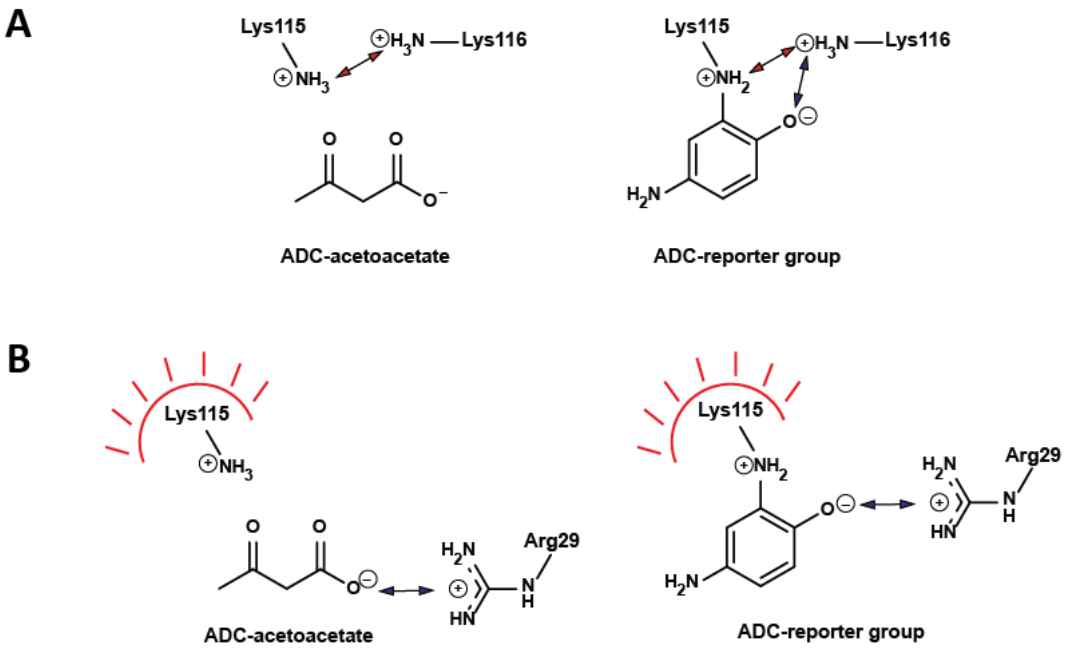


Figure 2. Schematic representation of Westheimer's electrostatic pK_a perturbation hypothesis (**A**). The red arrows represent electrostatic repulsion, and the blue arrows indicate electrostatic attraction. The 2009 structurally supported hydrophobic environment illustrates the proximity of Arg29 to orient the carboxylate group in plane with the subsequent imine (**B**) [5, 6].

function. However, the versatility of enzyme scaffolds means that these assumptions are often wrong [17], especially when the level of sequence identity falls below 40 percent [62]. The

acetoacetate decarboxylase-like superfamily (ADCSF; IPR010451, IPR023375, and IPR023653) is a prime example of this [2]. This superfamily contains about 2,000 enzymes, which were all assumed to catalyze the cleavage of acetoacetate to acetone and carbon dioxide. Although this conversion does appear to dominate the superfamily, it has been shown that there is significant chemical diversity within the ADCSF [2, 63]. There are too many sequences in even a small superfamily like the ADCSF to clearly see the relationships between them, so protein sequence similarity networks (SSN) were designed to visualize these large protein sequence data sets [18]. The aim of the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) is to separate superfamilies into isofunctional clusters based on overall sequence similarity, and to shed light on uncharacterized enzymes by observing their positions relative to well characterized ones. SSNs are determined via BLAST scores; they are more visually impactful than dendrograms and trees, which need computationally intense work for rigorous sequence alignments for large groups of sequences. The sequence similarity network for the ADCSF calculated at e = 10^{-10} is

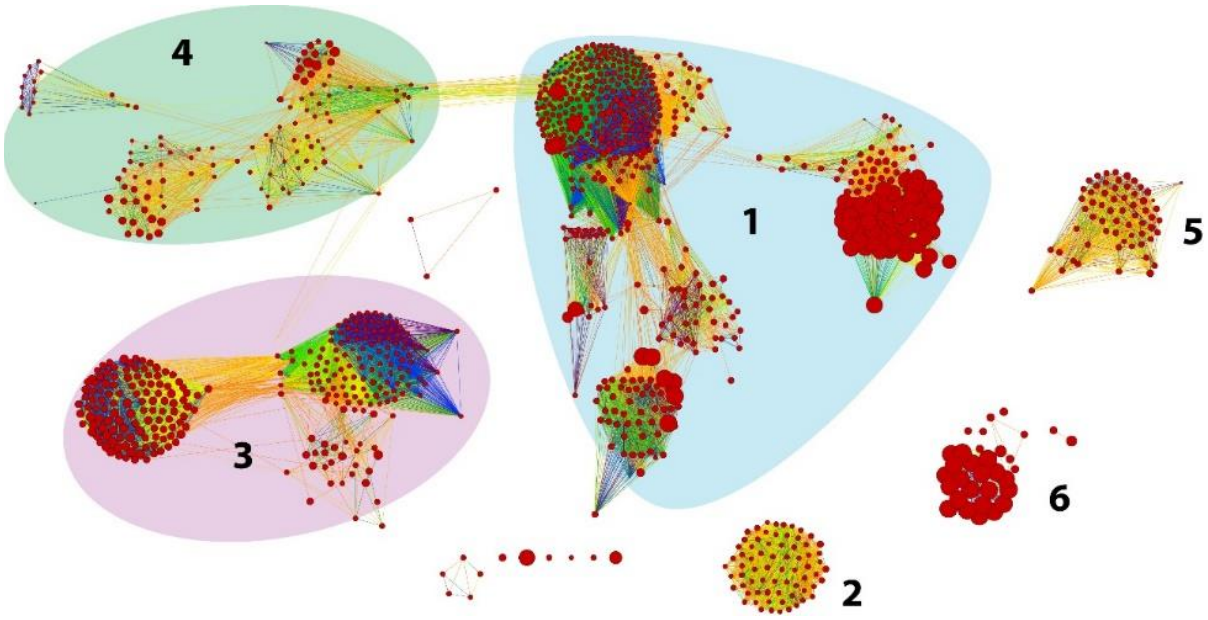


Figure 3. Sequence similarity network for ADCSF calculated with an e value cutoff of 10^{-10}. The clusters are based on manual investigation of active site residues of a limited number of active site sequences.

shown in Figure 3. The sizes of the nodes (red dots) reflect sequence length, and the edges (lines connecting the nodes) are colored according to sequence identity, with red corresponding to 0 percent identity, and purple to 100 percent. All of the sequences from Family V are more similar to each other than any other sequences in the diagram of Figure 3.

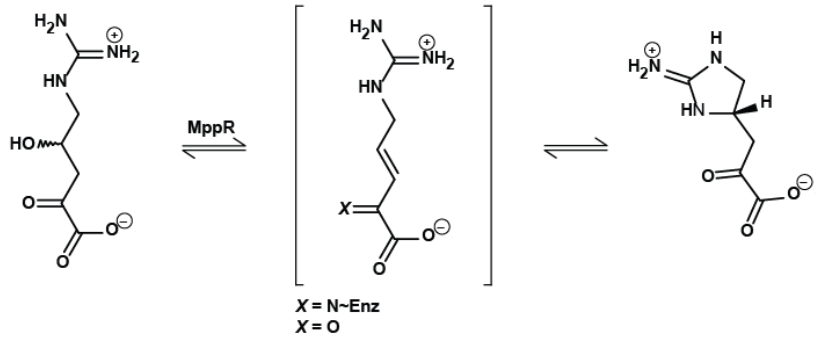


Figure 4. The role of MppR in L-Enduricididine production.

The first Family V enzyme to be studied was Mannopeptimycin R (MppR), an enzyme involved in the biosynthesis of the non-proteinogenic amino acid L-enduracididine, a necessary component of the antibiotics mannopeptimycin and enduracidin, both of which are effective against methicillin resistant staphylococcus aureus (MRSA).

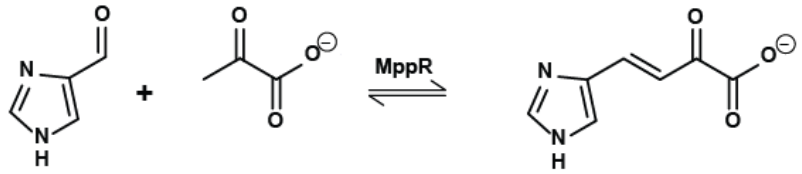


Figure 5. Aldolase-dehydratase reaction catalyzed by MppR.

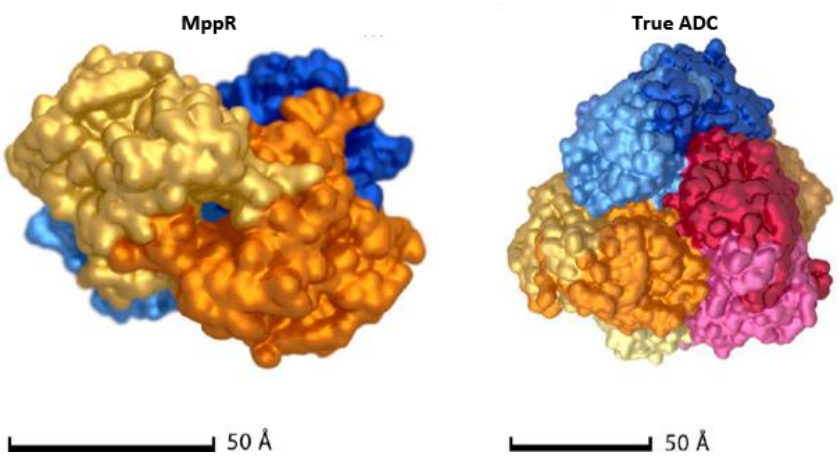


Figure 6. The quaternary structure comparison of MppR (Family V) and CaADC (Family I). PDB IDs are 4JM3[2] and 3BH2[5], respectively.

MppR from *S. hygroscopicus* (ShMppR) was observed in X-ray diffraction experiments to catalyze the cyclization of 2-oxo-4-hydroxy-5-guanidinovaleric acid (i.e., "4-hydroxy-ketoarginine") to give what would become, after hydrolysis from the enzyme, the keto form of the nonproteinogenic amino acid enduracididine, (Figure 4) [2]. MppR is proposed to take a highly oxidized arginine derivative and dehydrate the gamma hydroxyl to a double bond, before catalyzing the attack of a guanidinium nitrogen on the gamma carbon, to give the final iminoimidizolidine ring product (Figure 4) [2]. However, it can also catalyze the aldol condensation of pyruvate and imidazole-4-carboxaldehyde, (Figure 5). ShMppR functions via a mechanism that includes a Schiff base intermediate. This is similar to the Type I aldolases, which also use a lysine Schiff base in their mechanisms. Interestingly, all known Type I aldolases possess the TIM barrel fold. It

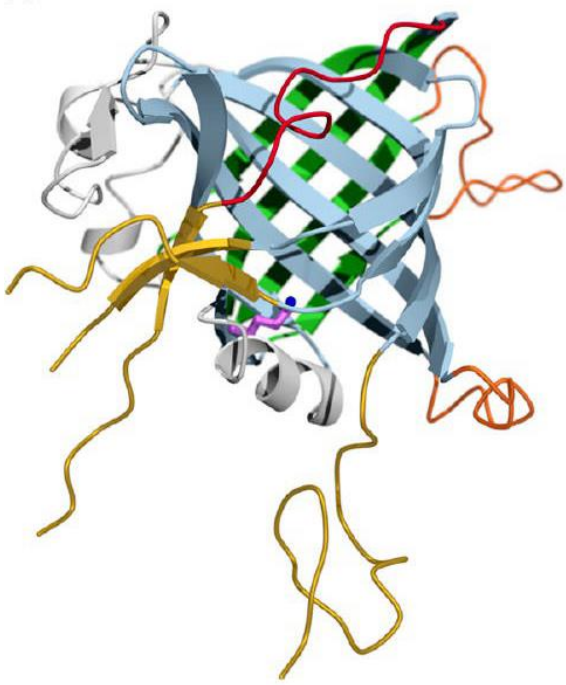


Figure 7. The fold of MppR is almost identical to that of ADC, rather than the TIM barrel fold, associated with Type I aldolases.

is not known if whether this is just the way evolution occurred, or if there is something unique about the TIM barrel that suits it for the aldolase reaction.

Although MppR is a tetramer and ADC is a homododecamer, there does not appear to be any functional significance to the quaternary structure. Despite the MppR protomer being almost identical to ADC, MppR does not perform any decarboxylation like that of typical ADCs [2]. Subtle differences in the active site are responsible for this. The adjacent glutamate in the active site, responsible for decarboxylase activity in CvADC [5], has been replaced with valine in MppR [2]. In addition, the arginine responsible for orienting the carboxylate group of acetoacetate in CvADC is occupied by the indole ring of tryptophan in MppR.

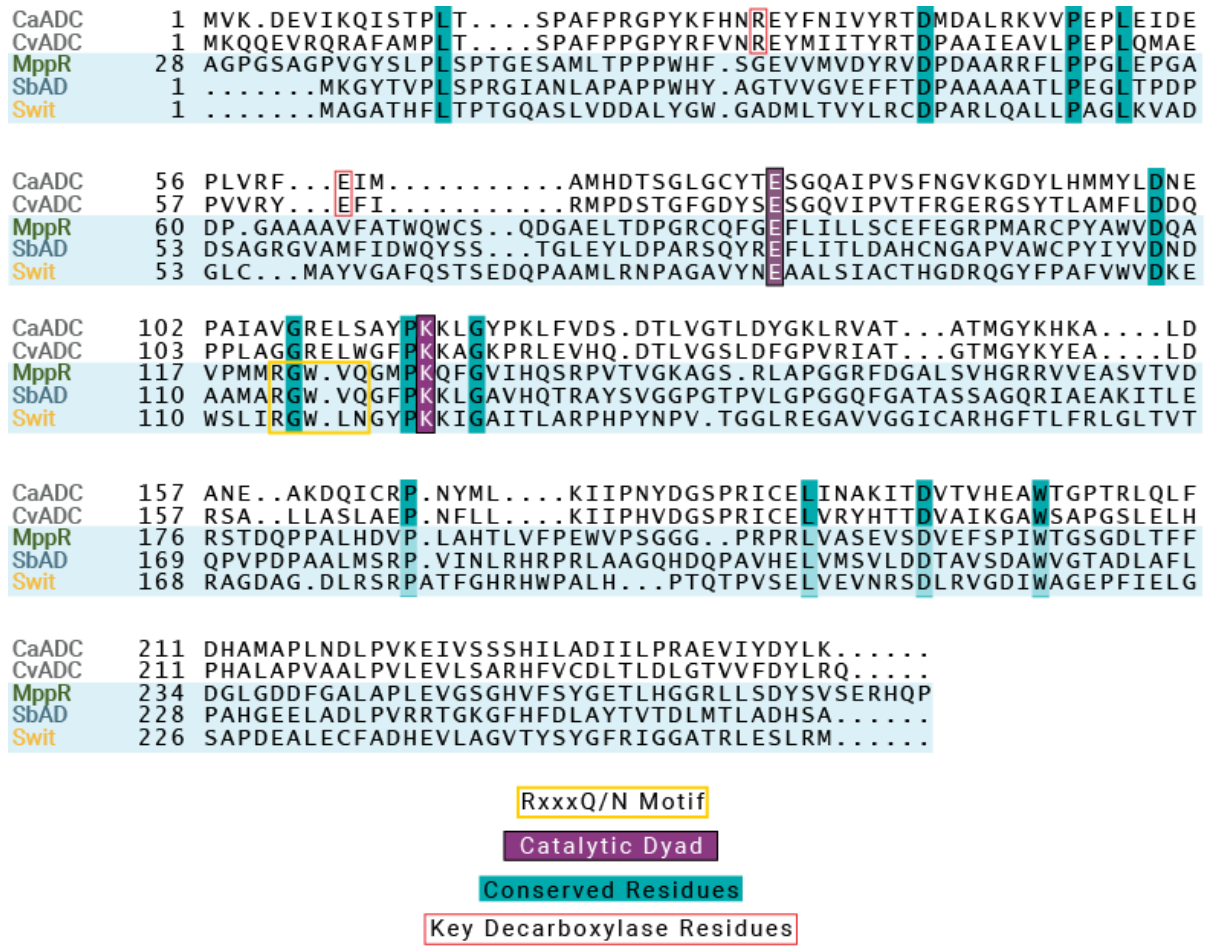


Figure 8. Sequence alignment excerpt comparing Family I and Family V (shaded blue).

Two hydrophobic residues in the active site of CvADC are polar in MppR, making a carboxylate subsite to bind α-keto acids (Figures 4 and 5).

An examination of multiple sequence alignments aids in explaining these differences in catalytic ability. The RxxxQ/N motif in Family V is not found in the Family I decarboxylases, and may indicate other members of the family. There are two key residues conserved between the two families, the lysine-glutamate catalytic dyad. The active site lysine and the likely glutamate acid/base catalyst are highlighted in purple in Figure 8. They are within hydrogen bonding distance of each other in the active site. Although this pair of residues and others are conserved, the residues responsible for decarboxylation are not retained in Family V. The arginine in Family I that aids in orienting the acetoacetate for nucleophilic attack (red box) is not conserved in Family V. Likewise, the glutamate that facilitates decarboxylation in Family I (red box) has different residues in Family V. The differences in these two key decarboxylase residues explains why MppR is not a decarboxylase, though the conservation of the active site lysine and glutamate hints that activity may still occur through a Schiff base mechanism.

Family V of the ADCSF is comprised of enzymes predicted to be involved in secondary metabolism, including the biosynthesis of antibiotics and the catabolism of aromatic hydrocarbons. To learn more about the diversity of chemistries performed in this family, we have been studying Sbi_00515 from *Streptomyces bingchenggensis* (SbAD) and Swit_4259 from *Sphingomonas wittichii* (Swit). Kinetic analyses of these enzymes demonstrate that neither decarboxylates acetoacetate or similar compounds. X-ray crystallography revealed that, although the quaternary structures are different, the overall folds of Sbi_00515 and Swit_4259 are almost indistinguishable from that of ADC, like with MppR.

Sbi_00515 catalyzes the aldol condensation and subsequent dehydration of pyruvate with a variety of aldehydes. It is quite specific for pyruvate, but reacts with a surprising number of unsaturated, aliphatic and aromatic aldehydes. I performed the protein purification, site-directed mutagenesis, and both steady state and transient kinetics on Sbi_00515 in order to determine the substrate specificity profile and develop a plausible mechanism. The kinetic work was supported by my X-ray crystallography, HPLC, and NMR work. I also performed single turnover kinetics and kinetic isotope effect experiments to further develop our initial mechanistic hypothesis. As it stands, our mechanism has the enzyme reacting with pyruvate to form a Schiff base, after which the nucleophilic enamine attacks the aldehyde's electrophilic carbonyl to form a new carbon-carbon bond. Details of the dehydration step are still unclear, but may involve substrate-assisted catalysis. The Schiff base with product is hydrolyzed to regenerate the enzyme. Although the biological function of Sbi_00515 is unknown, it may synthesize a non-canonical starter or extender unit of a nearby uncharacterized secondary metabolite biosynthesis cluster. *Streptomyces bingchenggensis*, the organism that produces Sbi_00515, is not available. Fortunately, there is a homolog from *Burkholderia terrae*, WQE_44883 (85% sequence identity), which will be used to investigate the enzyme's physiological role. The first step in this endeavor is to find the growth conditions under which the gene is expressed. Metabolomics experiments in the Shimazdu Lab for Advanced and Applied Analytical Chemistry at University of WI-Milwaukee may then be able to highlight molecules that are only produced when Sbi_00515 is expressed.

Swit_4259 is able to catalyze the aldol condensation with pyruvate and a limited range of aldehyde substrates, however it is orders of magnitude slower than Sbi_00515. It is much more

efficient catalyzing the hydration of unsaturated α-keto acids without retro-aldol cleavage. Its physiological function is also unknown, however, based on gene context, Swit_4259 is proposed to contribute to the degradation of aromatic hydrocarbons.

All of my work shows that Sbi_00515 and Swit_4259 are the first Schiff base forming aldolases that do not possess a TIM barrel fold. Although the tertiary structures are almost identical, these two enzymes exhibit different activities, likely due to subtle changes in the active site architecture. The significance of these enzymes is that they are the first evidence that the ADC fold is a versatile platform that can be adapted to perform different chemistries, expanding the knowledge of protein catalysis, and making these enzymes potential candidates for future protein engineering experiments.

1.5 References

[1] Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. (2008) Kemp elimination catalysts by computational enzyme design, *Nature 453*, 190-195.

[2] Burroughs, A. M., Hoppe, R. W., Goebel, N. C., Sayyed, B. H., Voegtline, T. J., Schwabacher, A. W., Zabriskie, T. M., and Silvaggi, N. R. (2013) Structural and functional characterization of MppR, an enduracididine biosynthetic enzyme from streptomyces hygroscopicus: functional diversity in the acetoacetate decarboxylase-like superfamily, *Biochemistry 52*, 4492-4506.

[3] Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St Clair, J. L., Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., and Baker, D. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction, *Science 329*, 309-313.

[4] Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., 3rd, Hilvert, D., Houk, K. N., Stoddard, B. L., and Baker, D. (2008) De novo computational design of retro-aldol enzymes, *Science 319*, 1387-1391.

[5] Ho, M. C., Menetret, J. F., Tsuruta, H., and Allen, K. N. (2009) The origin of the electrostatic perturbation in acetoacetate decarboxylase, *Nature 459*, 393-397.

[6] Gerlt, J. A. (2009) Acetoacetate decarboxylase: hydrophobics, not electrostatics, *Nat Chem Biol 5*, 454-455.

[7] Alcalde, M., Ferrer, M., Plou, F. J., and Ballesteros, A. (2006) Environmental biocatalysis: from remediation with enzymes to novel green processes, *Trends Biotechnol 24*, 281-287.

[8] Anastas, P., and Eghbali, N. (2010) Green chemistry: principles and practice, *Chem Soc Rev 39*, 301-312.

[9] Palomo, J. M., and Filice, M. (2015) New emerging bio-catalysts design in biotransformations, *Biotechnol Adv 33*, 605-613.

[10] Hansen, K. B., Hsiao, Y., Xu, F., Rivera, N., Clausen, A., Kubryk, M., Krska, S., Rosner, T., Simmons, B., Balsells, J., Ikemoto, N., Sun, Y., Spindler, F., Malan, C., Grabowski, E. J., and Armstrong, J. D., 3rd. (2009) Highly efficient asymmetric synthesis of sitagliptin, *Journal of the American Chemical Society 131*, 8798-8804.

[11] Bhosale, S. H., Rao, M. B., and Deshpande, V. V. (1996) Molecular and industrial aspects of glucose isomerase, *Microbiol Rev 60*, 280-300.

[12] Miller, T. R., Delcher, A. L., Salzberg, S. L., Saunders, E., Detter, J. C., and Halden, R. U. (2010) Genome sequence of the dioxin-mineralizing bacterium Sphingomonas wittichii RW1, *J Bacteriol 192*, 6101-6102.

[13] Buchholz, K., Kasche, V., and Bornscheuer, U. T. (2012) *Biocatalysts and enzyme technology*, 2nd ed., Wiley-Blackwell, Weinheim.

[14] Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001) Industrial biocatalysis today and tomorrow, *Nature 409*, 258-268.

[15] Xiao-Jing Luo, H.-L. Y. a. J.-H. X. (2012) Genomic Data Mining: An Efficient Way to Find New and Better Enzymes, *Enzyme Engg 1*, 1-4.

[16] NCBI. (2016) Basic Local Alignment Search Tool, https://blast.ncbi.nlm.nih.gov/Blast.cgi.

[17] Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies, *PLoS Comput Biol 5*, e1000605.

[18] Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., and Whalen, K. L. (2015) Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks, *Biochimica et biophysica acta 1854*, 1019-1037.

[19] Zhao, S., Sakai, A., Zhang, X., Vetting, M. W., Kumar, R., Hillerich, B., San Francisco, B., Solbiati, J., Steves, A., Brown, S., Akiva, E., Barber, A., Seidel, R. D., Babbitt, P. C., Almo, S. C., Gerlt, J. A., and Jacobson, M. P. (2014) Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks, *Elife 3*.

[20] Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., and Jensen, L. J. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration, *Nucleic acids research 41*, D808-815.

[21] Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S., and Karp, P. D. (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases, *Nucleic acids research 44*, D471-480.

[22] Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., Friedland, G. D., Huang, K. H., Keller, K., Novichkov, P. S., Dubchak, I. L., Alm, E. J., and Arkin, A. P. (2010) MicrobesOnline: an integrated portal for comparative and functional genomics, *Nucleic acids research 38*, D396-400.

[23] Davids, T., Schmidt, M., Bottcher, D., and Bornscheuer, U. T. (2013) Strategies for the discovery and engineering of enzymes for biocatalysis, *Curr Opin Chem Biol 17*, 215-220.

[24] Meyer, A. J., Ellefson, J. W., and Ellington, A. D. (2014) Library generation by gene shuffling, *Curr Protoc Mol Biol 105*, Unit 15 12.

[25] Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins, *Nat Rev Genet 16*, 379-394.

[26] Pritchard, L., Corne, D., Kell, D., Rowland, J., and Winson, M. (2005) A general model of error-prone PCR, *J Theor Biol 234*, 497-509.

[27] Rasila, T. S., Pajunen, M. I., and Savilahti, H. (2009) Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, Escherichia coli mutator strain, and hydroxylamine treatment, *Analytical biochemistry 388*, 71-80.

[28] Pavelka, A., Chovancova, E., and Damborsky, J. (2009) HotSpot Wizard: a web server for identification of hot spots in protein engineering, *Nucleic acids research 37*, W376-383.

[29] Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes, *Annu Rev Biophys Biomol Struct 29*, 291-325.

[30] Bohnuud, T., Luo, L., Wodak, S. J., Bonvin, A. M., Weng, Z., Vajda, S., Schueler-Furman, O., and Kozakov, D. (2016) A benchmark testing ground for integrating homology modeling and protein docking, *Proteins*.

[31] Bornscheuer, U. T. (2016) Protein engineering: Beating the odds, *Nat Chem Biol 12*, 54-55.

[32] Xiao, H., Bao, Z., and Zhao, H. (2015) High Throughput Screening and Selection Methods for Directed Enzyme Evolution, *Ind Eng Chem Res 54*, 4011-4020.

[33] Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., and Robins, K. (2012) Engineering the third wave of biocatalysis, *Nature 485*, 185-194.

[34] Becker, S., Hobenreich, H., Vogel, A., Knorr, J., Wilhelm, S., Rosenau, F., Jaeger, K. E., Reetz, M. T., and Kolmar, H. (2008) Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes, *Angew Chem Int Ed Engl 47*, 5085-5088.

[35] Fernandez-Alvaro, E., Snajdrova, R., Jochens, H., Davids, T., Bottcher, D., and Bornscheuer, U. T. (2011) A combination of in vivo selection and cell sorting for the identification of enantioselective biocatalysts, *Angew Chem Int Ed Engl 50*, 8584-8587.

[36] Chen, B., Lim, S., Kannan, A., Alford, S. C., Sunden, F., Herschlag, D., Dimov, I. K., Baer, T. M., and Cochran, J. R. (2016) High-throughput analysis and protein engineering using microcapillary arrays, *Nat Chem Biol 12*, 76-81.

[37] Colin, P. Y., Kintses, B., Gielen, F., Miton, C. M., Fischer, G., Mohamed, M. F., Hyvonen, M., Morgavi, D. P., Janssen, D. B., and Hollfelder, F. (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics, *Nat Commun 6*, 10008.

[38] Pavlova, M., Klvana, M., Prokop, Z., Chaloupkova, R., Banas, P., Otyepka, M., Wade, R. C., Tsuda, M., Nagata, Y., and Damborsky, J. (2009) Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate, *Nat Chem Biol 5*, 727-733.

[39] National Toxicology, P. (2011) 1,2,3-Trichloropropane, *Rep Carcinog 12*, 426-428.

[40] Winn, P. J., Ludemann, S. K., Gauges, R., Lounnas, V., and Wade, R. C. (2002) Comparison of the dynamics of substrate access channels in three cytochrome P450s reveals different opening mechanisms and a novel functional role for a buried arginine, *Proc Natl Acad Sci U S A 99*, 5361-5366.

[41] Bosma, T., Damborsky, J., Stucki, G., and Janssen, D. B. (2002) Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene, *Applied and environmental microbiology 68*, 3582-3587.

[42] Lutz, S. (2010) Beyond directed evolution--semi-rational protein engineering and design, *Curr Opin Biotechnol 21*, 734-743.

[43] van der Kamp, M. W., and Mulholland, A. J. (2013) Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology, *Biochemistry 52*, 2708-2728.

[44] Kries, H., Blomberg, R., and Hilvert, D. (2013) De novo enzymes by computational design, *Curr Opin Chem Biol 17*, 221-228.

[45] Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H., and Meiler, J. (2010) Practically useful: what the Rosetta protein modeling suite can do for you, *Biochemistry 49*, 2987-2998.

[46] Baker, D. (2010) An exciting but challenging road ahead for computational enzyme design, *Protein Sci 19*, 1817-1819.

[47] Althoff, E. A., Wang, L., Jiang, L., Giger, L., Lassila, J. K., Wang, Z., Smith, M., Hari, S., Kast, P., Herschlag, D., Hilvert, D., and Baker, D. (2012) Robust design and optimization of retroaldol enzymes, *Protein Sci 21*, 717-726.

[48] Giger, L., Caner, S., Obexer, R., Kast, P., Baker, D., Ban, N., and Hilvert, D. (2013) Evolution of a designed retro-aldolase leads to complete active site remodeling, *Nat Chem Biol 9*, 494-498.

[49] Lassila, J. K., Baker, D., and Herschlag, D. (2010) Origins of catalysis by computationally designed retroaldolase enzymes, *Proc Natl Acad Sci U S A 107*, 4937-4942.

[50] Rupp, B. (2010) *Biomolecular Crystallography: Principles, Practise, and Application to Structural Biology*, Garland Science, New York, NY.

[51] Rogers, P. C., J.-S.; Zidwick, M.J. (2006) *The Prokaryotes: Vol. 1: Symbiotic Associations, Biotechnology, Applied Microbiology*, Third ed., Springer Science+Business Media, Inc., New York, NY, USA.

[52] Jones, D. T., and Woods, D. R. (1986) Acetone-butanol fermentation revisited, *Microbiol Rev 50*, 484-524.

[53] Sreekumar, S., Baer, Z. C., Pazhamalai, A., Gunbas, G., Grippo, A., Blanch, H. W., Clark, D. S., and Toste, F. D. (2015) Production of an acetone-butanol-ethanol mixture from Clostridium acetobutylicum and its conversion to high-value biofuels, *Nat Protoc 10*, 528-537.

[54] Amador-Noguez, D., Brasg, I. A., Feng, X. J., Roquet, N., and Rabinowitz, J. D. (2011) Metabolome remodeling during the acidogenic-solventogenic transition in Clostridium acetobutylicum, *Applied and environmental microbiology 77*, 7984-7997.

[55] Hamilton, G. A. a. W., F. H. (1959) On the mechanism of the enzymatic decarboxylation of acetoacetate. I. , *Journal of the American Chemical Society 81*, 6332-6333.

[56] Frey, P. A., Kokesh, F. C., and Westheimer, F. H. (1971) A reporter group at the active site of acetoacetate decarboxylase. I. Ionization constant of the nitrophenol, *Journal of the American Chemical Society 93*, 7266-7269.

[57] Kokesh, F. C., and Westheimer, F. H. (1971) A reporter group at the active site of acetoacetate decarboxylase. II. Ionization constant of the amino group, *Journal of the American Chemical Society 93*, 7270-7274.

[58] Schmidt, D. E., Jr., and Westheimer, F. H. (1971) PK of the lysine amino group at the active site of acetoacetate decarboxylase, *Biochemistry 10*, 1249-1253.

[59] Highbarger, L. A., Gerlt, J. A., and Kenyon, G. L. (1996) Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the pKa of active-site lysine 115, *Biochemistry 35*, 41-46.

[60] Silverman, R. B. (2002) *The Organic Chemistry of Enzyme-Catalyzed Reactions*, Revised ed., Elsevier, San Diego, CA, USA.

[61] Brown, S. D., and Babbitt, P. C. (2014) New insights about enzyme evolution from large scale studies of sequence and structure relationships, *J Biol Chem 289*, 30221-30228.

[62] Furnham, N., Garavelli, J. S., Apweiler, R., and Thornton, J. M. (2009) Missing in action: enzyme functional annotations in biological databases, *Nat Chem Biol 5*, 521-525.

[63] Furnham, N., Dawson, N. L., Rahman, S. A., Thornton, J. M., and Orengo, C. A. (2015) Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies, *J Mol Biol*.

CHAPTER II: STRUCTURAL CHARACTERIZATION OF SbAD

2.1: Introduction

Analysis of MppR (mannopeptimycin R) from the acetoacetate decarboxylase-like superfamily (ADCSF) gave the first indication that this superfamily may contain enzymes that catalyze diverse reactions, and as such, is a potential source of useful biocatalysts [1]. Based on gene context, it is proposed that the MppR-like enzymes, Family V of the ADCSF, are involved in production of secondary metabolites. A BLAST search with MppR sequence returned a protein from a bacterial genome noted for its secondary metabolites: Sbi_00515 from *Streptomyces binchenggensis* BCW-1. The Gram-positive *S. bingchenggensis* has one of the largest genomes sequenced to date at about 12 million bp [2]. Much of the *S. bingchenggensis* genome is devoted to production of secondary metabolites. *S. bingchenggensis* contains at least 47 gene clusters linked to biosynthesis of identified or anticipated secondary metabolites [3]. These include milbemycins A3 and A4 that are used extensively to control parasitic worms in animals, 10 additional milbemycin analogues, the polyether antibiotic nanchangmycin, the antitumor agents bingchamide A and B, and at least 23 uncharacterized nonribosomal peptide, polyketide, or terpene biosynthetic clusters [2-8]. Sbi_00515 is currently annotated as an acetoacetate decarboxylase, yet like MppR, it is missing the same key arginine and glutamate residues responsible for decarboxylase activity in the ADCs, and is located near an uncharacterized polyketide biosynthetic cluster [2].

Sbi_00515 has 39% sequence identity to MppR from *Streptomyces hygroscopicus* (ShMppR), and 265 amino acids [9]. Based on sequence alignment to ShMppR, Sbi_00515 is likely

to have the same carboxylate binding site and lack of decarboxylase activity. Although both Family V enzymes retain the key lysine and likely acid/base catalyst glutamate that can form a Schiff base intermediate, it is possible that Sbi_00515 may not act in the same manner as MppR. Sequence alignment comparing the active site residues of MppR with Sbi_00515 shows that in two positions where MppR has Phe, Sbi_00515 has Tyr, and Glu in MppR is replaced with another Tyr in Sbi_00515. The change in residues of the likely active site of Sbi_00515 could influence substrate specificity or reaction(s) catalyzed. The *S. bingchenggensis* bacterium is not available for biological study due to intellectual property concerns, therefore its physiological activity will likely never be known. Therefore, the purpose of this study was to investigate the catalytic capacity of the MppR-like family specifically, and the ADCSF in general.

2.2 Materials and Methods

Cloning, Expression, and Purification of Sbi_00515.

The coding sequence of Sbi_00515 was optimized for expression in *Escherichia coli* and synthesized by GenScript Inc. (Piscataway, NJ). This synthetic gene was subcloned into the pE-SUMO$_{kan}$ expression vector (LifeSensors Inc., Malvern, PA) using primers containing BsaI and XbaI restriction sites (forward, 5'-GGTCTCAAGGTATGAAAGGTTATACGGTTCCG-3'; reverse, 5'-GCTCTAGATCATCATGCCGAATGGTCTGC-3'). The His$_6$-tagged SUMO−Sbi00515 fusion protein was expressed from *E. coli* BL21 Star (DE3) cells (Invitrogen Inc., Carlsbad, CA) carrying the pE-SUMO-Sbi_00515 plasmid. Cultures were grown in Luria-Bertani medium with 50 µg/mL kanamycin at 37 °C. At an OD_{600} of about 0.9, protein expression was induced with 0.4mM IPTG. The

temperature was reduced to 25 °C, and the cultures were grown overnight while being shaken at 250 rpm. Cells were harvested by centrifugation at 13000 x *g*, resuspended in 5 mL/g of buffer A [25 mM Tris (pH 8.0), 300 mM NaCl, and 10 mM imidazole] supplemented with 1 mg/mL hen egg lysozyme (Hampton Research) and frozen at −20 °C. Cells were lysed by thawing the resuspended cells for 2 hours at room temperature (RT), after which DNase I (Worthington Biochemical Corp., Lake-wood, NJ) was added to achieve a concentration of 0.1 mg/ml. The lysate was clarified by centrifugation at 39000 x *g* for 45 min and then applied to a 5 mL HisTrap column (GE Lifesciences,Piscataway, NJ) at a flow rate of 5 mL/min to isolate the His_6-SUMO−Sbi_00515 fusion protein. The protein was eluted by a four-step gradient of buffer B [25 mM Tris (pH 8.0), 300 mM NaCl, and 250 mM imidazole (5, 15, 50, and 100%)]. The His_6-SUMO−Sbi_00515 fusion protein eluted in the third and fourth steps and was ∼90% pure, as judged on Coomassie-stained sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS-PAGE) gels. Peak fractions were pooled and dialyzed overnight against 3.5 L of 25 mM Tris (pH 8.0) and 150 mM NaCl in the presence of about 3 μM SUMOprotease (LifeSensors Inc.). The dialysate was passed through the HisTrap column a second time to remove the cleaved His_6-SUMO tag as well as the protease. The resulting Sbi_00515 preparation was >95% pure. Selenomethionine-labed Sbi_00515 was purified using the same protocol, except that SelenoMethionine Medium Complete (Molecular Dimensions, Newmarket, Suffolk, U.K.) was used as the growth medium rather than Luria-Bertani broth and T7 Express Crystal cells (New England Biolabs, Ipswich, MA) were used in place of the BL21 Star (DE3) cells.

Crystallization, Structure Determination, and Model Refinement.

Initial crystallization conditions were identified by screening 45 mg/mL Sbi_00515 against the Index HT screen (Hampton Research). After optimization, diffraction-quality crystals were obtained by the hanging drop vapor diffusion method from 3.5−3.7 M potassium formate, 2−3% polyethylene glycol monomethyl ether 2000 (PEG MME 2k), and 100 mM BIS-Tris propane (pH 9.0). Drops contained 1−2 µL of a protein solution at 45 mg/mL and 1 µL of a crystallization solution. Crystals formed as plates that appeared after 3−4 days and grew to maximal dimensions of about 800 µm × 200 µm × 50 µm. Crystals of selenomethionine (SeMet)-substituted Sbi_00515 were grown using the same conditions. Crystals were cryoprotected with Paratone N (Hampton Research) and flashed-cooled by being plunged into liquid nitrogen. The structure of Sbi_00515 with pyruvate bound was obtained by transferring crystals of native Sbi00515 into 30 µL drops of a soaking solution containing 4.0M potassium formate, 3% PEG MME 2k, 100 mM BIS-Tris propane (pH 9.0), and 30 mM sodium pyruvate. After soaking overnight, crystals were treated with Paratone N and flash-cooled. X-ray diffraction data for SeMet Sbi_00515 and a higher-resolution native data set were collected at beamline 21-ID-D of the Life Science Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS). The Sbi_00515·pyruvate data set was collected using the rotating anode X-ray source at Marquette University. Data were processed with HKL2000 or MOSFLM and SCALA of the CCP4 Program Suite [10-14]. The structure of Sbi_00515 was determined by the single-wavelength anomalous diffraction (SAD) method using 1.9 Å-resolution data collected from a crystal of SeMet-substituted Sbi_00515 at 0.97921 Å, 4.0 eV above the tabulated K-edge wavelength for Se (0.97950 Å). The program autoSHARP was used to determine the Se substructure, which contained 16 of the 24 Se atoms in the asymmetric unit, and to calculate density-modified electron density maps. An initial model comprising about 75%

of the asymmetric unit contents was built automatically using the PHENIX package (phenix.autobuild) [15-17]. After iterative cycles of manual model building in COOT and maximum likelihood-based refinement using the PHENIX package (phenix.refine), ordered solvent molecules were added automatically in phenix.refine and culled manually in COOT [18, 19]. Hydrogen atoms were added to the model using phenix.reduce and were included in the later stages of refinement to improve the stereochemistry of the model [20]. Positions of H atoms were refined using the riding model with a global B factor. Regions of the model for translation-libration-screw (TLS) refinement were identified using phenix.find_tls_groups, and the TLS parameters were refined in phenix.refine. Once the refinement converged (e.g., for SeMet Sbi00515, R = 0.162 and R_{free} = 0.185), the model was validated using the tools implemented in COOT and PHENIX [21, 22]. Sections of the backbone with missing or uninterpretable electron density were not included in the final model. Side chains with poor or missing electron density were modeled in favored rotameric conformations. The B factors were allowed to refine without additional restraints, and the occupancies for atoms in all residues without alternate conformations were constrained to be 1.0.

The final, refined model of SeMet Sbi_00515, stripped of water molecules and H atoms, and with all B factors set to 20.0 $Å^2$, was used to determine the structures of Sbi_00515 "unliganded" (with formate bound) and with pyruvate bound by the difference Fourier method. A similar refinement protocol was used for both models presented here. In the case of the Sbi_00515·pyruvate structure, restraints for the link between Lys122 and pyruvate were generated with phenix.elbow [23]. Data collection and model refinement statistics are listed in

Table 2.1. Coordinates and structure factors have been deposited in the Protein Data Bank as entries 4ZBO and 4ZBT.

2.3 Results and Discussion

Overall Structure and Comparisons to CaADC and ShMppR.

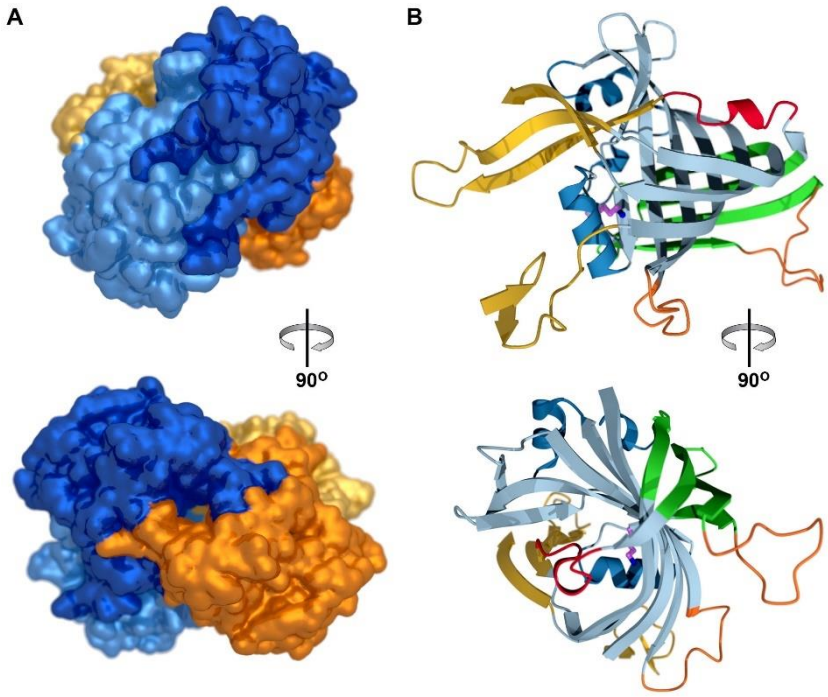


Figure 9. (**A**) Rendering of the solvent-accessible surface of Sbi00515 showing the extensive tetramer interface formed by the long, interdigitating loops (top). This interface is repeated on the back side with the orange and pale yellow protomers. The pair of dimers is held together (bottom) by short loops that appear to "clip" the two dimers together to form the tetramer. (**B**) The ribbon diagram showing a single protomer is oriented like the dark blue protomer in panel A. The main barrel of the β-cone fold is colored pale blue, with the "side barrel" colored green. The dimer interface loops are colored yellow, and the loops forming the interface between the dimers are colored orange. The section of the polypeptide that is colored red is the "active site loop" that may control access to the active site. The catalytic Lys122 is shown as a ball-and-stick representation with purple carbon atoms.

Sbi_00515 has a protomer weight is 28,145 Da and, like MppR, was shown to be tetrameric in solution using size exclusion chromatography. Sbi_00515 crystallized in space group $P2_12_12$ with four molecules in the asymmetric unit (unit cell dimensions a = 157.3Å, b = 123.5 Å, c = 53.1 Å, and $\alpha = \beta = \gamma$ =90°). The asymmetric unit contents are arranged in a tetramer that

matches closely the arrangement seen in ShMppR [1]. This tetramer is a dimer of dimers (Figure 9 A). The "dimer interface" is extensive, being comprised of interdigitating loops (Met1−Pro21 and Arg186−Glu203; yellow in Figure 9 B) that bury 2642 Å^2, or 20% of the protomer surface area. The interface between the dimers, which we call the "tetramer interface", is smaller, burying 1621 Å^2, or 12% of the protomer surface, and is also comprised of two surface loops (Ser68−Gln81 and Gln129−Gly164; orange in Figure 9 B). The homotetrameric quaternary structure shared by Sbi_00515 and ShMppR is markedly different from that of CaADC, which is a homododecamer.

Table 1. Crystallographic Data Collection and Model Refinement Statistics.

	SeMet Sbi_00515	native Sbi_00515	Sbi_00515·pyruvate
resolution (Å) (last shell)[a]	46.5−1.9 (1.97−1.90)	41.58−1.4 (1.45−1.40)	39.9−1.8 (1.86−1.80)
wavelength (Å)	0.97921	0.97896	1.54180
no. of reflections			
observed	1131019 (96476)	1626413 (91229)	285228 (8521)
unique	82988 (8183)	207128 (20393)	92300 (6104)
completeness (%)[a]	100.0 (100.0)	99.9 (99.5)	95.4 (64.0)
R_{merge} (%)[a,b]	0.145 (0.456)	0.071 (0.560)	0.102 (0.334)
multiplicity	13.6 (11.8)	7.9 (4.5)	3.1 (1.4)
⟨I/σ(I)⟩[a]	22.3 (5.9)	20.5 (2.2)	12.3 (1.7)
figure of merit (acentric)	0.283		
phasing power	1.102		
	Model Refinement		
no. of reflections in the work set	78722	207021	92221
no. of reflections in the test set	4185	2960	4631
R_{cryst} (R_{free})	0.162 (0.185)	0.154 (0.176)	0.163 (0.205)
no. of residues	1023	1034	1028
no. of solvent atoms	1166	1385	1289
no. of TLS groups	24	25	25
average B factor (Å^2)[c]			
protein atoms	15.9	21.0	21.4
ligand atoms	35.6	33.9	31.9
solvent	26.9	34.1	31.3
root-mean-square deviation			
bond lengths (Å)	0.005	0.009	0.011
bond angles (deg)	0.979	1.303	1.309
coordinate error (Å)	0.30	0.12	0.18
Ramachandran statistics (favored/allowed/outliers)	98.2/1.8/0	97.7/2.3/0	97.2/2.8/0

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

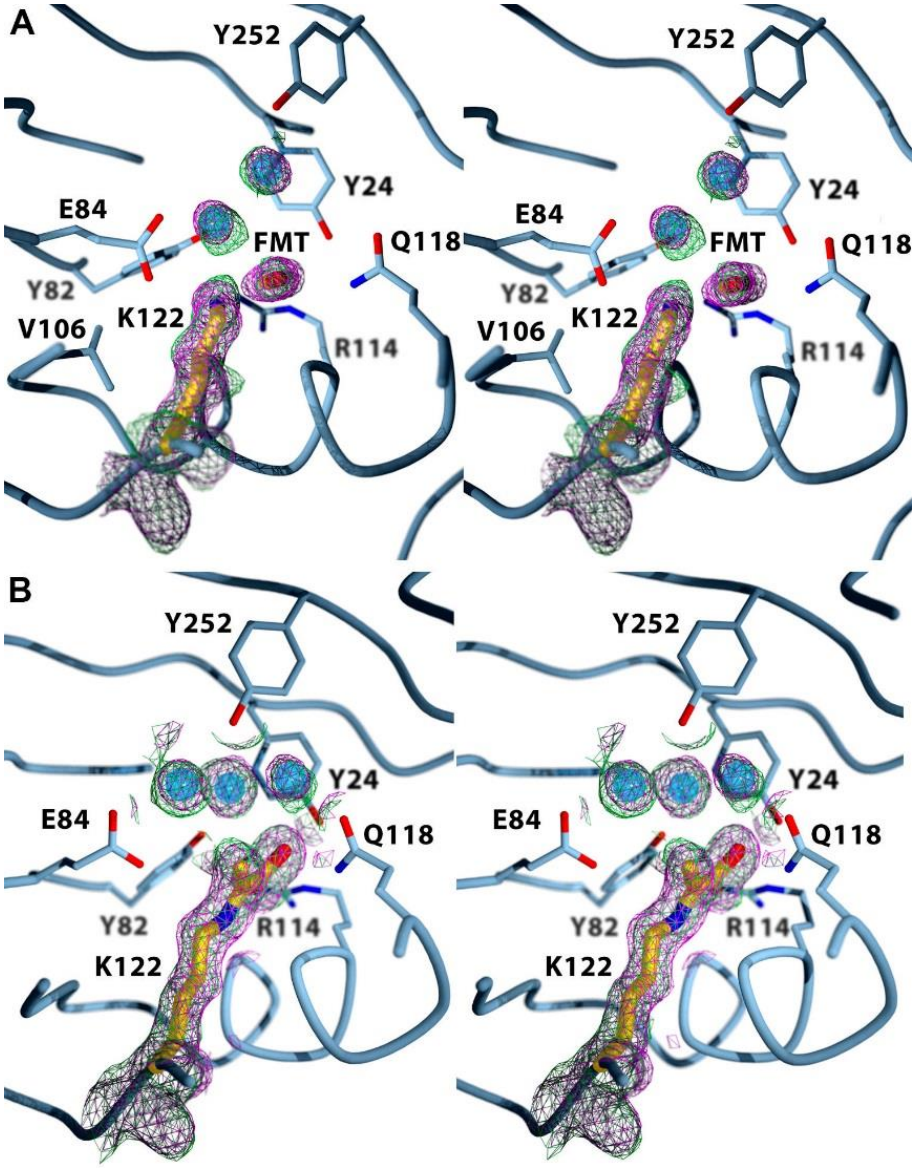


Figure 10. (**A**) Stereoview of the active site in the "unliganded" Sbi_00515 structure showing potassium formate from the crystallization solution bound (refer to Figure 11 for relevant hydrogen bonding interactions and their distances). The position of this formate ion shows that the carboxylate-binding site identified in *S. hygroscopicus* MppR is retained in Sbi_00515. The lower water molecule in this view (labeled "1" in Figure 11) bridges Tyr82 and Glu84, suggesting that both residues could participate in the Schiff base chemistry. The upper water molecule ("2" in Figure 11) is held in place by interactions with Tyr252, which is the major difference between the Sbi_00515 and ShMppR active sites. (**B**) The Sbi_00515·pyruvate Schiff base complex shows pyruvate covalently bound to the enzyme as indicated by continuous electron density between Lys122 and the α-carbon of pyruvate. The side chain of pyruvate is too short to unequivocally differentiate between the imine and enamine forms of the Schiff base. Note that water "1" has been displaced, water "2" remains in a similar position, and two new water molecules have been recruited so that Glu84, Gln118, and Tyr252 are all linked by an extensive hydrogen bonding network. In both panels, the experimental $2|F_o|-|F_c|$ electron density map (magenta) and the simulated annealing composite omit $2|F_o| - |F_c|$ electron density map (green), both contoured at 1.0σ, are drawn with a 2.0 Å radius around each atom of Lys122, formate or pyruvate, and select water molecules. This figure and subsequent figures showing crystallographic structures were rendered with the POVSCRIPT+[24] modification of MOLSCRIPT[25] and POVRAY.

In both Sbi_00515 and ShMppR, the trimerization platform of CaADC is occluded by the loops of the tetramer interface. This pair of sequence insertions accounts for the difference in quaternary structures between the classical ADCs and the MppR-like family. Thus far, there is no known functional consequence of the quaternary structure in ADCSF enzymes.

The tertiary structure of Sbi_00515 is essentially identical to those of both MppR and CaADC, with root-mean-square deviations (RMSDs) for all aligned Cα atoms of 1.4 and 1.7 Å, respectively. This is not surprising, given that all ADCSF proteins of known structure {PDB entries 3BH2, 3BH3[26], 4JM3[1], 3C8W [Joint Center for Structural Genomics (JCSG), unpublished], and 3CMB (JCSG, unpublished)} share the same β-cone or double-barrel fold first identified by Ho et al.[26] (Figure 9 B). In spite of the remarkable similarity of the overall folds, the active sites of Sbi_00515 and CaADC are different. Specifically, as in ShMppR, Arg29, the residue responsible

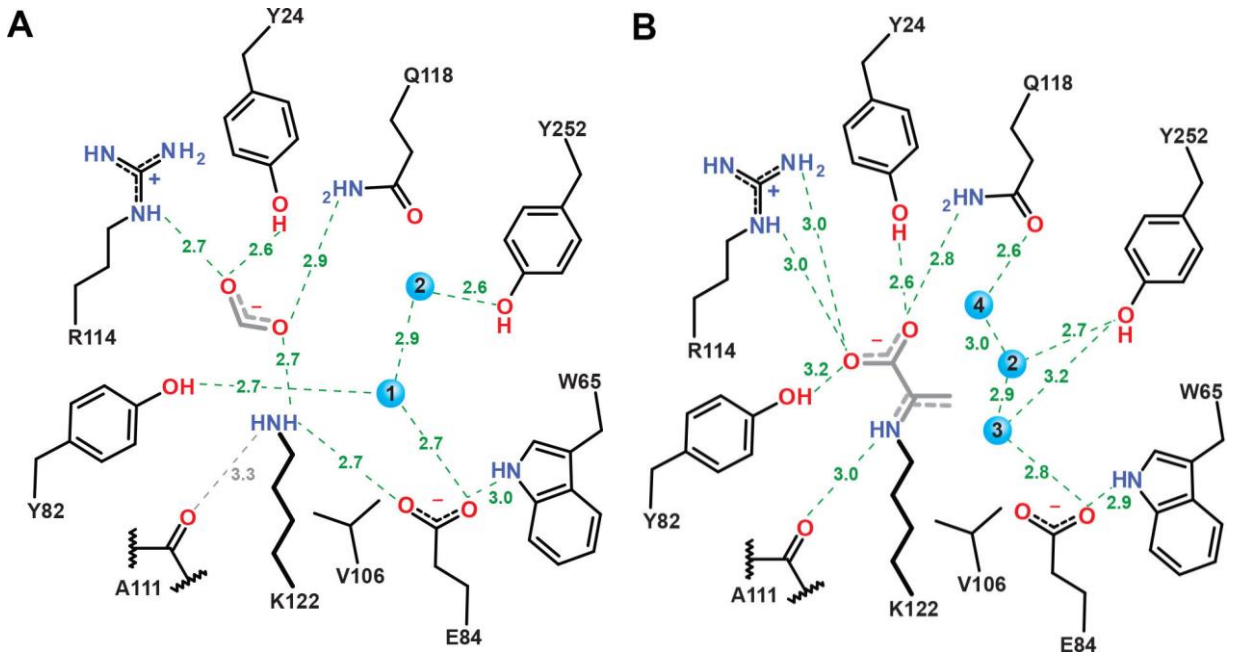


Figure 11. Schematic views of (**A**) the Sbi_00515·formate complex and (**B**) the Sbi_00515·pyruvate complex showing potential hydrogen bonding interactions and their distances (in Angstroms).

for orienting the substrate in CaADC, and Glu61, thought to electrostatically encourage the carboxylate group of acetoacetate to leave as CO_2, are missing in Sbi_00515. These two sequence changes alone would account for the loss of acetoacetate decarboxylase activity in the MppR-like enzymes.

The other major difference between CaADC and the MppR-like enzymes is the presence of the very polar "carboxylate-binding site" adjacent to the catalytic lysine residue (Lys122 in Sbi_00515). This site is marked in the "unliganded" structure of Sbi_00515 by a bound formate ion from the crystallization solution (Figures 10 A and 11 A). The ion is held by hydrogen bonding interactions between O1 of formate and the hydroxyl group of Tyr24 (2.6 Å) or the guanidinium group of Arg114 (2.7 Å), and between O2 of formate and the amide nitrogen atom of Gln118 (2.9 Å) or the amino group of Lys122 (2.7 Å). Two water molecules are also bound in close proximity to the catalytic Lys122 and the bound formate ion. One, labeled "1" in Figure 11 A, is bridging Tyr82 and Glu84 (equidistant at 2.7 Å), while the other, labeled "2," is held by interactions to water "1" (2.9 Å) and the hydroxyl group of Tyr252 (2.6 Å). The position of water molecule "1," only 3.3 Å from the amino group of Lys122, suggests that Tyr82 and Glu84 might both participate in the Schiff base chemistry of Lys122.

The active site of Sbi_00515 is similar to ShMppR, however a few changes in residues appears to alter the substrate specificity and reactivity of catalysis [1]. In Sbi_00515, Tyr252 replaces the Glu283 of MppR that is thought to help orient the guanidinium group of a hydroxylated arginine derivative for cyclization [1]. In addition, Tyr82 is a phenylalanine residue in MppR (Phe116), and Tyr24 of Sbi_00515 is also a phenylalanine in ShMppR (Phe58). All of the other residues are identical (Figure 12 A). Tyr24 of Sbi_00515 is especially interesting, because it

is a component of the carboxylate-binding site not present in ShMppR and alters the conformation of the Schiff base complex between pyruvate and Lys122.

Schiff Base Complex with Pyruvate.

Crystals of Sbi_00515 were soaked in a solution containing 30 mM sodium pyruvate, and the resulting covalent enamine/imine complex was trapped crystallographically (Figures 11 B and 12 B). The carboxylate moiety of pyruvate is bound in the carboxylate-binding site of the enzyme, making hydrogen bonding interactions with the hydroxyl groups of Tyr24 and Tyr82 (2.6 and 3.2 Å, respectively) and the amide nitrogen atom of Gln118 (2.8 Å), as well as a potential salt bridge with Arg114 (bidentate at 3.0 Å). Binding of pyruvate does not impact the positions of the side chains comprising the carboxylate-binding site [Tyr24, Arg114, and Gln118 (Figure 12 A)], indicating that this site is preconfigured to bind α-keto acids. This may explain why the Schiff base complex is sufficiently stable to be observed in the crystal structure. The residues closer to the catalytic center, Tyr82, Glu84, and Val106, experience small, but significant and concerted, changes in their positions. Tyr82 swings toward the catalytic lysine, moving the hydroxyl group by 0.6 Å, while Glu84 swings away from Lys122, shifting its carboxylate group by 1.4 Å. The distance between the nearest carboxylate O atom of Glu84 and the pyruvate methyl group is 3.9 Å. Given this arrangement, it is plausible to suggest that Glu84 might act as a general base catalyst to promote formation of the reactive enamine form of the Schiff base. Lastly, although it is some distance from the Lys122-Glu84 dyad, Tyr252 also experiences a 0.6 Å shift of its hydroxyl group toward Lys122, likely because of a rearrangement of the water structure around the catalytic

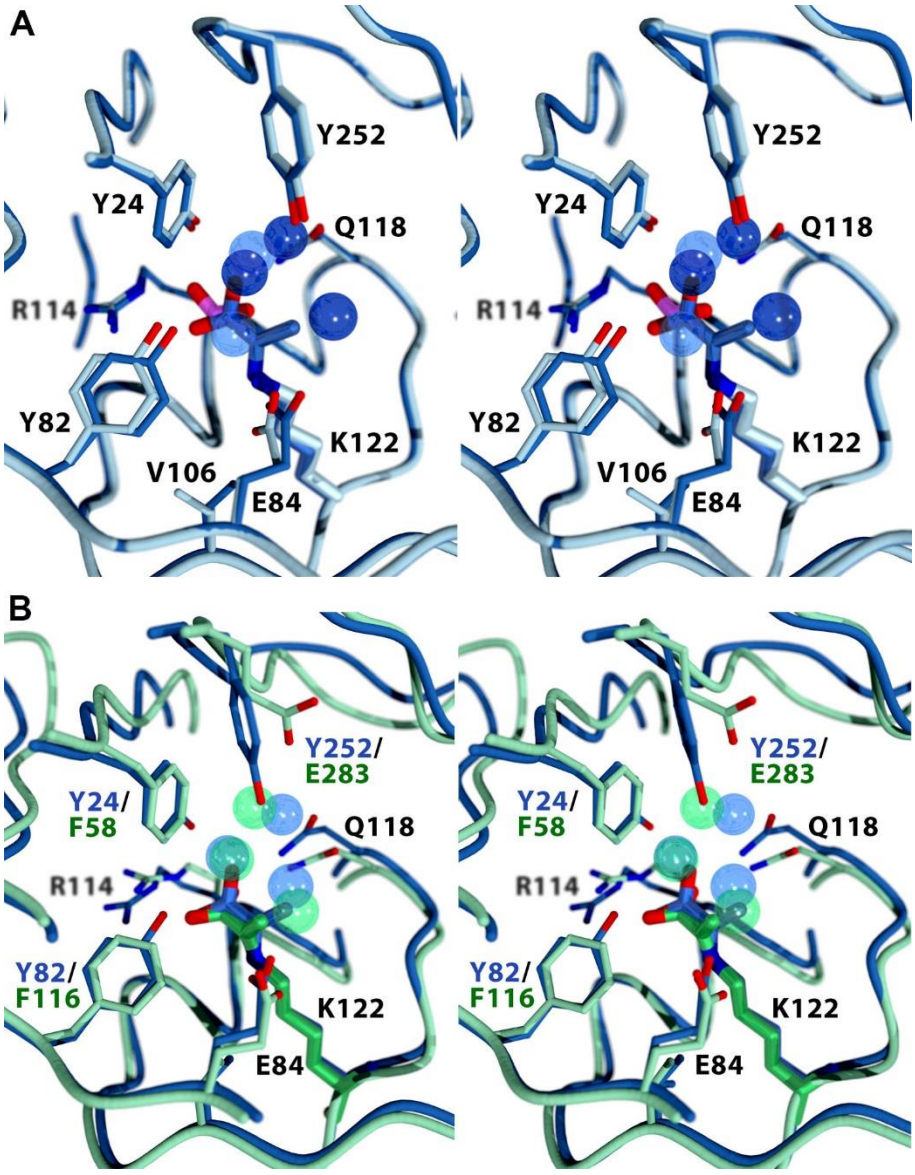


Figure 12. (**A**) Stereoview showing the complex of Sbi00515 and formate (pale blue model with purple formate ion) overlaid on the enzyme with pyruvate bound (dark blue model). Note the subtle changes in position that occur for residues Tyr82, Glu84, and Tyr252, while the residues comprising the carboxylate-binding site are essentially unperturbed. (**B**) MppR·pyruvate complex (green model) overlaid on the Sbi_00515·pyruvate complex (dark blue model). In spite of the high degree of structural similarity (1.7 Å rmsd for all aligned Cα atoms), the arrangement of residues in the carboxylate-binding site is subtly different because of the addition of Tyr24 in Sbi_00515. This change results in Sbi_00515 binding the pyruvate carboxylate with a monodentate interaction with Arg114 and an additional hydrogen bonding interaction with Tyr24, which holds the carboxylate in plane with the pyruvate methyl group (2.8 Å fromCH3 to carboxylate O). This change, coupled with the Glu→Tyr change at position 283/252, may account, in part, for the difference in catalytic activity between Sbi_00515 and MppR.

center that occurs upon formation of the Schiff base (Figure 12 B). In the pyruvate-bound structure, the water bridging Tyr82 and Glu84 is displaced (water "1" in Figure 11 A) and two

additional water molecules are recruited (labeled “3” and “4” in Figure 12 B) that create a chain connecting Glu84, Gln118, and Tyr252. The position of water molecule “2” moves along with Tyr252 by 0.8 Å, placing it 3.6 Å from the pyruvate methyl group. The position of this water is instructive, because when the incoming aldehyde substrate is placed so that the aldehyde O atom overlays water “2” and the *re* face of the carbonyl faces the pyruvate methyl group, the distance between the methyl and carbonyl carbons is approximately 3 Å.

It was somewhat surprising to find that although pyruvate occupies the same binding site in Sbi_00515 and ShMppR, the conformations of the Schiff base complexes are different. In ShMppR (PDB entry 4JMC), the pyruvate carboxylate group is nearly coplanar with the guanidinium of Arg148, making a clear salt bridge interaction. In Sbi_00515, the substitution of Tyr24 for Phe58 of ShMppR results in the carboxylate-binding site Arg and Gln residues taking different conformations. As a result, the pyruvate carboxylate in the Sbi_00515·pyruvate complex is rotated by about 30° relative to the ShMppR·pyruvate structure. The rotation of the carboxylate weakens or eliminates the salt bridge to Arg114 and creates a new interaction with Tyr24. This change also forces the carboxylate to nearly eclipse the methyl group (O2−C1−C2−C3 torsion angle of 14°). While it is possible that this nearly eclipsed conformation just happens to be the lowest-energy conformation in this complex, the strain inherent in this arrangement, the 2.8 Å distance between the carboxylate oxygen and the methyl carbon, and the potential loss of the salt bridge raise the possibility of substrate-assisted catalysis. It is possible that the carboxylate could abstract a proton from the former pyruvate methyl carbon during the transition from the presumed aldol intermediate to the unsaturated product. A substrate carboxylate is proposed to act as a general acid/base catalyst in sialic acid aldolase [27]. Sialic acid

aldolase is specific for pyruvate, and aldol condensation occurs between pyruvate, which acts as a proton donor and various aldehydes. The product of the condensation reaction is determined by the aldehyde acceptor, as sialic acid aldolase appears to lack strict stereochemical specificity, unlike most Type I aldolases [27, 28]. Despite the pyruvate specificity and enzymatic reversibility, a number of crystal structure illustrate that there are no acidic or basic residues near the reaction center, leading to the hypothesis that substrate-assisted catalysis must occur, as seen in Figure 13. [27, 29, 30]. In addition to sialic acid aldolase, other enzymes are thought to function with assistance from a substrate carboxylate as a general acid/base catalyst, such as *trans-o*-hydroxybenzylidenepyruvate hydratase-aldolase and D-4-deoxy-5-oxoglucarate dehydratase [31-33].

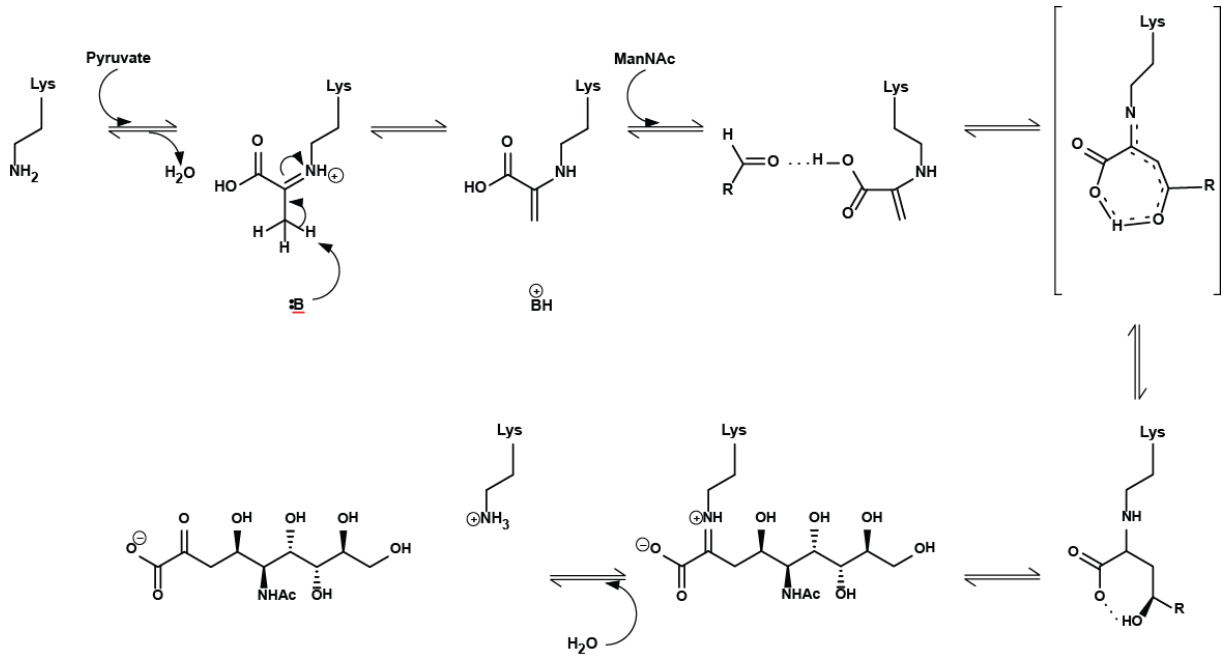


Figure 13. (**A**) Sialic acid proposed mechanism [27, 30].

2.4 Conclusions

The structures of the enzyme in complexes with formate and pyruvate point out possible roles of several active site amino acids. Both Tyr82 and Glu84 are near enough to the catalytic lysine residue to make contacts with the presumed hemiaminal intermediate during formation and breakdown of the Schiff base. Because only Glu84 is in a position to reach both Lys122 and the methyl group of pyruvate, and Tyr82 is not absolutely conserved in MppR-like enzymes, Glu84 is the most likely candidate for the general acid/base catalyst. The position of the water molecule held by Y252 in the presence and absence of pyruvate implies that this residue might orient the incoming aldehyde substrate for attack by the enamine nucleophile. Kinetic and structural characterizations of said active site residues and others are mutated in the following chapters to explore their effects on binding and catalysis.

2.5 References

[1] Burroughs, A. M., Hoppe, R. W., Goebel, N. C., Sayyed, B. H., Voegtline, T. J., Schwabacher, A. W., Zabriskie, T. M., and Silvaggi, N. R. (2013) Structural and functional characterization of MppR, an enduracididine biosynthetic enzyme from streptomyces hygroscopicus: functional diversity in the acetoacetate decarboxylase-like superfamily, *Biochemistry 52*, 4492-4506.

[2] Wang, X. J., Yan, Y. J., Zhang, B., An, J., Wang, J. J., Tian, J., Jiang, L., Chen, Y. H., Huang, S. X., Yin, M., Zhang, J., Gao, A. L., Liu, C. X., Zhu, Z. X., and Xiang, W. S. (2010) Genome sequence of the milbemycin-producing bacterium Streptomyces bingchenggensis, *J Bacteriol 192*, 4526-4527.

[3] Wang, X. J., Zhang, B., Yan, Y. J., An, J., Zhang, J., Liu, C. X., and Xiang, W. S. (2013) Characterization and analysis of an industrial strain of Streptomyces bingchenggensis by genome sequencing and gene microarray, *Genome 56*, 677-689.

[4] Xiang, W. S., Wang, J. D., Fan, H. M., Wang, X. J., and Zhang, J. (2008) New seco-milbemycins from Streptomyces bingchenggensis: fermentation, isolation and structure elucidation, *J Antibiot (Tokyo) 61*, 27-32.

[5] Xiang, W. S., Wang, J. D., Wang, X. J., and Zhang, J. (2007) Two new beta-class milbemycins from streptomyces bingchenggensis: fermentation, isolation, structure elucidation and biological properties, *J Antibiot (Tokyo) 60*, 351-356.

[6] Xiang, W. S., Wang, J. D., Wang, X. J., Zhang, J., and Wang, Z. (2007) Further new milbemycin antibiotics from Streptomyces bingchenggensis. Fermentation, isolation, structural elucidation and biological activities, *J Antibiot (Tokyo) 60*, 608-613.

[7] Wang, X. J., Wang, J. D., Xiang, W. S., and Zhang, J. (2009) Three new milbemycin derivatives from Streptomyces bingchenggensis, *J Asian Nat Prod Res 11*, 597-603.

[8] Zhang, B. X., Zhang, H., Wang, X. J., Wang, J. D., Liu, C. X., and Xiang, W. S. (2011) New milbemycins from mutant Streptomyces bingchenggensis X-4, *J Antibiot (Tokyo) 64*, 753-756.

[9] Mueller, L. S., Hoppe, R. W., Ochsenwald, J. M., Berndt, R. T., Severin, G. B., Schwabacher, A. W., and Silvaggi, N. R. (2015) Sbi00515, a Protein of Unknown Function from Streptomyces bingchenggensis, Highlights the Functional Versatility of the Acetoacetate Decarboxylase Scaffold, *Biochemistry 54*, 3978-3988.

[10] Otwinowski, Z. M., W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods in enzymology 276*, 307–326.

[11] Leslie, A. G. W. (1992) Recent changes to the MOSFLMpackage for processing film and image plate data, *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography 26*.

[12] Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Crystallogr D Biol Crystallogr 67*, 271-281.

[13] Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr D Biol Crystallogr 50*, 760-763.

[14] Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallogr D Biol Crystallogr 62*, 72-82.

[15] Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP, *Methods Mol Biol 364*, 215-230.

[16] Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., and Adams, P. D. (2008) Automated structure solution with the PHENIX suite, *Methods Mol Biol 426*, 419-435.

[17] Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr D Biol Crystallogr 64*, 61-69.

[18] Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P., and Adams, P. D. (2010) Joint X-ray and neutron refinement with phenix.refine, *Acta Crystallogr D Biol Crystallogr 66*, 1153-1163.

[19] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr D Biol Crystallogr 66*, 486-501.

[20] Word J. M., L. S. C., Richardson J. S., Richardson D.C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation, *J Mol Biol 285*, 1735-1747.

[21] Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr 66*, 12-21.

[22] Urzhumtseva L., A. P. V., Adams P. D., Urzhumtsev A. (2009) Crystallographic model quality at a glance., *Acta Cryst D65*, 297-300.

[23] Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation, *Acta Crystallogr D Biol Crystallogr 65*, 1074-1080.

[24] Fenn, T. D. R., D.; Petsko, G. A. (2003) POVScript+: A program for model and data visualization using persistence of vision ray-tracing, *J. Appl. Cryst. 36*, 944-947.

[25] Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *Journal of Applied Crystallography 24*, 946-950.

[26] Ho, M. C., Menetret, J. F., Tsuruta, H., and Allen, K. N. (2009) The origin of the electrostatic perturbation in acetoacetate decarboxylase, *Nature 459*, 393-397.

[27] Smith, B. J., Lawrence, M. C., and Barbosa, J. A. (1999) Substrate-Assisted Catalysis in Sialic Acid Aldolase, *J Org Chem 64*, 945-949.

[28] Stockwell, J., Daniels, A. D., Windle, C. L., Harman, T. A., Woodhall, T., Lebl, T., Trinh, C. H., Mulholland, K., Pearson, A. R., Berry, A., and Nelson, A. (2016) Evaluation of fluoropyruvate as nucleophile in reactions catalysed by N-acetyl neuraminic acid lyase variants: scope, limitations and stereoselectivity, *Org Biomol Chem 14*, 105-112.

[29] Huynh, N., Aye, A., Li, Y., Yu, H., Cao, H., Tiwari, V. K., Shin, D. W., Chen, X., and Fisher, A. J. (2013) Structural basis for substrate specificity and mechanism of N-acetyl-D-neuraminic acid lyase from Pasteurella multocida, *Biochemistry 52*, 8570-8579.

[30] Daniels, A. D., Campeotto, I., van der Kamp, M. W., Bolt, A. H., Trinh, C. H., Phillips, S. E., Pearson, A. R., Nelson, A., Mulholland, A. J., and Berry, A. (2014) Reaction mechanism of N-acetylneuraminic acid lyase revealed by a combination of crystallography, QM/MM simulation, and mutagenesis, *ACS Chem Biol 9*, 1025-1032.

[31] Eaton, R. W. (2000) trans-o-Hydroxybenzylidenepyruvate hydratase-aldolase as a biocatalyst, *Applied and environmental microbiology 66*, 2668-2672.

[32] Ferrara, S., Mapelli, E., Sello, G., and Di Gennaro, P. (2011) Characterization of the aldol condensation activity of the trans-o-hydroxybenzylidenepyruvate hydratase-aldolase (tHBP-HA) cloned from Pseudomonas fluorescens N3, *Biochimica et biophysica acta 1814*, 622-629.

[33] Sello, G., and Di Gennaro, P. (2013) Aldol reactions of the trans-o-hydroxybenzylidenepyruvate hydratase-aldolase (tHBP-HA) from Pseudomonas fluorescens N3, *Applied biochemistry and biotechnology 170*, 1702-1712.

CHAPTER III: IDENTIFICATION OF THE REACTION CATALYZED

3.1 Introduction

Previous studies have shown that, while ADCSF enzymes appear to be structurally homogeneous [e.g., CaADC (PDB entry 3BH2), *Chromobacterium violaceum* ADC (PDB entry 3BH3), *Legionella pneumophila* ADC (PDB entry 3C8W), *Methanoculleus marisnigri* ADC (PDB entry 3CMB), *Streptomyces hygroscopicus* MppR (PDB entry 4JM3), and Sbi_00515 (PDB entry 4ZBO, this study)], there is some diversity in terms of the substrates used and the chemistry catalyzed by members of this superfamily [1]. MppR from S. hygroscopicus (ShMppR), for instance, can catalyze the cyclization of 2-oxo-4-hydroxy-5-guanidinovaleric acid ("4-hydroxy-ketoarginine") as observed by X-ray crystallography [1]. After hydrolysis from the enzyme, the compound is the keto form of the nonproteinogenic amino acid enduracididine. While the precise role of ShMppR in the biosynthesis of enduracididine remains unclear, it is apparent that very little is known about the degree of substrate and reaction diversity within the ADCSF.

Not surprisingly, the divergence of ShMppR and Sbi_00515's catalytic activities from the classical acetoacetate decarboxylases like CaADC is associated with a number of amino acid substitutions in the active site. Chief among these are the loss of an Arg side chain proposed to orient the substrate for decarboxylation in CaADC, loss of the "second Glu" side chain thought to promote decarboxylation in CaADC, and an increase in polarity in the neighborhood of the catalytic Lys in ShMppR and Sbi_00515 [2, 3]. This latter change results in the formation of an α-carboxylate-binding site [1]. Comparative genome analysis of ShMppR and Sbi_00515 and closely related homologues showed that there exists a subset of ADCSF enzymes that share this same

pattern of amino acid substitutions in their putative active sites. These so-called MppR-like enzymes comprise one of seven sequence clusters or families within the ADCSF (Family V, Figure 3).

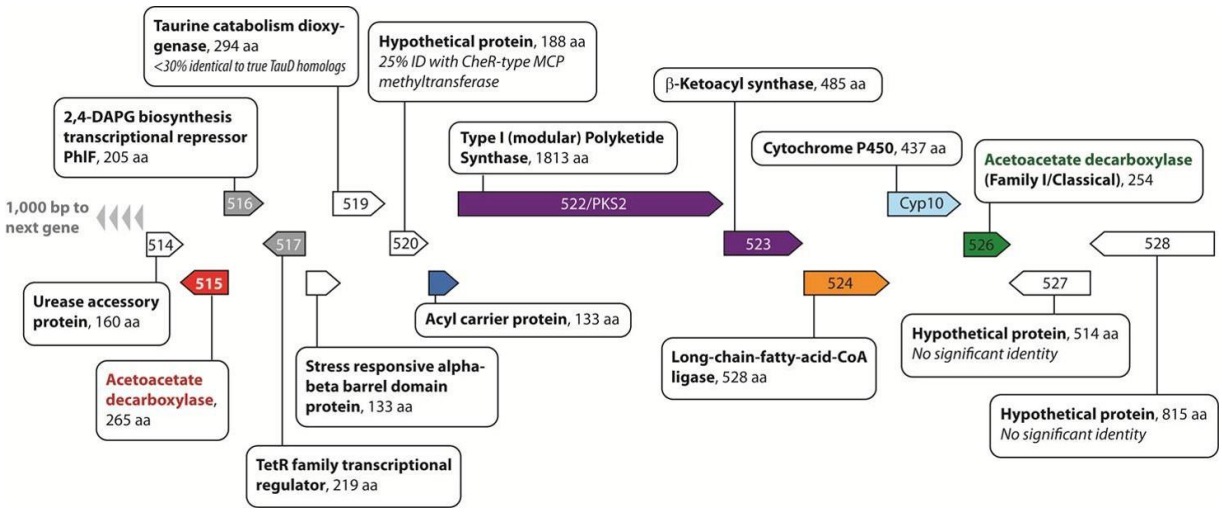


Figure 14. Gene content of Sbi_00515, in red as 515. The bubbles are annotations from the NCBI database. Often in bacteria, genes for a process are located near one another, with the Type I polyketide synthase (522/PKS2) in purple of primary interest.

Sbi_00515 was chosen for study on the basis of its sequence similarity to ShMppR (39% sequence identity) and its gene context, which suggests a potential role in an uncharacterized polyketide biosynthetic cluster (Figure 14). It is predicted that the polyketide biosynthetic cluster produces a compound relevant as an anti-tumor compound, BE-14106 (Figure 15) [4]. Attempts to obtain samples of live cells or genomic DNA were unsuccessful because of intellectual property concerns. As such, the precise biological function of Sbi_00515 cannot currently be verified. Although Sbi_00515 and ShMppR have both been shown by X-ray crystallography to bind pyruvate, the carboxylate group of pyruvate is bound in different conformations in each enzyme

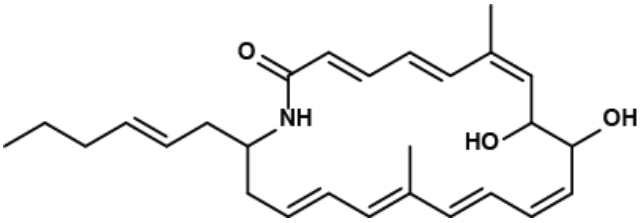


Figure 15. BE-14106, the predicted product of Type I polyketide synthase (522/PKS2) of Figure 14.

[3]. Other subtle active site differences between ShMppR and Sbi_00515 result in different positions for conserved waters. This hints that Sbi_00515 and ShMppR may not have the same activities, despite the fact that both are expected to retain the Schiff base mechanism of acetoacetate decarboxylase based on their active site residues.

3.2 Materials and Methods

Materials

Aldehyde substrates and sodium pyruvate were sourced from Sigma-Aldrich.

Cloning, Expression, and Purification of Sbi_00515.

The protein was previously cloned and expressed as described in Chapter II [3].

Preparation of Benzylidenepyruvate (1).

(E)-2-Oxo-4-phenylbut-3-enoicacid (Scheme 1, **1**) was prepared by a modification of the technique of Reimer [5] . A suspension of benzaldehyde (5.22 g, 49.2 mmol), sodium pyruvate (5.42 g, 49.3 mmol), and KOH (1.41 g,25.2 mmol) in 15 mL of methanol and 15 mL of water was warmed from < 10 to 35 °C and held at that temperature for 1.5 hours, over which time the starting materials dissolved, and then the product crystallized. Filtration, rinsing with 50 mL of cold methanol, and drying *in vacuo* gave 5.25 g of a yellow solid. A solution of 5.18 g of the yellow solid in 150 mL of water was filtered and acidified with HCl (2.0 M, 180 mL), and the resulting precipitate was isolated by filtration and then dried by addition and rotary evaporation of 50mL

of toluene *in vacuo* to yield 2.91 g of product as an orange solid (34% yield). The ^{1}H and ^{13}C NMR spectra of this product were identical to those reported previously [6].

Preparation of 3-(2-furyl)acrolenepyruvate (22).

6-(furan-2-yl)-2-oxohexa-3,5-dieneoate (Scheme 4, **22**) synthesis began with 0.4243 g (3.47 mmol) of 3-(furan-2-yl)acrylaldehyde was combined with 0.420 g (3.86 mmol) of sodium pyruvate and 0.1300g (3.25 mmol) of sodium hydroxide in a dry round bottom flask. Once combined, 50 mL of methanol was added to the reaction mixture and the flask was attached to a reflux condenser, then flushed with nitrogen. The reaction mixture was allowed to reflux for 23 hours, at which point TLC had indicated the total consumption of the starting aldehyde and the reflux was stopped. The reaction mixture was allowed to cool to room temperature then chilled to precipitate the product. The precipitate was filtered and the filtrate acidified by bubbling CO_2 through until a pH of 8.0 was attained. The filtrate was concentrated via rotary evaporation, dried in a desiccator for multiple days and combined with the precipitate afforded 0.8824 g of a bright yellow solid (124%).

The product was placed into an Erlenmeyer flask where upon 20 mL of dichloromethane, 4 mL of trimethylamine and 8 mL of distilled H_2O was added. With stirring, CO_2 was bubbled through the mixture until the trimethylamine and aqueous phases became homogenous. This was transferred to a separatory funnel and the product was further extracted by two more 20 mL additions of dichloromethane. The organic phases were combined, dried over $MgSO_4$ and concentrated by rotary evaporation to afford a dark amber oil. Once dried, the sodium salt was formed by suspending the oil in ethyl acetate and first washing with 1 M HCl until the organic

phase indicated a pH of 2.0. The organic layer was then separated and a 1 M methanolic solution of sodium hydroxide was added until a pH of 8.0 was attained. This was then concentrated by rotary evaporation and dried in a desiccator for two days to afford 0.5430 g (69% isolated) of the sodium salt verified by ESI-LCMS and ^{1}H NMR.

Preparation of cinnamylidenepyruvate (23).

Potassium (3E,5E)-2-oxo-6-phenylhexa-3,5-dienoate (Scheme 4, **23**) was preparped from 3.292 g (24.91 mmol) cinnamaldehyde, 2.804 g (25.48 mmol) sodium pyruvate, and 0.1123 g (2.808 mmol) NaOH were combined with 25 mL CH_3OH before being placed under an atmosphere of N_2 and heated to reflux for 17 h. The volume was observed to be heterogeneous with yellow solid, but TLC showed that the starting aldehyde was still present (25% CH_3CN/CH_2Cl_2 R_f: 0.85) and concentrated to dryness with rotary evaporation. 0.4030 g (10.08 mmol) NaOH and 46 mL CH_3OH were placed into the reaction vessel and heated at reflux for 3 h before the starting aldehyde was no longer visible by TLC. The yellow solid was collected by vacuum filtration, rinsed with two 100 mL portions CH_3OH, and placed under vacuum overnight to afford 2.083 g (37% crude yield). CO_2 was bubbled into the filtrate until it reached pH 8.0 before being concentrated to dryness with rotary evaporation before being combined with 150 mL 18Ω H_2O and 75mL triethylamine (TEA). CO_2 was bubbled into the mixture while on ice until it became homogeneous before being extracted with five 50 mL CH_2Cl_2. The organic portions were combined, dried over $MgSO_4$, filtered, concentrated with rotary evaporation, and placed under vacuum overnight to yield 2.050 g of an auburn oil. The oil was combined with 100 mL EtOAc, heated to boiling, decanted from the remaining undissolved semisolid, and cooled to room temperature. The supernatant was combined with 1.643 g (9.013 mmol) potassium 2-ethylhexanoate (K-2EHA)

dissolved in 30 mL EtOAc which produced a precipitate. The solid was collected by centrifugation, washed with two 30 mL portions EtOAc, and placed under vacuum overnight to remove residual solvent to afford 0.433 g of crude cinnamylidenepyruvate as a yellow solid. Initially, 2.045 g of the isolated yellow solid was dissolved in 185 mL 2.4 M triethylammonium bicarbonate (TEABC) and extracted with four 50 mL portions CH_2Cl_2. The organic extracts were combined, dried over $MgSO_4$, filtered, concentrated with rotary evaporation, and placed under vacuum overnight to yield 1.0963 g of an auburn oil. The oil was combined with 50 mL EtOAc and stirred at room temperature before the supernatant was decanted and mixed with 0.8073 g (4.429 mmol) K-2EHA producing a solid upon mixing. The solid was collected by centrifugation, washed with three 30 mL portions EtOAc, and dried under vacuum to afford 0.7836 g of crude cinnamylidenepyruvate as a yellow solid. Cinnamylidenepyruvate was crude at 1.118 g, was recrystallized from 55 mL CH_3OH and 100 mL EtOAc, boiled down to 1/3 of the total volume, and stored in a -10 ˚C freezer overnight. The resulting solid was collected by centrifugation and dried under vacuum to yield 0.914 g cinnamylidenepyruvate, as a yellow solid (15% yield, 99.4% cinnamylidenepyruvate, 0.6% by mass pyruvate, mp = 204 ˚C (darkens to tan), 228 ˚C (shrinks, darkens to brown).

^{1}H NMR (300MHz, D_2O): δ7.63 (d, J=8.1Hz, 2H); 7.51 (q_{AB}, J=15.6Hz, 9.6Hz, 1H); 7.48-7.41(m, 3H); 7.20-7.07 (m, J=15.6Hz; 8.7Hz, 2H); 6.40 (d, J=15.6Hz, 1H)

^{13}C NMR (300MHz, D_2O): δ198.3; 173.0; 151.7; 144.7; 136.4; 130.6; 129.7; 128.4; 127.4; 126.0

ESI-MS (DUIS, + ion): 519 ($2M + K^+$, 100%); 279 ($M + K^+$, 96%)

ESI-MS (DUIS, - ion): 441 ($2M^- + K^+$, 89%); 201 (M^-, 100%)

Preparation of 4-nitrocinnamylidenepyruvate (24).

Potassium (3E,5E)-6-(4-nitrophenyl)-2-oxohexa-3,5-dienoate (Scheme 4, **24**) synthesis began with 1.456 g (8.222 mmol) *trans*-4-nitrocinnamaldehyde and 1.351 g (12.28 mmol) sodium pyruvate, suspended in 25 mL. NaOH, 0.1133 g (2.832 mmol), was dissolved in 25 mL CH_3OH, added to the reaction mixture, and quantitatively transferred with 10 mL CH_3OH. The mixture was placed under an atmosphere of N_2 and heated at reflux for 3.5 h before the starting aldehyde was no longer visible by TLC (25% CH_3CN/CH_2Cl_2 R_f: 0.87). The mixture was allowed to cool to room temperature, concentrated with rotary evaporation, and placed under vacuum overnight to afford 2.133 g of a brown solid. The crude product, 2.132 g, was added to 120mL 18Ω H_2O and 100mL TEA. CO_2 was bubbled into the mixture until it became homogeneous, extracted with six 100 mL portions CH_2Cl_2. The organic extracts were combined, dried over $MgSO_4$, filtered, concentrated with rotary evaporation, and placed under vacuum to yield 2.372 g of a dark brown solid. The dark brown solid was 1.539 g and was mixed with 140 mL EtOAc. That was heated to boiling before the supernatant was decanted, cooled to room temperature. Potassium 2-ethylhexanoate, 1.357 g, was dissolved in 50mL EtOAc was added to the supernatant, producing a precipitate. The solid was collected by centrifugation, washed with two 40 mL portions EtOAc, and dried under vacuum to produce 0.2123 g of the crude product which was recrystallized from 12 mL CH_3OH and 15 mL EtOAc. After concentrating to 10 mL total volume the mixture brought to room temperature and placed on ice for several minutes before collecting the solid by centrifugation and drying under vacuum to yield 0.1257 g (5% yield) *trans*-4-nitrocinnamaldehyde (98% pure, 2% pyruvate mp = 198˚C (darkens from brown to dark brown).

^{1}H NMR (300MHz, D_2O): δ8.21 (d, J=8.4Hz, 2H); 7.72 (d, J=8.7Hz, 2H); 7.46 (q_{AB}, J=15.3Hz, 8.4Hz, 1H); 7.20 (m, 2H); 6.46 (d, J=15.6Hz, 2H)

^{13}C NMR (300MHz, D_2O): δ198.9; 173.6; 150.9; 149.0; 144.1; 142.4; 132.3; 129.9; 129.0; 125.8

MS (DUIS, + ion): 324 ($M + K^+$, 89%), 150 ($C_8H_7NO_2 + H^+$, 100%)

MS (DUIS, - ion): 531 ($2M^- + K^+$, 65%); 493 ($2M^- + H^+$, 69%); 246 (M^-, 38%); 202 ($M^- - CO_2$, 100%);

Steady State Enzyme Kinetics.

All 0.5 mL kinetic assays were conducted in triplicate at 25 °C in 50 mM Bis-Tris (pH 7.0) for Table 2. The kinetic assays observed in Table 3 and 4 were also at 25 °C, but in 100 mM Bis-Tris pH 6.5 on a TgK DX2 stopped-flow instrument. The initial velocity of Sbi_00515-catalyzed hydrolysis of **1** was measured directly by monitoring the decrease in absorbance at 330 nm. Likewise, initial velocities of Sbi_00515-catalyzed condensation of pyruvate with various aldehyde substrates was also monitored directly. Extinction coefficients and optimal wavelengths for monitoring the reactions are listed in Table 2. The extinction coefficients were determined empirically by comparing appropriate peaks in the proton NMR spectra of the various enone products produced in the condensation reactions with the peak for the α-proton of tryptophan (added to the NMR samples as a standard) to estimate the concentration of the product. The tryptophan stock concentration was calculated from the absorbance at 280 nm and the known extinction coefficient (5540 M^{-1} cm^{-1}) [7]. The k_{cat} and K_M values were determined from the initial velocity data using the equation $v_0 = V_m[A]/(K_M + [A])$, where [A] is the concentration of the aldehyde or benzylidenepyruvate substrate, v_0 is the initial velocity, V_m is the maximal velocity, and K_M is the Michaelis constant.

High-Performance Liquid Chromatography (HPLC) Analysis of Sbi00_515 Reaction Mixtures.

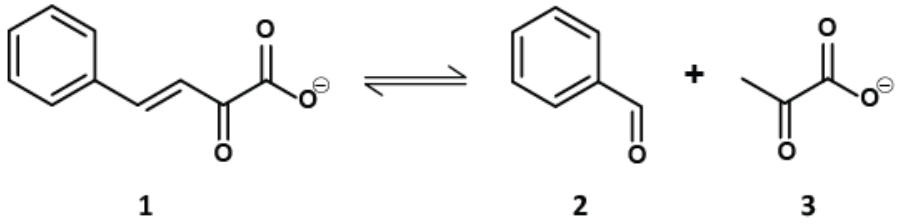
Reaction mixtures (500 µL) containing 12.3 mM **1** in 50 mM sodium phosphate (pH 7.4) and 5.7 µM enzyme, as well as authentic standards of **1** and **2**, were separated using a 50 mm × 2.0 mm, 2.5 µm Synergy Polar RP column (Phenomenex) on an Agilent 1220 HPLC system. The gradient ran from 5 to 50% acetonitrile in a 0.1% TFA/water mixture over 8 min at a flow rate of 0.5 mL/min. Analytes were detected by absorbance at 260 nm. All samples were incubated for 50 min at RT before being passed through 10k molecular weight cutoff centrifugal filter devices to remove the enzyme.

^{1}H NMR Analysis of Sbi_00515 Reaction Mixtures.

Reaction mixtures (500 µL) contained 2.8 µM Sbi_00515 and 12.5mM **1** or 50mM **2** with 50 mM pyruvate in 50 mM sodium phosphate buffer (pH 7.4). The reaction mixtures were incubated at RT for 80 min (**1**) or 35 min (**2** with pyruvate), after which the samples were evaporated to dryness in a CentriFan centrifugal evaporator. The residues were resuspended in 500 µL of D_2O. NMR spectra were recorded at RT on a Bruker DRX500 500 MHz spectrometer equipped with a BBI probe. All data sets consisted of 128 scans. Standard samples of **1** (12.5 mM) and the mixture of **2** with pyruvate (50 mM each) were treated in the same way as the enzymatic reaction mixtures.

3.3 Results and Discussion

Sbi_00515 was first assayed for acetoacetate decarboxylase activity as described by Highbarger et al [8]. Because Sbi_00515 shares the same active site residue changes relative to CaADC that were observed in ShMppR, it was not surprising that it did not react with acetoacetate. As ShMppR was shown to react with pyruvate, a number of α-keto acids, including glyoxylate, pyruvate, oxaloacetate, and α-ketoglutarate, were tested as substrates for Sbi_00515.



Scheme 1. Sbi_00515 catalyzed reaction of benzylidenepyruvate (**1**).

While glyoxylate, pyruvate, and α-ketoglutarate were observed to form Schiff base complexes in X-ray diffraction experiments, and the enzyme binds pyruvate with a K_D of 5.7 ± 0.6 mM as measured by fluorescence titration, none of these were decarboxylated by the enzyme as judged from HPLC analysis of reaction mixtures.

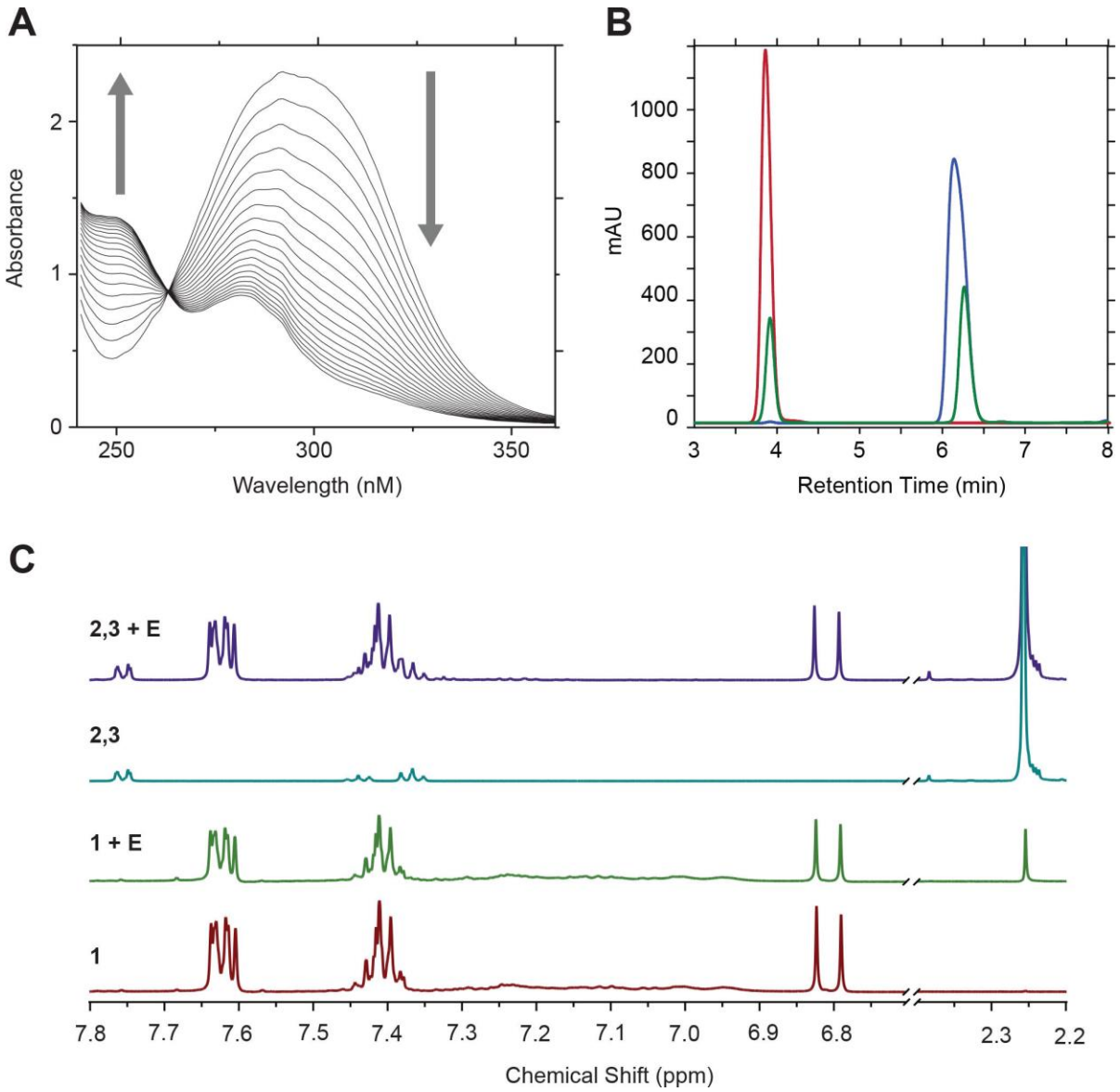
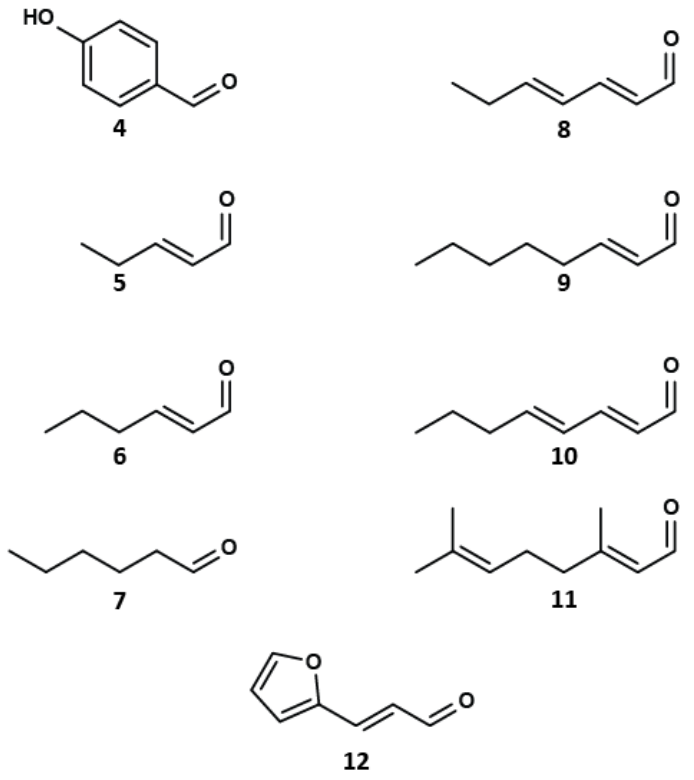


Figure 16. (**A**) Sbi00515-catalyzed cleavage of 1 observed by UV−vis spectroscopy. The spectra are separated by 4.5 min and thus cover a 90 min period. The 500 µL reaction mixture contained 100 mM **1** and 12.8 µM Sbi_00515 in 10 mM MES (pH 6.0). The reaction was initiated by the addition of enzyme. (**B**) HPLC analysis of reaction mixtures and authentic benzaldehyde (**2**) suggests that Sbi_00515 catalyzes the hydrolysis of **1** to give **2** and pyruvate (**3**). Starting material (**1**, blue trace) and benzaldehyde, the putative product (**2**, red trace), are very well resolved, eluting at 6.3 and 3.9 min, respectively. A sample containing 12.3 mM **1** and 5.7 µM Sbi_00515 (green trace) clearly shows a loss of starting material with corresponding accumulation of a product that also elutes at 3.9 min under these conditions. (C) Proton NMR spectra of reaction mixtures and standards confirm that the products of the reaction between Sbi00515 and 1 are 2 and pyruvate. The authentic starting material (**1**, red trace) gives a spectrum that is distinct from that of benzaldehyde (cyan trace). The doublet at 6.8 ppm is characteristic of the unsaturated bond. The methyl protons of pyruvate give a singlet at 2.25 ppm (cyan trace). After incubation with the enzyme, pyruvate is clearly present (green trace). Given 50 mM pyruvate, the reaction can be run in reverse (purple trace), where the doublet at 6.8 ppm confirms the production of **1**. Together, these data confirm that Sbi_00515 has in vitro aldolase-dehydratase activity.

The similarity of the Sbi_00515 structure to that of ShMppR suggested that it should be able to react with α-keto acids. Sbi_00515 was tested for activity with benzylidenepyruvate [**1** (Scheme 1)], as any reaction resulting in modification of the double bond (e.g., hydration) would disrupt the chromophore and be accompanied by a change in the UV−vis absorbance spectrum.



Scheme 2. Aldehydes used in substrate screen I (Table 2).

UV−vis spectroscopy (Figure 16 A) showed that incubation of **1** with Sbi_00515 resulted in the time dependent loss of absorbance at 300 nm and a concomitant increase in absorbance at 254nm. The structure of **1** and the rise of a species absorbing at 254 nm suggested that **1** was being hydrolyzed to give benzaldehyde (**2**) and pyruvate [**3** (Scheme 1)]. Comparison of the reaction mixtures with pure pyruvate and benzaldehyde by reverse phase HPLC analysis confirmed that these were the products of the reaction (Figure 16 B). This was confirmed further by ^{1}H NMR spectroscopy (Figure 16 C). These data clearly indicate that Sbi_00515 was able to hydrate the double bond of **1** and catalyze the retro-aldol cleavage to give benzaldehyde and pyruvate. This same reaction has been observed in the *trans*-*o*-hydroxybenzylidenepyruvate hydratase-aldolase from

Pseudomonas fluorescens, though this enzyme appears, from secondary structure prediction, to have the TIM barrel fold typical of Schiff base-forming aldolases [9, 10].

Since these end-point assays were conducted over relatively long time periods (30−120 min), it was possible that the reaction observed was cryptic and did not represent the physiological activity of the enzyme. To address this question, both the hydrolysis and condensation reactions were characterized by steady state enzyme kinetics. Though the reactions are clearly catalytic, the modest values of the pseudo-second-order rate constants of 54, 62, and 100 M^{-1} s^{-1} when **1**, **2**, and **3** are varied, respectively (Table 2), suggested either that these compounds are nonideal substrates that participate in a facsimile of the native reaction for this enzyme or that they induce nonphysiological chemistry that is only evidence of reaction promiscuity. To clarify this point, a number of additional aldehyde substrates (Scheme 2) were screened for activity with saturating pyruvate (Table 2). Interestingly, 4-hydroxybenzaldehyde (**4**) has a pseudo-second-order rate constant comparable to that of **2**; however, the K_M value is approximately 200-fold less than that of **2**, and k_{cat} is more than 300-fold slower. This may indicate that Sbi_00515 binds **4** more tightly, but in an orientation that is less favorable for catalysis. Moving from the aryl aldehyde **2** to aliphatic aldehydes **5**−**11** results in a marked improvement in K_M with small decreases in k_{cat}, such that the catalytic efficiency improves by as much as 2 orders of magnitude. There is a clear trend of decreasing K_M with increasing chain length going from 6.6 ± 0.4 mM for **5** to 35.0 ± 2.0 μM for **11**. There is no clear relationship between substrate size and the k_{cat} values.

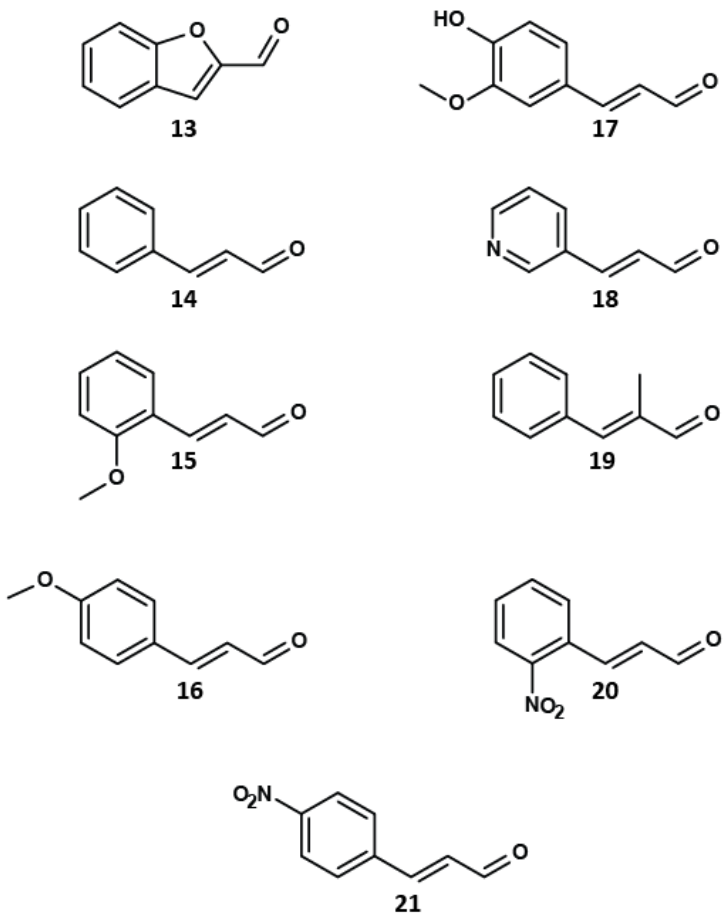
Table 2. Steady State Kinetic Parameters Measured for the Reaction of 2.0 μM Sbi_00515 with the Enone Substrate **1** in the Absence of Pyruvate or in the Presence of Saturating Concentrations of Pyruvate (50mM) and Various Aldehyde Substrates (see Schemes 1 and 2), Subsaturating Concentrations of **2** (50mM) and Pyruvate (**3**), or Saturating Concentrations of **12** (10mM) and Pyruvate[a]

				product	
substrate	k_{cat} (sec-1)	K_M (mM)	k_{cat}/K_M (M^{-1} sec^{-1})	λ_{mon} (nm)	ε (M^{-1} cm^{-1})
benzylidenepyruvate **(1)**	1.2 ± 0.2	22.3 ± 7.0	5.4 ± 1.9 x 10^1	300	24159
benzaldehyde **(2)**	3.2 ± 0.2	51.5 ± 6.1	6.2 ± 0.8 x 10^1	300	11426
4-hydroxybenzaldehyde **(4)**	0.01 ± 0.001	0.4 ± 0.1	2.5 ± 4.0 x10^1	360	1217
trans-2-pentenal **(5)**	1.2 ± 0.1	6.6 ± 0.4	1.8 ± 0.1 x 10^2	290	26021
trans-2-hexenal **(6)**	12.9 ± 0.6	1.8 ± 0.1	7.2 ± 0.5 x 10^3	290	8576
hexanal **(7)**	no reaction	–	–	–	–
2,4-heptadienal **(8)**	0.90 ± 0.04	0.3 ± 0.02	3.0 ± 0.2 x 10^3	360	2432
trans-2-octenal **(9)**	7.8 ± 0.3	0.4 ± 0.04	2.0 ± 0.2 x 10^4	330	447
2,4-octadienal **(10)**	7.2 ± 0.2	0.7 ± 0.1	1.0 ± 0.1 x 10^4	380	1613
3,7-dimethyl-2,6-octadienal **(**citral, **11)**	0.28 ± 0.01	0.04 ± 0.002	7.0 ± 0.4 x 10^3	330	3393
trans-3-(2-furyl)acrolein **(12)**	22.9 ± 0.8	1.2 ± 0.08	1.9 ± 0.1 x 10^4	390	10062
pyruvate **(3)**[b]	1.8 ± 0.1	17.2 ± 2.8	1.0 ± 0.2 x10^2	300	–
pyruvate **(3)**[c]	22.5 ± 1.2	11.8 ± 1.6	1.9 ± 0.3 x 10^3	390	–

[a]As described in the text, the poor binding and limited solubility of **2** made it impossible to fully saturate the enzyme. The kinetic parameters apply to the substrate in column 1. The extinction coefficients given are for benzylidenepyruvate itself (row 1) and the condensation products of pyruvate with the aldehyde substrates tested (remaining rows). [b]Because of the high K_M value for **2**, reaction mixtures were not saturated with respect to this substrate. [c]Reaction kinetics of pyruvate with **12** were nearly saturated (the concentration of **12** was about 10-fold greater than the measured K_M value).

The complete lack of activity with **7**, as judged by HPLC analysis shows that the unsaturated bond adjacent to the aldehyde group is required for the observed catalysis. There also does not appear

to be a relationship between the kinetic values for **9** and **10**, suggesting that the chemistry is taking place between the aldehyde and pyruvate. The greater the degrees of freedom of **9** due to its lack of a second double bond is negligible. The tunnel from the surface to the active site may force **9** to adopt a conformation that mimics the shape of **10**. The β-methyl of **11** is likely responsible for the slower k_{cat} compared to **10** and **12**, as the protruding functional group sticks out into the tunnel of the active site.



Scheme 3. Aldehydes used in substrate screen II (Table 3).

The aromatic aldehyde 3-(2-furyl)acrolein (**12**) was chosen because it combines a short, unsaturated alkyl chain with an aromatic group. This substrate was significantly more efficient than any of the other substrates tested with a k_{cat} value of 22.9 ± 0.8 s^{-1} and a K_M value of 1.2 ± 0.1 mM, for a pseudo-second-order rate constant of 1.9×10^4 M^{-1} s^{-1}. It is interesting to note that the diene substrate **8** reacts 20-fold more slowly than **12** despite the high degree of similarity of their structures. The ability of the furan ring to accept a hydrogen bond and/or stack with an aromatic side chain in the active site may result in **12** binding in a more favorable orientation for catalysis.

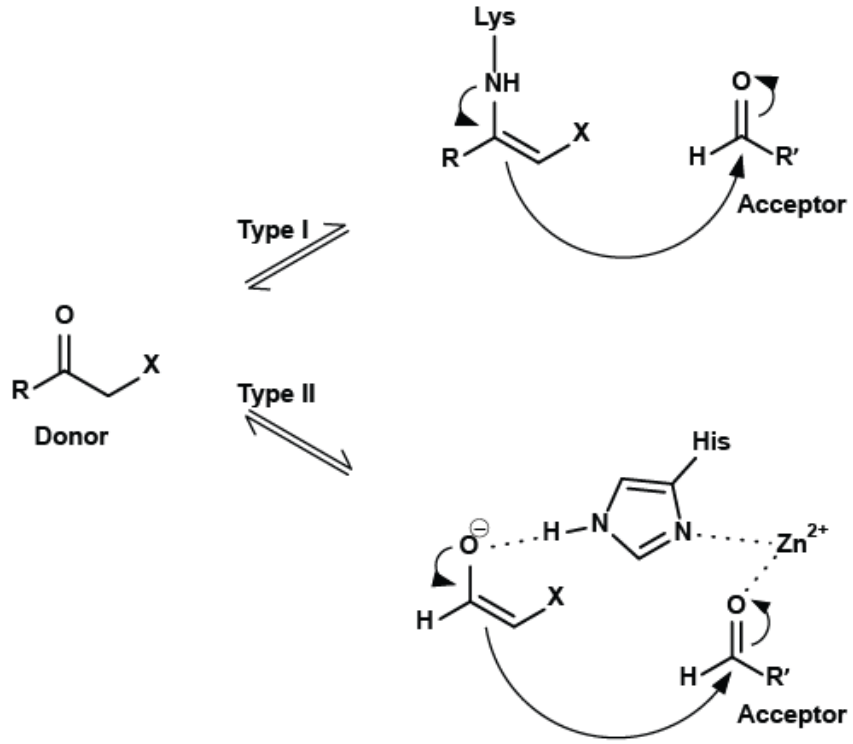
Table 3. Steady State Kinetic Parameters Measured for the Reaction of 0.1 μM Sbi_00515 in the Presence of Saturating Concentrations of Pyruvate (50mM) and Various Aldehyde Substrates (see Scheme 3), or Saturating Concentrations of **14** (1mM) and Pyruvate.

				product
substrate	k_{cat} (sec^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} sec^{-1})	ε (M^{-1} cm^{-1})
2-benzofurancarboxaldehyde **(13)**[a]	7.3 ± 0.2	699.4 ± 45.7	1.0 ± 0.1 x 10^4	10749
trans-cinnamaldehyde **(14)**	28.3 ± 0.4	52.9 ± 1.6	5.4 ± 0.2 x 10^5	17050
2-methoxycinnamaldehyde **(15)**	15.7 ± 0.4	100.2 ± 4.0	1.6 ± 0.1 x 10^5	11908
4-methoxycinnamaldehyde **(16)**	3.3 ± 0.1	13.1 ± 0.9	2.5 ± 0.2 x 10^5	46035
3-methoxy-4-hydroxycinnamaldehyde **(17)**	1.7 ± 0.1	117.2 ± 19.6	1.5 ± 0.3 x 10^4	34851
3-(3-pyridyl)acrolein **(18)**	398.2 ± 2.3	191.6 ± 2.1	2.08 ± 0.03 x 10^6	5117
α-methyl-*trans*-cinnamaldehyde **(19)**	2.0 ± 0.1	426.9 ± 58.8	4.7 ± 0.7 x 10^3	12624
trans-2-nitrocinnamaldehyde **(20)**	24.7 ± 0.6	30.6 ± 3.0	8.1 ± 0.8 x 10^5	11963
trans-4-nitrocinnamaldehyde **(21)**	10.2 ± 0.1	3.1 ± 0.1	3.3 ± 0.2 x 10^6	49616
pyruvate **(3)**[b]	22.3 ± 0.4	2914 ± 181	7.7 ± 0.5 x 10^3	–

The λ_{mon} (nm) for each substrate was 360, except for [a], which was 380 nm. [b]Reaction kinetics of pyruvate with 14 were nearly saturated (the concentration of 14 was about 20-fold greater than the measured K_M value).

Finally, the relatively robust activity with **12** does suggest that the aldol condensation and dehydration activity of Sbi_00515 is real, and not a spurious reaction resulting from the presence of a reactive enamine in the active site. It should also be noted that the aldol condensation activity was observed only with pyruvate. Other α-keto acids that were tested, such as glyoxylate, 2-oxobutyrate, oxaloacetate, and α-ketoglutarate, did not react with any of the aldehyde substrates used in this study. Like the Type I lysine-dependent aldolases, Sbi_00515 is specific for its donor substrate, pyruvate, and promiscuous with respect to the acceptor, aldehyde substrate.

An estimated 1 in 10 enzymes adopt the TIM barrel fold [11], or $(\beta/\alpha)_8$, fold including Type I and II aldolases. Type I aldolases have an essential active site lysine residue that is involved in



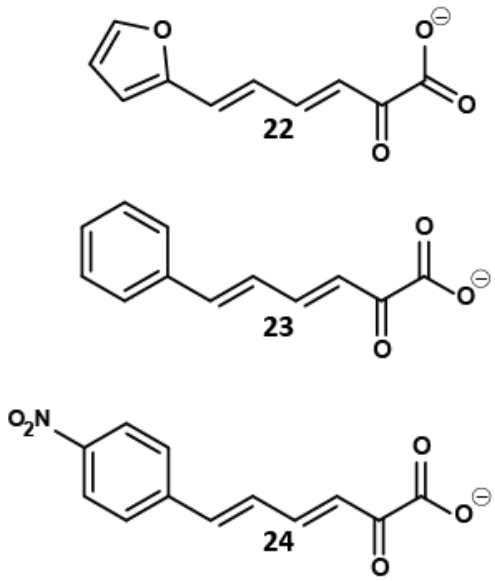
Scheme 4. The general mechanisms for Type I and II aldolases.

Schiff base formation with a donor substrate, pyruvate [12]. Type II aldolases function with a metal ion, typically Zn^{+}, as a cofactor [13]. The general mechanisms for Type I and II aldolases are shown in Scheme 4. The Type I sialic acid aldolase, N-acetylneuraminate lyase (NAL), catalyzes the reversible aldol cleavage of N-acetylneuraminic acid (Neu5Ac), the most abundant of sialic acids, to form pyruvate and N-acetyl-D-mannosamine (ManNAc) [12]. Sialic acid aldolase, as mentioned in chapter 2, Figure 13, uses an active site lysine to form a Schiff base to C2 of an α-keto acid moiety of the substrate [12]. The WT structure of *Pasteurella multocida* NAL shows the that the active site tyrosine (Tyr136) lies almost parallel to the Schiff base intermediate, and is involved in the hydrogen bonding network with the substrate [12]. Tyr136 is proposed to take a hydrogen from the active site lysine, and transfer it to the carbonyl (C2) of pyruvate with the release of a water molecule. The structure, PDB ID 4IMD, shows a conserved water molecule 2.7 angstroms from Tyr 136 and 2.8 Å from the Schiff base intermediate. Although the study by Huynh *et al*. calls the pyruvate-bound structure the Schiff base intermediate, there is no way to tell at 2.10 angstrom resolution whether pyruvate is bound as

the imine or enamine [12]. Another study uses crystal structures, as well as quantum mechanics/molecular mechanics (QM/MM) and site-directed mutagenesis to test the catalytic importance of the active site tyrosine [14]. This time, an *E. coli* NAL was used, and demonstrated Y137F had about 570-fold lower k_{cat}/K_M, with approximately the same K_M values, and no activity with Y137A. QM/MM sampled molecular dynamics, and in all cases observed a spontaneous proton transfer from the active site Tyr137 to the ManNAc aldehyde oxygen [14]. QM/MM also proposed the use of water in the breakdown of the Schiff base and the substrate/product, and reprotonate the Tyr137 phenolate [14].

Adding a bulkier aromatic group than the furan ring of **12** continued to decrease K_M values by an order of magnitude (compare values in Table 2 to 3). Comparing **12** and **13**, the k_{cat} values are approximately equal, while the K_M values differ significantly. The combination of an unsaturated aliphatic aldehyde with a phenyl group is a better match to the shape of the funnel-like active site. In contrast, active sites of the true ADCs are long, narrow tunnels, reflecting their smaller substrate [2]. The lower K_M values in Table 3 suggest that the increased bulk allows for more extensive van der Waals contacts between the aldehyde and the active site funnel. In addition, the aromatic group and *trans* double bond of the aliphatic portion of these aldehydes matches the shape of the active site and creates the potential for π-π interactions with aromatic side chains lining the funnel. While none of the *cis* counterparts of this series of aldehydes were tested for activity, one might speculate that they may not bind as well, since a *cis* compound would require a spontaneous *cis-trans* isomerization before it could be accommodated in the active site.

Trans-cinnamaldehyde (**14**) was a favorable scaffold for investigating the effects of substituents on the phenyl ring, as cinnamaldehyde analogs are commercially available. The K_D for *trans*-cinnamaldehyde was 45.8 ± 4.1 μM. It would have been more advantageous to vary the same substituent around the phenyl ring to control for effects due to the identity of the functional group, however the compounds in Scheme 3 were those available for purchase. The *ortho* or *para* positions were more efficient substrates than those in the *meta* position. The *para* position for **16** and **21** had a smaller k_{cat}, but a tighter K_M than **15** and **20**, leading to a higher pseudo-second-order rate constant. When **16** and **21** are modeled into the active site using the pyruvate-bound structure as an anchor for the pyruvate moiety, both of these compounds more completely filled the active site than **14**, and **21** had a tighter K_D at 14.36 ± 1.32 μM. The more extensive van der Waals contacts with these para-substituted cinnamaldehydes likely accounts for the lower K_M values. The k_{cat} values for **16** and **21** may be smaller than **15** and **20**, which contain a methoxy and nitro group, respectively, in the *ortho* position, due to their positions further into the active site than the *para* positions. Therefore, the contacts are in the tunnel lining the active site, not with bulk solvent like the *para* versions, which could effectively shield them from any additional van der Waals interactions (**15**) or hydrogen bonding (**20**). Substituents in the *meta* position appear to interfere with binding, as evidenced by the the 16-fold slower k_{cat} of **17**. The *para* substituent of **16** fully fills the active site in modeling, the *meta* position of **17** juts out into the hydrophobic residues lining the active site. Surprisingly, 3-(3-pyridyl)acrolein (**18**) has a much faster turnover number than the rest of the substrates in Table 3, however the K_M is almost 200 μM. This could be due to **18** having a lone pair of electrons on the aromatic ring, which could influence how it binds in the active site. The α-methyl group of **19** clashes with the



Scheme 5. Compounds used for retro-aldol cleavage (Table 4).

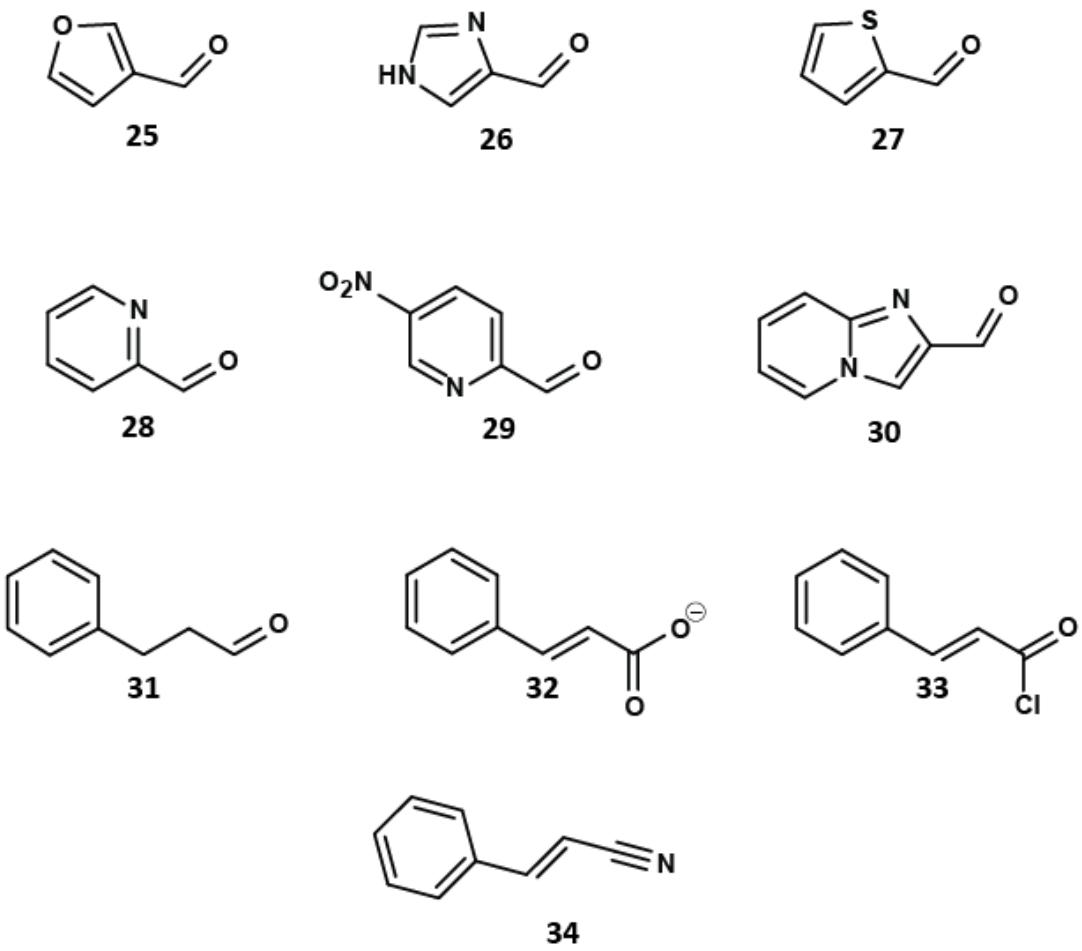
active site, resulting in lower k_{cat} and specificity constant. Based on the values from Table 3.2 for *trans*-4-nitrocinnamaldehyde, **21** is a highly efficient substrate, though not necessarily the physiological one. **14** was used with pyruvate for steady state kinetics in Table 3.2 due to better solubility. The steady state values for pyruvate from Table 3 increased slightly for the pseudo-second-order rate constant, decreased for K_M, yet maintained the same k_{cat} compared to the values in Table 2. Perhaps there is a limit to the turnover number, and the subtle changes in apparent binding affinity and specificity constant reflect the preference for aromatic, unsaturated aliphatic groups for aldehyde acceptors.

After substrates **2**-**21** were tested for the synthetic direction, specific substrates (**12**, **14**, **21**) were chosen to combine with pyruvate to observe the reversibility of Sbi_00515 (**22**, **23**, **24**), as seen in Scheme 5. The compounds for catabolic catalysis are in Scheme 5, and steady state kinetics observed in Table 4.

Table 4. Steady State Kinetic Parameters Measured for the Reaction of 0.1 µM Sbi_00515 in the Presence of **22**, **23**, or **24**.

				product
substrate	k_{cat} (sec^{-1})	K_M (µM)	k_{cat}/K_M (M^{-1} sec^{-1})	ε (M^{-1} cm^{-1})
3-(2-furyl)acrolenepyruvate **(22)**[a]	0.21 ± 0.02	52.9 ± 8.3	4.0 ± 0.7 x 10^3	10,062
cinnamylidenepyruvate **(23)**	1.57 ± 0.04	24.9 ± 1.7	6.3 ± 0.5 x 10^4	17,050
4-nitrocinnamylidenepyruvate **(24)**	0.68 ± 0.04	8.5 ± 0.9	8.0 ± 1.0 x 10^4	49,616

The λ_{mon} (nm) for each substrate was 360, except for [a], which was 390 nm.



Scheme 6. Non-reactive compounds for Sbi_00515 with saturating pyruvate (**3**).

The benzyl groups of **23** and **24** are an order of magnitude greater in their specificity constants than that of the furyl ring of **22**. The better K_M values of **23** and **24** compared to **22** suggest the former fills out the active site more fully, by forming more π-π stacking and van der Waal interactions. **24** has a slower turnover rate than **23**, which may be due to better K_M values. Nevertheless, **22**-**24** are all 1-2 orders of magnitude lower in pseudo-second-order rate constants when contrasted against their synthetic, anabolic counterpart reaction. If there is excess pyruvate in the bacterium [15], that would drive catalysis in the synthetic direction.

Numerous other small compounds were tested in saturating, 50mM pyruvate that demonstrated no activity with Sbi_00515 (Scheme 6). First, other benzaldehyde-like compounds were tried. **25**-**27** may have been influenced by the electronegative atoms within the ring to explain why no reaction with pyruvate and Sbi_00515 occurred. There also could have been unfavorable or unreactive binding in the active site. Benzaldehyde derivatives **28** and **29** did not react. This is likely due to benzaldehyde being a modest substrate to begin with, and the additional electronegative nitrogens could have had an inductive effect, bound to a functional group in the active site to prevent catalysis, or not being able to bind Sbi_00515 without greater than mM quantities. Compound **30** was tested to compare to the 2-benzofurancarboxaldehyde compound (**13**), and the fact that a nitrogen in the ring system could behave, as 3-(3-pyridyl)acrolein does. **30**-**34** were tested as a confirmation that only an unsaturated aliphatic aldehyde would reaction. Hydrocinnamaldehyde, **30**, has too many degrees of freedom to undergo catalysis. There is no unsaturation in the aliphatic portion to keep **30** in a conformation near the aldehyde to direct it towards the pyruvate in the active site for catalysis. **32** – **34** confirmed that an aldehyde was the only functional group able to react with pyruvate in Sbi_00515.

3.4 Conclusions

The steady state kinetic data presented here show that Sbi_00515, a protein of unknown function from *S. bingchenggensis*, has *in vitro* pyruvate aldolase-dehydratase activity and is not an acetoacetate decarboxylase as it is currently annotated. Based on this activity, it is proposed

that Sbi_00515 be renamed to *Streptomyces bingchenggensis* Aldolase-Dehydratase, SbAD. This is the first Schiff base-forming aldolase known to use a fold other than the TIM barrel. Given the proximity of the gene encoding SbAD to an uncharacterized polyketide biosynthetic cluster and the strong preference for an unsaturated aldehyde substrate, it is tempting to speculate that SbAD might synthesize an unusual starter or extender unit for the polyketide. However, more detailed genetic and biological studies must be conducted before anything substantive can be concluded in this regard. Other lysine dependent (TIM barrel) aldolases have been used in protein engineering and as templates to generate enzymes that catalyze reactions not found in nature, as mentioned in the introduction of chapter II [16, 17]. The discovery of SbAD and the versatility of the ADC fold may influence how proteins are engineered, but more mechanistic studies need to be done. Numerous work has been done on TIM barrels, as mentioned previously, for engineering, however the ADC fold has not been considered in these endeavors. The funnel of the active site, lined with hydrophobic residues, could be engineered with bulkier groups to narrow the site, or changed to polar groups and accept for hydrogen bonding substrates. ADC could function as a group to hold potential active sites of designed transition states as *de novo* work, however it remains to be tested.

3.5 References

[1] Burroughs, A. M., Hoppe, R. W., Goebel, N. C., Sayyed, B. H., Voegtline, T. J., Schwabacher, A. W., Zabriskie, T. M., and Silvaggi, N. R. (2013) Structural and functional characterization of MppR, an enduracididine biosynthetic enzyme from streptomyces hygroscopicus: functional diversity in the acetoacetate decarboxylase-like superfamily, *Biochemistry 52*, 4492-4506.

[2] Ho, M. C., Menetret, J. F., Tsuruta, H., and Allen, K. N. (2009) The origin of the electrostatic perturbation in acetoacetate decarboxylase, *Nature 459*, 393-397.

[3] Mueller, L. S., Hoppe, R. W., Ochsenwald, J. M., Berndt, R. T., Severin, G. B., Schwabacher, A. W., and Silvaggi, N. R. (2015) Sbi00515, a Protein of Unknown Function from Streptomyces bingchenggensis, Highlights the Functional Versatility of the Acetoacetate Decarboxylase Scaffold, *Biochemistry 54*, 3978-3988.

[4] Wang, X. J., Zhang, B., Yan, Y. J., An, J., Zhang, J., Liu, C. X., and Xiang, W. S. (2013) Characterization and analysis of an industrial strain of Streptomyces bingchenggensis by genome sequencing and gene microarray, *Genome 56*, 677-689.

[5] Reimer, M. (1931) Benzalpyruvic Acid Dibromide, *J. Am. Chem. Soc. 53*, 3147−3149.

[6] Zhu, L., Meng, Q., Fan, W., Xie, X., and Zhang, Z. (2010) Direct asymmetric hydrogenation of 2-oxo-4-arylbut-3-enoic acids, *J Org Chem 75*, 6027-6030.

[7] Mach, H., Middaugh, C. R., and Lewis, R. V. (1992) Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins, *Analytical biochemistry 200*, 74-80.

[8] Highbarger, L. A., Gerlt, J. A., and Kenyon, G. L. (1996) Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the pKa of active-site lysine 115, *Biochemistry 35*, 41-46.

[9] Ferrara, S., Mapelli, E., Sello, G., and Di Gennaro, P. (2011) Characterization of the aldol condensation activity of the trans-o-hydroxybenzylidenepyruvate hydratase-aldolase (tHBP-HA) cloned from Pseudomonas fluorescens N3, *Biochimica et biophysica acta 1814*, 622-629.

[10] Sello, G., and Di Gennaro, P. (2013) Aldol reactions of the trans-o-hydroxybenzylidenepyruvate hydratase-aldolase (tHBP-HA) from Pseudomonas fluorescens N3, *Applied biochemistry and biotechnology 170*, 1702-1712.

[11] Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P.,

Weaver, D. S., and Karp, P. D. (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases, *Nucleic acids research 44*, D471-480.

[12] Huynh, N., Aye, A., Li, Y., Yu, H., Cao, H., Tiwari, V. K., Shin, D. W., Chen, X., and Fisher, A. J. (2013) Structural basis for substrate specificity and mechanism of N-acetyl-D-neuraminic acid lyase from Pasteurella multocida, *Biochemistry 52*, 8570-8579.

[13] Silverman, R. B. (2002) *The Organic Chemistry of Enzyme-Catalyzed Reactions*, Revised ed., Elsevier, San Diego, CA, USA.

[14] Daniels, A. D., Campeotto, I., van der Kamp, M. W., Bolt, A. H., Trinh, C. H., Phillips, S. E., Pearson, A. R., Nelson, A., Mulholland, A. J., and Berry, A. (2014) Reaction mechanism of N-acetylneuraminic acid lyase revealed by a combination of crystallography, QM/MM simulation, and mutagenesis, *ACS Chem Biol 9*, 1025-1032.

[15] Roy, S. O., and Packard, T. T. (1998) NADP-Isocitrate dehydrogenase from Pseudomonas nautica: kinetic constant determination and carbon limitation effects on the pool of intracellular substrates, *Applied and environmental microbiology 64*, 4958-4964.

[16] Althoff, E. A., Wang, L., Jiang, L., Giger, L., Lassila, J. K., Wang, Z., Smith, M., Hari, S., Kast, P., Herschlag, D., Hilvert, D., and Baker, D. (2012) Robust design and optimization of retroaldol enzymes, *Protein Sci 21*, 717-726.

[17] Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J. L., Althoff, E. A., Zanghellini, A., Dym, O., Albeck, S., Houk, K. N., Tawfik, D. S., and Baker, D. (2008) Kemp elimination catalysts by computational enzyme design, *Nature 453*, 190-195.

CHAPTER IV: ELUCIDATING THE MECHANISM OF SbAD

4.1 Introduction

Steady state kinetics can indirectly probe the order of addition of substrates and release of reactants and products however transitory intermediate information after the substrate(s) bind is not observable by this technique [2]. The pre-steady state, with half-times of less than one second, is the initial period during which intermediates form, and lasts until the steady state is reached [3-5]. The most widely used method of studying fast reactions is stopped flow [3]. Figure 17 illustrates the basic set-up of the instrument. Rapid spectral acquisition via stopped flow can be done by diode-array or a photomultiplier tube (PMT) [2]. Diode-array mode can detect and integrate all of the light for the exposure time to record a snapshot of the spectrum of the reaction solution by using a high-intensity xenon lamp, whereas PMT mode observes a reaction at a specific wavelength [2]. The diode-array mode has a loss in time resolution, but covers a range of wavelengths, whereas

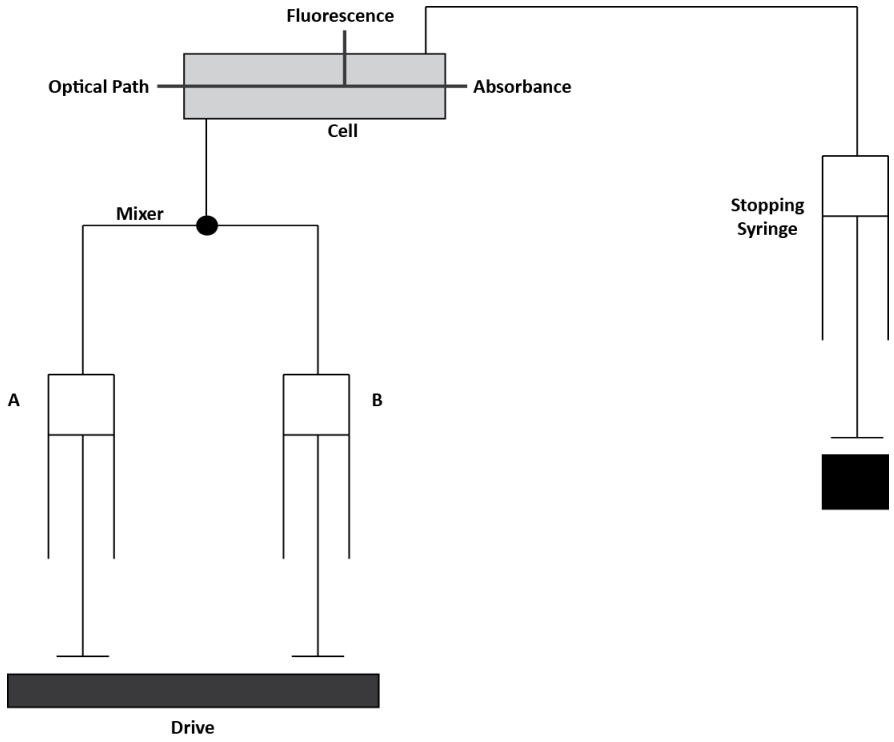


Figure 17. Simplified schematic of a stopped flow apparatus, adapted from TGK Scientific [1]. The drive syringes are labeled **A** and **B**.

a specified wavelength in PMT mode has excellent time resolution. Arrays may not always be able to show parts of the spectrum of interest, but may show intermediates one might not see at a single wavelength (*e.g.* if the reaction were being monitored at an isosbestic point). The reaction is started by pushing the plungers of the two drive syringes simultaneously to cause the reactants to mix. The mixture is then forced through the observation cell, and into the stopping syringe [3]. The mechanical stop activates the detecting and recording system, as well as prevents further mixing. The time between the start of mixing and the start of recording, for instance, absorbance, is known as the dead time, during which the reaction cannot be observed. For most instruments it is usually on the order of a 1–2 milliseconds [3, 7]. Steady state kinetics forms the basis of transient state kinetics [8].

Most experiments are done under pseudo-first order conditions in order to simplify a reaction. The benefit of this in transient kinetics is the direct measurement of reactions at the enzyme active site in order to determine the thermodynamic (equilibrium) and kinetic foundations for enzymatic specificity and efficiency [9, 10]. Even under pseudo-first-order conditions, the data can deviate from a simple, single exponential. One such deviation is a burst phase at the earliest time points, which is indicative of a build-up of intermediate on the way to the establishment of the steady state [2]. Burst kinetics may indicate a covalent intermediate in the reaction, or that product release is the rate-limiting step in the mechanism. The magnitude of the intercept of the burst is typically proportional to the enzyme concentration [3]. There may also be deviations from simple, monophasic data that can arise due to contributions from more than one intermediate during a time course [5]. One method to resolve such ambiguities is single-turnover kinetics. The enzyme in excess over substrate concentration allows for direct

observation of the conversion of substrates to intermediates and products in a single pass of the reactions through the enzymatic pathway [9]. Reactant can be titrated, or chemical agent is added to stop the reaction after one turnover [11]. Although this method can directly determine rate constants, an excess of enzyme is not a trivial matter, as depending on the binding constants of the substrates used, this might require a large or even unrealistic concentration of protein.

The goal of data fitting in transient kinetics is to establish the reaction pathway by investigating the concentration dependence on the rate of the reaction [9]. Unlike steady state kinetics, which fit initial rate data via simple linear regression, pre-steady state kinetic data are fit to integrated equations from simple (pseudo-first order) rate equations. This involves fitting the data to an equation fitting the time dependence of the reaction. Solutions of differential equations describing the pseudo-first order reaction will be a sum of exponential terms, with a single exponential as an example:

$$Y = Ae^{-kt} + C$$

where *A* is the amplitude, *k* is the rate of the reaction, *t* is the time, and *C* is the end point [9]. Nonlinear regression methods are used in an iterative approach for best fit, including the minimization of the sum of square error between the fitted and raw data [9]. Throughout the data fitting process, overfitting the data is a cause for concern. Programs that solely use singular value decomposition may rely too heavily on Chi^2 values, instead of global fitting, as found by the work of K. Johnson [11-14]. Computer simulations of a proposed kinetic mechanism is another important method to determine the validity of the rate constants obtained experimentally from transient (as well as steady state) kinetics. A model of the proposed mechanism is fit to rate constants (and

other values in global fitting), and allowed to digitally proceed. The comparison of the modelled mechanism and experimental values indicate the accuracy of the predicted mechanism, whereby overfitting is again cause for misleading rate constants [11].

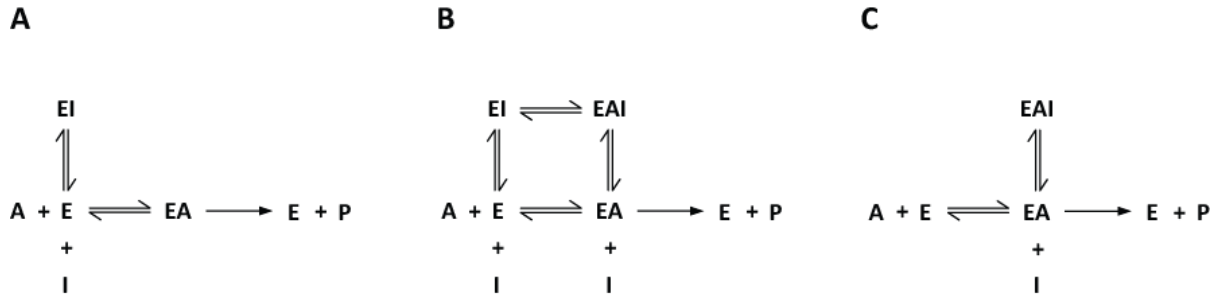


Figure 18. Simplified examples of inhibition, where A is substrate, E is enzyme, I is inhibitor, and P is product. (**A**) Competitive inhibition: Apparent increase in K_M, no change in V. (**B**) Mixed inhibition: apparent V and V/K_M are decreased. (**C**) Uncompetitive inhibition: Apparent V is decreased, V/K_M unchanged. Not pictured is rare, pure non-competitive inhibition: apparent V decreased, K_M unchanged [3].

Steady state and transient kinetics are complementary techniques. For instance, a transient single turnover kinetic experiment may give clues about the order of addition of a two substrate reaction, but not conclusive results. In that case, initial velocity studies with inhibitors is another approach to distinguish between possible mechanisms [2]. Different approaches can be used, depending whether the enzyme exhibits substrate or product inhibition, allosteric regulation, tight-binding/slow-binding inhibition, or whether an inhibitor binds irreversibly or reversibly in the active site. In reversible inhibition, different types of inhibition are possible, based on the classic definitions: competitive, mixed, uncompetitive, or pure noncompetitive inhibition [2, 3, 15]. Illustrations of the terminology are in Figure 18; this is necessary as even within the field there is not good agreement as to what these terms actually mean. Plots of the initial velocity experiments shed light on the type of inhibition used based on fitting the experimental data [16, 17].

One of the most powerful physical organic chemistry tools in mechanistic enzymology is the kinetic isotope effect (KIE), as it can describe what bonds have been broken, formed, or rehybridized during a rate-limiting step [3, 18]. A KIE can be defined as "a change in the rate or equilibrium constant of a reaction upon substitution of a heavy atom for a light one at or adjacent to the position of bond cleavage in a molecule undergoing reaction."[2] Isotopes are isosteric and isoelectronic as they reflect changes in vibrational frequencies of reactants going to products, and therefore do not alter the mechanism of a reaction, just the observed rate constant of a sensitive step [2]. A primary isotope effect occurs when a bond is made or broken to the isotopic (heavy) atom during a reaction, and a secondary isotope effect occurs when the bonding to the isotopic atom changes, but no bonds were broken or made [19, 20]. A secondary KIE is largely due to hybridization changes, whereas a primary KIE mainly arises from the difference in mass on vibrational energy levels. If the heavy atom is attached to the bond undergoing cleavage, it is known as α, whereas if it is in a position attached to an atom adjacent to the one undergoing cleavage, it is known as β. When the bonds being made, broken, or rehybridized involve hydrogen/deuterium, the effect is relatively large and can be measured. KIEs with other atoms have been studied, but the effects are small and difficult to quantify [18]. Isotope effects are given as the ratio of rate constants with light and heavy atom substitutions, for instance k_H/k_D. The ratio is said to be "normal" if it is above unity, or one, and "inverse" if it is less than one. There are three ways to measure isotope effects: direct comparison of initial rates, equilibrium perturbation, and the most sensitive method, internal competition, as it follows the changes in the isotopic ratio as the reaction proceeds [2, 21]. Like all kinetic measurements, careful attention needs to be paid to the proposed step at hand, as overfitting data can affect analysis, or intrinsic

isotope effects on V_{max} values may be masked by slower chemical steps that are not isotope-sensitive [20, 22-24].

The techniques explored above are done in a buffer of specific pH, since enzymes are profoundly affected both structurally and functionally. Most enzymes possess a number of ionizable groups distributed across their surface and in the active site [3]. A pH-rate profile can illustrate the required state of protonation of groups required for binding and/or catalysis, but may also reflect the stability of the enzyme over the pH range [2]. Thus, pH-rate data are notoriously difficult to interpret. The estimated pKa value from a plot of log (V) versus pH, for example can be tested by mutating the residue in question to an amino acid that cannot ionize, and repeat the pH-rate profile for verification. Even in this case, though, the mutation might alter the environment of the active site and thus shift the pK_a of a different residue; the observed effect of the mutation may not be due to the pK_a of the mutated residue. A single or double

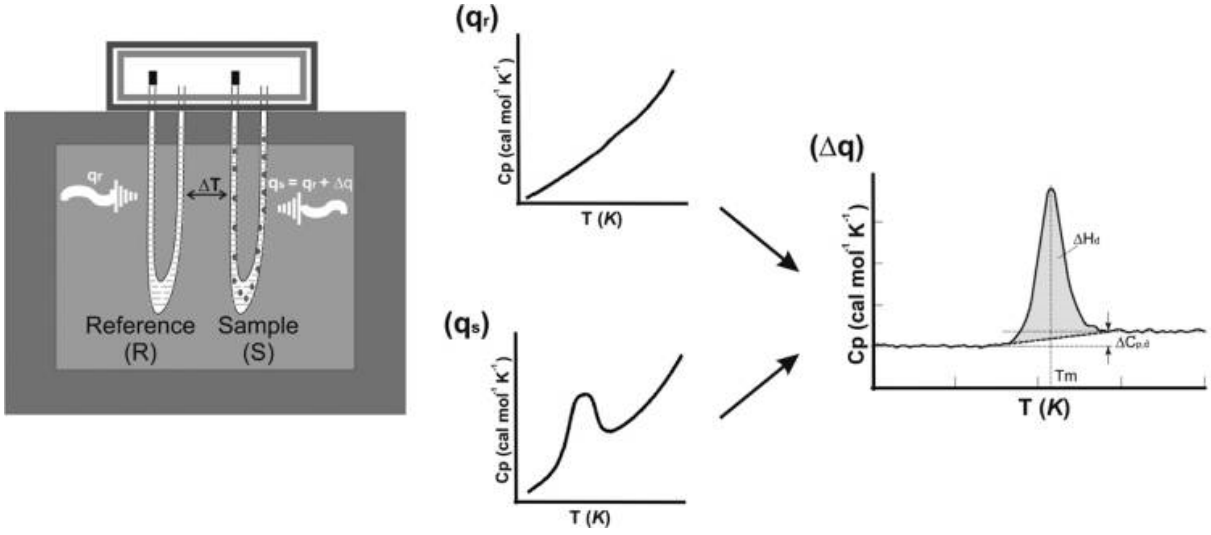


Figure 19. The image is from Gill *et al.*[6] for a DSC experiment. ΔT is the amount of heat required to raise the temperature by the same amount. q_s and q_r are the heat flows for the sample and reference cell, where q_s requires more heat than q_r due to the contribution of protein in the cell, Δq. The scan for q_r is subtracted from q_s to the resultant Δq. ΔH_d is the change in enthalpy, $\Delta C_{p,d}$ is the change in heat capacity, T_m is the transition and melting point, and $_d$ is denatured.

mutant in the active site of an enzyme can hint towards its involvement in catalysis and/or binding in a reaction. Sometimes, a mutant can not only drastically change the activity of an enzyme, it may alter substrate specificity, improve trace activities, or lead to a new reaction mechanism [25]. Steady state and transient kinetics with active site (or other important site) mutants can further illustrate key residues involved in the mechanism. X-ray crystallography is another important tool for the characterization of active site mutants. A crystal structure of a mutant form of the protein may show new binding contacts, or trap an intermediate by slowing down or inhibiting the mechanism.

Mutant forms of a protein may not crystallize in the same conditions as WT, leading to complications in enzyme stability. One method to quickly assess buffers, additives, and/or ligands for stabilization is by a thermal shift assay, also known as a differential scanning fluorimetry or a ThermoFluor™ assay [26]. The goal of this technique is to establish a connection between temperature, and a specific physical property; in this case, the denaturation and aggregation of protein with increasing temperature to find conditions that stabilize the protein. This rapid and inexpensive technique uses a real-time PCR instrument with the fluorescent SYPRO orange dye (λ_{ex} 470 nm/λ_{em} 570 nm). SYPRO orange fluoresces upon binding to hydrophobic portions of the protein, which become more available as the temperature increases and the protein begins to denature. Monitoring the fluorescence emission as a function of temperature gives a change in the emission of the dye that indicates the point where 50% of the protein molecules have denatured: the melting temperature (T_m).

Once buffers, potentially small molecule additives, have been optimized, one can directly determine the thermodynamics of protein unfolding in more detail by using differential scanning

calorimetry (DSC). DSC has numerous applications from chemistry to pharmacology and nanoscience, however one of the focuses in biochemistry, is protein folding [6]. When a macromolecule unfolds, a change in heat capacity is observed. The heat capacity of a folded protein is higher, because it takes energy to surpass the phase change of folded to unfolded protein [27]. The increasing disorder of the protein as the temperature rises produces the heat capacity in the aqueous solution [6]. The simplest DSC experiment consists of a reference cell, filled with only the buffer, and a sample cell, containing the macromolecule in the exact same buffer. The temperatures of both cells are simultaneously raised. The amount of excess heat absorbed or released by the molecule in the sample requires the input energy to match the temperature of the sample to that of the reference [6]. Therefore, more energy is required to bring the sample cell, containing the protein in buffer, to the same temperature as the reference. The power compensation signal is recorded as the calorimetric output [28]. Figure 19 shows a schematic of a DSC experiment and data analysis. The enthalpy can also be calculated in an indirect, non-calorimetric manner by van't Hoff analysis [27]. This takes an equilibrium constant (K) so that the standard free energy change, ΔG°, is zero so that:

$$\Delta G°(T) = -RT \ln K(T)$$

where R is the universal gas constant and T is the temperature in Kelvin, and:

$$\Delta G°(T) = \Delta H°(T) - T\Delta S°(T)$$

where ΔH° is the change in enthalpy and ΔS° is the change in entropy. Combining the equations produces

$$\ln K(T) = \left(\frac{-\Delta H(T)}{R}\right)\left(\frac{1}{T}\right) + \left(\frac{\Delta S°(T)}{R}\right)$$

and rearrangement gives the van't Hoff equation

$$\frac{d \ln K(T)}{d\,(1/T)} = \frac{-\Delta H°(T)}{R}$$

If the enthalpy change derived from the van't Hoff analysis agrees with the value determined by DSC, then the denaturation process of the protein can be approximated by a two state model, which states the protein exists in two forms, folded and unfolded [28]. If the van't Hoff enthalpy value is less than the DSC value, an intermediate during unfolding is likely to have formed [29].

SbAD has undergone ThermoFluor™, DSC, and van't Hoff analysis to more thoroughly understand its stability. Based on substrate specificity and functional analysis, cinnamaldehyde and 4-nitrocinnamaldehyde with saturating pyruvate have been used for study in the synthetic direction, and cinnamylidenepyruvate and 4-nitrocinnamylidenepyruvate for the catabolic direction. The proposed mechanism (Figure 20), is based on the observed aldolase-hydratase activity, and the conserved Schiff base-forming residues from the true ADCs. First, pyruvate binds in the active site. A carbinolamine is formed on the way to a Schiff base intermediate. The aldehyde is oriented by Tyr252 before nucleophilic attack to the carbonyl carbon of the aldehyde. An aldol intermediate is formed, as well as another Schiff base, and then hydrolysis from the enzyme occurs for product release.

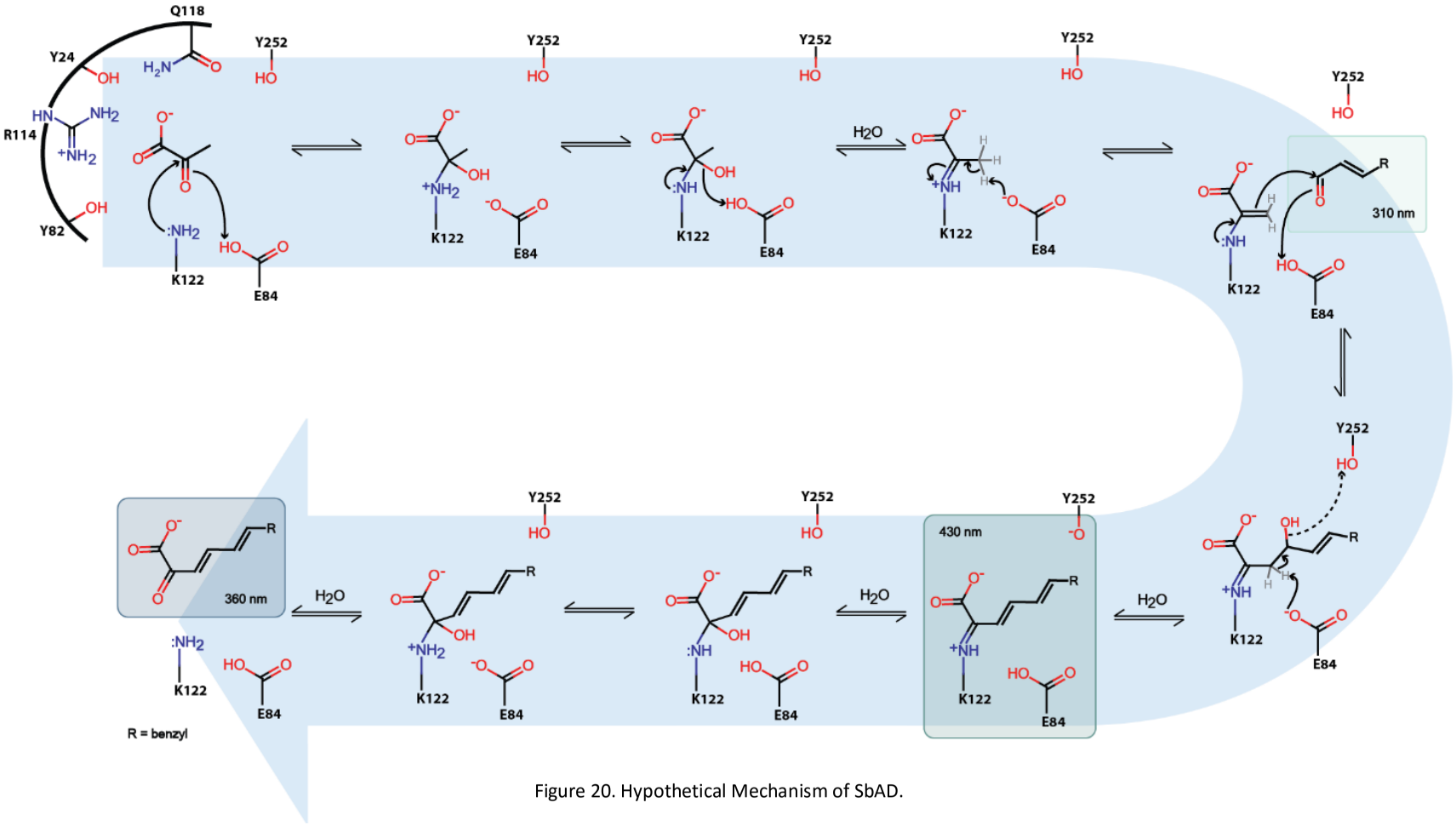


Figure 20. Hypothetical Mechanism of SbAD.

4.2 Materials and Methods

Materials

Cinnamaldehyde, 4-nitrocinnamaldehyde, 2,3-butanedione, 4-phenyl-3-buten-2-one and sodium pyruvate were sourced from Sigma-Aldrich.

Cloning, Expression, and Purification of Sbi_00515.

The protein was previously cloned and expressed as described in Chapter II [30].

Preparation of cinnamylidenepyruvate.

The compound was previously synthesized as described in Chapter III.

Preparation of 4-nitrocinnamylidenepyruvate.

The compound was previously synthesized as described in Chapter III.

Transient Kinetics.

Photodiode array data were collected on a TgK DX2 stopped flow instrument under single turnover conditions. A solution containing saturating, 500 µM, wild-type SbAD was mixed with an equal volume of 17 µM cinnamaldehyde and 50 mM sodium pyruvate (all concentrations represent concentrations in the optical cell after mixing). Reactions were monitored at wavelengths ranging from 317 – 703 nm at 0.8 – 0.9 nm intervals, at 5.0 °C for a total of 1.49 seconds in 50 mM BIS-TRIS pH 7.0. The same conditions were used to study the reverse reaction, where 17 µM cinnamylidenepyruvate was mixed with 500 µM SbAD WT.

Transient state kinetics were also done under pseudo-first order conditions to probe the order of addition. SbAD WT was 491 µM, and saturating pyruvate was 60 mM. The cinnamaldehyde concentrations varied were 10, 15, 20, 25, 30, 35, and 40 µM final, and buffer used was 50 mM BIS-Tris pH 7.0. All reactions were at 5.0 °C. Wavelengths were monitored at 360 nm and 430 nm; all reactions were done in triplicate on a 1.0 second time scale. All fitting was done in OriginPro 2016 Student Version 64Bit.

Kinetic isotope effects with substrates $^{1}H_3$ and $^{2}H_3$ pyruvate were done at 5.0 °C in single turnover conditions. Pyruvate was saturating, 60 mM. Cinnamaldehyde concentrations were 20, 25, 30, 35, and 40 µM final. Experiments were performed on a TgK DX2 stopped flow instrument with 491 µM SbAD WT in 50 mM BIS-Tris pH 7.0. Wavelengths were monitored at 360 nm and 430 nm; all reactions were done in triplicate on a 1.0 second time scale. All fitting was done in OriginPro 2016 Student Version 64Bit.

Steady State Burst Kinetics.

Time courses for the reactions with varying concentrations of SbAD WT with saturating 200 mM pyruvate and 500 µM 4-nitro-cinnamaldehyde were performed at 4.0 °C in 50 mM BIS-Tris pH 6.5.

Steady State Inhibition Kinetics.

All assays are the average of triplicate runs. The kinetic assays were done at 25 °C in triplicate in 100 mM Bis-Tris pH 6.5 on a TgK DX2 stopped-flow instrument. The initial velocity of Sbi_00515-catalyzed hydrolysis of **1** was measured directly by monitoring the decrease in absorbance at 360 nm. Likewise, initial velocities of SbAD-catalyzed condensation of pyruvate

with cinnamaldehyde and either inhibitor **4** or **7** was also monitored directly. All fitting was done in GraphPad Prism 5. The equations used to fit steady state kinetic data by non-linear regression were as follows:

Non-competitive	$Y=(V_{max}*X/(1+[I]/Ki))/(X+K_M)$
Uncompetitive	$V_{maxApp}=V_{max}/(1+I/AlphaK_i)$
	$K_{MApp}=K_M/(1+I/AlphaK_i)$
	$Y=V_{maxApp}*X/(K_Mapp+X)$
Mixed	$V_{maxApp}=V_{max}/(1+I/(Alpha*K_i))$
	$K_{MApp}=K_M*(1+I/K_i)/(1+I/(Alpha*K_i))$
	$Y=_{VmaxApp}*X/(K_{MApp}+X)$
Competitive	$K_{MObs}=K_M(1+[I]/K_i)$
	$Y=V_{max}*X/(K_{MObs}+X)$

Steady State Mutant Kinetics

Assays were done with 0.1 µM SbAD and mutants in 50 mM Bis-Tris pH 6.5 on a TgK DX2 stopped-flow instrument. The initial velocity of SbAD-catalyzed hydrolysis of **3** and aldolase-dehydratase activity of **1** and **2** were measured directly by monitoring in absorbance at 360 nm. The k_{cat} and K_M values were determined from the initial velocity data using the equation $v_0 = V_m[A]/(K_M + [A])$, where [A] is the concentration of the aldehyde or benzylidenepyruvate substrate, v_0 is the initial velocity, V_m is the maximal velocity, and K_M is the Michaelis constant.

Fluorescence Titration for K_D Determination

All work was done on a Hitachi F-4500 fluorimeter with a fluorescence cuvette in triplicate. The cuvette was rinsed between each use. 1.0 – 3.0 μM SbAD WT, and mutants were titrated with substrate at λ_{ex} 280nm for tryptophan excitation and monitored from 300 – 600 nm for λ_{em} in quenching for tryptophan with substrate. A plot of fractional saturation versus concentration was fitted against a model equation ($Y = X / (X + K_D)$) to determine K_D. SbAD WT with sodium pyruvate (**2**) and 4-nitrocinnamaldehyde was done in 100 mM MOPS pH 7.0, and cinnamaldehyde (**1**) in 50 mM BIS-Tris pH 6.8. All fitting was done in OriginPro 2016 Student Version 64Bit and Kaleidagraph.

pH Rate Profile

A tri-buffer system was used containing 100 mM sodium acetate, MES, and TRIS for pH ranging from 4.0 to 9.0 by 0.5 intervals. All kinetic assays were conducted in triplicate at 25 °C with 0.1 μM SbAD WT and E84Q on a TgK DX2 stopped-flow instrument. The initial velocity of SbAD-catalyzed aldolase-dehydratase activity of **1** and **2** were measured directly by monitoring in absorbance at 360 nm. The k_{cat} and K_M values were determined from the initial velocity data using the equation $v_0 = V_m[A]/(K_M + [A])$, where [A] is the concentration of the cinnamaldehyde, v_0 is the initial velocity, V_m is the maximal velocity, and K_M is the Michaelis constant. The plots of log(Initial Rate) versus $[H^+]$ to determine pK_a from the equation $\log(C / (1 + (K_2/H) + (H/K_1)))$, where K_1 and K_2 are the dissociation constants for the ionization of groups that must be unprotonated and protonated, respectively, for catalysis, and C is the pH-independent value of the kinetic parameter of interest [31].

Differential Scanning Calorimetry (DSC)

DSC was performed on a TA Instruments Nano DSC with 9 – 12 mg/mL enzyme in 100 mM MOPS pH 7.0. Experiments were done in triplicate after an initial run to obtain an idea of appropriate temperature range.

Crystallization, Structure Determination, and Model Refinement

SbAD Y252F was previously crystallized as described in Chapter II. Crystals of SbAD Y252F were soaked in 5 mM 4NCP for two hours before being flash cooled. Molecular replacement from the SbAD WT (PDB ID 4ZBO) was used to determine phases.

Size Exclusion Chromatography

The solution molecular weight of SbAD WT, D107N, D109A, and D109N was estimated by gel filtration chromatography using a Sephadex S-200 analytical column on an AKTA FPLC. The column was calibrated using standard proteins including equine heart cytochrome C (12.4 kDa), bovine erythrocyte carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), yeast alcohol dehydrogenase (150 kDa), and sweet potato β–amylase (200 kDa). To define the standard curve, each protein was dissolved at 10mg/mL in the running buffer and the separation was performed on 500 µL of the standard mixture at 25 °C. A sample of the SbAD variants (10 mg/mL) was run immediately afterward using the same column and identical conditions.

4.3 Results and Discussion

No intermediates were detected from the progress curves obtained from steady state kinetics, which is logical because it is cumulative of all the steps in a mechanism. Single turnover

transient kinetics were done to detect, if any, intermediates could be detected in the reaction. Excess SbAD and pyruvate with limiting cinnamaldehyde revealed an increase and concomitant decrease at 430 nm, previously unseen in other UV-Vis analyses (Figure 21). The appearance of cinnamylidenepyruvate, the product of the reaction, can be monitored at 340 – 360nm.

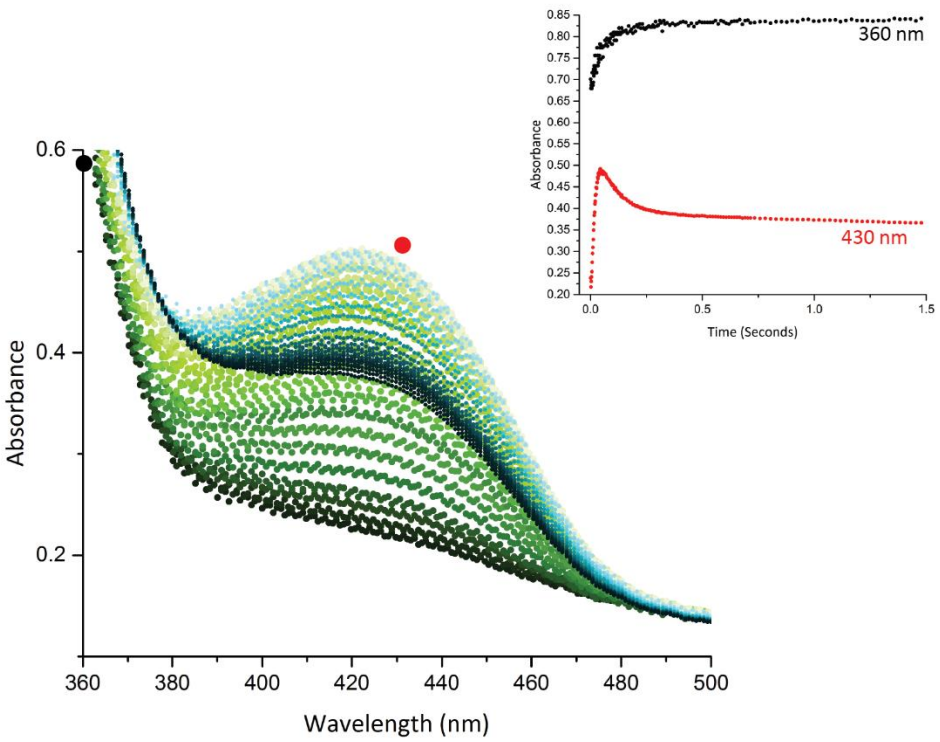


Figure 21. Cinnamylidenepyruvate product formation occurs at 360 nm. There is a rise at 430 nm at the same time as product is released. The decrease and steady state of the longer wavelength indicate an intermediate decrease, and a slow decrease in the steady state as substrate (cinnamaldehyde) is used.

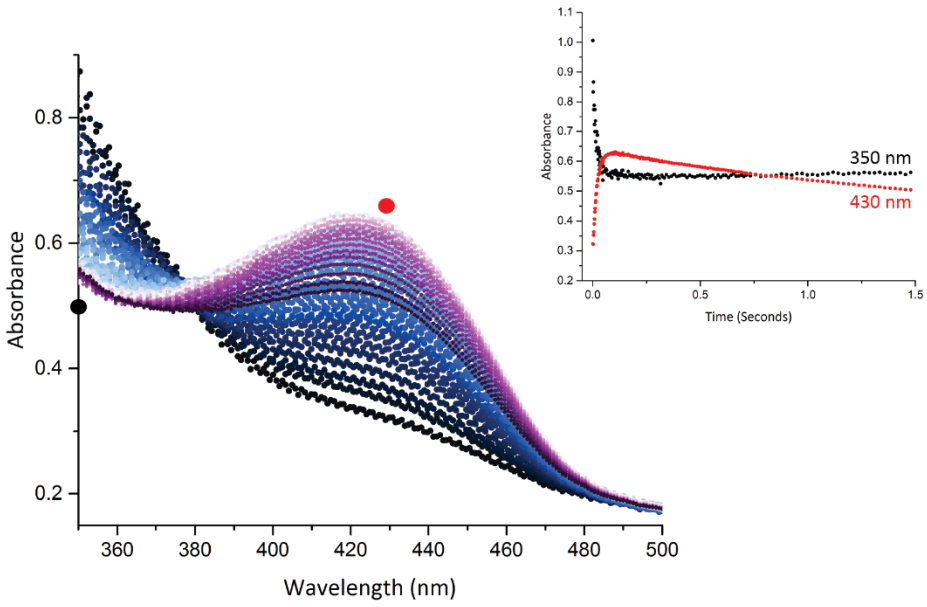
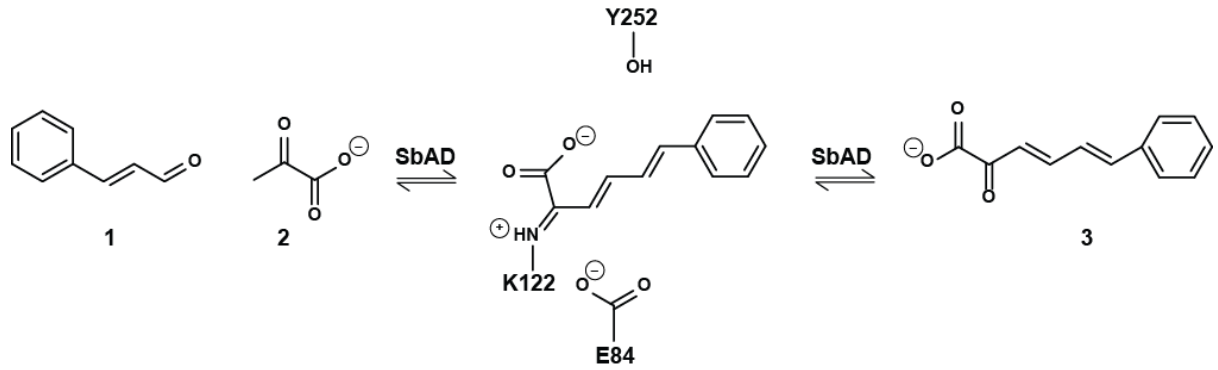


Figure 22. Cinnamylidenepyruvate product decreases drastically at 350 nm while an increase at 430 nm occurs. Once steady state is reached for 350 nm, it remains linear, while 430 nm indicates an intermediate with the decreased slope.

The reversible reaction was also monitored with excess SbAD and limiting cinnamylidenepyruvate. Again the rise and subsequent fall at 430 nm occurred (Figure 22).



Scheme 7. Proposed intermediate at 430 nm.

The longer wavelength spectra indicate a more conjugated intermediate, which is proposed in Scheme 7. Inspection at 430 nm revealed a burst of absorbance before attainment of the steady state rate, as observed in Table 5.

Table 5. Burst kinetics were observed with saturating pyruvate (200mM) and 4-nitrocinnamaldehyde (500 µM).

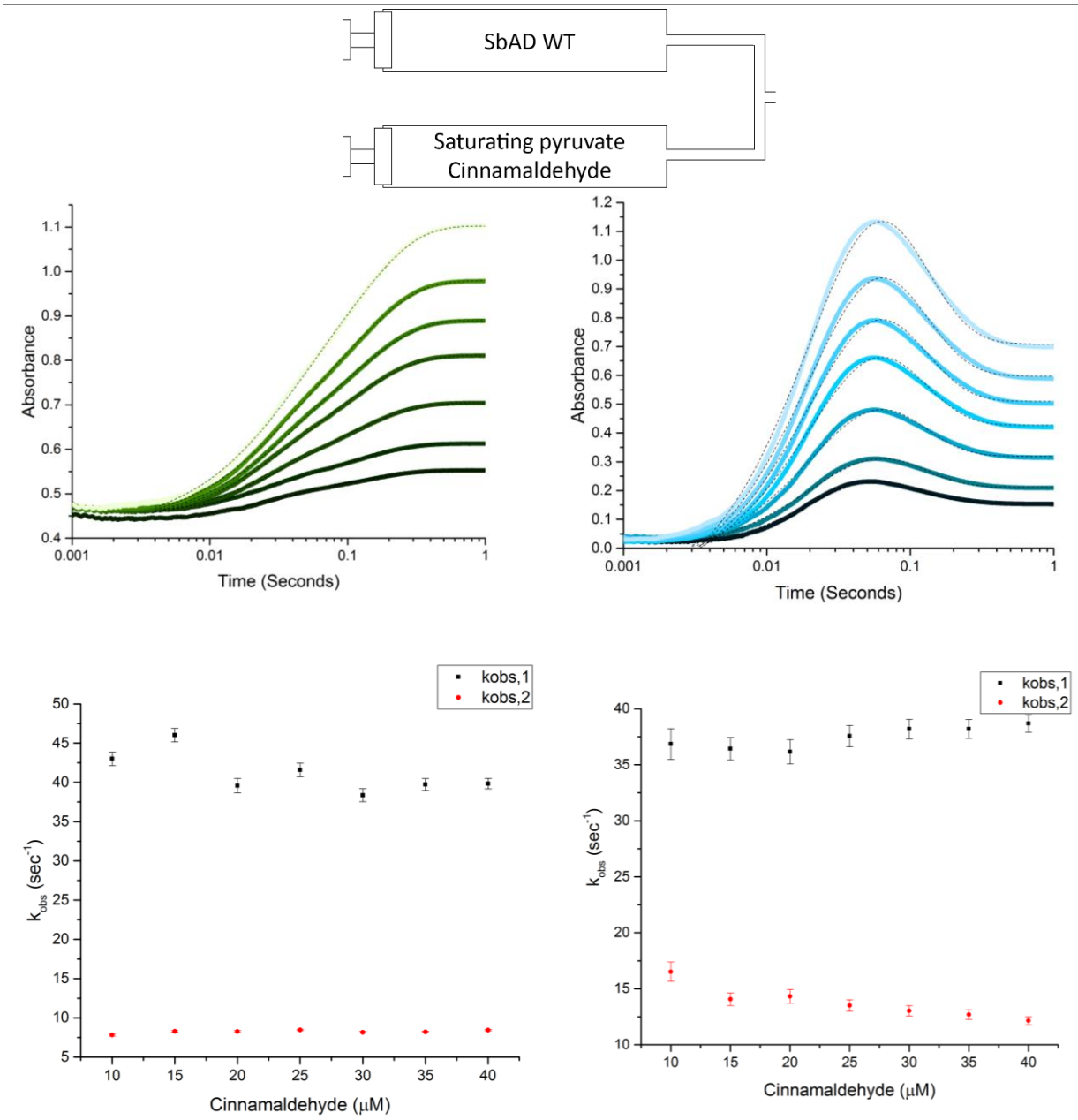
	$y = -A \exp^{(-kt)} + mx + C$		
[SbAD WT] (µM)	1.25	2.5	5.0
Amplitude (A)	0.029 ± 0.002	0.063 ± 0.001	0.136 ± 0.003
Burst rate (k, s^{-1})	262.4 ± 10.7	206.8 ± 2.9	237.1 ± 3.0
Steady state rate (m, A s^{-1})	0.002 ± 1.0 x 10^{-4}	0.004 ± 1.0 x 10^{-4}	0.009 ± 1.8 x 10^{-4}

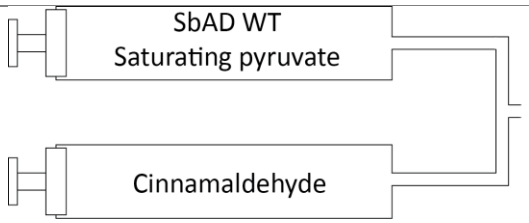
The amplitude doubles as the enzyme concentration doubles, confirming the reaction observed is enzyme catalyzed, and not spurious in solution, likewise with the steady state rate. The burst rate does not appear to vary with enzyme concentration. This indicates that a rate limiting step, likely product release, is building up before attainment of the steady state is reached.

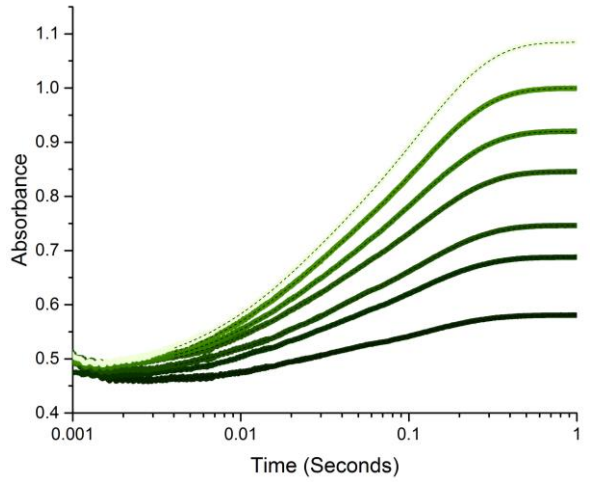
Single turnover transient kinetics were repeated: it was anticipated that in order for cinnamaldehyde to bind and react, pyruvate needed to be bound to the active site lysine first. As

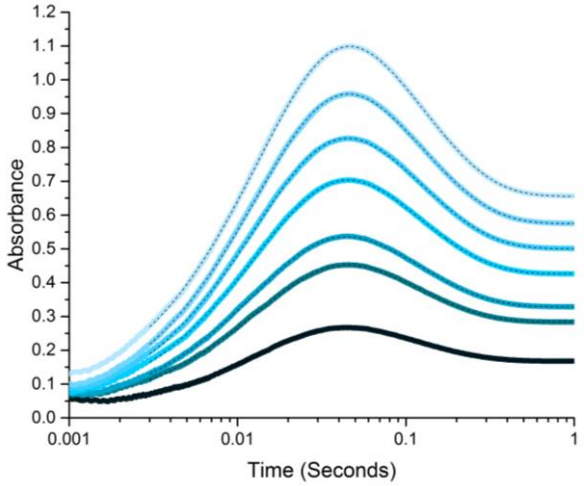
illustrated in Figure 23, two syringes of the stopped flow apparatus were filled with various orders of SbAD and substrate. If cinnamaldehyde were to bind first, it appears that pyruvate would not be able to fit into the narrow active site tunnel. Based on the data from Figure 23, that clearly did not occur, and the reaction proceeded whether or not cinnamaldehyde was incubated with enzyme.

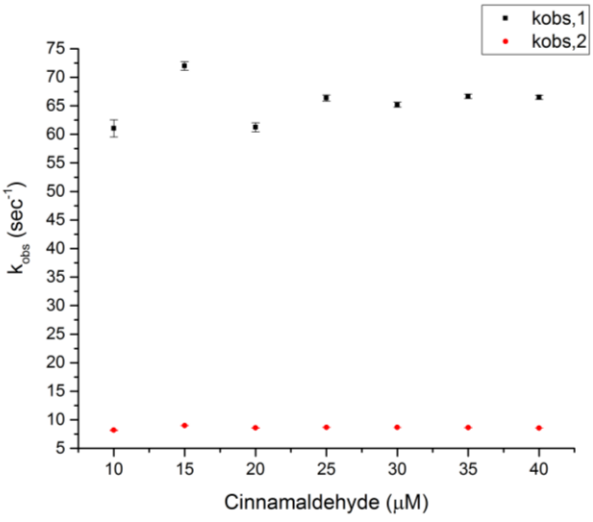
Figure 23. Order of addition reactions were used to determine if there was an observable binding order between pyruvate and cinnamaldehyde for SbAD WT. The syringe schemes represent the syringes from stopped flow, located above the data. The left panel in green is from 360 nm, following product formation. The right panel, with blue traces, monitored 430 nm, is intermediate formation and loss. In both traces, the dashed black line is the fit to the data, interpretation of which is underneath the figures.

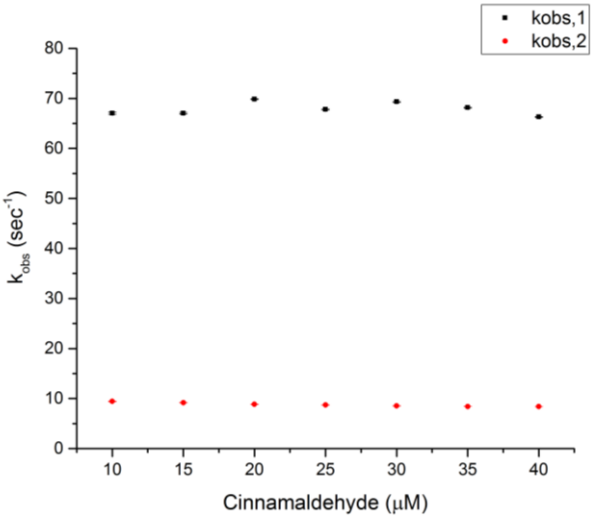


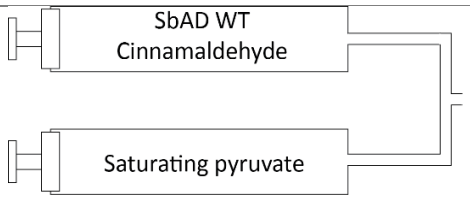

SbAD WT
Saturating pyruvate
Cinnamaldehyde

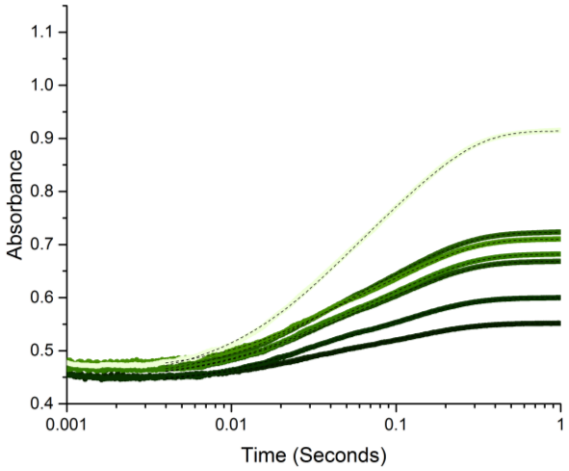

Absorbance
Time (Seconds)

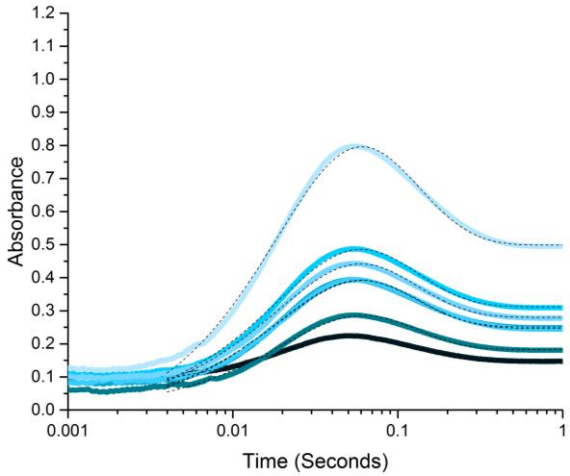

Absorbance
Time (Seconds)

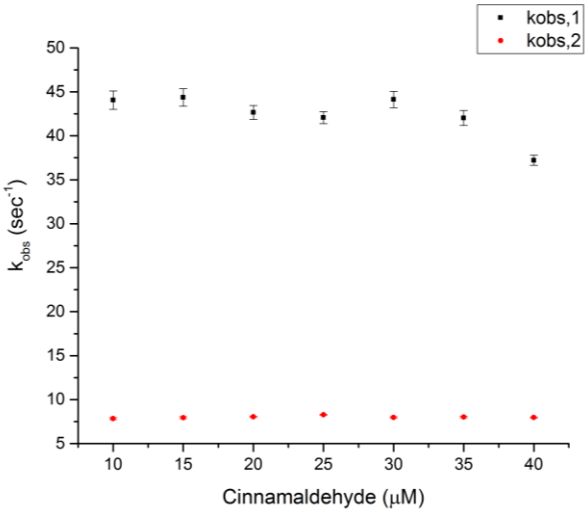

kobs,1
kobs,2
kobs (sec-1)
Cinnamaldehyde (μM)

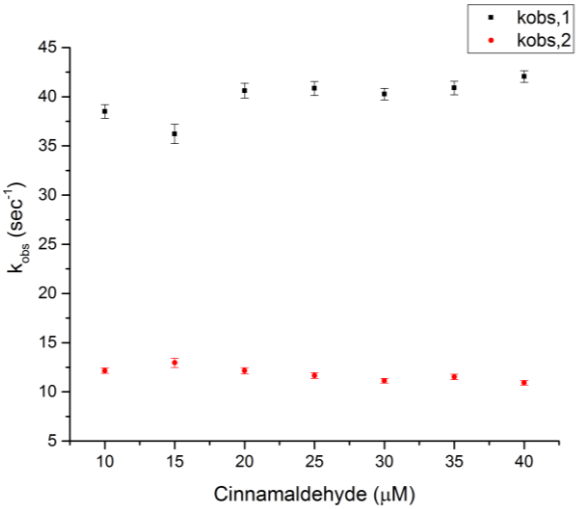

kobs,1
kobs,2
kobs (sec-1)
Cinnamaldehyde (μM)

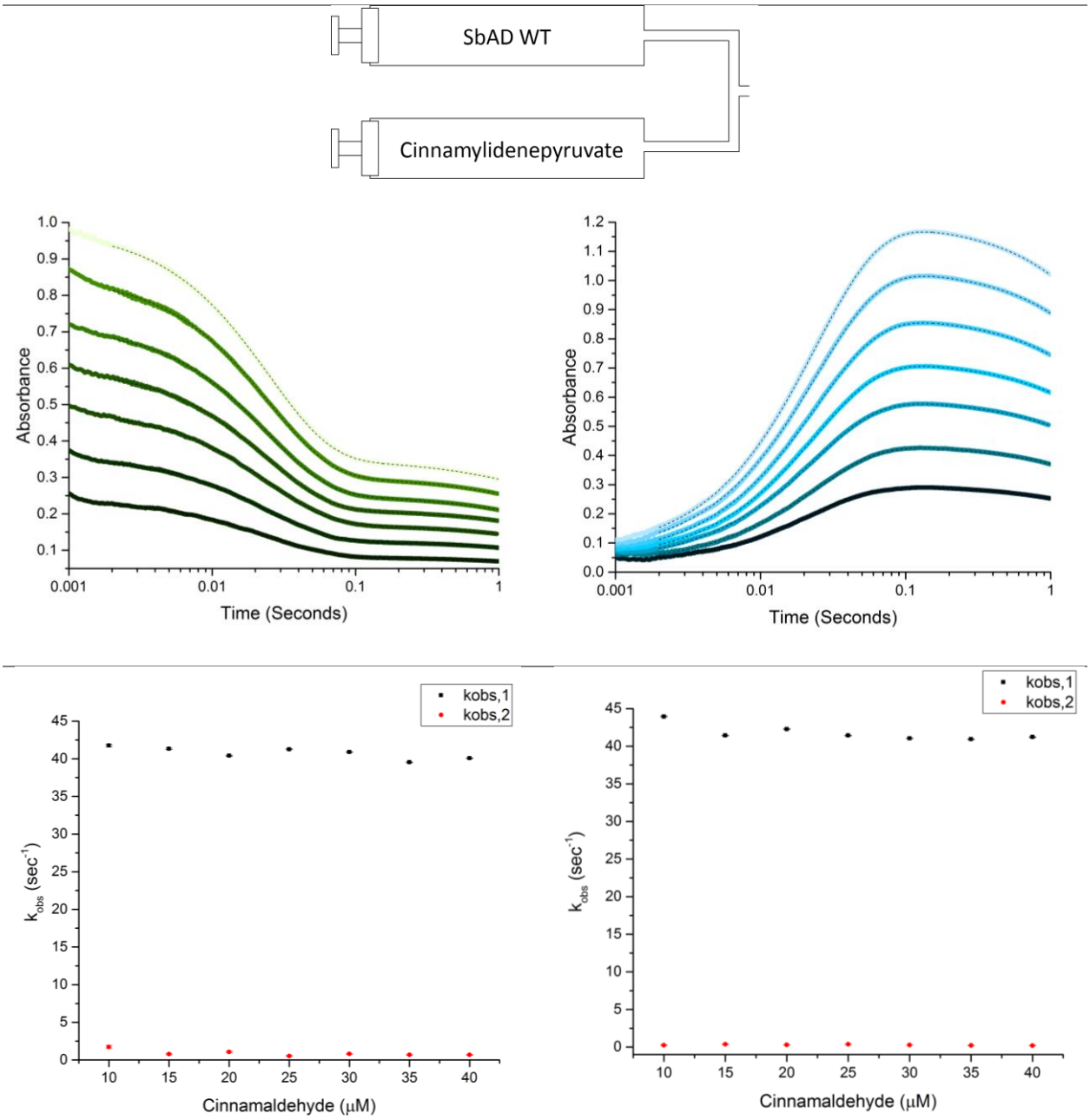

SbAD WT
Cinnamaldehyde
Saturating pyruvate


Absorbance
Time (Seconds)

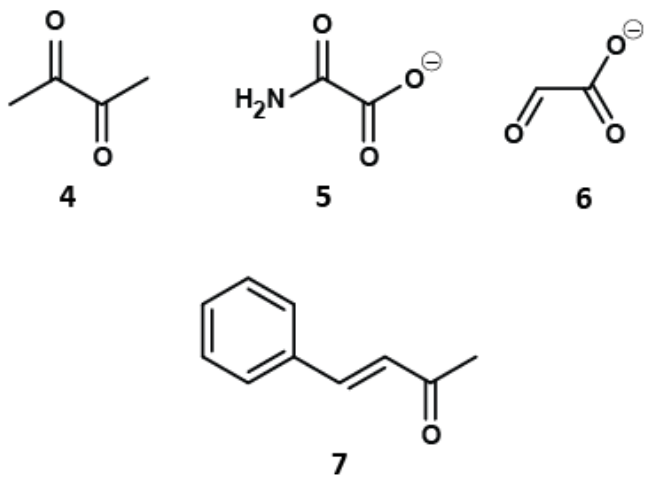

Absorbance
Time (Seconds)


kobs,1
kobs,2
kobs (sec-1)
Cinnamaldehyde (μM)


kobs,1
kobs,2
kobs (sec-1)
Cinnamaldehyde (μM)



Based on all of the order of addition experiments in Figure 23, there is no dependence on concentration for k_{obs}. This implies that if there are any concentration dependent steps, they occur during the dead time of the stopped flow instrument, or that the substrate-dependent binding step is invisible. There is an approximately 20 sec^{-1} advantage to having SbAD incubated with pyruvate previous to cinnamaldehyde mixing than when SbAD is by itself or combined with



Scheme 8. Inhibitors tested for SbAD.

cinnamaldehyde before pyruvate addition, suggesting that it is more efficient to have pyruvate bound first, before cinnamaldehyde. The reaction of SbAD with CMP and SbAD mixed against pyruvate and cinnamaldehyde had the about the same $k_{obs,1}$, showing that the reversibility of SbAD occurs on a similar time scale. Interestingly, the turnover number for steady state kinetics at room temperature was 28.3 ± 0.4 sec^{-1} in 50 mM BIS-Tris pH 6.5, whereas the stopped flow values were recorded at 5 °C in the same buffer, but at pH 7.0. Temperature has an effect on the reaction rate, the discrepancy in rates from temperature likely has to do with steady state being the culmination of all steps in the reaction, whereas single turnover is recorded for transient species. Overall, the synthetic direction with incubated pyruvate provides the reaction with a k_{obs} of 65 sec^{-1}, and about 45 sec^{-1} if SbAD is alone, then mixed with substrates, or if SbAD is incubated with cinnamaldehyde before reacting with pyruvate. The retro-aldol breakdown is about 45 sec^{-1} as well.

Table 6. K_D values for cinnamaldehyde (**1**) and pyruvate (**2**), compared to substrate analogs **4**-**7** as determined by fluorescence titration.

Compound	K_D
Cinnamaldehyde (**1**)	58.18 ± 6.74 μM
4-phenyl-3-buten-2-one (**7**)	42.34 ± 12.78 μM
Pyruvate (**2**)	6.48 ± 0.59 mM
2,3-butanedione (**4**)	12.24 ± 1.67 mM
Oxamate (**5**)	97.04 ± 10.60 mM
Glyoxylate (**6**)	> 500 mM

Though the single turnover order of addition kinetics hinted that pyruvate may be more efficient if bound first, it didn't answer the question of whether or not pyruvate has to bind first for a reaction to occur with SbAD. Inhibition kinetics

addressed this question by competing substrate analogs with pyruvate and cinnamaldehyde. The candidate substrates are shown in Scheme 8, and K_D values in Table 6. The substrate analog for cinnamaldehyde (**1**) was 4-phenyl-3-buten-2-one (**7**), as the nucleophilic attack by the pyruvate enamine on the ketone of **7** cannot occur due to the added bulk of the methyl group. Other options for sodium pyruvate were available, however 2,3-butanedione (**4**) had the most similar binding constant. This suggests that the methyl group of pyruvate may also be important to binding. The best inhibition models and the associated K_i values are given in Table 7.

Table 7. Inhibition kinetics were used determine the order of addition of pyruvate and cinnamaldehyde in the active site of SbAD.

Compound of interest	Varying inhibitor	Inhibition model	K_i
pyruvate	2,3-butanedione	Non-competitive	36.38 ± 2.45 mM
cinnamaldehyde	2,3-butanedione	Mixed	8.12 ± 1.66 mM
pyruvate	4-phenyl-3-buten-2-one	No Effect	–
cinnamaldehyde	4-phenyl-3-buten-2-one	Competitive	258.6 ± 77.39 μM

Based on Table 7, it appears that pyruvate likely binds first, as 4-phenyl-3-buten-2-one has no effect on catalysis when pyruvate is varied. In addition, pyruvate was surprisingly non-competitive with 2,3-butanedione, suggesting that although cinnamaldehyde or 2,3-butanedione may already be in the active site, pyruvate is still able to bind. Cinnamaldehyde was competitive with **7**, which shows that after pyruvate is bound, the aldehyde acceptor is needed next for catalysis, consistent with the proposed mechanism.

Kinetic isotope effects were also done to tease apart the mechanism of SbAD. $^1H_3$ pyruvate and $^2H_3$ pyruvate were observed via single turnover to see the effect deuterated substrate had on the transient kinetics with cinnamaldehyde going in the synthetic direction to

form cinnamylidenepyruvate (CMP). The data are shown in Figure 24, where the end points have

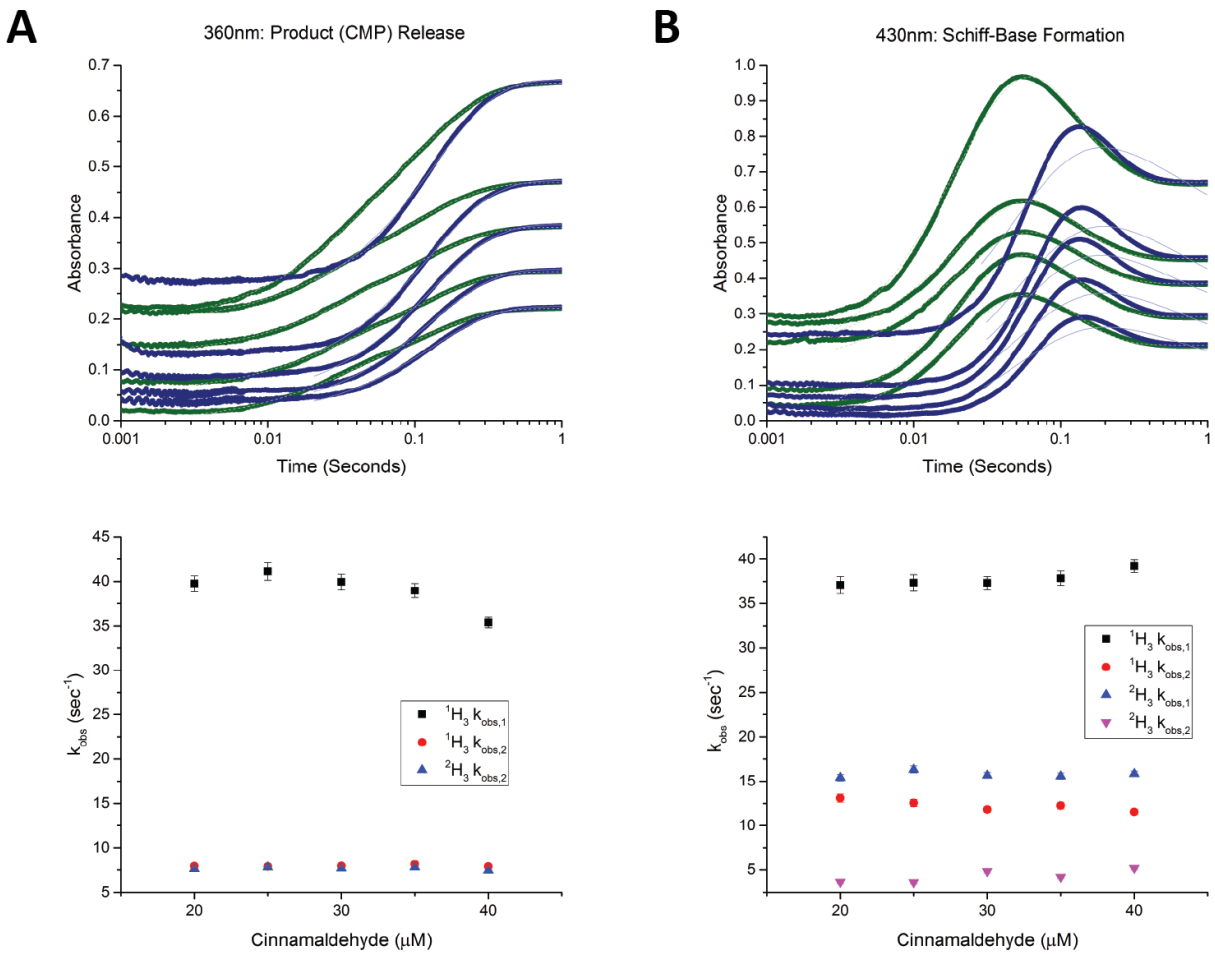


Figure 24. KIE of $^1H_3$ and $^2H_3$ of pyruvate, where $^1H_3$ is in green with a dashed fit line, and $^2H_3$ is in blue with a solid fit line. Underneath are the k_{obs} associated with 360 and 430 nm.

been artificially moved to match one another for comparison purposes. The data at 360nm for $^1H_3$ (Figure 24 A, green) is best modeled by a two exponential fit, whereas the data with $^2H_3$ (Figure 24 A, blue) shows a lag, and then appears to speed up such that only the second k_{obs} from $^1H_3$ is comparable. These results are consistent with a secondary, inverse effect, consistent with the proposed mechanism where the enamine goes from sp^2 to sp^3 during nucleophilic attack on the aldehyde substrate. The 430 nm data (Figure 24 B), however, show a primary effect, as the rates for $^2H_3$ (Figure 24 B, blue) are slower in $k_{obs,1}$ and $k_{obs,2}$. The fits do not agree well with the

data for $^2H_3$ pyruvate (Figure 24 B, blue), suggesting that there is more going than can be modeled by a simple two exponential equation.

In addition to transient kinetics, steady state kinetics were also done with KIE. First, the reaction of **1** and $^2H_3$ pyruvate in the synthetic direction was studied to see if there would be any difference in the steady state rates, Table 8. The values are largely equivalent, which is not surprising, as the steady state rate is all of the accrued steps in the reaction. In the transient kinetic experiments, after 1 second, largely equivalent k_{obs} are obtained. We also analyzed the reaction for a solvent isotope effect. Phosphate buffer was used, as it could be controlled for pH via the Henderson-Hasselbalch equation rather than titrated to pH with deuterated HCl or NaOH. The results for the reaction in the catabolic direction, of **3** going to **1** and **2**, in proteo- and deutero buffer are seen in Table 9. In this preliminary experiment, the k_{cat} shows a clear difference between the reaction rates between the proteo and deuterated solvent, suggesting the importance of water in the retro-aldol cleavage reaction.

Table 8. Steady state kinetics comparing saturating $^1H_3$ and $^2H_3$ pyruvate with cinnamaldehyde, monitored at 360 nm at 25 °C. Reactions were done in triplicate.

	$^1H_3$ pyruvate	$^2H_3$ pyruvate
$k_{cat,cinn}$ (sec^{-1})	9.0 ± 1.1	10.4 ± 1.0
$K_{M,cinn}$ (μM)	23.0 ± 8.1	40.1 ± 6.6
$k_{cat,cinn}/K_{M,cinn}$ ($M^{-1}sec^{-1}$)	3.9 x 10^5	2.6 x 10^5

Table 9. Steady state kinetics comparing saturating ^1H and ^2H phosphate buffer with CMP catalysis, monitored at 360 nm at 25 °C. Reactions were done in triplicate.

	^1H 50 mM PO_4 pH 7.0	^2H 50 mM PO_4 pD 7.4
$k_{cat,cinn}$ (sec^{-1})	0.35 ± 0.01	0.06 ± 0.001
$K_{M,cinn}$ (μM)	18.8 ± 2.5	14.3 ± 1.4
$k_{cat,cinn}/K_{M,cinn}$ ($M^{-1}sec^{-1}$)	1.9 x 10^4	4.2 x 10^3

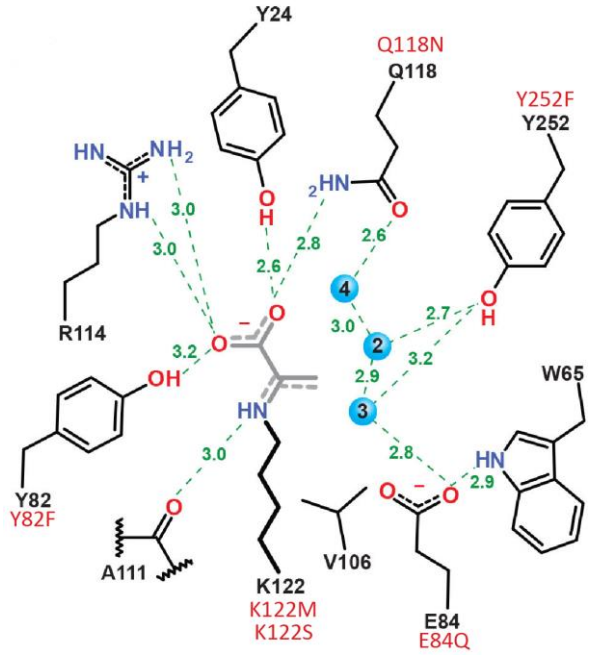


Figure 25. Active site residues of interest that were mutated to observe the effect of steady state kinetics. This is the active site pyruvate-bound structure seen in Chapter II.

Another important component of elucidating the mechanism of SbAD has been mutant analysis. Point mutations were generated at select positions within the active site, then these SbAD variants were characterized by steady state kinetics, fluorescence titration of substrates to test effects on binding, and X-ray crystallography. Figure 25 illustrates the residues that were mutated in the active site. We hypothesized that a change of Try82, a component of the carboxylate binding site, would decrease k_{cat} and increase K_M, as it does contribute a potential hydrogen bonding interaction to the pyruvate carboxylate group. We also hypothesized that mutating the active site lysine would result in an inactive enzyme. Likewise, if Glu84 is the only acid/base catalyst, there should be no activity if it is mutated to a residue that cannot play this role. Tyr252 was the residue bound to the same conserved water in both the formate- and pyruvate-bound structures. We postulated that Tyr252 might be involved in binding the incoming aldehyde substrate and/or orienting it for attack by the enamine. Mutating Tyr252 to Phe removes any potential non-covalent interaction with the aldehyde and should have a noticeable effect on catalysis. Lastly, Gln118 was mutated to Asn. Another Family V enzyme, Swit_4259 (Swit; see Chapter V), naturally has Asn at this position, and appears differ from SbAD in terms of its function. We speculated that mutating Gln118 to Asn would result in an SbAD variant that

Table 10. Steady state kinetics for active site mutants of SbAD at 25°C.

SbAD	$k_{cat,cinnamaldehyde}$ (sec^{-1})	$K_{M,cinnamaldehyde}$ (μM)	k_{cat}/K_M (M^{-1} sec^{-1})
WT	23.00 ± 0.38	38.43 ± 2.03	5.98 ± 0.33 x 10^5
E84Q	2.52 ± 0.02	38.88 ± 1.10	6.48 ± 0.19 x 10^4
Q118N	0.214 ± 0.006	52.86 ± 5.20	4.05 ± 0.41 x 10^3
K122M	No Activity		
Y252F	1.45 ± 0.01	5.42 ± 0.67	2.67 ± 0.33 x 10^5
Y82F Y252F	1.20 ± 0.02	7.22 ± 0.75	1.66 ± 0.17 x 10^5
E84Q Y252F	No Activity		

behaved more like Swit. Double mutants were also created in an effort to detect cooperative residues. Tyr82Phe/Tyr252Phe and Tyr82Phe/Glu84Gln could affect binding and catalysis. If that is the case for the double mutants, the steady state kinetics should have lower k_{cat} and higher K_M than the individual mutants. Glu84Gln/Tyr252Phe was based on two possible key catalytic residues in the active site, and are predicted to render the enzyme inactive if they are both needed for catalysis. The effects of the mutations on the steady state parameters and relevant binding constant(s) in the synthetic direction with cinnamaldehyde and pyruvate are given in Table 10 and 11, respectively. All experiments were performed with near-saturating pyruvate (50 mM) at 25 °C. It is clear from Table 10 that Glu84 is not the sole acid/base catalyst, as there is only a 10-fold reduction in activity in the Glu84Gln mutant.

Table 11. K_D values for SbAD WT and active site mutants

SbAD	Sodium Pyruvate (mM)	Cinnamaldehyde (μM)
WT	6.48 ± 0.59	58.18 ± 6.74
Y82F	1.87 ± 0.27	81.83 ± 8.72
E84Q	5.28 ± 1.06	56.58 ± 3.24
Q118N	5.36 ± 1.01	73.89 ± 6.66
K122M	11.29 ± 1.79	87.29 ± 9.57
Y252F	5.50 ± 0.86	41.74 ± 2.98
E84Q Y252F	54.69 ± 5.99	93.24 ± 9.25

Table 12. Comparison of SbAD WT and Y82F in steady state kinetics at 4 °C with 4-nitrocinnamaldehyde (4NC) and 4-nitrocinnamylidenepyruvate (4NCP).

	SbAD WT	SbAD Y82F
$k_{cat,4NC}$ (sec^{-1})	6.3 ± 0.1	13.9 ± 0.3
$K_{M,4NC}$ (μM)	2.9 ± 0.7	11.8 ± 1.0
$k_{cat}/K_{M,4NC}$ (M^{-1} sec^{-1})	2.2 ± 0.5 x 10^6	1.2 ± 0.1 x 10^6
$k_{cat,4NCP}$ (sec^{-1})	1.9 ± 0.5	1.9 ± 0.2
$K_{M,4NCP}$ (μM)	82.3 ± 29.1	75.1 ± 9.2
$k_{cat}/K_{M,4NCP}$ (M^{-1} sec^{-1})	2.3 ± 1.0 x 10^4	2.5 ± 0.4 x 10^4

The Glu118Asn shows a dramatic decrease in k_{cat}, while the K_M is largely unaffected, suggesting that is plays a catalytic role, at least indirectly, and is not there solely for substrate binding. As expected, Lys122Met had no activity because there is no amine available for the Schiff base chemistry. The Tyr252Phe variant exhibited a tighter K_M than WT, however its k_{cat} was slower. This may be due to the inability to properly orient the aldehyde group for catalysis. All of the values are from steady state kinetics, however, which do not articulate if the activity is due to binding or catalysis. Tyr82Phe was not shown in Table 10 because it was measured with different substrates and at a different temperature. Table 12 compares WT and Tyr82Phe reacting with 50 mM pyruvate and varying concentrations of 4-nitrocinnamaldehyde at 4°C in 100 mM MOPS pH 7.0. The values for the reverse reaction, breakdown of 4-nitrocinnamylidenepyruvate (4NCP) under the same conditions

Table 13. Steady state kinetics for active site mutants of SbAD in the catabolic direction.

SbAD	$k_{cat,CMP}$ (sec^{-1})	$K_{M,CMP}$ (μM)	k_{cat}/K_M (M^{-1} sec^{-1})
WT	1.14 ± 0.05	15.42 ± 1.95	7.39 ± 0.99 x 10^4
E84Q	0.56 ± 0.02	27.13 ± 1.91	2.06 ± 0.16 x 10^4
Q118N	0.37 ± 0.03	16.57 ± 3.33	2.23 ± 0.48 x 10^4
K122M		No Activity	
Y252F	0.74 ± 0.02	4.82 ± 0.43	1.53 ± 0.14 x 10^5
Y82F Y252F	0.53 ± 0.04	26.42 ± 4.56	2.0 ± 0.38 x 10^4
E84Q Y252F		No Activity	

are also given. The Tyr82Phe mutant and WT SbAD are almost identical. It appears that Y82 is not a necessary part of the carboxylate-binding site, and it plays no discernable catalytic role. Table 13, the breakdown direction of cinnamylidenepyruvate (CMP) **3**, the mutants largely follow SbAD WT. Either the loss of conjugation at 360 nm is occurring from the complete breakdown of **3**, or it could be that **3** has bound to the enzyme, lost conjugation, but cannot complete the catabolic process.

The K_D values of cinnamaldehyde and pyruvate were measured to more fully understand mutations upon substrate binding (Table 11). Based on the activity of Gln118Asn, it was proposed that the lack of activity was to do impairment of substrate binding. However, the binding constants for both pyruvate and cinnamaldehyde are identical. Therefore, Gln118Asn may play a catalytic role, perhaps keeping the substrate carboxylate for substrate assisted catalysis, similar to sialic acid aldolase, mentioned in Chapter II [32]. Alternatively, Glu118 might keep another catalytic residue in position for catalysis. The mutants with no activity in steady state kinetics also have K_D values essentially identical to WT. The K_D of Lys122Met has only about doubled in value, suggesting that the carboxylate-binding site is responsible for the bulk of pyruvate binding, at least, and that the measured values are not dominated by the covalent attachment to Lys122. In the Glu84Asn/Tyr252Phe double mutant, the K_D for pyruvate is nearly 10-fold higher, though the K_D for cinnamaldehyde is about the same. The K_D for 4NCP is not in Table 11 because Lys122Met was used to prevent catalysis from occurring. 4NCP with Lys122Met was 14.44 ± 3.52 µM. Lys122Met was a reliable source to approximate K_D as its values with pyruvate an cinnamaldehyde resembles SbAD WT.

Most of the kinetics were done between pH 6.5 and 7.0, however a formal pH rate profile remained to be done. The goal was to find approximate pK_as for the active site residues that could be participating in the reaction. The results for the WT enzyme are shown in Figure 26 A. The results from fitting WT were pK_as of 4.09 and 8.95. The fit does not represent what has been measured, indicating that there are more complex reactions going on than simple proton transfer. The 4.09 pK_a indicated that perhaps the Glu84 mutant was the largely responsible acid/base catalyst, so Glu84Gln also had the pH rate profile done. The pK_as from Glu84Gln were 4.37 and 7.89. It was expected to show different activity if E84 was the only acid/base catalyst, but the fit looks like WT, thereby concluding again that there are more complicated reactions occurring than a pH rate profile can answer.

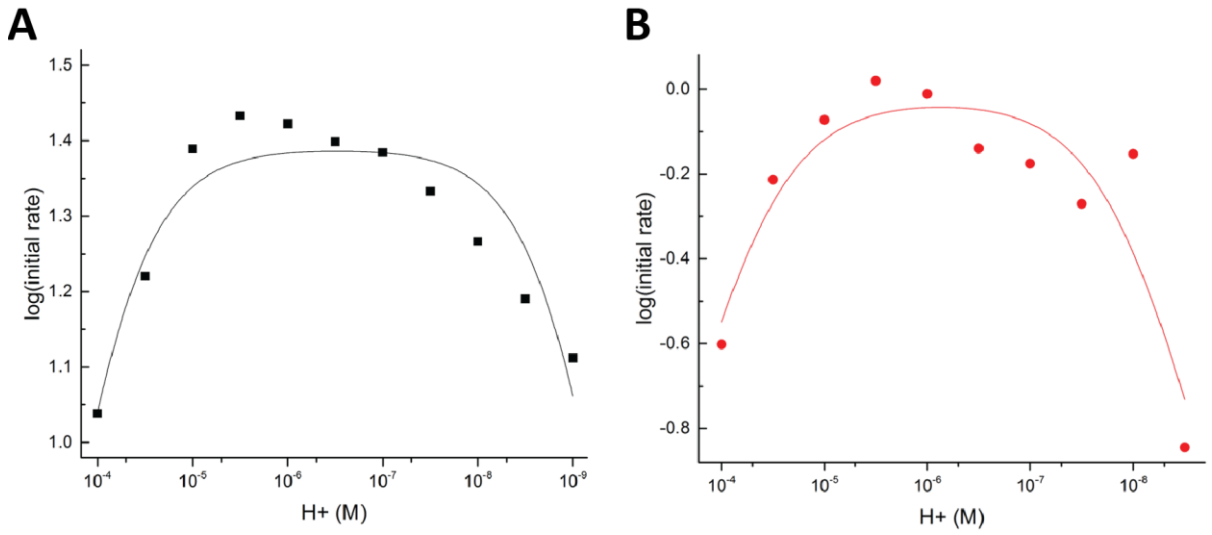


Figure 26. pH rate profile for SbAD WT (**A**) and SbAD E84Q (**B**). Both follow the same overall shape from fitting, concluding that E84 is not the sole acid/base catalyst in the reaction, but does indeed react more slowly than WT.

During the mutagenesis process, it was proposed that E84Q could be impaired due to perturbation of the structure. Therefore, DSC was done to test the stability of this mutant (Figure 27).

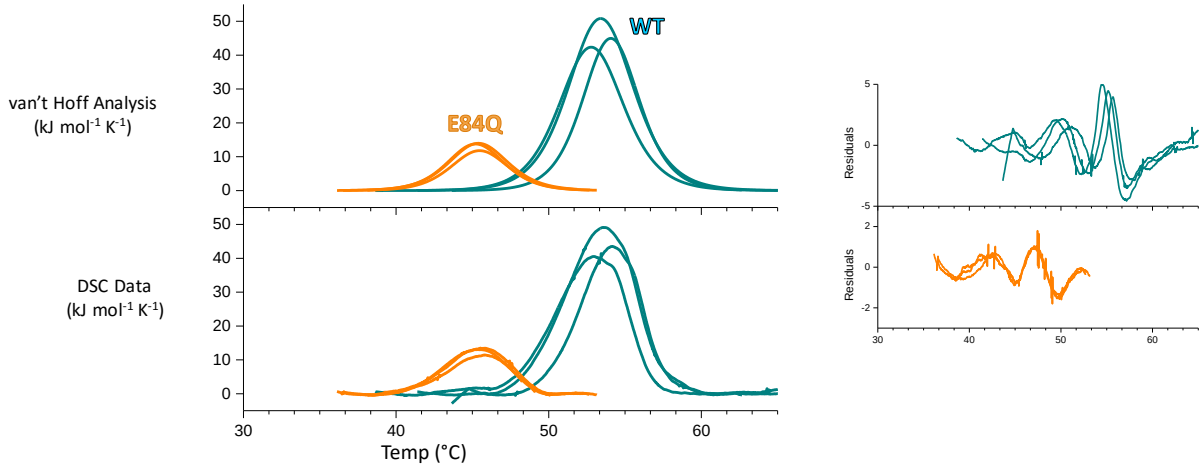


Figure 27. DSC for WT and E84Q. The raw data is compared to the van't Hoff analysis.

Table 14. Results and analysis of DSC for SbAD WT and active site mutant, where ΔH_{vH} is the van't Hoff analysis.

SbAD	T_m (°C)	ΔH (kJ mol^{-1})	ΔS (kJ mol^{-1} K^{-1})	ΔH_{vH} (kJ mol^{-1})
WT	53.4 ± 0.6	241.2 ± 37.0	0.74 ± 0.1	654.3 ± 55.1
E84Q	45.4 ± 0.1	62.2 ± 6.0	0.20 ± 0.0	666.8 ± 11.2

A clear difference is seen for the T_m of Glu84Gln compared to WT, as seen in Table 14. Since the $\Delta H°$ values derived from DSC are smaller than the van't Hoff analysis within the error estimates, it appears that there is a two-state transition between folded and unfolded SbAD. There is no information based on ΔH_{vH} to determine if the proteins formed oligomers before aggregation or intermediates on the way to unfolding.

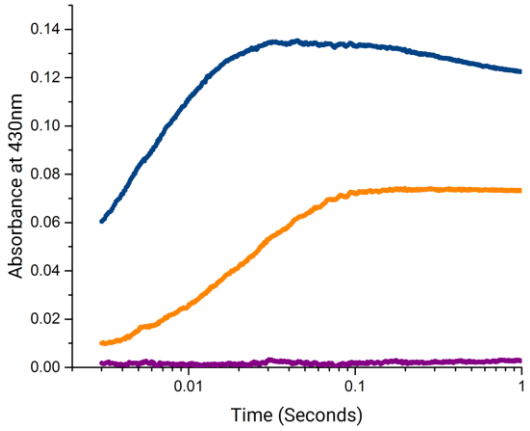


Figure 28. 50 µM 4NCP reacted with 10 µM SbAD WT, Y252F, and E84Q in 50 mM BIS-Tris pH 6.5, 5 °C.

All of the values quoted thus far are from steady state kinetics, however, which makes it difficult to confidently assess each residue's role in SbAD catalysis. Select active site mutants were looked at by transient kinetics under single turnover conditions. At 430 nm (Figure 28), it is

apparent that, while the intermediate begins to decay in the WT enzyme by about 100 milliseconds, it remains stable in Tyr252Phe well past 1 second (see Appendix). The Glu84Gln mutant appears to be completely dead on this time scale, but the reaction is occurring as evidenced by the steady state kinetics (see above). The stability of the proposed Schiff base intermediate in the Tyr252 active site suggested that it would be a good candidate for crystallization.

Table 15. Crystallographic Data Collection and Model Refinement Statistics.

	SbAD Y252F·4NCP
resolution (Å) (last shell)[a]	50.00-1.90 (1.97-1.90)
completeness (%)[a]	99.8 (99.4)
R_{merge} (%)[a,b]	0.110 (0.494)
multiplicity	19.8 (4.0)
⟨I/σ(I)⟩[a]	7.0 (6.0)
	Model Refinement
R_{cryst} (R_{free})	0.166 (0.192)
average B factor (Å²)[c]	
protein atoms	22.2
solvent	32.9
root-mean-square deviation	
bond lengths (Å)	0.005
bond angles (deg)	0.953
Ramachandran statistics (favored/allowed/outliers)	98.2/1.8/0

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

The condensation product was trapped in all four active sites, with waters excluded from the active site tunnel, as seen in Figure 29. The red loop marks the surface of the protein, and mouth of the active site. On either side of the active site loop, there are two openings to this narrow, funnel-shaped active site, and the Tyr252Phe mutant is visible (red label in Figure 29). Aside from the carboxylate-binding site, the rest of the active site is hydrophobic and binding of

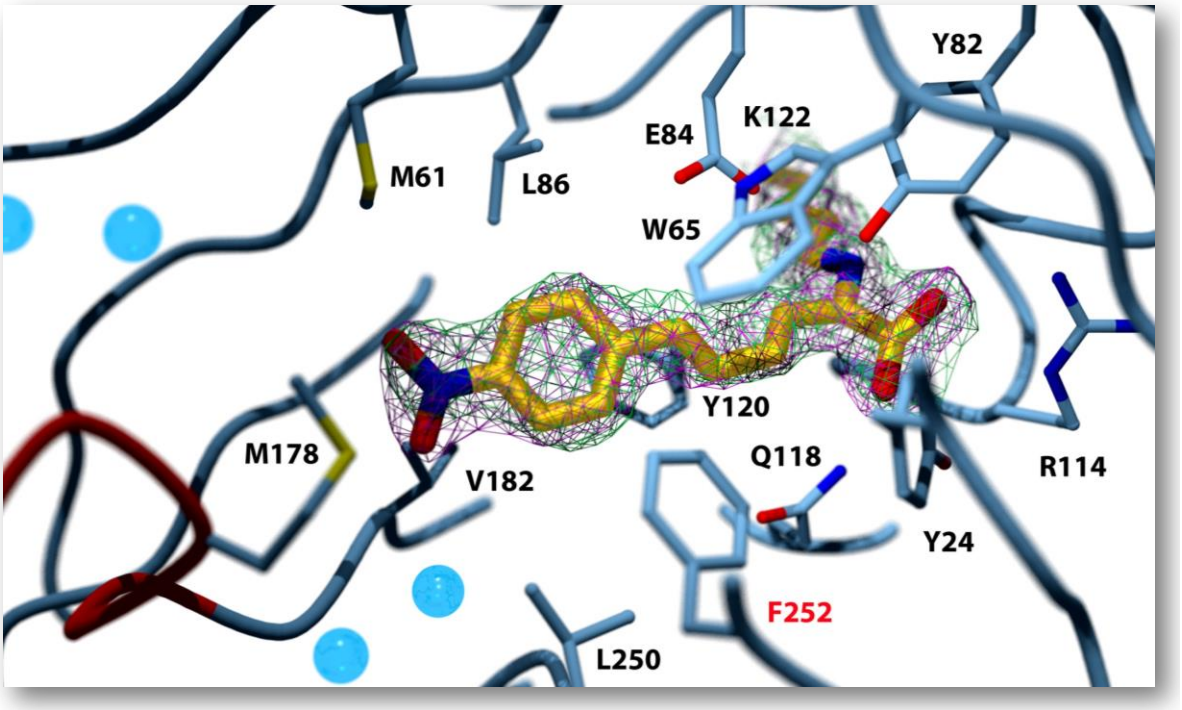


Figure 29. Crystal structure of SbAD Y252F with 4NCP bound in the active site. $2|F_o|-|F_c|$ electron density is in magenta and SA omit electron density is in green. Both are contoured at 1.2σ.

the product is ruled by Van der Waal interactions. It may be that this complex is trapped because there is no hydrogen bonding group or water to facilitate hydration of the double bond. Thus, Tyr252 would appear to have a catalytic role. The torsion angles of the bound product (Figure 30 and Table 16) are also intriguing.

Table 16. Values of the dihedral were averaged over the four active sites in the crystallographic asymmetric unit.

Torsion angles	Average (degrees)	Range (degrees)
1	169.9 ± 3.0	166.4 – 173.3
2	163.7 ± 5.1	156.2 – 167.8
3	164.8 ± 2.8	161.6 – 167.9
4	**82.4 ± 12.8**	**72.2 – 100.5**
5	175.2 ± 2.5	171.5 – 176.8
6	34.4 ± 13.1	19.1 – 48.0

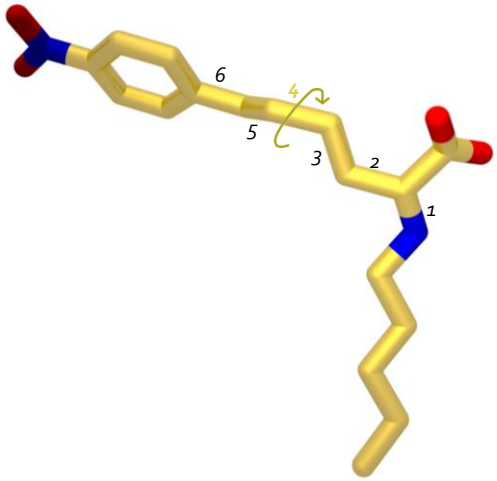


Figure 30. Torsion angles of the 4NCP complex.

The torsion angle labeled as 4 is near 90 °, suggesting that it contributes to substrate destabilization by breaking up the conjugation from the ring system, and making it easier to add water across the double bond that would lead to an aldol intermediate before retro-aldol cleavage. The Schiff base complex with 4NCP is the intermediate thought to be seen at 430nm. If the conjugation is broken like this when the protein is in solution, then this Schiff base should not be seen spectroscopically.

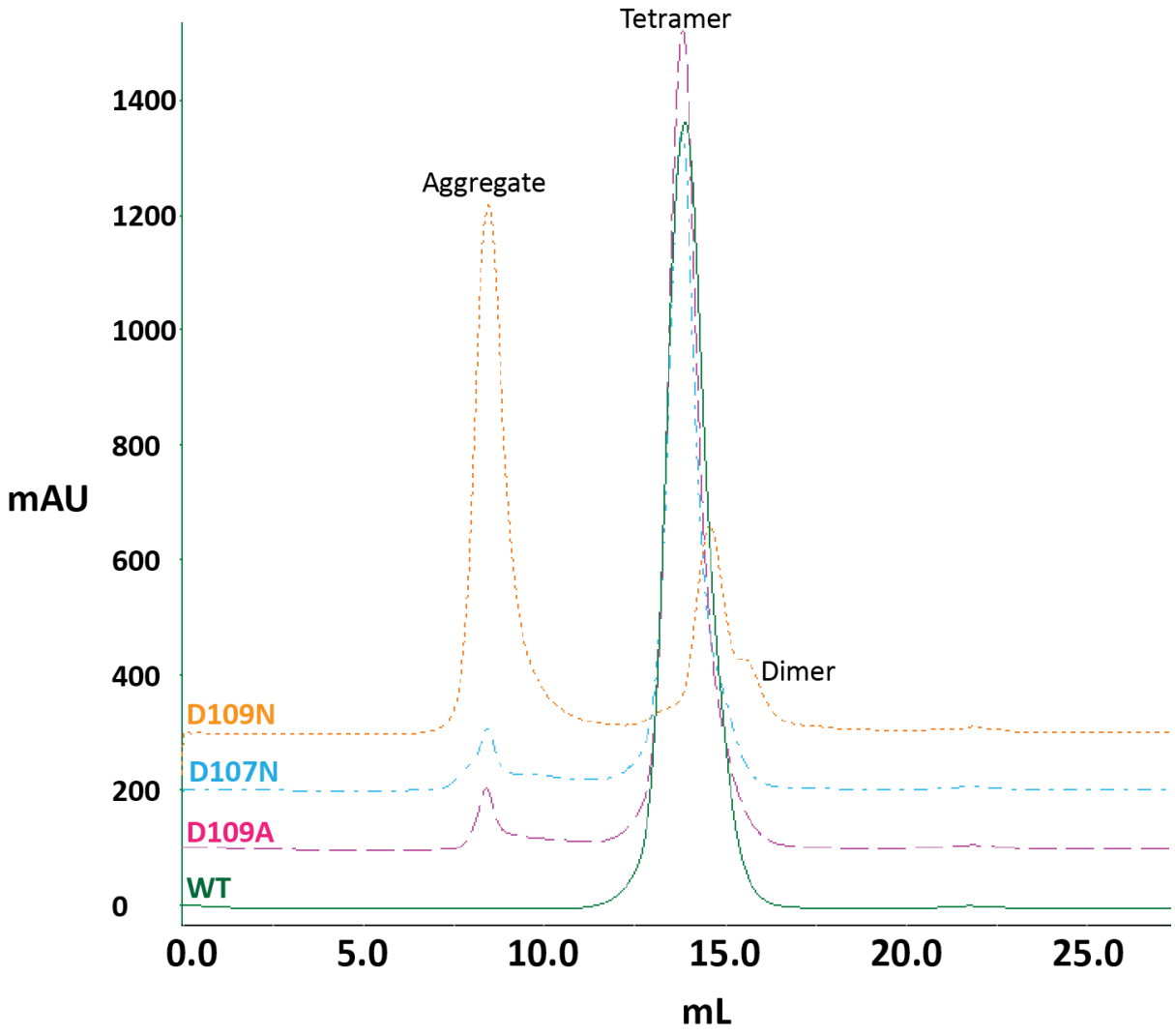


Figure 31. Size exclusion results. Each trace has been artificially shifted by 100 mAu for clarity.

In addition to active site mutants, protomer interface mutants were also explored. They were initially predicted to have perturbed pK_a values due to Theoretical Microscopic Titration

Curves (THEMATICS) and Partial Order Optimum Likelihood (POOL) analysis [33]. THEMATICS uses the sequence of a protein in a crystal structure to search for ionizable residues, along with the probability that a residue participates in binding or catalysis from POOL to determine hypothetical pK_as. The program accurately predicted the perturbed Lys122 in the active site, however it also suggested Asp107 and Asp109 also have perturbed pK_a values. As a means to explore this possibility, Asp107 was mutated to Asn, and Asp109 to Asn and Ala. All three mutants were analyzed by gel filtration, which indicated the presence of significant aggregation (Figure 31). All mutants and WT were produced, then incubated at 4 °C for 24 hours before gel filtration analysis. The Asp109Asn variant showed multiple peaks in the chromatogram, indicating that it exists as aggregate, as well as a mixture of other oligomeric states. The molecular weight of the species in the first peak was estimated to be 74.6 kDa, between a dimer and trimer, while the second peak appears to be a dimer. Attempts were made to determine the structure of the Asp109Asn variant, however the crystals were poor and did not diffract well enough. Characterization of Asp109Asn by DSC shows that it is much less stable than WT and Glu84Gln (Figure 31).

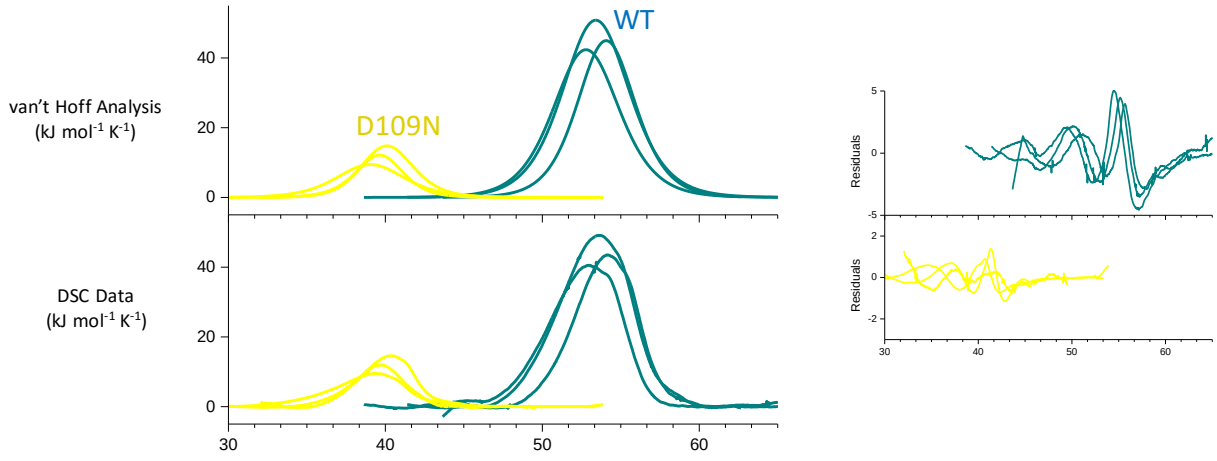


Figure 32. DSC for WT and D109N. The raw data is compared to the van't Hoft analysis.

Table 17. Results and analysis of DSC for SbAD WT and protomer mutant, where ΔH_{vH} is the van't Hoff analysis.

SbAD	T_m (°C)	ΔH (kJ mol^{-1})	ΔS (kJ mol^{-1} K^{-1})	ΔH_{vH} (kJ mol^{-1})
WT	53.4 ± 0.6	241.2 ± 37.0	0.74 ± 0.1	654.3 ± 55.1
D109N	39.6 ± 0.5	54.5 ± 10.2	0.17 ± 0.0	709.1 ± 137.8

Again, the van't Hoff analysis is larger, within error, of the calculated of ΔH as determined by experiment, suggesting a simplified two-state model of either the protein is folded, or not. The steady state kinetics and selected binding constants shed light on to the instability. Table 18 compares the steady state kinetics of the protomer mutants with cinnamaldehyde (**1**) and saturating pyruvate (**2**) compared to WT, as well as the breakdown of cinnamylidenepyruvate (**3**). Asp107Asn surprisingly had a higher turnover number than WT. It is not entirely clear why this is, as modeling in Asn at the 107 position does not appear to disturb the hydrogen bonding distances. Asp109Ala, and more to Asp109Asn, have slower k_{cat} and bigger K_M in the synthetic direction, possibly due to a slight possible conformational shift.

Table 18. The anabolic and catabolic steady state kinetics of protomer interface mutants.

SbAD	$k_{cat,cinnamaldehyde}$ (sec-1)	$K_{M,cinnamaldehyde}$ (μM)	k_{cat}/K_M (M^{-1} sec^{-1})	$k_{cat,CMP}$ (sec-1)	$K_{M,CMP}$ (μM)	k_{cat}/K_M (M^{-1} sec^{-1})
WT	23.00 ± 0.38	38.43 ± 2.03	5.98 ± 0.33 x 10^5	1.14 ± 0.05	15.42 ± 1.95	7.39 ± 0.99 x 10^4
D107N	47.89 ± 0.81	27.62 ± 1.67	1.73 ± 0.11 x 10^6	1.82 ± 0.03	11.44 ± 0.66	1.59 ± 0.10 x 10^5
D109A	5.86 ± 0.12	56.15 ± 3.29	1.04 ± 0.06 x 10^5	0.64 ± 0.03	19.70 ± 2.54	3.25 ± 0.45 x 10^4
D109N	0.151 ± 0.006	11.09 ± 2.19	1.36 ± 0.27 x 10^4	0.33 ± 0.02	11.98 ± 1.68	2.75 ± 0.04 x 10^4

Asp109Ala behaves much like WT, the truncated residue does not greatly affect its steady state values. Overall, Table 18 indicates that the protomer mutant Asp109Asn is an inactive protein,

such that the concentration of active enzyme is much less than the concentration of the total enzyme since the K_M values are equivalent to WT, and k_{cat} is slower than WT.

4.4 Conclusions

The proposed mechanism of Figure 20 has been confirmed throughout this chapter as a likely method of catalysis. Pyruvate binds to the enzyme first, for a sequential, ordered reaction. An imine with pyruvate and the active site lysine, then attacks the aldehyde carbon from the nucleophilic enamine. The Schiff base with product bound is proposed to be observed at 430nm, however the crystallographic data from Tyr252Phe suggests the break in conjugation with the ring system would result in lack in observable absorbance at that longer wavelength. It is possible torsion aids breakdown in the retro-aldol process, and was slow enough to occur from the mutant that it was able to be trapped. Hydrolysis of product from the enzyme is likely to be rate limiting based on the burst kinetics and single turnover work. The protomer interface mutants show structural instability and do not point to cooperativity in the enzyme. SbAD demonstrates aldolase-dehydratase ability with the ADC overall fold, and illustrates the subtle differences between true ADCs and MppR by catalyzing different reactions, but still retaining a Schiff base mechanism.

4.5 References

[1] (2016) Stopped-Flow, TGK Scientific, http://www.hi-techsci.com/techniques/stoppedflow/.

[2] Cook, P. F., Cleland, W. W. (2007) *Enzyme Kinetics and Mechanism*, Garland Science Publishing, New York, NY.

[3] Cornish-Bowden, A. (2012) *Fundamentals of Enzyme Kinetics*, Fourth ed., Wiley-Blackwell, Weinheim, Germany.

[4] Silverman, R. B. (2002) *The Organic Chemistry of Enzyme-Catalyzed Reactions*, Revised ed., Elsevier, San Diego, California.

[5] Fisher, H. F. (2005) Transient-state kinetic approach to mechanisms of enzymatic catalysis, *Acc Chem Res 38*, 157-166.

[6] Gill, P., Moghadam, T. T., and Ranjbar, B. (2010) Differential scanning calorimetry techniques: applications in biology and nanoscience, *J Biomol Tech 21*, 167-193.

[7] Hiromi, K. (1979) *Kinetics of Fast Enzyme Reactions: Theory and Practice*, Kondansha Ltd., Tokyo, Japan.

[8] Cornish-Bowden, A. (2015) One hundred years of Michaelis–Menten kinetics, *Perspectives in Science 4*, 3-9.

[9] Johnson, K. A. (1992) *Transient-State Kinetic Analysis of Enzyme Reaction Pathways*, Thrid ed., Academic Press, Inc., San Diego, CA.

[10] Pollard, T. D., and De La Cruz, E. M. (2013) Take advantage of time in your experiments: a guide to simple, informative kinetics assays, *Mol Biol Cell 24*, 1103-1110.

[11] Fisher, H. F. (2016) Approaches to the solution of coupled multiexponential transient-state rate kinetic equations: A critical review, *Arch Biochem Biophys 603*, 54-63.

[12] Johnson, K. A. (2009) Fitting enzyme kinetic data with KinTek Global Kinetic Explorer, *Methods in enzymology 467*, 601-626.

[13] Johnson, K. A., Simpson, Z. B., and Blom, T. (2009) FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data, *Analytical biochemistry 387*, 30-41.

[14] Johnson, K. A., Simpson, Z. B., and Blom, T. (2009) Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data, *Analytical biochemistry 387*, 20-29.

[15] Strelow, J., Dewe, W., Iversen, P. W., Brooks, H. B., Radding, J. A., McGee, J., and Weidner, J. (2004) Mechanism of Action assays for Enzymes, In *Assay Guidance Manual* (Sittampalam, G. S., Gal-Edd, N., Arkin, M., Auld, D., Austin, C., Bejcek, B., Glicksman, M., Inglese, J., Lemmon, V., Li, Z., McGee, J., McManus, O., Minor, L., Napper, A., Riss, T., Trask, O. J., and Weidner, J., Eds.), Bethesda (MD).

[16] (2015) Key Concepts: Enzyme Inhibition, GraphPad Prism, http://www.graphpad.com/guides/prism/6/curve-fitting/index.htm?reg_key_concepts_enzyme_inhibition.htm.

[17] Copeland, R. A. (2005) *Evaluation of Enzyme Inhibitors in Drug Discovery*, Wiley.

[18] Anslyn, E. V., Dougherty, D. A. (2006) *Modern Physical Organic Chemistry*, University Science Books, Sausalito, California.

[19] Cleland, W. W. (2006) *Enzyme Mechanisms from Isotope Effects*, Taylor & Francis Group, LLC, Boca Raton, Florida.

[20] Kohen, A., Limbach, H-H. (2006) *Isotope Effects in Chemistry and Biology*, Taylor & Francis Group, LLC, Boca Raton, Florida.

[21] Palfey, B. A., and Fagan, R. L. (2006) Analysis of the kinetic isotope effects on initial rates in transient kinetics, *Biochemistry 45*, 13631-13640.

[22] Fitzpatrick, P. F. (2006) *Isotope Effects from Partitioning of Intermediates in Enzyme-Catalyzed Hydroxylation Reactions*, Taylor & Francis Group, LLC, Boca Raton, Florida.

[23] Northrop, D. B. (1975) Steady-state analysis of kinetic isotope effects in enzymic reactions, *Biochemistry 14*, 2644-2651.

[24] Fisher, H. F. (2013) The application of transient-state kinetic isotope effects to the resolution of mechanisms of enzyme-catalyzed reactions, *Molecules 18*, 8230-8242.

[25] Toscano, M. D., Woycechowsky, K. J., and Hilvert, D. (2007) Minimalist active-site redesign: teaching old enzymes new tricks, *Angew Chem Int Ed Engl 46*, 3212-3236.

[26] Huynh, K., and Partch, C. L. (2015) Analysis of protein stability and ligand interactions by thermal shift assay, *Curr Protoc Protein Sci 79*, 28 29 21-14.

[27] Bruylants, G., Wouters, J., and Michaux, C. (2005) Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design, *Curr Med Chem 12*, 2011-2020.

[28] Spink, C. H. (2008) Differential Scanning Calorimetry, *Methods in Cell Biology 84*, 115-141.

[29] Marky, L. A., and Breslauer, K. J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves, *Biopolymers 26*, 1601-1620.

[30] Mueller, L. S., Hoppe, R. W., Ochsenwald, J. M., Berndt, R. T., Severin, G. B., Schwabacher, A. W., and Silvaggi, N. R. (2015) Sbi00515, a Protein of Unknown Function from Streptomyces bingchenggensis, Highlights the Functional Versatility of the Acetoacetate Decarboxylase Scaffold, *Biochemistry 54*, 3978-3988.

[31] Gadda, G., and Fitzpatrick, P. F. (2000) Mechanism of nitroalkane oxidase: 2. pH and kinetic isotope effects, *Biochemistry 39*, 1406-1410.

[32] Smith, B. J., Lawrence, M. C., and Barbosa, J. A. (1999) Substrate-Assisted Catalysis in Sialic Acid Aldolase, *J Org Chem 64*, 945-949.

[33] Somarowthu, S., and Ondrechen, M. J. (2012) POOL server: machine learning application for functional site prediction in proteins, *Bioinformatics 28*, 2078-2079.

CHAPTER V: STRUCTURAL AND FUNCTIONAL ANALYSIS OF SWIT_4259

5.1 Introduction

Some of the most dangerous and wide-spread pollutants on earth are aromatic hydrocarbons like dioxins, polychlorinated biphenyls (PCBs), herbicides, and pesticides. These compounds are recalcitrant in the environment due to their low solubility in water, tendency to adsorb to solids, and dissimilarity to naturally occurring compounds (e.g. halogenation). Thus, there is significant interest in natural or engineered bioremediation pathways capable of metabolizing these contaminants to harmless organic compounds [1]. Some microorganisms, particularly bacteria, have evolved the ability to consume aromatic hydrocarbons as carbon and energy sources [2-4]. *Sphingomonas wittichii* RW1 is a Gram-negative bacterium that was first isolated from contaminated sediments in the Elbe River in Germany, and it is known for its ability to degrade a wide variety of aromatic substrates, including dioxins and PCBs [5, 6]. The *S. wittichii* RW1 genome thus contains a large number of aromatic hydrocarbon catabolism gene clusters, only a handful of which have been characterized.

Relative to other species, the *S. wittichii* genome is enriched for TonB-dependent receptors (TDRs; COG1629), dehydrogenases (COG1028), and phenylpropionate dioxygenases (COG4638) [7]. In fact, the *S. wittichii* RW1 genome contains the most TonB-dependent receptors of any genome sequenced to date [8]. This is interesting, because the TonB system is normally associated with uptake of iron-bound siderophores, but the majority of TDR genes in *S. wittichii* RW1 are not in operons associated with iron transport. TonB-dependent receptors have been implicated in the transport of other compounds, including sugars [9], alginate [10], and aromatic

hydrocarbons (AH) [11]. This strongly suggests that many of the TDR-containing gene clusters in *S. wittichii* may actually be cryptic AH-degrading operons.

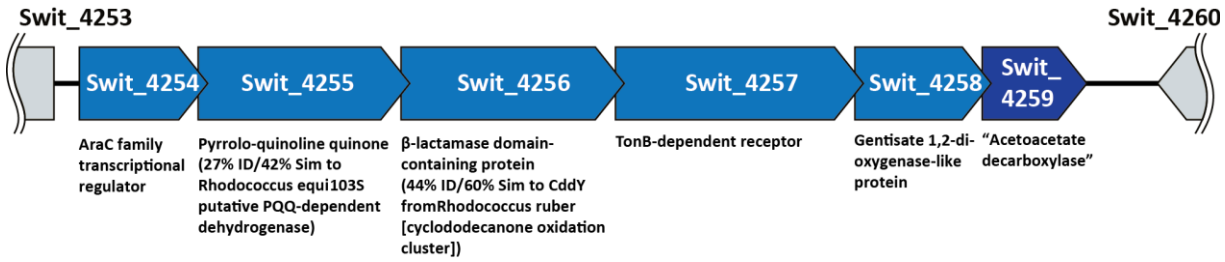


Figure 33. Gene context of *S. wittichii* Swit_4259.

One of these TDR-containing, putative AH-degrading gene clusters in *S. wittichii* (Figure 33) is particularly interesting for its inclusion of an acetoacetate decarboxylase-like enzyme, Swit_4259. Swit_4259 (SwADC) has 18% sequence identity to the prototypical *Clostridium acetobutylicum* acetoacetate decarboxylase (AAB53232), placing it in the acetoacetate decarboxylase-like superfamily (ADCSF). The ADCSF is a poorly-characterized group of enzymes with significant diversity in both substrate and reaction specificities [12]. The characterized acetoacetate decarboxylases (*e.g.* the *C. acetobutylicum* enzyme) comprise the majority of ADCSF sequences and are designated Family I. The reactive lysine residue that forms a Schiff base intermediate is conserved in all but one ADCSF family (Family II). SwADC belongs to ADCSF Family V, whose members, based on gene context, are primarily involved in secondary metabolism.. MppR from *Streptomyces hygroscopicus* [13] was shown to catalyze both aldol condensation followed by dehydration of pyruvate and imidazole-4-carboxaldehyde as well as cyclization of 4-hydroxy-2-keto arginine to produce 2-keto-enduracididine [12], as seen in chapter II, though it is not known if these activities are physiologically relevant. Herein we describe the tertiary structure of SwADC as well as a preliminary characterization of its enzymatic activity.

5.2 Materials and Methods

Cloning, expression, and purification of SwADC.

The gene at locus Swit_4259 encoding the protein identified by GenBank accession code YP_001264737 [7] was optimized for expression in *E. coli* and synthesized by GenScript (Piscataway, NJ). The SwADC gene was amplified using oligonuculeotide primers SwADC-pSUf (5'-GGTCTCAAGGTATGGCTGGTGCTACGCAT-3'), which introduces a Bsa I site, and SwADC-pSUr (5'-GCTCTAGATCATCACATACGCAGGGATTCCA-3'), which introduces a stop codon and Xba I site. Purified PCR products were digested with Bsa I and Xba I and ligated into the pE-SUMO$_{kan}$ expression vector (LifeSensors Inc, Malvern, PA) to generate pSUMO-SwADC. This plasmid expresses SwADC with His$_6$-tagged small ubiquitin-like modifier (SUMO) fused to the N-terminus [14]. The sequence of pSUMO-SwADC was confirmed by sequencing (GenWiz Inc). BL21 Star (DE3) *E. coli* cells (Invitrogen) were transformed with pSUMO-SwADC and used to inoculate 20 ml cultures of Luria-Bertani broth supplemented with 50 μg/mL kanamycin. These starter cultures were grown overnight at 37 °C with agitation at 250 rpm. Large-scale cultures (1 L) were inoculated from the starter cultures and grown at 37 °C/225 rpm until they reached an OD$_{600\ nm}$ of about 1.0. The temperature of the incubator was reduced to 25 °C and expression of SwADC was induced by addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 400 μM. Cultures were grown overnight at 25 °C. Cells were harvested from the expression culture by centrifugation at 7400 × *g* for 30 min and resuspended in Buffer A (25 mM Tris, pH 8.0, 300 mM NaCl, and 10 mM imidazole) at a rate of 5 mL of buffer per gram of cell pellet. Lysozyme was added to a final concentration of 1 mg/mL and the mixture was stored overnight at −20 °C. The cell suspension was thawed at RT and allowed to stand for 2 hours, after which

deoxyribonuclease I (DNase I, Worthington Biochemical Inc.) was added at a rate of 0.5 mg per 100 ml of lysate. The lysate was clarified by centrifugation at 39000 × *g* for 45 min.

Immobilized metal affinity chromatography (IMAC) purification of SwADC was performed in two steps. The clarified lysate was applied to a 5 mL HisTrap column (GE Life Sciences, USA) and eluted using a 4-step gradient from 0 to 5, 15, 50, and 100% Buffer B (25 mM Tris, pH 8.0, 300 mM NaCl, and 250 mM imidazole). SwADC eluted primarily at 50% Buffer B (about 125 mM imidazole), but a significant amount eluted at the 100% Buffer B step (about 250 mM imidazole). The fractions from both elution peaks containing SwADC, as judged by Coomassie-stained SDS-PAGE gels, were pooled. SUMO protease (LifeSensors Inc) was added to a final concentration of about 5 µM [15]. The pooled eluate was dialyzed overnight at 4 °C in Buffer C (25 mM Tris, pH 8.0, 150 mM NaCl) to remove imidazole. In the second step of the purification, the same HisTrap column was used to remove the cleaved His_6-SUMO fusion partner and the His_6-tagged SUMO protease. The resulting protein was > 95% pure as judged by SDS-PAGE. A 10 mL HiTrapDesalting column (GE Life Sciences, USA) was used to exchange the buffer to 25 mM Bis-Tris pH 7.1 (titrated with saturated iminodiacetic acid). The concentration of SwADC was estimated by the absorbance at 280 nm (ε_{280} = 46,660 $M^{-1}cm^{-1}$). The final yield of purified SwADC was approximately 15 mg per L of culture.

Size exclusion chromatography.

The solution molecular weight of SwADC was estimated by gel filtration chromatography using a BioBasic SEC-300 (250mm × 4.6 mm) column (Thermo Scientific, USA) on an Agilent 1220 Compact LC. The mobile phase consisted of 25 mM Tris, pH 8.0, 300 mM NaCl at a flow rate of

0.5 ml/min with UV-Vis detection at 280 nm. The column was calibrated using standard proteins including equine heart cytochrome C (12.4 kDa), bovine erythrocyte carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), yeast alcohol dehydrogenase (150 kDa), and sweet potato β–amylase (200 kDa). To define the standard curve, each protein was dissolved at 10 mg/mL in the running buffer and the separation was performed on 250 µL of the standard mixture at 25 °C. A sample of SwADC (6 mg/mL) was run immediately afterward using the same column and identical conditions.

Crystallization, data collection and structure determination.

Preliminary crystallization conditions were identified by sparse matrix screening using the Index HT, PEG/Ion HT, and PEGRx HT screens (Hampton Research, USA) at 21 °C. Crystals appeared after 1 to 2 days in a variety of conditions. The initial crystals ranged in shape from individual rods to what appeared to be bundles of parallel needles. The single, rod-shaped crystals diffracted very poorly (d_{min} ~5 Å), but the "bundles of rods" diffracted to a maximum resolution of ~2.0 Å. Diffraction-quality crystals were grown by the hanging-drop vapor diffusion method. Drops were formed by mixing 1-2 µl of SwADC (37 mg/ml) in 20 mM phosphate pH 7.0 with 1 µl of well solution containing 15 % PEG 3,350 and 0.1-0.2M magnesium formate. Crystals were flash-cooled by plunging them in liquid nitrogen, without a cryoprotectant. Diffraction data were collected from an SwADC crystal at the Life Sciences-Collaborative Access Team, beamline 21-ID-G equipped with a MAR 300 CCD detector, 50 x 50 µm beam at a wavelength of 0.97857 Å. A total of 440 frames were collected from ω = 150 ° to 370 ° with an oscillation range of 0.5 °. Diffraction data were integrated and scaled using MOSFLM [16, 17] and SCALA [18] as implemented in

the CCP4 suite version 6.3 [19]. Initial phases were obtained by molecular replacement in PHASER [20] using the structure of the related ADCSF enzyme MppR (PDB ID 4JM3) as the starting model. Iterative rounds of model building and refinement were performed using PHENIX [21, 22] and COOT [23, 24]. Structures were validated using the tools available in the PHENIX package [25, 26]. The data processing and model refinement statistics are summarized in Table 19.

Acetoacetate decarboxylase activity assay.

SwADC was assayed for acetoacetate decarboxylase activity by monitoring the disappearance of the enolate form of acetoacetate spectrophotometrically at 270 nm ($\Delta\varepsilon$ = 24.6 M^{-1} cm^{-1}) as described [27]. Reactions were run at 25 °C in 50 mM phosphate buffer, pH 6.5, 2 mM EDTA, on a Thermo Scientific Evolution 300 spectrophotometer. The reaction mix containing 20 mM lithium acetoacetate in the assay buffer was used as the blank. The reaction was initiated by the addition of 10 µM SwADC in the same buffer.

Determination of binding constants.

There is one tryptophan (Trp24) in the active-site pocket close to Lys122 (about 7 Å between Nε of Lys122 and the indole ring of Trp24) which is sensitive to the binding of small molecules in the active site. Potential substrates can be titrated to determine the dissociation constant (K_D) of each compound. The affinity of SwADC for a number of α-keto-acids such as pyruvate and α–keto glutarate was tested. Alanine was used as a control compound since it cannot react with the active site lysine. The fluorescence signal of Trp excited at 280nm was monitored using TgK Scientific stopped flow with 20 µM SwADC or Hitachi F-4500 fluorimeter with 5 µM SwADC and

varied substrate concentrations ranging from 1 mM to 1 M, all in in 100 mM Bis-Tris pH 6.5. The change in fluorescence signal was also tested using 2-oxo-adipic acid only in 5 mM concentration owing to the expense of the compound.

Production of 2-hydroxymuconic acid (2HM) (Scheme 9, **1**)

The synthesis of 2-hydroxymuconic acid was based on that of Metanis *et al* [28]. A solution of t-BuOK was prepared by dissolving potassium metal (1.92 g) in t-BuOH (20 mL) and refluxing until homogenous. Then, 8 mL of the t-BuOK solution (excess) was combined with 2 mL of ether and stirred at 0° C for 15 minutes. A solution of diethyl oxalate was made by mixing 1.35 mL of diethyl oxalate and 1 mL of ether. This solution was cooled to 0 °C and added to the t-BuOK solution via syringe. The reaction mixture was allowed to stir for another 15 minutes. A solution of ethyl crotonate was made by mixing 1.2 mL of ethyl crotonate with 1 mL ether. This solution was cooled to 0° C and added via syringe. The mixture was allowed to stir for 30 minutes and then placed stored at 4° C overnight to precipitate the potassium salt. Potassium salt was filtered and rinsed with hexanes and dried under vacuum to yield 1.907 g of the potassium salt. The potassium salt was dissolved in 25 mL DI H_2O and 1 mL of glacial acetic acid was added. Resulting precipitate was collected by filtration and dried under vacuum to yield 1.3209 g of the diester. The diester was dissolved in a NaOH solution (2 M, 20 mL) and left at room temperature overnight. The solution was then acidified to pH 2 with HCl (12 M) and then placed on an ice bath for 1 hour. Precipitate was collected by filtration and dried under vacuum to yield 339.5 mg of diacid.

Preparation of 3-(2-furyl)acrolenepyruvate (Chapter III, Scheme 4, **22**)

The compound was synthesized as described in Chapter III.

Steady-state kinetics and product identification.

Based on our hypothesis that SwADC is an aldolase-dehydratase acting in an uncharacterized hydrocarbon degradation pathway, the activity of SwADC was tested with potential substrates including 2-oxo-3-hexenedioate (OHD), and 3-(2-furyl)acrolein and pyruvate. The assays were performed with 0.5 – 1 μM SwADC in 50 mM BIS-Tris pH 6.5, and various substrate concentrations. In aldol condensation reactions, sodium pyruvate was saturating (50 mM)

The enzyme 4-oxalocrotonate tautomerase (4-OT) was used to produce OHD from 2-hydroxy muconic acid (2HM). Reactions contained 50 mM sodium phosphate buffer pH 6.5, 1 μM 4-OT, and 100 μM 2HM in a final volume of 0.5 ml. The production of 2-oxo-3-hexenedioate was monitored at 236 nm (ε = 6580 M^{-1} cm^{-1}) or in UV-Vis spectra from 220 to 340 nm. Each assay was incubated without SwADC until the absorbance at 236 nm reached a maximum. At this point, SwADC was added to the mixture and the loss of absorbance at 236 nm was monitored. The assay was performed with 100 μM 2HM and different concentrations of SwADC (1, 2, or 4 μM). The assay also was performed using 5 μM SwADC and concentrations of 2HM ranging from 0 to 200 μM. The activities of the K122A, K122M, and Y120F variants were tested using the same protocol.

5.3 Results and Discussion

Overall structure of SwADC.

SwADC is a 28.8 kDa protein that, according to size exclusion chromatography, exists as a homotetramer in solution. The protein crystallized from 15% PEG 3,350 and 0.1 – 0.2M magnesium formate, forming long, hexagonal rods on the order of 0.5 x 0.05 x 0.05 mm. These

crystals diffracted X-rays to ~2.0 Å-resolution. Numerous, various proposed substrates have been soaked into SwADC crystals, as well as co-crystallized, however there has been no defining electron density detected from either method. Diffraction data were collected at the Life Science Collaborative Access Team (LS-CAT) beamline 21-ID-G at the Advanced Photon Source, Argonne National Laboratory. The data were processed in MOSFLM [29, 30] and scaled in SCALA [31]. The structure of SwADC was determined by molecular replacement using the structure of a SbAD protomer (PDB ID 4ZBO) as the search model (see Table 19 for data collection and refinement statistics). SwADC crystals belonged to space group C2 with 4 molecules in the asymmetric unit. As shown in Figure 34, the overall fold of the SwADC protomer is nearly identical to those of the prototypical Family I ADCs, as well as the recently described Family V enzymes like SbAD from *Streptomyces bingchenggensis* and MppR from *S. hygroscopicus*, which are not decarboxylases [12, 32]. This fold, dubbed the β-cone [33] or double barrel fold [34], consists of a large, antiparallel, 13-stranded β-sheet that forms the main barrel and folds back on itself to form the side barrel. Unlike prototypical ADCs that are active as

Table 19. Crystallographic Data Collection and Model Refinement Statistics.

	SwADC
resolution (Å) (last shell)[a]	38.39 – 1.95 (2.02 – 1.95)
wavelength (Å)	0.9786
no. of reflections	
observed	334092 (27317)
unique	73627 (7275)
completeness (%)[a]	100 (100)
R_{merge} (%)[a,b]	0.07611 (0.1883)
multiplicity	4.5 (3.8)
⟨I/σ(I)⟩[a]	13.92 (7.06)
Model Refinement	
no. of reflections in the work set	73600 (7271)
no. of reflections in the test set	3688 (349)
R_{cryst} (R_{free})	0.1719 (0.2068)
no. of residues	1025
no. of TLS groups	27
average B factor (Å²)[c]	
protein atoms	35.94
solvent	35.84
root-mean-square deviation	
bond lengths (Å)	0.012
bond angles (deg)	1.28
Ramachandran statistics (favored/allowed/outliers)	98.5/1.5/0

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

homododecamers, SwADC has a tetrameric quaternary structure similar to SbADC and MppR (Figure 34).

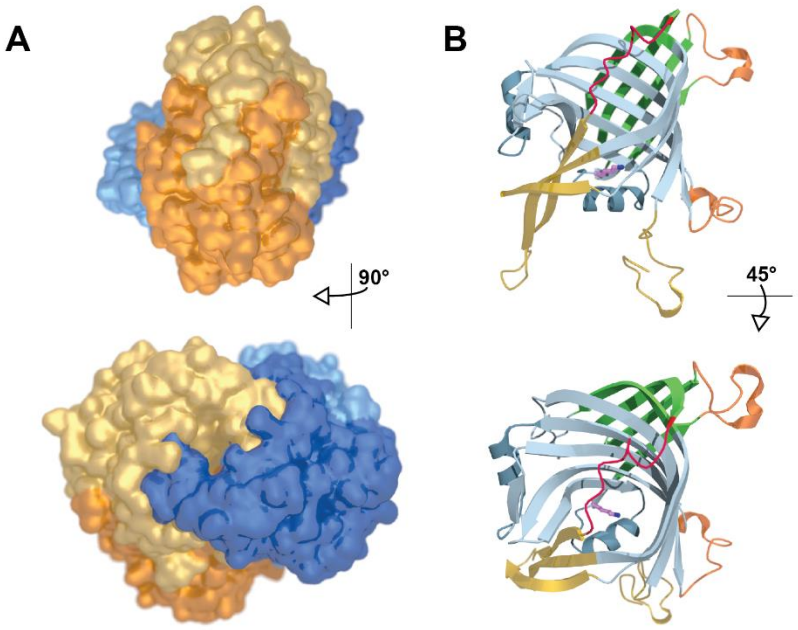


Figure 34. (**A**) Rendering of the solvent-accessible surface of SwADC showing the extensive tetramer interface formed by the long, interdigitating loops (top). This interface is repeated on the back side with the blue and pale blue protomers. The pair of dimers is held together (bottom) by short loops that appear to "clip" the two dimers together to form the tetramer. (**B**) The ribbon diagram showing a single protomer is oriented like the yellow protomer in panel A. The main barrel of the β-cone fold is colored pale blue, with the "side barrel" colored green. The dimer interface loops are colored yellow, and the loops forming the interface between the dimers are colored orange. The section of the polypeptide that is colored red is the "active site loop" that may control access to the active site. The catalytic Lys122 is shown as a ball-and-stick representation with purple carbon atoms.

Although the tetramer in Figure 34 does not contain a ligand, only conserved waters, chains A and C, represented by the pale yellow and pale blue protomers, had different active site conformations than chains B and D, shown as orange and dark blue. Interestingly, the tetramer for SwADC is a dimer of dimers. The two dimer pairs from each dimer are different, as opposed to the two dimers differing from one another. Figure 35 illustrates these different active site conformations, with the corresponding schematic representations and the relevant distances provided in Figure 36.

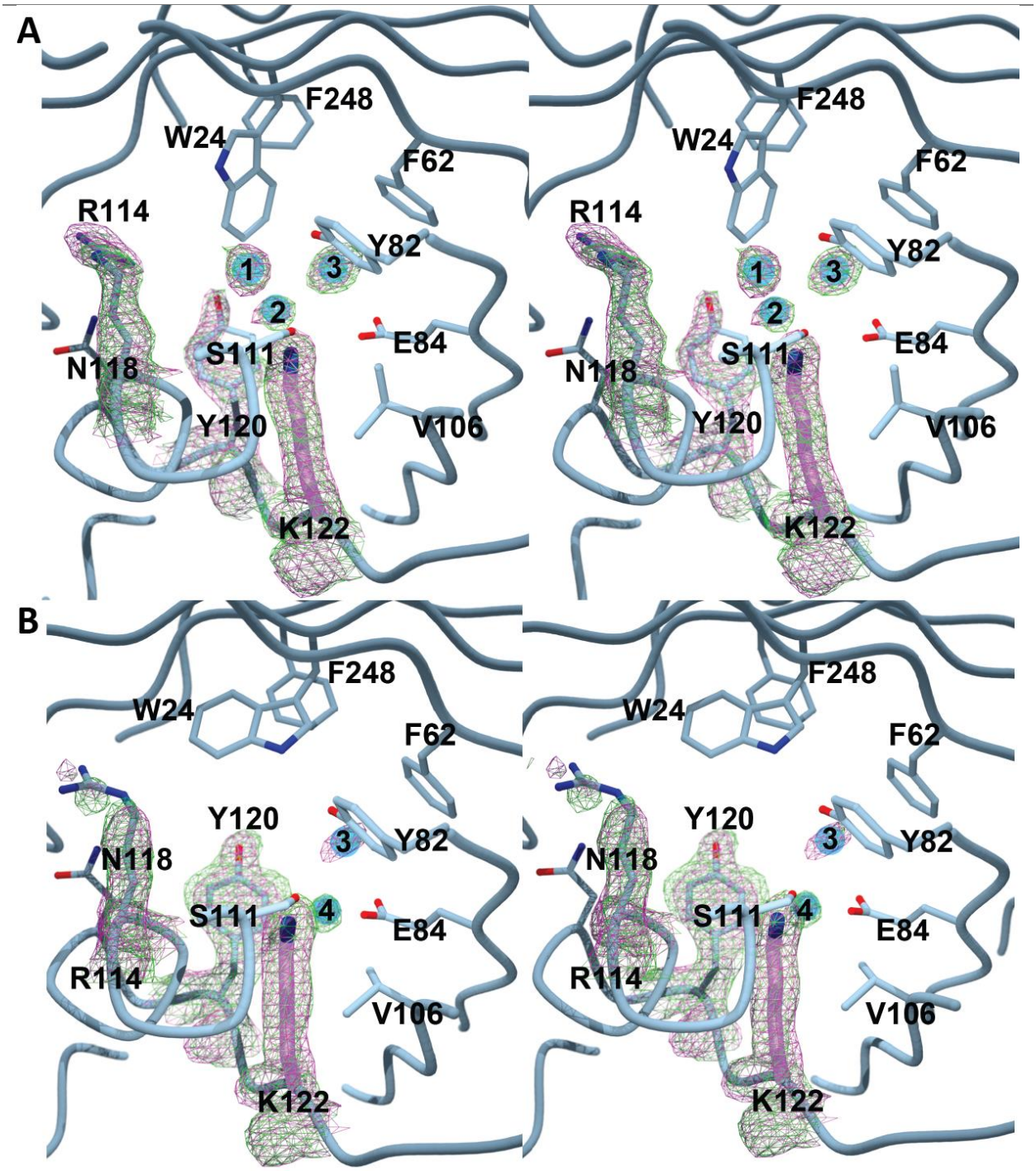


Figure 35. The two different active sites of SwADC, represented by (**A**) and (**B**), respectively. (**A**) Stereoview of the active site in the SwADC structure showing conserved waters (refer to Figure 36 for relevant hydrogen bonding interactions and their distances). (**B**) One putative new water molecule has been recruited by Glu84. In both panels, the experimental $2|F_o|-|F_c|$ electron density map (magenta) and the simulated annealing composite omit $2|F_o| - |F_c|$ electron density map (green), both contoured at 1.0σ, are drawn with a 2.0 Å radius around each atom of Lys122 and select water molecules. This figure and subsequent figures showing crystallographic structures were rendered with the POVSCRIPT+[35] modification of MOLSCRIPT[36] and POVRAY.

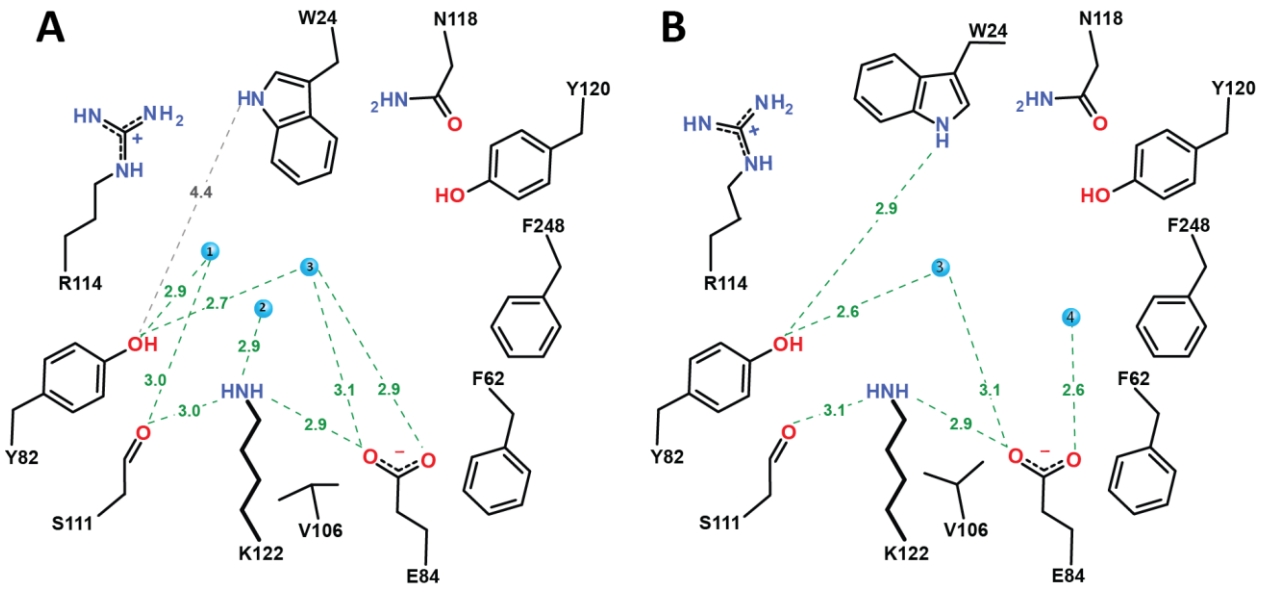


Figure 36. Schematic views of (**A**) one of the active sites of SwADC and (**B**) the other conformation showing potential hydrogen bonding interactions and their distances (in Angstroms) in green.

Chains A and C (Figures 35 A and 36 A, conformation A) has three ordered water molecules bound. In chains B and D (Figures 35 B and 36 B, conformation B) water molecule 3 from conformation A has been retained, water molecules 1 and 2 have been lost, and a new water molecule (labelled "4" in Figures 35 B and 36 B) is bound to Glu84. However, the biggest difference between the two active site conformations is that in conformation A, Arg114 exists in a single conformation as evidenced by the well-defined electron density, and Tyr120 samples multiple conformations such that much of the associated electron density is missing. In conformation B, the opposite is true. Tyr120 takes a single conformation and the guanidinium group of Arg114 is disordered. These two residues are not connected by hydrogen bonding or van der Waals interactions of their side chains, nor do they communicate through a water-mediated hydrogen bond network. In addition to this, the indole of Trp24 flips about 90° between conformations A and B, and Val106 rotates about 180°. Perhaps all these changes are based on a concerted change in the tetrameric structure, such that one version of the active site is better

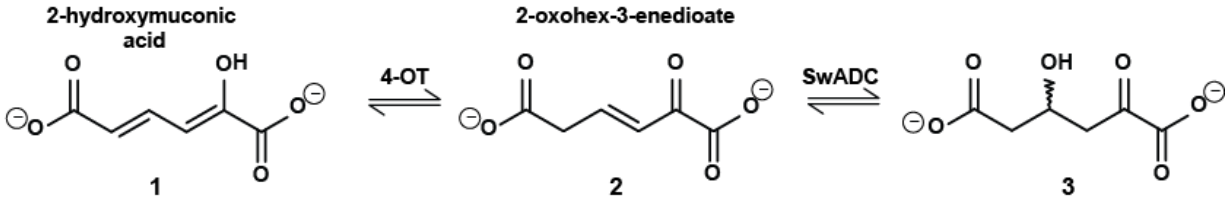
configured for catalysis than the other. However, the conformations of the active sites are lowest energy forms of the protein under crystallization conditions, and so these subtle differences may not be physiologically relevant. Thus, SwADC may be a cooperative enzyme, though our preliminary kinetics on admittedly poor substrates do not suggest that this is the case.

Functional characterization. The biochemical function of SwADC remains elusive, though there are indications that it can function as a pyruvate aldolase-dehydratase. As a Family V enzyme of the ADCSF, the binding constants of pyruvate and other α-keto acids were measured (Table 20). The K_D of pyruvate is twice as high as that for SbAD. Surprisingly, given the number of rotatable bonds, 2-oxo-adipate has a somewhat lower K_D, suggesting that the enzyme might have a preference for longer, aliphatic side chains on the α-keto acid (Table 20). Based on the space available in the active site and the array of residues lining the pocket, it seemed that a 6-8 carbon long di-carboxylic acid might be a viable substrate.

Table 20. Binding constants for α-keto acids with SwADC.

α-keto acid	**K_D (mM)**
pyruvate	10.1 ± 0.9
2-oxoglutarate	11.0 ± 1.4
2-oxoadipate	5.9 ± 0.7

Scheme 9. 2-hydroxymuconic acid (2HM) reacts with 4-oxalocrotonate tautomerase (4-OT) to produce 2-oxohex-3-enedioate (OHD). SwADC then reacts to form an aldol.



Accordingly, 2-oxohex-3-enedioate (generated by 4-oxalocrotonate tautomerase from 2-hydroxymuconic acid) was tested for activity with SwADC (Scheme 9). When 100 µM OHD was incubated with 10 µM SwADC, there was a steady decrease in the absorbance due to cleavage or

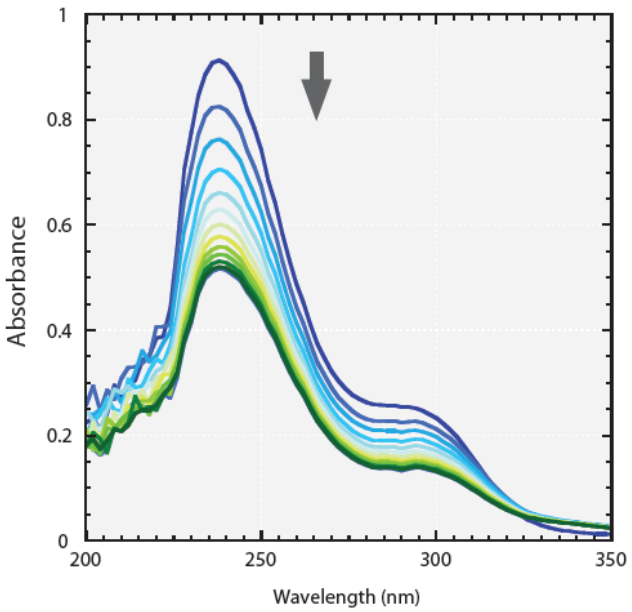


Figure 37. 100 µM 2HM was treated with 1 µM 4-OT for 20 min before 10 µM SwADC was added. Scans were recorded every 4 minutes at 25 °C. The λ_{300nm} is remaining 2HM, and λ_{236nm} is OHD.

loss of conjugation of OHD (Figure 37). The rate of decrease was dependent on both the concentration of OHD and the concentration of enzyme (data not shown). The steady state kinetics of this reaction were analysed for the wild type, as well as three active site mutants: Tyr120Phe, Tyr120Phe/Phe248Tyr, and Asn118Glu (Table 20). The mutants were selected for their possible catalytic and/or binding roles based on homology to SbAD. Tyr120 is one of the residues that is disordered in one conformation, but not the other. Perhaps Tyr120 needs to undergo conformational change in order for a reaction to occur. To test this idea, Tyr120 was mutated to a Phe, as there is a Phe in this position in SbAD. Phe248 was mutated to Tyr as it matches the residue in SbAD, which is Tyr252. Phe248Tyr could not be obtained in soluble form, however the double mutant Tyr120Phe/Phe248Tyr expressed and purified well, and the pair of mutations make this variant more like SbAD. Lastly, Asn118 was mutated Glu, again to match SbAD.

Table 20. Steady state kinetics for SwAD WT and active site mutants with OHD, monitored at 236 nm.

SwADC	$k_{cat,\ OHD}$ (sec^{-1})	$K_{M,\ OHD}$ (µM)	k_{cat}/K_M ($M^{-1}\ sec^{-1}$)
WT	0.010 ± 0.001	66.4 ± 16.3	151 ± 40
Y120F	0.042 ± 0.006	281.1 ± 67.2	149 ± 42
Y120F F248Y	0.018 ± 0.002	340.4 ± 57.5	53 ± 11
N118Q	No activity	-	-

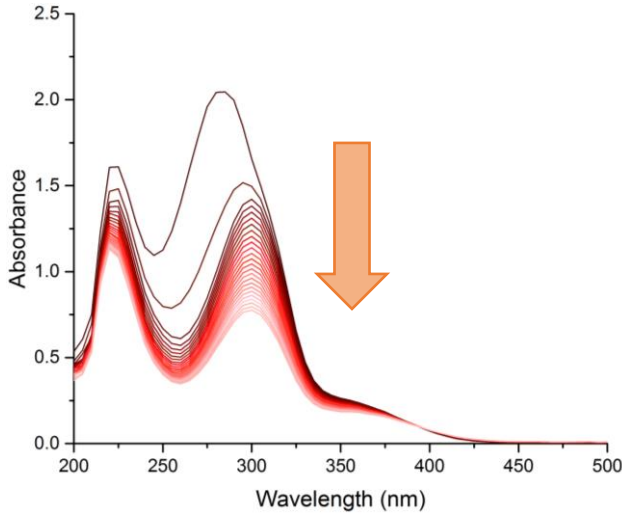
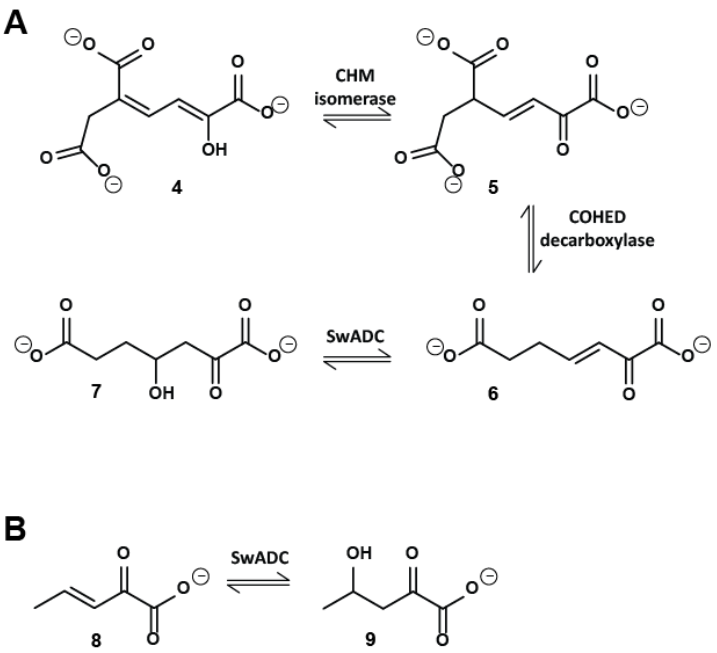


Figure 38. 10 µM Swit reacted with 2-oxo-3-heptenedioic acid for 2 minute scans at 30 °C.

In other crystal structures (data not shown), Arg114 appears to wave around the active site, while Asn118 has continuous electron density. SbAD, however, naturally has Gln at this position, and continuous electron density for Arg114 and Glu118. We proposed that the longer side chain of Glu for Asn118Glu may fix Arg114 in place, altering binding or catalysis. Based on the results in Table 20, Tyr120Phe is about 4 times faster at turnover, however the K_M is higher. These two factors compensate to attain the same pseudo-second order rate constant. Perhaps the substrate binds Tyr120Phe in a different manner resulting in a faster reaction. The double mutant had the lowest k_{cat}/K_M, however all of the SwADC variants reacted quite slowly with OHD. The K_M values suggest that OHD does not bind terribly well. A seven-carbon compound like oxo-heptene-dioate, rather than the six-carbon OHD, could conceivably fill the active site more completely and be a more efficient substrate.

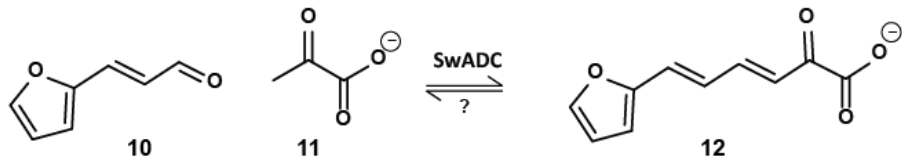
To that end, 5-carboxymethyl-2-hydroxymuconic acid (CHM) was isomerized by CHM isomerase, to yield 5-(carboxymethyl)-2-oxo-3-hexene-1,6-dioic acid (COHED). COHED was then decarboxylated to 2-oxo-3-heptenedioic acid. When 2-oxo-3-heptenedioic acid was incubated with SwADC, there was a time-dependent reduction in absorbance at all wavelengths, similar to what was observed with OHD (Figure 38). The reaction was not characterized by steady state kinetics as both CHM, CHM isomerase, and COHED decarboxylase were difficult to produce. Efforts focused instead on identifying the product of the reaction with OHD by mass spectroscopy



Scheme 10. The proposed reaction in (**A**) is observed in Figure 38. The reaction in (**B**) has been confirmed by mass spectroscopy as well.

(MS) and nuclear magnetic resonance (NMR) (Appendix II). Based upon MS data, it appears unlikely that SwADC performs retro-aldol cleavage of OHD. Instead, it appears to simply hydrate the double bond to give the aldol. Vinyl pyruvate (**8**, Scheme 10 B) was also treated with SwADC and the reaction mixture analyzed by MS. As with OHD, only the aldol product could be detected; there was no evidence of other products. There was also no evidence of the aldol forms of OHD or **8** in standards that had not been incubated with SwADC. The NMR data were inconclusive owing to the complicated spectra arising from the mixture of compounds present. No cleavage products of either substrate were detected by reverse phase HPLC.

Based on the close relationship between SwADC and SbAD, cinnamaldehyde and pyruvate were used to test for aldol condensation activity in SwADC. SwADC was not able to catalyze the aldol condensation with these substrates, likely because the aromatic group of cinnamaldehyde is too bulky and/or lipophilic for the SwADC active site. To test this, 3-



Scheme 11. Reaction tested for SwADC based on the results of SbAD.

(2-furyl)acrolein was used as the acceptor substrate (Scheme 11). Both the condensation and the retro-aldol cleavage (catabolic) reactions were characterized by steady state kinetics (Table 21). It is difficult to say if the breakdown direction stops at the aldol, or a full retro-aldol cleavage occurs. The aldol is difficult to detect via MS, and pyruvate does not appear on NMR spectra or MS data for the catabolic process yet the fact that the enzyme catalyzes the aldol condensation reaction between pyruvate and **10** to give the enone suggests that it should be able to catalyze the retro-aldol reaction as well.

Table 21. Steady state kinetics for Scheme 11. SwADC was 0.5 – 1 μM. Reactions took place at 25 °C in 50 mM BIS-Tris pH 6.5 and monitored at 390 nm.

SwADC	$k_{cat,\mathbf{10}}$ (sec^{-1})	$K_{M,\mathbf{10}}$ (μM)	k_{cat}/K_M (M^{-1} sec^{-1})	$k_{cat,\mathbf{12}}$ (sec^{-1})	$K_{M,\mathbf{12}}$ (μM)	k_{cat}/K_M (M^{-1} sec^{-1})
WT	0.084 ± 0.006	2.99 ± 0.65	2.8 ± 0.6 x 10^4	0.09 ± 0.02	42.02 ± 13.99	2.1 ± 0.9 x 10^3
N118Q	0.068 ± 0.005	3.64 ± 0.82	1.9 ± 0.4 x 10^4	No Activity		
Y120F F248Y	0.041 ± 0.003	1469.8 ± 252.2	27.9 ± 5.2	No Activity		

The wild type SwADC and the two mutants tested all functioned in the aldol condensation reaction (synthetic direction), with the Asn118Glu variant turning over at about the same rate as the WT, and the Tyr120Phe/Phe248Tyr variant only slightly slower. Likewise, the WT and Asn118Glu SwADC had the same Michaelis constant, indicating that the mutation likely did not affect substrate binding. The Y120F/F248Y double mutant has a K_M value for **10** almost 500-fold higher than that of the WT enzyme. Since it was not possible to test both single mutants, it is not clear whether this effect is due to Tyr120, Phe248, or the combination of the two. It is clear, however, that the double mutant does not bind **10** well. Interestingly, only the WT enzyme catalyzed the retro-aldol cleavage (catabolic direction) of **12**. Again, since the signal loss at 390 nm was the method of detection, it is not possible to tell if the reaction stopped at the aldol, which would eliminate the conjugation that gives rise to the absorbance signal. Interestingly,

neither of the mutants were active. Perhaps they do not hold the substrate in the correct orientation for reaction, or their K_M is greater than 10 mM.

In any case, the function of SwADC is certainly distinct from those of the classical ADCs, as well as the Family V enzymes SbAD, which is an aldolase-dehydratase [32], and MppR, which likely catalyzes the cyclization of a highly oxidized arginine derivative [12]. Comparisons of the active sites shed light on the reasons for this (Figures 39 and 40). The residues responsible for Schiff base chemistry are conserved: the active site lysine and the likely acid/base catalyst glutamate that is within hydrogen bonding distance. In SwADC numbering, Arg114, Asn118, and Val106 are the same between SwADC and MppR. However, there are significant differences, as observed in Figure 39.

Figure 39. Stereoview comparing MppR and SwADC. The green peptide chain is MppR, and the blue is SwADC. The active site lysine and likely acid/base catalyst glutamate are labelled according to SwADC, and are conserved in both proteins, along with R114 and N118. There are numerous differences, however observed throughout the rest of the model. For instance, the switching of Trp and Phe on the top portion, and the Met in MppR. Although the carboxylate-binding site appears to be conserved, other residues in the active site aside from those that contribute to forming a Schiff base intermediate, are not.

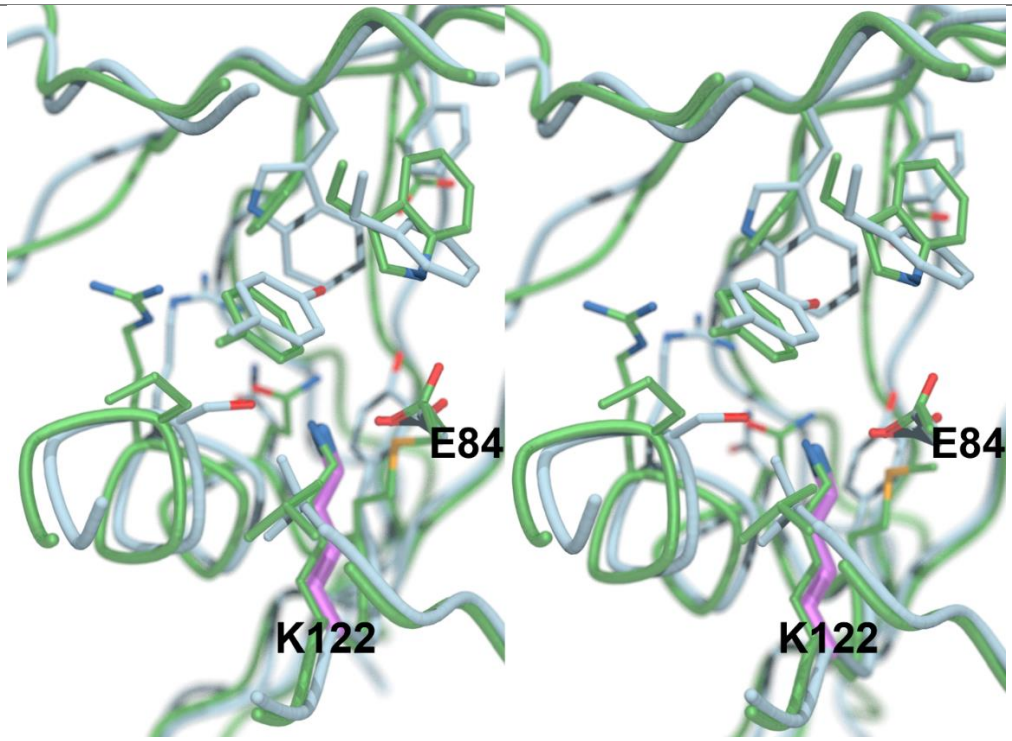
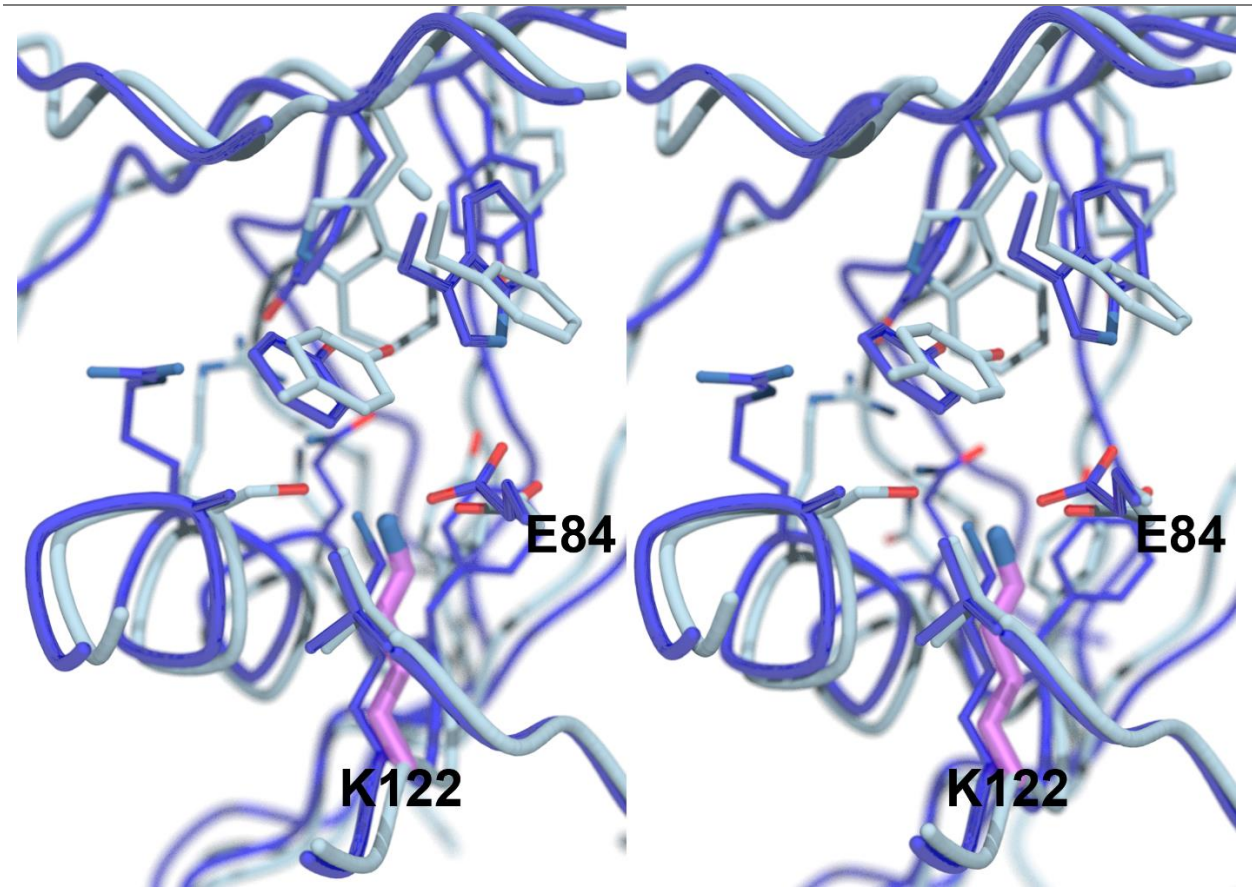


Figure 40. Stereoview comparing SbAD and SwADC. The dark blue peptide chain is SbAD, and the pale blue is SwADC. The active site lysine and likely acid/base catalyst glutamate are conserved in both proteins, along with Arg114. There are numerous differences, however observed throughout the rest of the model. Although the carboxylate-binding site appears to be conserved, other residues in the active site aside from those that contribute to forming a Schiff base intermediate, are not.



A comparison of SwADC and SbAD (Figure 40) shows K122 and E84 are conserved, contributing to Schiff base intermediate in the mechanism. Similar to Figure 39, although Arg114 is conserved, it is facing in the opposite direction. We can now see the difference near Arg114 for the Asn118 in SwADC and Glu118 in SbAD. Mutating position 118 in either enzyme alters the steady state kinetics, and suggests a role in catalysis and/or substrate binding. The tertiary structures of SwADC and the known decarboxylases are extremely similar (RMSD <2.0Å for all common Cα atoms). The loss of decarboxylase activity is the result of a series of key amino acid substitutions

and subtle structural rearrangements. Thus, there is significant reaction diversity even within a single family of the ADCSF.

5.4 Conclusions

While SwADC is very weak with the substrates tested, it can catalyze the aldol condensation and subsequent dehydration of pyruvate and a number of aldehydes to yield the corresponding enones. However, MS data points towards hydration only in the catabolic direction, noted by the hydrations of vinyl pyruvate (**2**) and OHD (**8**). Future work will focus on characterizing the other members of the proposed catabolic gene cluster with the goal of identifying the physiological substrate for this catabolic pathway.

5.5 References

[1] Dwyer, R. U. H. a. D. F. (1997) Biodegradation of Dioxin-Related Compounds: A Review, *Biorem. J. 1*.

[2] Monna, L., Omori, T., and Kodama, T. (1993) Microbial degradation of dibenzofuran, fluorene, and dibenzo-p-dioxin by Staphylococcus auriculans DBF63., *Appl. Environ. Microbiol. 59*.

[3] Aly, H. A. H., Huu, N. B., Wray, V., Junca, H., and Pieper, D. H. (2008) Two angular dioxygenases contribute to the metabolic versatility of dibenzofuran-degrading Rhodococcus sp. strain HA01., *Appl. Environ. Microbiol. 74*.

[4] Li, Q., Wang, X., Yin, G., Gai, Z., Tang, H., Ma, C., Deng, Z., and Xu, P. (2009) New metabolites in dibenzofuran cometabolic degradation by a biphenyl-cultivated Pseudomonas putida strain B6-2., *Environ. Sci. Technol. 43*.

[5] Wittich RM, W. H., Sinnwell V, Francke W, Fortnagel P. . (1992) Metabolism of dibenzo-p-dioxin by Sphingomonas sp. strain RW1., *Appl Environ Microbiol. 58*.

[6] Nam, I. H., Hong, H. B., and Schmidt, S. (2014) Is the biotransformation of chlorinated dibenzo-p-dioxins by Sphingomonas wittichii RW1 governed by thermodynamic factors?, *J Microbiol 52*, 801-804.

[7] Miller TR, D. A., Salzberg SL, Saunders E, Detter JC, Halden RU. (2010) Genome sequence of the dioxin-mineralizing bacterium Sphingomonas wittichii RW1., *J Bacteriol. 192*, 6101.

[8] Miller, T. R., Delcher, A. L., Salzberg, S. L., Saunders, E., Detter, J. C., and Halden, R. U. (2010) Genome sequence of the dioxin-mineralizing bacterium Sphingomonas wittichii RW1, *J Bacteriol 192*, 6101-6102.

[9] Blanvillain, S., D. Meyer, A. Boulanger, M. Lautier, C. Guynet, N. Denance, J. Vasse, E. Lauber, and M. Arlat. . (2007) Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria., *PLoS One 2*.

[10] Postle, K., and Kadner, R. J. (2003) Touch and go: tying TonB to transport, *Mol Microbiol 49*, 869-882.

[11] Nguyen, T. P., Helbling, D. E., Bers, K., Fida, T. T., Wattiez, R., Kohler, H. P., Springael, D., and De Mot, R. (2014) Genetic and metabolic analysis of the carbofuran catabolic pathway in Novosphingobium sp. KN65.2, *Appl. Microbiol. Biotechnol.*

[12] Burroughs, A. M., Hoppe, R. W., Goebel, N. C., Sayyed, B. H., Voegtline, T. J., Schwabacher, A. W., Zabriskie, T. M., and Silvaggi, N. R. (2013) Structural and functional characterization of MppR, an enduracididine biosynthetic enzyme from streptomyces hygroscopicus:

functional diversity in the acetoacetate decarboxylase-like superfamily, *Biochemistry 52*, 4492-4506.

[13] Magarvey, N. A., Haltli, B., He, M., Greenstein, M., and Hucul, J. A. (2006) Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens, *Antimicrob. Agents Chemother. 50*, 2167-2177.

[14] Wang, H., Xiao, Y., Fu, L., Zhao, H., Zhang, Y., Wan, X., Qin, Y., Huang, Y., Gao, H., and Li, X. (2010) High-level expression and purification of soluble recombinant FGF21 protein by SUMO fusion in Escherichia coli, *BMC Biotechnol 10*, 14.

[15] Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D., and Butt, T. R. (2004) SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins, *J Struct Funct Genomics 5*, 75-86.

[16] Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta crystallographica. Section D, Biological crystallography 67*, 271-281.

[17] Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography 26*.

[18] Evans, P. (2006) Scaling and assessment of data quality, *Acta crystallographica. Section D, Biological crystallography 62*, 72-82.

[19] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta crystallographica. Section D, Biological crystallography 67*, 235-242.

[20] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J. Appl. Crystallogr. 40*, 658-674.

[21] Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta crystallographica. Section D, Biological crystallography 66*, 213-221.

[22] Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P., and Adams, P. D. (2010) Joint X-ray and neutron refinement with phenix.refine, *Acta crystallographica. Section D, Biological crystallography 66*, 1153-1163.

[23] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta crystallographica. Section D, Biological crystallography 60*, 2126-2132.

[24] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta crystallographica. Section D, Biological crystallography 66*, 486-501.

[25] Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Headd, J. J., Moriarty, N. W., Richardson, J. S., Richardson, D. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2010) phenix.model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics, *J. Appl. Crystallogr. 43*, 669-676.

[26] Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta crystallographica. Section D, Biological crystallography 66*, 12-21.

[27] Highbarger, L. A., Gerlt, J. A., and Kenyon, G. L. (1996) Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the pKa of active-site lysine 115, *Biochemistry 35*, 41-46.

[28] Metanis, N., Keinan, E., and Dawson, P. E. (2005) A designed synthetic analogue of 4-OT is specific for a non-natural substrate, *Journal of the American Chemical Society 127*, 5862-5868.

[29] Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Crystallogr D Biol Crystallogr 67*, 271-281.

[30] Leslie, A. G. W. (1992) Recent changes to the MOSFLMpackage for processing film and image plate data, *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography 26*.

[31] Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallogr D Biol Crystallogr 62*, 72-82.

[32] Mueller, L. S., Hoppe, R. W., Ochsenwald, J. M., Berndt, R. T., Severin, G. B., Schwabacher, A. W., and Silvaggi, N. R. (2015) Sbi00515, a Protein of Unknown Function from Streptomyces bingchenggensis, Highlights the Functional Versatility of the Acetoacetate Decarboxylase Scaffold, *Biochemistry 54*, 3978-3988.

[33] Ho, M. C., Menetret, J. F., Tsuruta, H., and Allen, K. N. (2009) The origin of the electrostatic perturbation in acetoacetate decarboxylase, *Nature 459*, 393-397.

[34] Tagaki, W., and Westheimer, F. H. (1968) Acetoacetate decarboxylase. The molecular weight of the enzyme and subunits, *Biochemistry 7*, 895-900.

[35] Fenn, T. D. R., D.; Petsko, G. A. (2003) POVScript+: A program for model and data visualization using persistence of vision ray-tracing, *J. Appl. Cryst. 36*, 944-947.

[36] Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *Journal of Applied Crystallography 24*, 946-950.

CHAPTER VI: CONCLUSIONS

Despite the retention of an overall tertiary fold and a Schiff base intermediate in their mechanisms, there is significant diversity in terms of substrate specificity and the reactions catalyzed among enzymes of Family V of the ADCSF. MppR was the first Family V enzyme to be characterized. It was shown to catalyze the cyclization of 4-hydroxy-2-ketoarginine, as well as aldol condensation of 4-imidazolecarboxaldehyde and pyruvate with subsequent dehydration to yield the enone [1]. The latter of these two reactions is also catalyzed by SbAD [2], an aldolase-dehydratase from *Streptomyces bingchenggensis* that will accept a range of aldehyde substrates and SwADC, an enzyme of unknown function from *Sphingomonas wittichii*. Although all three enzymes *can* catalyze the aldol condensation reaction, it is only the primary reaction for SbAD. In addition, there are differences in the substrate preferences in all cases. Subtle differences in the residues lining the active sites of these 3 enzymes accounts for their different reactions and substrate specificities. SbAD is most efficient with aldehydes having a bulky, substituted-aromatic group, like 4-nitrocinnamaldehyde, whereas SwADC was only observed to catalyze the reaction with the smaller, more polar 3-(2-furyl)acrolein. Transient kinetic studies of the reaction of cinnamaldehydes with pyruvate catalyzed by SbAD, together with KIE analysis, allowed us to propose a mechanism for this ADCSF enzyme. The order of addition was not conclusive from transient kinetics alone, so inhibitors were used to determine that the mechanism is ordered sequential with pyruvate likely binding first. Mutation of residues in the active site of SbAD identified the key residues for catalysis. Binding constants measured for pyruvate and cinnamaldehyde show that none of the residues tested had a significant effect on substrate binding, suggesting that substrate binding is driven more by van der Waals forces and size/shape

complementarity between the substrate and active site. Tetramer interface mutants also suggest that SbAD may undergo a conformational change during catalysis, though there is no evidence of cooperativity in any of the kinetics data. The enzyme Swit_4259 from *Sphingomonas wittichii* (SwADC) also maintained the overall fold of ADC, but appears to function as a hydratase or aldolase dehydratase, as evidence for the retro-aldol reaction is lacking. Based on the collected results, SwADC appears to violate the principle of reversibility, whereby the reaction catalyzed in one direction should also be catalyzed in the reverse. All three enzymes in Family V lack the key decarboxylase residues found in Family I: the arginine that orients the acetoacetate in the active site for nucleophilic attack, and the second glutamate, whose negative charge aids in decarboxylation. The loss of those key residues, but retention of the lysine nucleophile and likely acid/base glutamate, might lead one to believe that the Family V enzymes would have similar substrate specificities and activities. However, even these three incredibly similar enzymes illustrate the variety of reactions catalyzed in Family V alone. All three enzymes are also found in organisms noted for secondary metabolism, and based on gene context, may aid in those biosynthetic (SbAD) or aromatic hydrocarbon degradation (SwADC) processes.

While much has been learned about SbAD and SwADC, several outstanding questions remain. The biological function of SbAD is still unknown. Studying a close homolog from another bacterium might offer clues to its biological function. To this end, we have identified two species with SbAD homologs having greater than 50% sequence identity: *Burkholderia terrae* (WQE_44883, 57%) and *Actinobacteria bacterium OK006* (OK006_7770, 91%). Real time PCR could be used to identify growth conditions for one or both species that result in expression of the SbAD homologs. Comparative metabolomics experiments using the instrumentation in the

Shimadzu Laboratory for Advanced and Analytical Chemistry could identify compounds that are only present when these genes are expressed. Then, performing the same experiment in knock-out strains, we should be able to identify a small number of compounds that might be associated with the SbAD homolog. Another outstanding question regarding catalysis in SbAD is the number of protons moving in the reaction. This might be answered with a proton inventory experiment, which can detect the coupling of several proton transfers in a single step [3]. Since the pH rate profile has been done, a pD profile would be needed next. The pH/pD at which the reactions in H_2O and D_2O give the same turnover number would be optimal for the proton inventory, detecting how many protons are in flight during turnover. The methods to investigate SbAD can also be applied to SwADC. SwADC requires more substrate specificity experiments, transient kinetics to see if there is any burst phase in the reaction, and single turnover to gather rate constants. Another method to detect an intermediate could be to incubate the enzyme with substrate, use $NaBH_4$ to reduce the imine to an amide, and then use mass spectroscopy to determine what has bound. In the case of an isomer bound, a chiral column for HPLC may be able to distinguish the compound based on signal, retention time, and comparison to known compounds. Overall, Family V contains enzymes that catalyze a variety of reactions; it may be that the only factor common to most ADCSF enzymes is the Schiff base chemistry. This work has contributed to our knowledge of the structure-function relationship in proteins and could potentially increase the reliability of function annotations for ADCSF sequences in genome sequence databases.

6.1 References

[1] Burroughs, A. M., Hoppe, R. W., Goebel, N. C., Sayyed, B. H., Voegtline, T. J., Schwabacher, A. W., Zabriskie, T. M., and Silvaggi, N. R. (2013) Structural and functional characterization of MppR, an enduracididine biosynthetic enzyme from streptomyces hygroscopicus: functional diversity in the acetoacetate decarboxylase-like superfamily, *Biochemistry 52*, 4492-4506.

[2] Mueller, L. S., Hoppe, R. W., Ochsenwald, J. M., Berndt, R. T., Severin, G. B., Schwabacher, A. W., and Silvaggi, N. R. (2015) Sbi00515, a Protein of Unknown Function from Streptomyces bingchenggensis, Highlights the Functional Versatility of the Acetoacetate Decarboxylase Scaffold, *Biochemistry 54*, 3978-3988.

[3] Anslyn, E. V., Dougherty, D. A. (2006) *Modern Physical Organic Chemistry*, University Science Books, Sausalito, California.

APPENDIX I: SbAD

Genomic context of the gene encoding Sbi_00515.

Gene Product	Length (aa)		Annotation	Closest characterized homolog	Identity (%)
Sbi_00514	160	+	Urease accessory protein	UreD, Streptomyces mirabilis	79
Sbi_00515	265	-	Acetoacetate decarboxylase	MppR, Streptomyces hygroscopicus	39
Sbi_00516	205	+	2,4-DAPG biosynthesis transcriptional repressor PhlF	Mycofactocin system transcriptional regulator, Rhodococcus sp. AD45	36
Sbi_00517	219	-	TetR family transcriptional regulator	None	~
Sbi_00518	133	+	Stress responsive alpha-beta barrel domain protein	None	~
Sbi_00519	294	+	Taurine catabolism dioxygenase	Alpha-ketoglutarate-dependent taurine dioxygenase, Burkholderia caribensis MBA4	46
Sbi_00520	188	+	Hypothetical protein	FcoT, type III thioesterase, Mycobacterium tuberculosis	41
Sbi_00521	133	+	Acyl carrier protein	MelD protein, Melittangium lichenicola	34
Sbi_00522	1813	+	Modular polyketide synthase	None	~
Sbi_00523	485	+	Beta-ketoacyl synthase	Ketosynthase-acyltransferase di-domain from module CurL of the curacin A polyketide synthase, Moorea producens 3L	41
Sbi_00524	524	+	Long chain fatty acid-CoA ligase	Dicarboxylate-CoA ligase PimA, Streptomyces sp. AA4	54
Sbi_00525	437	+	Cytochrome P450	Cytochrome P450 eryF, Saccharopolyspora erythraea	36
Sbi_00526	254	+	Acetoacetate decarboxylase	Acetoacetate decarboxylase, Clostridium acetobutylicum ATCC 824	56
Sbi_00527	514	-	Hypothetical protein	None	~
Sbi_00528	528	-	Hypothetical protein	None	~

Figure S1. Proton NMR spectrum of a solution containing 12.5 mM **1** in 50 mM sodium phosphate buffer (D_2O), pH 7.4.

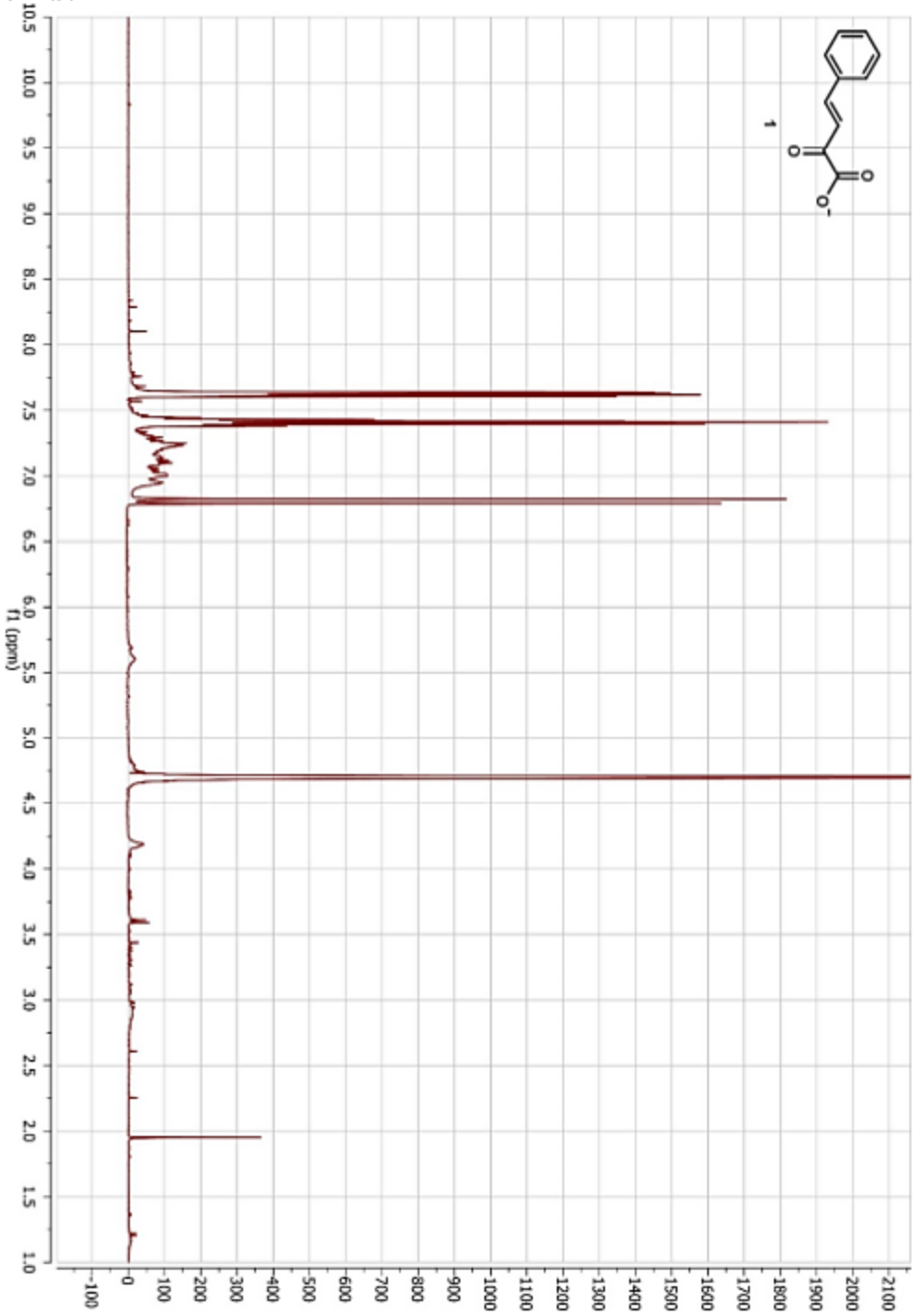


Figure S2. Proton NMR spectrum of a reaction mixture containing initially 12.5 mM **1** and 2.8 µM Sbi00515 in 50 mM sodium phosphate buffer (D_2O), pH 7.4. The reaction (500 µl) was carried out in 50 mM sodium phosphate buffer (H_2O), pH 7.4, for 80 min, after which the sample was evaporated to dryness in a CentriFan centrifugal evaporator and resuspended in 500 µl of D_2O.

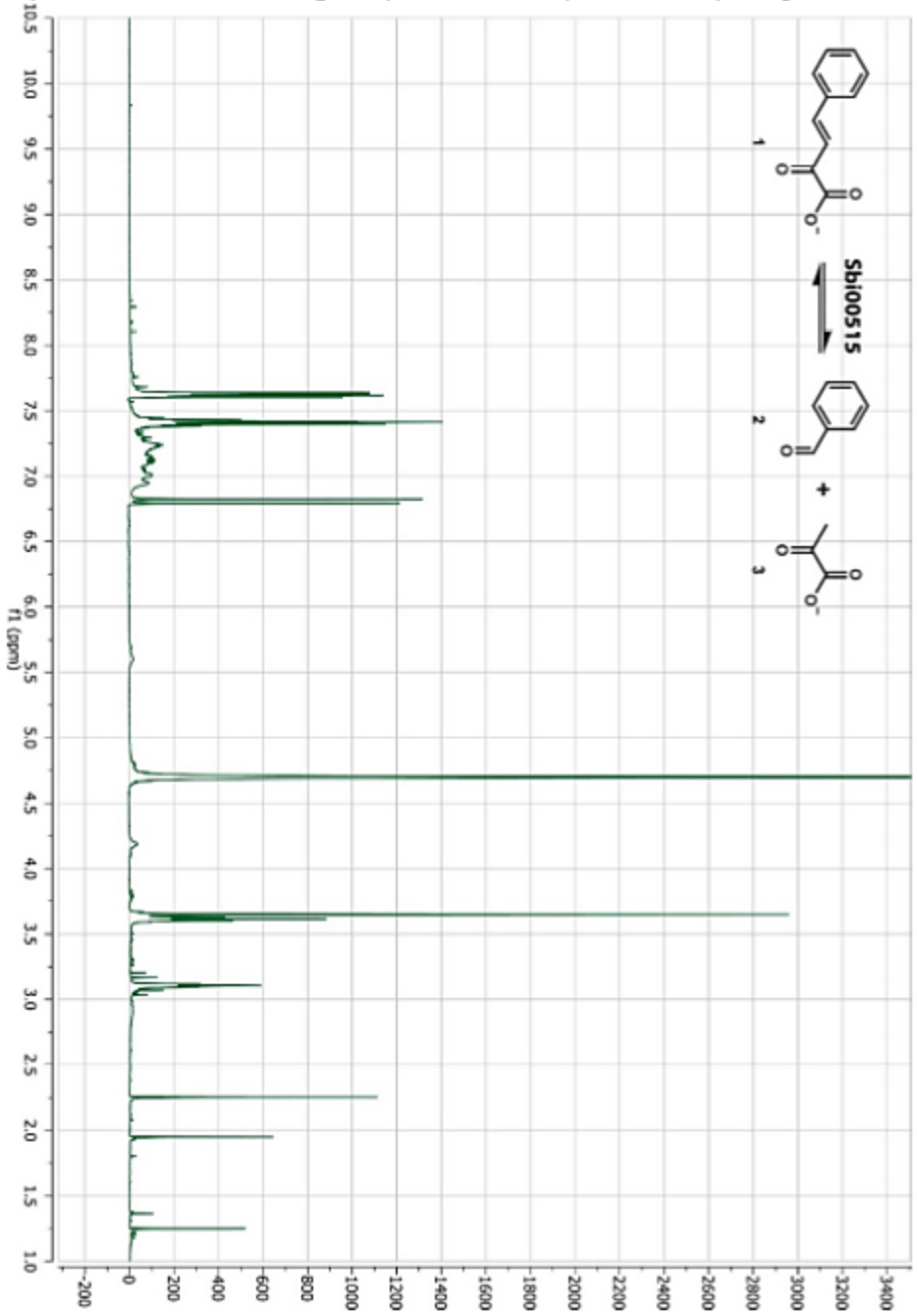


Figure S3. Proton NMR spectrum of a solution containing 50 mM **2** and 50 mM **3** in 50 mM sodium phosphate buffer (D_2O), pH 7.4.

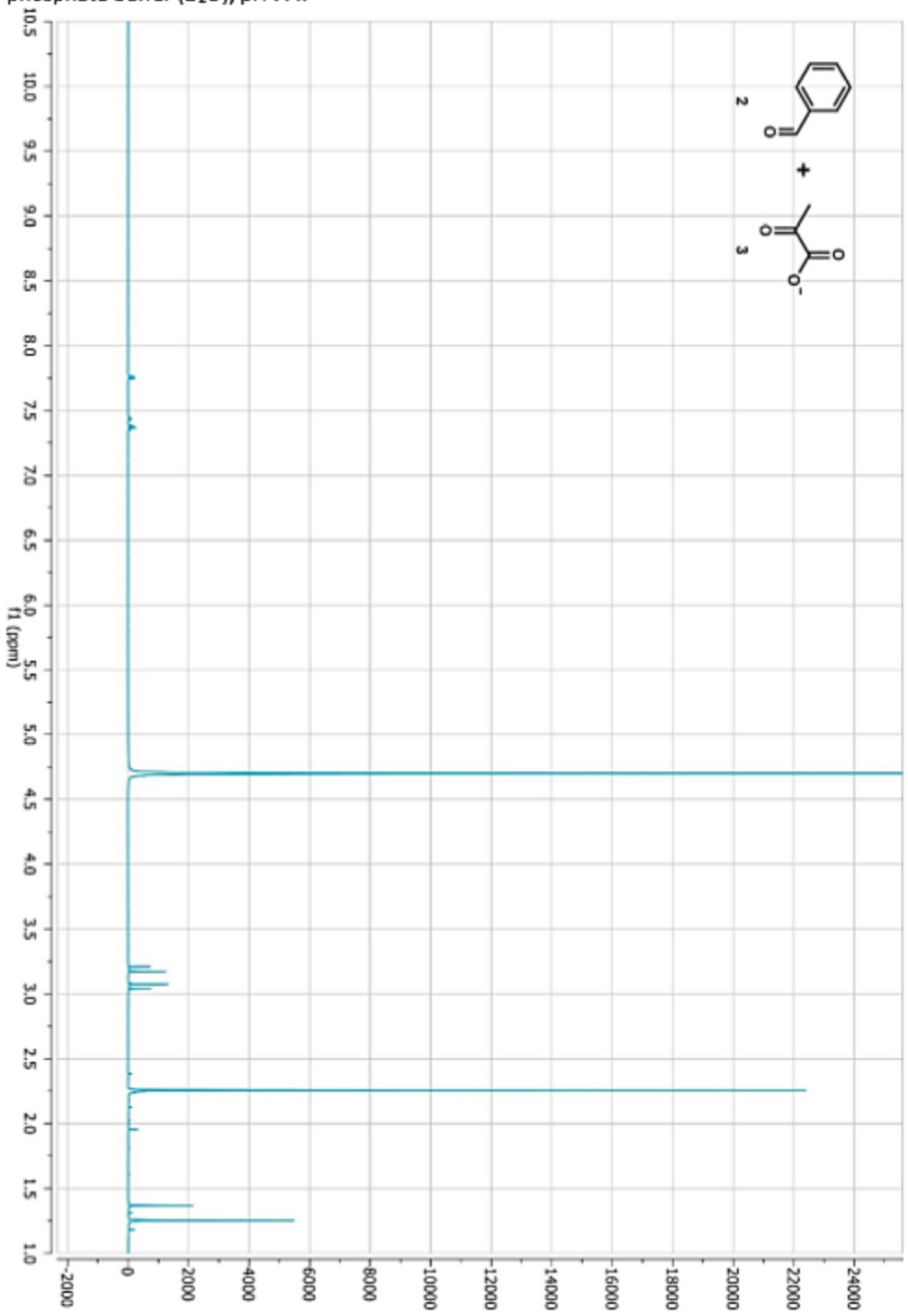
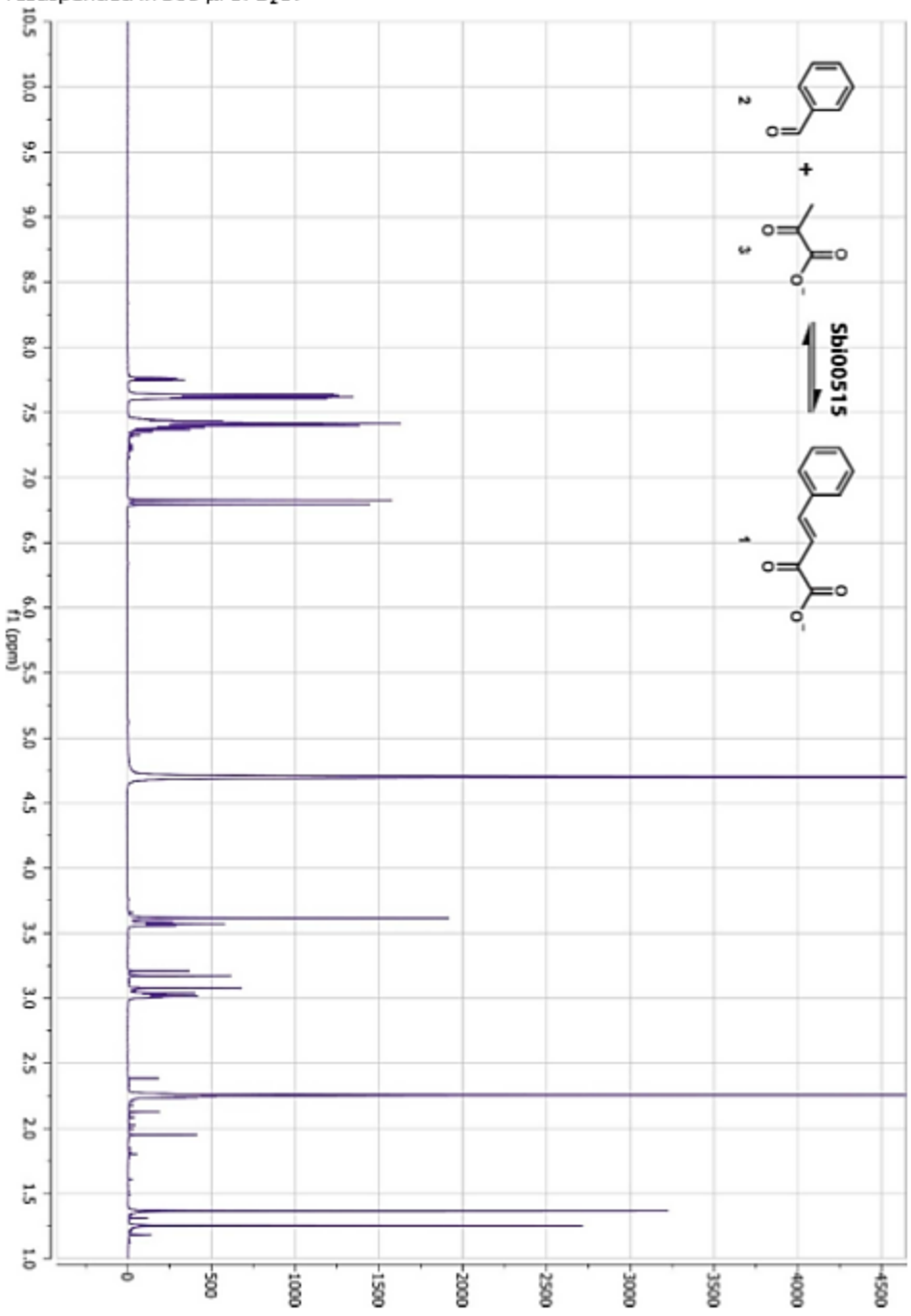
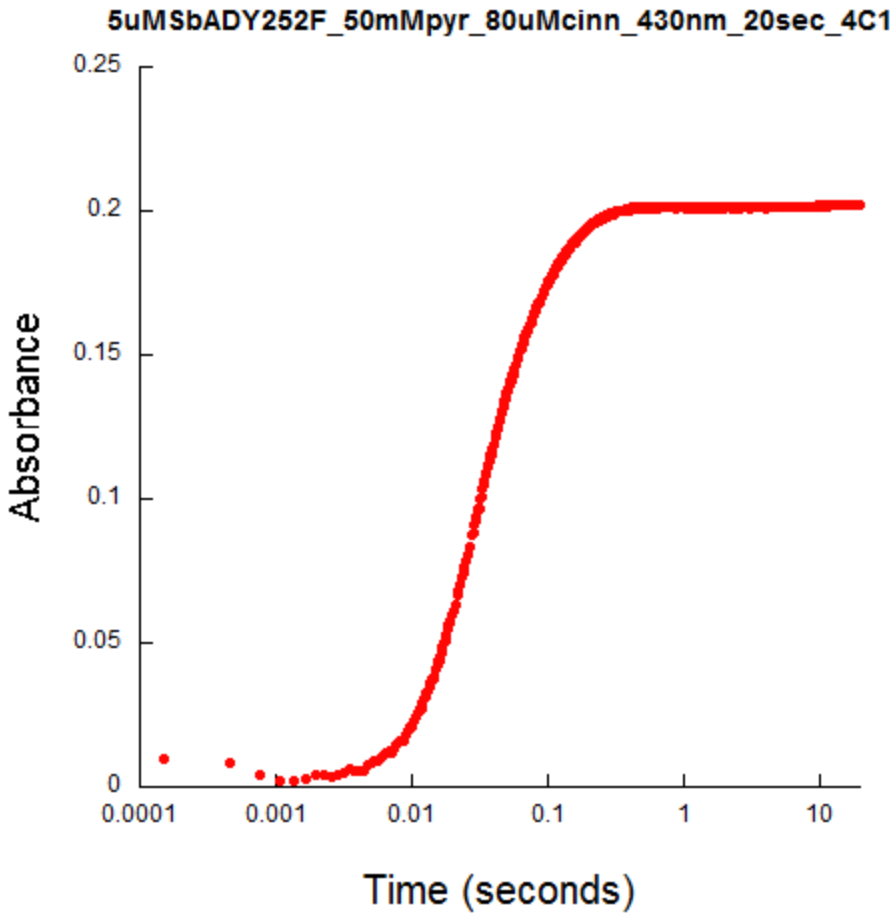


Figure S4. Proton NMR spectrum of a reaction mixture (500 µl) containing initially 50 mM 2, 50 mM 3, and 2.8 µM Sbi00515. The reaction was carried out in 50 mM sodium phosphate buffer (H_2O), pH 7.4, for 35 min, after which the sample was evaporated to dryness in a CentriFan centrifugal evaporator and resuspended in 500 µl of D_2O.




5uMSbADY252F_50mMpyr_80uMcinn_430nm_20sec_4C1
Absorbance
0.25
0.2
0.15
0.1
0.05
0
0.0001
0.001
0.01
0.1
1
10
Time (seconds)

APPENDIX II: SwADC

(LC)MS; Negative polarity Swit WT

Sample	m/z	Intensity
Buffer (20mM Ammonium acetate pH 7.0)	87.15	226
Buffer (20mM Ammonium acetate pH 7.0) [intensity > 300]	105.20	980
	189.10	340
	245.45	440
	301.80	400
	361.45	550
2mM Sodium pyruvate standard	87.20	2865
2mM 2KP standard	113.20	4000
2KP + Swit WT	87.15	558
	113.20	2100
	131.20	3500
2KP + Swit Y120F (ZM -80°C stock)	87.20	365
	113.20	2400
	131.20	5000
2mM 2-HM	113.20	4000
	157.15	16000
	173.15	89
2mM 2-HM + 4-OT	113.20	4000
	131.20	400
	157.15	10672
	173.15	117
	175.15	390
2mM 2-HM + 4-OT + Swit WT	87.20	700
	113.20	1500
	131.20	4500
	157.15	3000
	173.15	213
	175.15	6152
2mM 2-HM + 4-OT + Swit Y120F (ZM -80°C stock)	86.95	369
	113.15	1357
	131.20	1000
	157.15	4364
	173.10	265
	175.10	7960
2mM CHM	215.15	44602
	217	4000
	230.20	168
2mM CHM + CHMI	195.05	11600
	215.15	725
	217	3129
	231.15	875
2mM CHM + CHMI + Swit WT	195.05	15500
	215.15	2141

	217.05	3548
	231.15	1162
	187.15	646
	189.10	2678
2mM CHM + CHMI + Swit Y120F	189	2000
	195.05	22000
	215	2000
	217	4100
	231.15	1500
2mM CHM + CHMI + COHED	169.90	401
	213	665
	111.10	6600
	171.15	3200
	189	800
	195.05	13100
	217	2615
2mM CHM + CHMI + COHED + SbAD WT	217	3000
	171.15	3800
	195.05	18200
	189.05	2327
	231.1	800
2mM CHM + CHMI + COHED + SbAD Y120F	171.05	6200
	189.05	4100
	195.05	22800
	217.05	4000
	231.15	1600
2mM CHM + CHMI + COHED + OHEDH	171.15	6200
	189.05	6200
	195.05	15500
	215.10	2100
	217.05	4100
	231.10	1500

10 µL injections.

CURRICULUM VITAE

Lisa Susannah Mueller

I am thrilled to be on the threshold of my career in structural biology and eager to progress. I am interested in studying the structures and functions of proteins involved in disease states and natural product biosynthesis.

EDUCATION

University of Wisconsin-Milwaukee
Fall 2010 – Summer 2016

Graduate Student in Nicholas R. Silvaggi Laboratory
Focus on X-ray crystallography, steady state/transient enzyme kinetics, kinetic isotope effects, proton NMR, and protein expression, purification from recombinant *E. coli* as well as site-directed mutagenesis to relevant enzymes in the acetoacetate decarboxylase-like super family

UW-Milwaukee
Fall 2005 – Spring 2010

Bachelors of Science in Biochemistry, Calculus-Based Coursework

- Graduated magna cum laude, maintained Dean's Honor Roll GPA requirements 3.75 or higher
- Undergraduate research included cell/tissue culture, western blotting, gel filtration chromatography, atomic absorption and mass spectroscopy

PUBLICATIONS

Further Writing
2015 – Present

Two Manuscripts in Preparation

ACS Biochemistry
2015

Sbi00515, a Protein of Unknown Function from *Streptomyces bingchenggensis*, Highlights the Functional Versatility of the Acetoacetate Decarboxylase Scaffold
Mueller, L. S. et al. Biochemistry **54**(25): 3978-3988.

CONFERENCES

ORAL PRESENTATIONS

Illinois Institute of Technology, IL
September 2015

35th Midwest Enzyme Chemistry Conference

Philadelphia, PA
July 2015

65th Annual American Crystallographic Association Meeting

- Etter Early Career Symposium

POSTER PRESENTATIONS

UW-Milwaukee
April 2015
Department of Chemistry and Biochemistry Research Symposium

Galveston, TX
January 2015
24th Enzyme Mechanisms Conference

Northwestern University, IL
September 2014
34th Midwest Enzyme Chemistry Conference

Madison, WI
May 2014
36th Steenbock Symposium: Enzyme Structure and Function – In Honor of the Life of W. W. Cleland

UW-Milwaukee
May 2013
Department of Chemistry and Biochemistry Research Symposium

University of Illinois-Chicago, IL
October 2013
32nd Midwest Enzyme Chemistry Conference

Coronado, CA
January 2013
23rd Enzyme Mechanisms Conference

Boston, MA
July 2012
62nd Annual Crystallographic Association Meeting

UW-Milwaukee
May 2012
Department of Chemistry and Biochemistry Research Symposium

SEMINARS

UW-Milwaukee
October 2013
Biophysical-Associated Laboratories Biweekly Meeting
"Structural and Functional Analysis of an Acetoacetate Decarboxylase-*Like* Protein from *Streptomyces bingchenggensis*"

UW-Milwaukee
January 2012
Department of Chemistry and Biochemistry
"Enzymes in Organic Synthesis: Engineering Biocatalysts for Industrial and Pharmaceutical Applications"

AWARDS

UW-Milwaukee
2012-2015
Advanced Opportunity Program Fellowship
Stipend and tuition remission during academic semesters to support economically disadvantaged and under-represented populations in graduate study

UW-Milwaukee
April 2015
Keith Hall Award for Excellence in Graduate Research
First place poster presentation, Department of Chemistry and Biochemistry Research Symposium

UW-Milwaukee
May 2013

UW-Milwaukee Graduate School Research Award
Third place poster presentation, Department of Chemistry and Biochemistry Research Symposium

UW-Milwaukee
May 2012

Keulks Award for Graduate Student Research
Second place poster presentation, Department of Chemistry and Biochemistry Research Symposium

PROFESSIONAL DEVELOPMENT

Denver, CO
July 2016

66th Annual American Crystallographic Association Meeting
Co-chair of Diversity and Inclusion Session: talks to focus on successful strategies for inclusion, retention, and stereotype threat through training, mentoring, or research; engaging diverse populations through outreach using crystallography

Argonne National Laboratory
June 2015

Advanced Photon Source, Life-Sciences Collaborative Access Team Data Collection Workshop
Hands-on X-ray crystallography workshop, focused on resolving twinning and optimal data collection and model refinement techniques

UW-Milwaukee
2013 – 2014

Department of Chemistry and Biochemistry Graduate Student Council President
Led a diverse team that organized the departmental research symposium, guided new graduate students during orientation, and represented graduate student biochemistry division at faculty meetings

UW-Milwaukee
August 2014, September 2015

Panelist for Incoming Graduate Students

Illinois Institute of Technology, IL
March 2014

6th Yao Yuan Biotech-Pharma Symposium: Academic – Industry Connections

UW-Milwaukee
March 2014

Scientific Process Judge for Badger State Science and Engineering Fair for High School Students

UW-Milwaukee
Spring 2014

Center for Student Involvement: Coaching, Organization, Development, and Engagement (CODE) Program
Weekly meetings on leadership techniques and successful strategies in team projects

Oak Ridge National Laboratory; Argonne National Laboratory August 2013	**National School on Neutron and X-ray Scattering** Hands-on training of instrumentation, data collection, and analysis of triple-axis spectrometer, small angle neutron scattering, nano-scale ordered materials diffractometer, and liquids reflectometer at Oak Ridge; X-ray fluorescence microscopy, small angle X-ray scattering, coherent X-ray diffraction at Argonne
UW-Milwaukee October 2012	**Research Computing Boot Camp** Introduction to UNIX and parallel computing
UW-Milwaukee Fall 2012	**Responsible Conduct of Research Seminar Series** Weekly lectures and discussions about ethics in research; research grant provided by NSF required training
UW-Milwaukee 2010 – 2012	**Teaching Assistant for Lecturer Gloria Freschl: Introductory General, Organic, and Biochemistry for Nursing Students** Discussion and laboratory instructor for fall, spring, and summer academic semesters, as well as weekly tutoring hours for all introductory level chemistry courses offered
UW-Milwaukee Summer 2011	**Project Assistant** Organized and cleaned teaching laboratory stock rooms

MEMBERSHIPS

- American Chemistry Society
- American Crystallographic Association
- American Society for Biochemistry and Molecular Biology
- Biophysical Society
- Milwaukee Area Structural Biology Interest Group
- Protein Society

LANGUAGES

- English – native
- French – limited working proficiency