The Dissertation Committee for Jamison Parker Huddleston certifies that this is the approved version of the following dissertation:

On the Reactions of *trans*-3-Chloroacrylic Acid Dehalogenase and a *cis*-3-Chloroacrylic Acid Dehalogenase Homologue, Cg10062: Mechanistic and Evolutionary Implications

Committee:

Christian P. Whitman, Supervisor

Kenneth A. Johnson

Sean M. Kerwin

Walter L. Fast

David W. Hoffman

On the Reactions of *trans*-3-Chloroacrylic Acid Dehalogenase and a *cis*-3-Chloroacrylic Acid Dehalogenase Homologue, Cg10062: Mechanistic and Evolutionary Implications

by

Jamison Parker Huddleston, B.S.Bioch.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2015

Dedication

To my wife, Angela, my parents Cathie, Jon, and Leanne, my immediate and extended family members and my in-laws, Dave and Ellen. Their relentless love and support have helped make this possible.

Acknowledgements

This dissertation is the compilation of work performed over the last seven years and was made possible by the help of many people and organizations. This research was supported by the National Institute of Health, the Robert A. Welch Foundation, and a Continuing Fellowship from the College of Pharmacy. I want to express my gratitude to my advisor, Dr. Christian Whitman. His guidance, encouragement, and patience have been instrumental in my development as a graduate student. I want to thank my committee members, specifically, Dr. Kenneth Johnson. I was afforded the opportunity to learn and work with a world renowned kineticist. The chance to work with someone of Ken's caliber is extremely rare and without his help, my research profile would be very different. I also want to specifically thank two of my colleagues who have helped me greatly over the years, Dr. Gottfried Schroeder and Dr. William Johnson Jr. As a new graduate student I had very little lab experience and essentially no research experience, Dr. Schroeder encouraged me and taught me how to be a research scientist. Together, we learned many different things, helped each other, and became good friends. Without him, it would have been much more challenging to accomplish as much. Dr. Johnson has also been crucial in my success as graduate student. Bill's experience and, more impressively, his intuition have been invaluable time saved and have helped focus me. I also want to thank my other lab mates and fellow students that have helped me over the years. "Teamwork makes the dream work" − Maxwell

On the Reactions of *trans*-3-Chloroacrylic Acid Dehalogenase and a *cis*-3-Chloroacrylic Acid Dehalogenase Homologue, Cg10062: Mechanistic and Evolutionary Implications

Jamison Parker Huddleston, Ph.D

The University of Texas at Austin, 2015

Supervisor: Christian P. Whitman

The tautomerase superfamily (TSF) provides an excellent model system to study enzyme specificity, catalysis, and divergent evolution. *trans*-3-Cholroacrylic acid dehalogenase (CaaD), *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD), and malonate semialdehyde decarboxylase (MSAD) are three TSF members that catalyze the final reactions in the degradation of the nematocide, 1,3-dichloropropene. All three enzymes have the TSF characteristic β-α-β fold and catalytic amino terminal proline (Pro-1). Both CaaD and *cis*-CaaD dehalogenate their respective isomers of 3-chloroacrylic acid yielding malonate semialdehyde. Subsequently, MSAD decarboxylates malonate semialdhyde resulting in acetaldehyde and CO_2. Their catalytic and substrate specificities are exquisite considering they share three key and positionally conserved residues. As part of an effort to understand how such specificity evolved, a pre-steady-state kinetic analysis of CaaD was carried out. Alongside a similar study on *cis*-CaaD, a

fluorescent mutant of CaaD was constructed that had minimal kinetic differences from the wild-type. The mutant was validated as an accurate fluorescent reporter of change in enzyme state that allowed for the reaction to be followed using stopped-flow methods. Stopped-flow fluorescence, rapid chemical quench data and ultraviolet spectroscopy were globally fit by computational simulation. The fit resulted in a kinetic mechanism for CaaD affording detailed information about the reaction, including measuring the rate of product release, the rate of chemistry, a previously unknown partially rate-limiting step associated with a conformational change, and the definition of binding constants for both products (MSA and Br^-). In addition to the dehalogenation reaction, the reaction of the fluorescent mutant with a mechanism-based inhibitor, 3-bromopropiolate, was characterized. The values for the apparent rate of inhibition and potency were defined and estimates were determined for the values of the rate of chemistry and the release of bromide. The information gathered during these inhibition experiments was used to further refine the CaaD dehalogenation mechanism eliminating ambiguities present in the initial data set.

Finally, the reactions of a *cis*-CaaD homologue, Cg10062 from *Corynebacterium glutamicum* were characterized. Cg10062 shares high sequence similarity (53%) and the same six critical active site residues as *cis*-CaaD, but Cg10062 has poor *cis*-CaaD activity. Moreover, Cg10062 dehalogenates both 3-chloroacrylic acid isomers. The reactions of Cg10062 with propiolate, 2-butynoate, and 2,3 butadienoate were

investigated. Cg10062 functions as a hydratase/decarboxylase using propiolate generating malonate semialdehyde and acetaldehyde. Cg10062 catalyzes a hydration-dependent decarboxylation of propiolate as exogenously added malonate semialdehyde is not decarboxylated. With 2,3 butadienoate and 2-butynoate, Cg10062 functions as a hydratase and yields only acetoacetate. Mutations to the activating residues Glu114 and Tyr103 produced a range of results from a reduction in wild-type activity to a switch of activity. Possible intermediates for the hydration and decarboxylation products can be trapped as covalent adducts to Pro-1 when $NaCNBH_3$ is incubated with certain combinations of substrate and mutant enzymes. Three mechanisms are presented to explain these findings along with the strengths and weaknesses of each mechanism in terms of being able to account for experimental observations.

Table of Contents

List of Tables

List of Figures

List of Schemes

Chapter 1: Introduction

1.1 Evolution of an Enzyme

Why enzymology? For thousands of years, we have made advantageous use of enzymes in various processes. About 150 years ago, humans began to study these enzymes. The Danish chemist Christian Hansen is credited with the first preparations of a purified enzyme for industrial use.[1] In 1874, he extracted rennet, an enzyme found in the stomach of mammals, from dried calves' stomachs using saline solution.[1] Rennet is a group of protease-like enzymes used in the process of creating cheese from milk. The use of rennet for this purpose is estimated to date back about 8,000 years.[2] Enzymes also play a critical role in other old processes as baking bread and the production of alcohol by fermentation.[1] The enzymes used in these two processes are both found in yeast. In 1876, Wilhelm Kuhne coined the term "enzyme" to describe the isolation of the ferments from the host organism.[1] The word "enzyme" literally translates as "in yeast", derived from the Greek "en" meaning "in" and "zyme" meaning yeast.[1] Early enzymology was founded on a desire to understand the simple transformations that were observed in these processes.

In 1894, the study of enzymes and how they work accelerated because of a hypothesis provided by Emil Fischer.[1] Fischer proposed a model explaining enzyme specificity and turnover. This became known as the "lock and key" model (Figure 1.1).

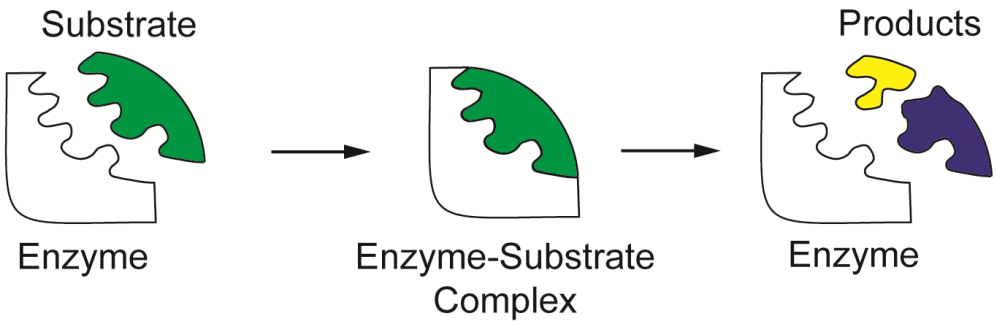


Figure 1.1: Emil Fischer's Lock and Key Model

Some form of the "lock and key" model is still found today as an elementary introduction to enzymes and their properties. Although a simplistic view of an enzyme, this model asks a core question in enzymology: how does an enzyme develop a "lock" which is a match for a "key"? This question remains a topic of many heated discussions.

A simple answer to this question, 'the enzyme evolved the lock to fit the key', prompts another question: how does the evolution of an enzyme occur? One simple method for an enzyme to evolve is by mutation and selection.[3] A point mutation occurs when the coding DNA is altered and, as a consequence, the amino acid sequence is altered. If the point mutation provides a selective advantage for the host, the host will have a greater chance of survival and more opportunity to pass its genes to another generation. This is essentially the theory of evolution proposed by Charles Darwin[4], but applied at the microscopic level. However, beneficial point mutations are rare; while most often point mutations are silent or deleterious. Point mutations in the DNA that lead to loss of regulatory control over a protein's expression levels, gene duplication events, and horizontal gene transfer can all play a role in the evolution of an enzyme.[3]

Enzymes usually function as part of a metabolic pathway. Metabolic pathways use multiple enzymes and generally operate in sequence to achieve a specific end product. These end products are frequently essential to the function and life of the

organism. Metabolic pathways are vital in all known forms of life. Several of these pathways, including those from carbohydrate, lipid, and energy metabolism, are maintained, in some form, across all three domains of life.[5] Thinking about the function of a single enzyme can be difficult, but understanding that function as a part of a metabolic pathway with multiple enzymes and how the enzyme and the pathway has evolved can be overwhelming. Fortunately, there is a strong foundation of past research to help.

To begin, consider the evolution of a single enzyme. The ability to evolve or "evolvability" of an enzyme refers to the likelihood of a particular enzyme to develop and maintain a new activity.[3] A highly evolvable enzyme suggests that it can develop a new activity rather easily. The resulting activity can be recruited to carry out its new task for the benefit of the host organism. The evolvability of a particular enzyme is primarily determined by two aspects: stability and potential for functional innovation.[3] Simply stated, the more stable and functionally malleable a particular enzyme is, the higher the chance it can quickly evolve a new activity that will benefit the host.

Stability means potentially critical point mutations will not destabilize the overall protein. A new function could require several mutations, so that a stable backbone scaffold is critical. Stability does come at a cost: flexibility in an enzyme, especially around the active site, can play an important role in an enzyme's functionality.[3] Movements of loop domains or closures of cap domains can be very important for protein activity. Maintaining a balance between stability and flexibility is critical for a scaffold with high evolvability.

The ability of an enzyme to alter its function, or functional innovation, also plays a critical role in the evolvability of an enzyme.[3,6] At the extremes of functional innovation lies opposing methods for evolving a new activity. One possibility is for an

enzyme to alter its substrate specificity while keeping the chemistry the same. Many enzymes are believed to follow this route including, malate and lactate dehydrogenase and the protease enzymes, trypsin and chymotrypsin.[7] Alternatively, an enzyme can alter the chemistry it performs while maintaining substrate binding.[8] It has been suggested that the primary choice of Nature for the evolution of an enzyme is to alter the substrate specificity because having the right chemistry in place is more difficult than binding an alternate substrate.[7,8] Many enzymes that diverged from a common evolutionary ancestor likely used a combination of these two methods to optimize binding of alternative substrates and the chemistry that is being carried out.

One way Nature can get an evolutionary "head start" is to use a promiscuous enzyme. A promiscuous enzyme is one that shows interactions or activities with alternate substrates. Promiscuity is observed in two distinct forms: catalytic promiscuity and substrate ambiguity.[3,9] Catalytic promiscuity refers to the ability of an enzyme to process different substrates to give different products. The reactions generally share a common intermediate or a mechanistic strategy. Substrate ambiguity refers to the ability of an enzyme to catalyze the same chemical reaction on series of different substrates. The role promiscuity plays in developing enzymes with new functions is one of the topics addressed by the work in this dissertation.

There are two main theories to explain the evolution of metabolic pathways. One theory involves enzyme recruitment.[10] Enzyme recruitment is the process by which a catalyst with a new activity is utilized by the host for a purpose that was not originally evolved.[3,10] A new activity can be acquired by a beneficial genetic alteration (e.g., point mutation). If this activity provides a selective advantage to the host organism, the mutation has a higher probability of being passed on to the next generation. In 1976, R. A. Jensen proposed a refinement of the concept to explain metabolic expansion.[10]

Jenson suggested that ancient enzymes might have had poor substrate specificity allowing them to process a greater variety of substrates. A relatively small number of enzymes with high promiscuity would enable a simple organism to operate at maximum function. Presumably, however, these enzymes would catalyze reactions poorly and provide the driving force for improvement.

Jensen proposed that by coupling gene duplication events with random point mutations, an ancient, non-specific, promiscuous enzyme could improve its catalytic efficiency and evolve to catalyze one specific reaction.[3,10] The product of this new reaction, if stable, can then become the target for another promiscuous enzyme which could then undergo another evolution event. Recruiting these new enzymes to process a series of chemical reactions as a part of a metabolic pathway would then become a relatively straightforward task. When new potential substrates become available to an organism, enzymes with promiscuous activities could then be improved and new metabolic pathways develop. If this is the preferred method of metabolic expansion, one would expect different enzymes to compose a metabolic pathway. In general, this is what is found in metabolic pathways: not all the enzymes in the pathway are similar in sequence and structure. Many contemporary enzymes exhibit some promiscuity.[3]

There is one caveat to this theory of evolution of metabolic pathways: it requires the existence of at least a few ancient, promiscuous enzymes. The origin of these enzymes is not explained by Jensen.[10] One possibility for the origin of these initial enzymes was suggested by N. H. Horowitz, in 1945, as part of his retrograde theory of evolution of metabolic pathway.[11] Although Horowitz's retrograde theory is less favored due to a growing body of evidence against it, it does provide an explanation for the origins of the ancient enzymes.

Horowitz's theory of retrograde evolution is based on another theory proposed by two independent evolutionary theorists, J. B. S. Haldane and A. I. Oparin, in the 1930s.[11,12] The "Oparin-Haldane theory" suggested that early forms of life lived in a "rich, primordial soup" full of all the necessary components for life.[12] Horowitz suggested that this type of environment would allow for early organisms to be mostly heterotrophic, and these organisms could develop a few enzymes which would carry out simple one step biosynthetic processes.[11] As life began to thrive, the "rich, primordial soup" became depleted. This applied pressure on organisms to develop the ability to synthesize essential compounds. By changing the genome through gene duplication or gene fusion or by gaining new genes through horizontal gene transfer, the small number of enzymes available to make these compounds could quickly multiply. This sets the stage for the available enzymes to evolve according to Jensen's theory of metabolic recruitment. A picture of what these "ancestral" enzymes might have looked like can be formed by studying the enzymes in the context of an enzyme superfamily.

1.2 THE ENZYME SUPERFAMILY

During the 1990's, advances in protein structure crystallography, genome sequencing, polymerase chain reaction (PCR) techniques, and computational power allowed researchers to quickly acquire a lot of detailed information about thousands of enzymes. It quickly became obvious that many enzymes with similar sequences share similar structures.[13-17] However, enzymes that share both sequence and structure similarity do not always possess the same functionality (e.g. mandelate racemase and muconate lactonizing enzyme are two examples).[15] These observations stimulated the development of new terminology to describe these groups of enzymes. The difficulty

was developing a method to classify groups of enzymes which accounted for both the similarities in structure and function. Homologous enzymes are defined as those that descend from a common ancestor and therefore are structurally similar.[15] The term "superfamily", as defined by J. A. Gerlt, P. C. Babbit, and others, is used to describe a group of homologous enzymes that catalyze the same chemical reaction with different substrate specificity or that catalyze a different overall reaction, but share a common mechanistic feature enabled by a conserved active site architecture. [13-15]

Conservative estimates predict that there are around a few thousand protein folds in nature.[14] This relatively small number suggests that in more complex species (species with 10,000 or more theoretical gene products), enzymes with similar structures must carry out different reactions.[14] The enolase superfamily (ESF) is one of the most functionally diverse superfamilies known to date, catalyzing at least 14 different reactions. The name comes from the glycolytic enzyme, enolase. Mandelate racemase (Figure 1.2) and muconate lactonizing enzyme, are two members of the ESF.[16,17] All the reactions catalyzed by this superfamily share a common mechanistic strategy. A general base at the active site abstracts the α–proton of the carboxylate moiety on the substrate to form an enolate intermediate, which is stabilized by a divalent metal ion.[13-17] The active sites of the enolase superfamily members are positioned inside of a $(\beta/\alpha)_8$-barrel with another domain acting as a cap (Figure 1.2).[13,17]

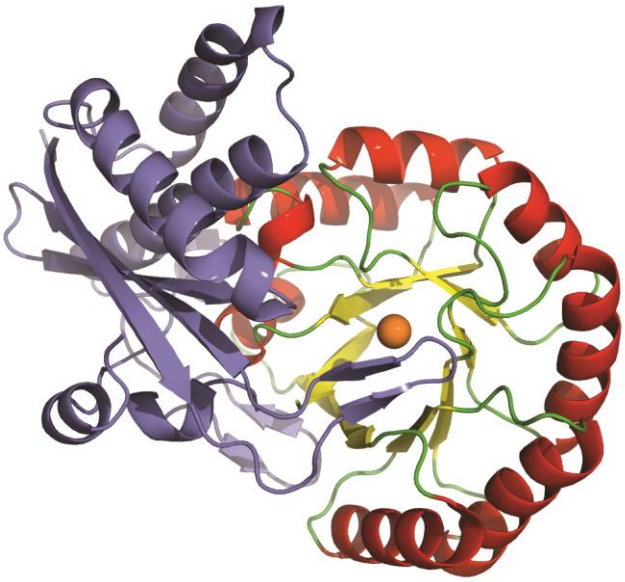


Figure 1.2: Structure of Mandelate Racemase, an ESF Member

The crystal structure of mandelate racemase. The α-helices and the β-sheets of the $(\beta/\alpha)_8$-barrel are colored in red and yellow, respectively. Shown in blue is the cap domain. The active site metal, manganese (Mn^{2+}), is shown as an orange sphere. PDB code: 2MNR

The barrel is "hard-wired" to carry out the chemistry, although the exact chemistry varies among superfamily members, leading to its diversity. The cap may be responsible for the substrate specificity and/or the exclusion of bulk solvent from the active site.[16,17] This structural design modulates the active site, so that new activities can be developed and optimized rather easily utilizing the same scaffold and same general reaction strategy. The β/α-barrel fold is observed in several enzyme superfamilies, lending credence to its evolvability.

Studying the evolution of enzymes and their promiscuity, achieves multiple aims. The results have implications for understanding antibiotic resistance. The results also

assist protein engineering efforts and design, which have a range of applications from bioremediation to general synthesis of pharmaceuticals and other important chemicals.[3] Understanding how enzymes evolve can be used to design more diverse molecules. These molecules include antibiotics that might evade resistance, better pharmaceuticals with limited side effects, and environmentally friendly compounds.

In the tautomerase superfamily (TSF), a diverse set of chemical reactions is observed. All reactions are catalyzed by enzymes that use the same simple structural motif (β-α-β) and an amino-terminal proline.[18,19] The reactions include tautomerization, hydration, dehalogenation, and decarboxylation. There are recent reports that more complex reactions are catalyzed such as carbon-carbon bond formation.[20] Members of the TSF might have undergone several evolution events from a progenitor enzyme. Gene duplication and gene fusion events are evident in the family. An example of this is present in a comparison of the sequence lengths. The sequences are either "short" or "long" in length. The "long" sequences likely arose from a fusion of two "short" sequences because they contain two attached units of the β-α-β structural motif.

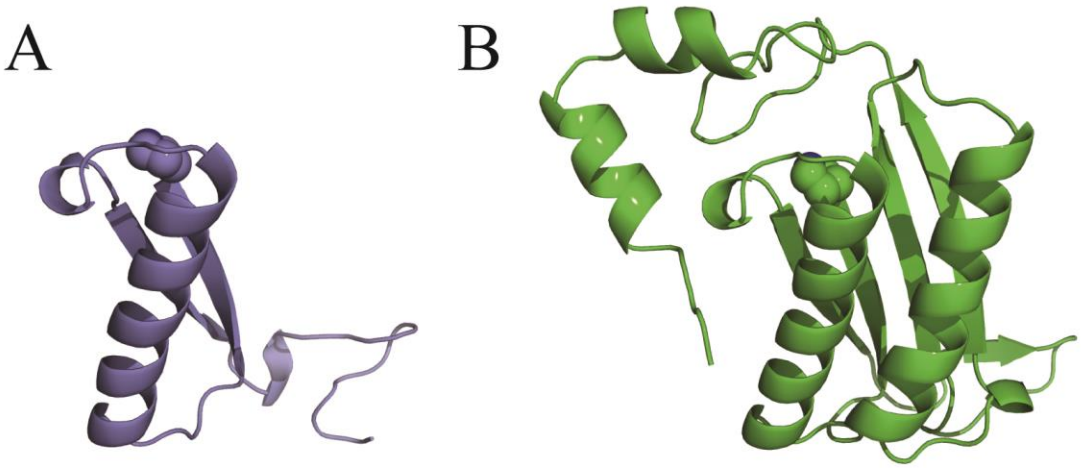


Figure 1.3: β-α-β Structure, the Characteristic Motif of the Tautomerase Superfamily

The crystal structures of the two enzymes in the tautomerase superfamily. A) The "short" chain structure of 4-oxalocrotonate tautomerase consisting of a single β-α-β unit (PDB code: 4OTA). B) The "long" chain structure of *cis*-3-chloroacrylic acid dehalogenase consisting of two linked β-α-β units (PBD code: 2FLZ). Shown in spheres in both is the catalytic amino-terminal proline.

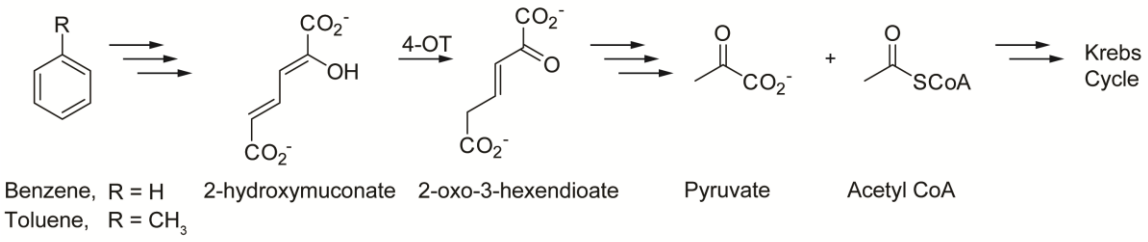
Several enzymes in the TSF show a high degree of catalytic promiscuity, most notably 4-oxalocrotonate tautomerase (4-OT). These observations make the tautomerase superfamily a good model system to study the mechanism of enzyme evolution, specifically the features defining enzyme specificity and catalysis. Tautomerase superfamily members have been studied extensively. The work in this dissertation focuses on three members of the TSF.

1.3 Overview of the Tautomerase Superfamily

The tautomerase superfamily (TSF) is a collection of structurally homologous enzymes that are characterized by two features, a β–α–β protein fold and a catalytic amino-terminal proline.[18,19] There are five known subfamilies in the superfamily, which are titled for the first characterized member: 4-oxalocrotonate tautomerase (4-OT), 5-(carboxymethyl)-2-hydroxymuconate isomerase (CHMI), macrophage migration inhibitory factor (MIF), *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD), and malonate semialdehyde decarboxylase (MSAD).[19]

4-OT, from *Pseudomonas putida* mt-2, was first identified as a part of the *meta*-fission pathway, which is responsible for the degradation of simple aromatic hydrocarbons.[19] In this pathway, benzene or toluene is first converted to catechol or a catechol derivative by a series of enzymatic reactions. Subsequently, catechol or the catechol-derivative can be processed through the *meta*-fission pathway (Scheme 1.1).[19] The end products of the pathway are pyruvate and acetaldehyde. The latter is converted to acetyl-CoA. Both products can be shuttled into the Kerbs cycle.

Scheme 1.1: *meta*-Fission Pathway for Degradation of Aromatic Hydrocarbons



4-OT catalyzes the 1,5-keto-enol tautomerization of 2-hydroxymuconate to 2-oxo-3-hexendioate.[18] 4-OT consists of monomers made up of 62 amino acids in length and exists as a hexamer (Figure 1.4A). The amino-terminal proline, which is critical for efficient catalysis, has an unusually low pK_a value of 6.4.[18,19] In solution, the secondary

amine of proline has a pK_a of about 10.6. This low pK_a value in 4-OT allows it to function as a catalytic base (under physiological conditions at pH 7.3). The lone pair on Pro1 of 4-OT is proposed to help shuttle a proton in the 1-5 keto-enol tautomerization reaction.[18,19]

In addition to 4-OT's biological tautomerization reaction, the enzyme shows a high degree of catalytic promiscuity. It has been shown that 4-OT catalyzes at least 5 other reactions including hydration that can result in dehalogenation[19], decarboxylation, enamine formation,[20] isomerization,[21] and aldol condensation[20]. It has also been shown that single point mutations in the active site can improve the efficiency of these reactions. Most of the known reactions catalyzed by 4-OT appear to be a result of the unique pK_a of the amino-terminal proline.[20]

Another enzyme in the TSF, CHMI participates in the degradation of aromatic amino acids as part of the homoprocatechuate pathway in *Escherichia coli* strain C.[18] CHMI consists of monomers that are 125 amino acids in length and exists as a trimer (Figure 1.4B). CHMI functions similar to 4-OT, but uses 5-(carboxymethyl)-2-hydroxymuconate as its substrate. MIF is a pro-inflammatory cytokine where each monomer is made up 115 amino acids and is a trimer. MIF is unique among the known superfamily members because it only contains the catalytic Pro1 and no other conserved active site residues.[19] MIF catalyzes the tautomerization of phenylenolpyruvate to phenylpyruvate, but this activity has no known biological relevance.[19] Pro1 in MIF and CHMI act as a general base catalyst in these reactions.[19]

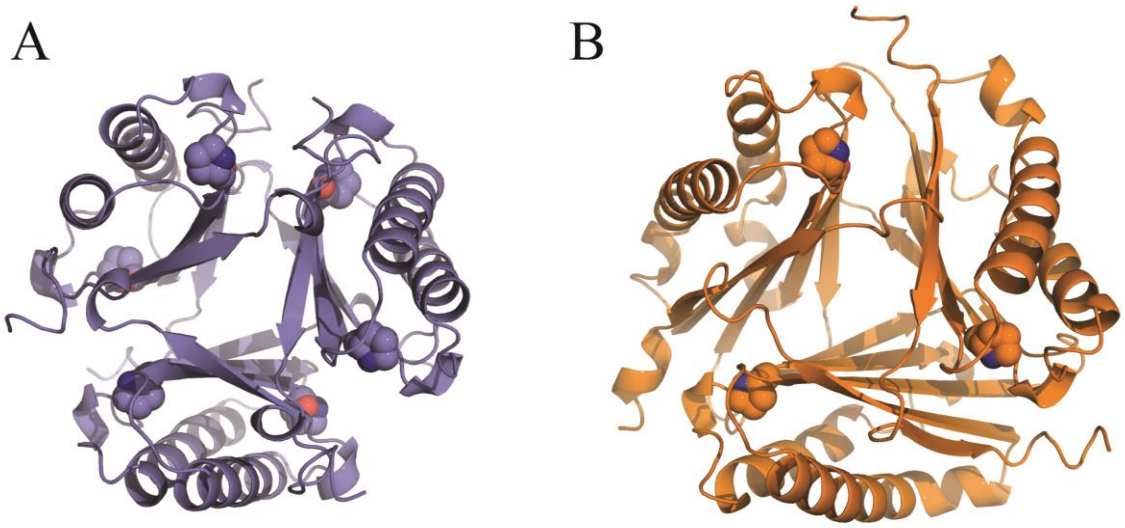


Figure 1.4: 4-OT and CHMI Oligomers

A) The 4-OT hexamer. Shown in spheres are the six amino-terminal prolines. B) The CHMI trimer. Shown in spheres are the three amino-terminal proline residues.

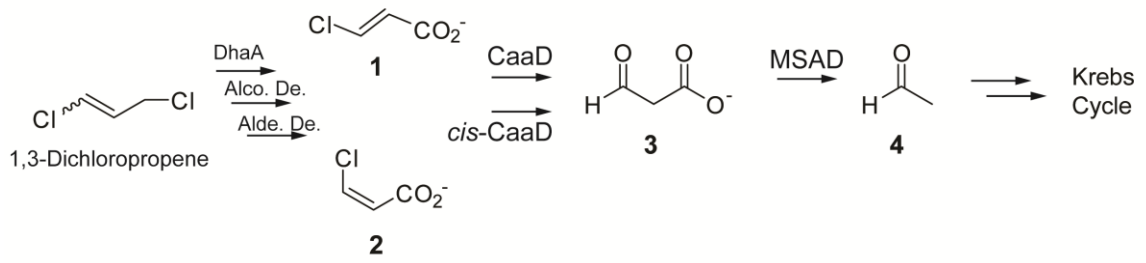
The amino-terminal proline can also act as a general acid catalyst. Three enzymes that are part of the 1,3-dichloropropene catabolic pathway, CaaD, *cis*-CaaD (Figure 1.3B), and MSAD, all use Pro1 as a general acid catalyst. These three enzymes and a homologue, designated Cg10062, and the reactions they catalyze will be the focus of the work presented in this dissertation, as they represent model systems to better understand enzyme specificity, functionality, promiscuity, and the roles they play in the evolution of enzymes.

1.4 1,3-Dichloropropene Catabolic Pathway

1,3-Dichloropropene is a synthetic compound introduced into the environment in the 1940's as the active ingredient of Shell D-D® and Telone II®.[22] These products were

marketed as nematicides used to protect crops from nematodes. During the late 1960's and early 1970's, soil bacteria were identified that degraded 1,3-dichloropropene.[22] In 1998, Janssen *et al* identified and characterized a specific pathway for its degradation (Scheme 1.2) in *Pseudomonas cichorii* 170.[22]

Scheme 1.2: 1,3- Dichloropropene Catabolic Pathway



In the first step, 1,3-dichloropropene is converted to 3-chloroallyl alcohol (not shown) by a hydrolytic dehalogenase, known as DhaA, which catalyzes the dehalogenation reaction using a covalent intermediate between the enzyme and substrate.[22] An active site aspartate acts as a nucleophile and attacks the substrate causing the displacement of a halide ion.[23] Water hydrolyzes the covalent intermediate to yield free enzyme and the haloalcohol. DhaA is structurally unrelated to the two other dehalogenases in the 1,3-dichloropropene pathway.

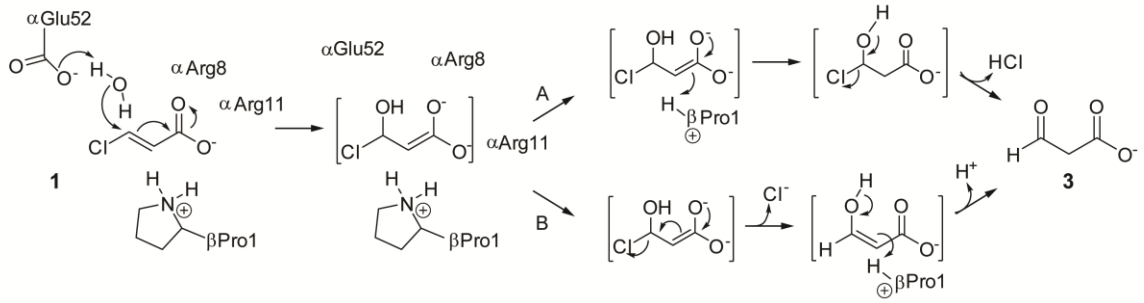
After the formation of the haloalcohol, an alcohol dehydrogenase and an aldehyde dehydrogenase oxidize the alcohol moiety to yield *trans*- and *cis*-3-chloroacrylates (**1** and **2**, respectively).[22] The enzymes, *trans*-3-chloroacrylic acid dehalogenase and *cis*-3-chloroacrylic acid dehalogenase process their respective isomers to yield malonate semialdehyde (**3**).[22,24,25] Malonate semialdehyde decarboxylase processes **3** to yield acetaldehyde (**4**).[26] Subsequently, **4** can be shuttled into the Krebs Cycle (although the exact mechanism is not known).

1.5 *trans*-3-CHLOROACRYLIC ACID DEHALOGENASE

trans-3-Chloroacrylic acid dehalogenase (CaaD) catalyzes the hydration of **1** to **3**. CaaD is a heterohexamer composed of three α subunits and three β subunits.[18,24] One can consider CaaD as trimer of heterodimers. Sequence analysis, mutational analysis, and the relatively small subunits (70 and 75 amino acids, respectively) of CaaD placed it into the 4-OT family of the tautomerase superfamily.[24] Both the α and β subunits contain an N-terminal proline. However, only the βPro1 is catalytically active.[24]

A single active site of CaaD is comprised of residues from one α subunit and one β subunit with secondary residues from a second, adjacent α subunit. In addition to βPro1, CaaD contains three other catalytically crucial residues: two arginine residues, αArg8 and αArg11, and a glutamate residue, αGlu52. Direct ^{15}N NMR titration studies of CaaD using ^{15}N-labeled enzyme and pH rate profiles showed that the pK_a of the βPro1 is 9.2, a relatively normal pK_a for the secondary amine, but 3 orders of magnitude higher than that of 4-OT (6.4).[27] This led to the conclusion that the N-terminal βPro1 of CaaD acts as a general acid catalyst as it would be mostly cationic at biological pH. The results of kinetic, mutagenesis, mass spectrometry, and X-ray crystallography studies suggested the current catalytic mechanism for CaaD, shown in Scheme 1.3.[24,27]

Scheme 1.3: Proposed Mechanism for CaaD using *trans*-3-Chloroacylate

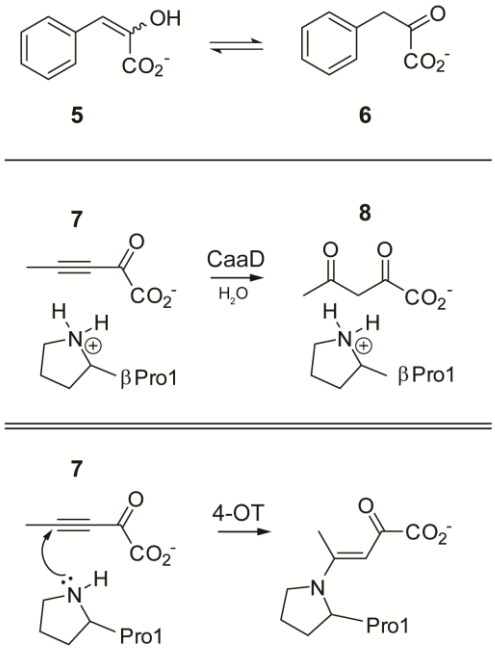


The two arginines help bind and position the substrate, **1** via its carboxylate moiety. The glutamate residue deprotonates the active site water molecule and the water molecule undergoes a Michael addition to **1** at the C3 position to form an enolate intermediate. Formation of this intermediate and stabilization is facilitated by the two arginine residues. Two plausible pathways can form the product. In path A, the proline donates a hydrogen atom to the C2 position forming an unstable halohydrin, which can then directly expel the halide to form **3**. In path B, the intermediate undergoes an α,β elimination, losing the halide and forming the enol form of **3**. The enol is then tautomerized to form **3**. CaaD also catalyzes the ketonization of phenylenolpyruvate (**5**) to phenylpyruvate (**6**) .[28] It can be inferred from this observation that CaaD likely follows path B. Therefore, path B is preferred over path A in the mechanism for CaaD.

Like 4-OT, CaaD also shows promiscuity. It catalyzes the tautomerization of phenylenolpyruvate, but it does not ketonize 2-hydroxymuconate.[18,29] It will also catalyze the hydration of 2-oxo-3-pentynoate (2-OP, **7**) to form acetopyruvate (**8**). Conversely, **7** completely inactivates 4-OT by forming a covalent adduct with Pro1 (Scheme 1.4).[18,29] This dichotomy of reactions with **7** reflects the difference in the pK_a values for the Pro1 residues of CaaD and 4-OT.

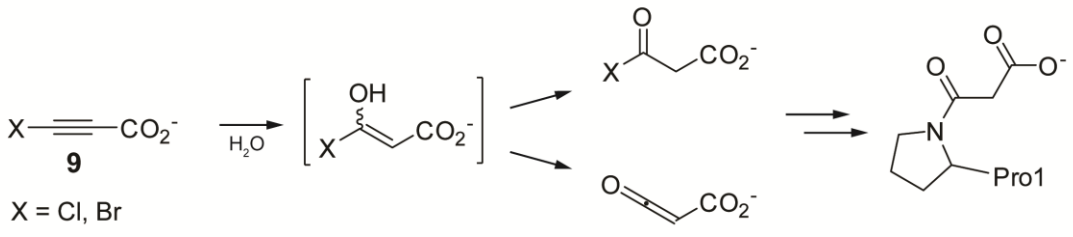
The higher pK_a of Pro1 in CaaD makes it cationic at physiological pH and allows for the hydration of **7**.[19] On the contrary, the lower pK_a of Pro1 in 4-OT allows the Pro1 to function as a nucleophile and attack C4 of **7**, causing it to become covalently bonded to the enzyme.[19] These pK_a differences in Pro1 may, in part, explain why CaaD performs a hydration reaction instead of a tautomerization reaction. This also illustrates the ability of **7** to be used as a mechanistic probe for the pK_a of Pro1 in the TSF. From these results one can conclude that although their structures are similar, CaaD and 4-OT catalyze very different reactions.

Scheme 1.4: Reactions of 2-oxo-3-Pentynoate (7) with CaaD and 4-OT



Another set of acetylene compounds, 3-chloro- and 3-bromopropiolate (**9**), were designed to be used as mechanistic probes of hydration activity.[29] The halopropiolates covalently and irreversibly modify CaaD and similar enzymes in the family at the Pro1 position like **7** does with 4-OT. The mechanism of inhibition is not known definitely, but the likely modes of inhibition are outlined in Scheme 1.5.[29]

Scheme 1.5: Proposed Mechanism of Inhibition by 3-Halopropiolate



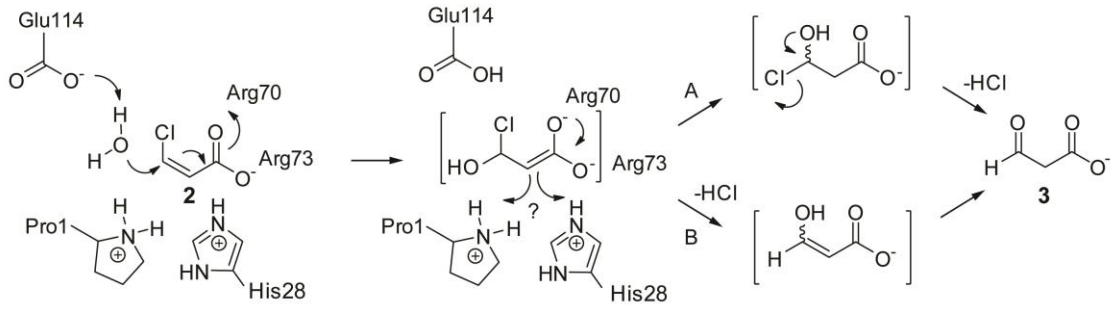
The halopropiolates are proposed to act like classic suicide inhibitors. The initial addition of water catalyzed by the enzyme generates two possible highly reactive species that can irreversibly add to βPro1. The reactive species is unknown. A crystal structure with the covalently modified βPro-1 residue of CaaD was solved to 2.3 Å. This structure was used in the docking studies presented in Chapter 2.

1.6 *cis*-3-CHLOROACRYLIC ACID DEHALOGENASE

The TSF member, *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD), also participates in the degradation of 1,3-dichloropropene.[22,25] Like CaaD, *cis*-CaaD catalyzes the hydration of **2** to yield **3**. However, unlike CaaD, *cis*-CaaD is a homotrimer composed of three 149 amino acid monomers.[25] The active site of *cis*-CaaD has the same critical residues found in CaaD: Pro1, Arg70, Arg73, and Glu114. *cis*-CaaD contains two two critical residues that are not found in CaaD: His28 and Tyr103.[25] A single *cis*-CaaD monomer contains most of the critical residues that compose a single active site (5 out of 6). However, Tyr103 is located on a β-sheet from the adjacent monomer.[30] Based on structural and mutagenesis studies, the proposed mechanism is the same as that proposed for CaaD (Scheme 1.6). The source of the proton that is donated at C2 is less clear as two potential proton donors are available (Pro1 and His28).

The additional residues are proposed to act as follows. His28 might help the two Arg residues (Arg70 and Arg73) in substrate binding, polarization, and enolate stabilization. Tyr103 is positioned to assist Glu114 in water activation. *cis*-CaaD also shows catalytic promiscuity using **5** and **7**.[28,31] However, *cis*-CaaD processes these substrates significantly less efficiently than CaaD. *cis*-CaaD is also covalently modified by **9** at Pro1.[25]

Scheme 1.6: Proposed Mechanism for *cis*-CaaD using *cis*-3-Chloroacrylate (2)



To uncover more details about the mechanism of *cis*-CaaD, the reaction of *cis*-CaaD with *cis*-3-bromoacrylate (**10**) was subjected to a thorough steady-state and pre-steady-state kinetic analysis.[32,33] *cis*-CaaD processes **10** and **2** with similar steady-state parameters, but the released bromide ion is more accurately quantified than the chloride ion.[32] *cis*-CaaD contains a tryptophan residue, Trp-101, at the back of the active site pocket.[30] This tryptophan residue enables the reactions of *cis*-CaaD to be monitored using stopped-flow fluorescence by following the native enzyme fluorescence at 340 nm.[33] Stopped-flow methods provide accurate data with minimal background about the reaction at very early time points during the reaction. The data gathered in this pre-steady-state phase of the reaction can provide information about individual rate constants for each kinetically significant step of the reaction.[33] Similarly, rapid-quench experiments can provide information about individual rates constants and potentially, the rate-limiting step in catalysis.[33,34] In contrast, a steady-state kinetic analysis provides only k_{cat} and K_m, which are a function of the individual rate constants and only give vague information about the overall reaction.[33] Global data fitting by simulation allows all of the data to be simultaneously fit to a single kinetic model.[35,36] Fitting data in this manner defines the individual rate constants for each observable step without any simplifying

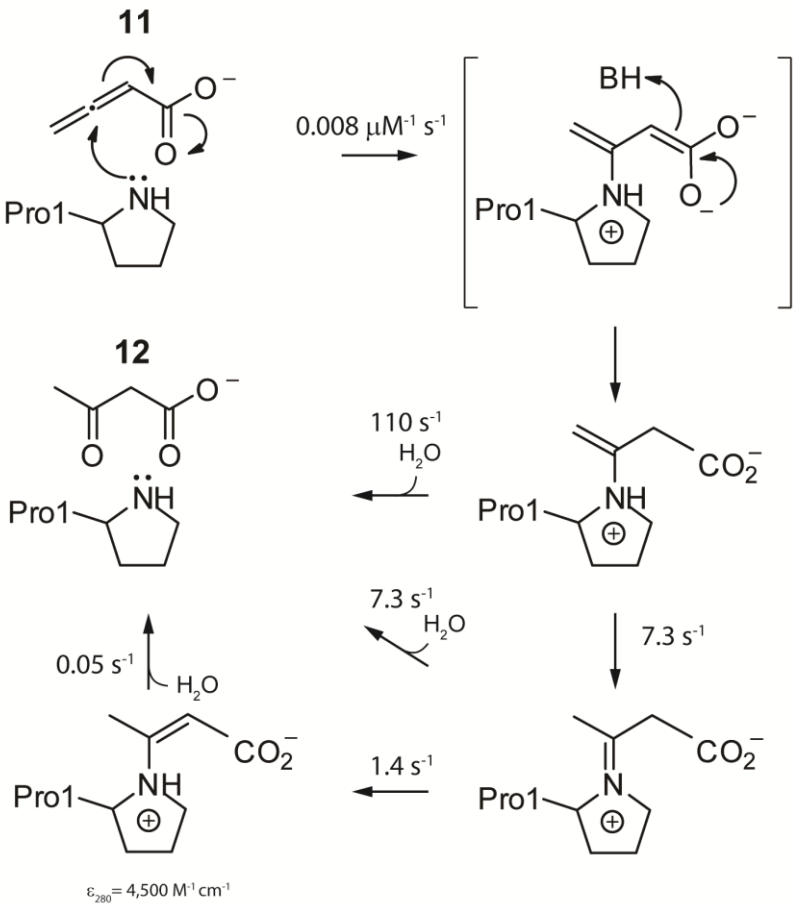
assumptions.[35-37] A more detailed discussion of data fitting by simulation can be found in Chapter 3. On the basis of these experiments (and data fitting), a kinetic model for the catalysis of **10** by *cis*-CaaD was proposed and published.[32] However, the proposed model fails to account for data acquired after publication. Simple refinement of the old model could not account for the new data. This led to an investigation of alternative substrates with *cis*-CaaD, which might provide a simpler model and possibly lead to a correct model (i.e., accounts for all of the data) for *cis*-CaaD with **10**.

In the course of the investigation of *cis*-CaaD with the alternative substrate, 2,3-butadienoate (**11**), an allene compound, a previously unknown reaction mechanism for *cis*-CaaD was uncovered.[38] Like **10**, *cis*-CaaD adds water to **11** to yield acetoacetate (**12**). The steady-state parameters with **11** are similar to those of **10**.[38] However, the stopped-flow fluorescence traces show a rate that is at least 20-fold less than k_{cat} as measured by steady-state kinetic experiments. It is not possible to account for this data with an in-line kinetic model. However, a kinetic model that uses multiple pathways to yield the formation of a single product can account for both the observed rate of turnover (k_{cat}) and the slow formation of the intermediate.

Along with the stopped-flow experiments, trapping studies were carried out with **11** and *cis*-CaaD.[38] These studies showed that *cis*-CaaD is covalently modified by **11** exclusively at the Pro1 position when the reducing agent, sodium cyanoborohydride, is added. Moreover, turnover studies showed that *cis*-CaaD is covalently modified by **11** (in the presence of $NaCNBH_3$) with 1.3 equivalents per active site.[38] This result suggested that *cis*-CaaD processes **11** almost exclusively through a covalent catalytic mechanism. The covalent enamine intermediate was also directly observed by UV spectroscopy after deconvolution of time-dependent wavelength spectra using singular value decomposition (Scheme 1.7).[38] With these results in hand, a kinetic model was

proposed and used to globally fit the full set of kinetic data. This resulted in the estimates of six individual rate constants, along with a fluorescence factor for the enzyme and an extinction coefficient for a secondary enamine.[38] The data resulting from this study led to the proposal of a different type of mechanism by which *cis*-CaaD processes **11** to **12** (Scheme 1.7). These studies also uncovered mechanistic details of *cis*-CaaD, which were previously unknown. We believe this covalent catalysis mechanism might occur during the catalytic cycle of *cis*-CaaD using **10**.

Scheme 1.7: Proposed Mechanism for *cis*-CaaD using 2,3-Butadienoate (11)

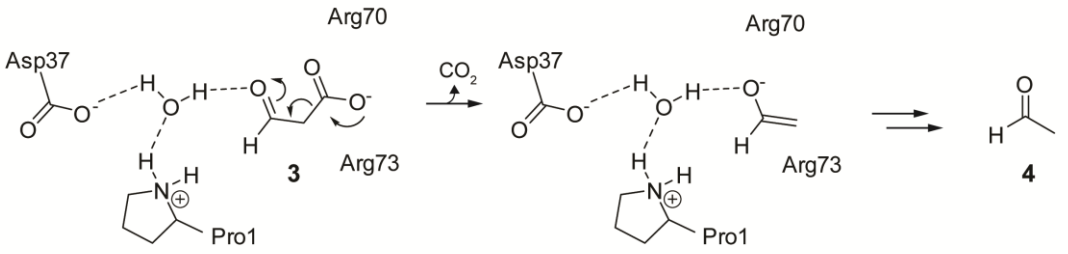


Despite having very similar active site architectures, *cis*-CaaD and CaaD can only dehalogenate their respective isomers. To date, there is no known compound that the enzymes process equivalently, except for **9**, which results in covalent modification and inhibition. A fascinating question is, "how do CaaD and *cis*-CaaD achieve such exquisite specificity?" Following the kinetic approach outlined above, the kinetic mechanism of CaaD was investigated and the results are reported in Chapters 2 and 3.

1.7 MALONATE SEMIALDEHYDE DECARBOXYLASE

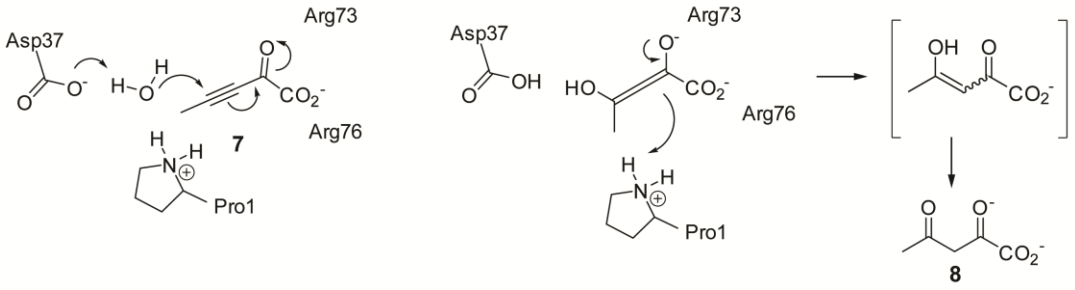
Malonate semialdehyde decarboxylase (MSAD) is the last enzyme in the 1,3 dichloropropene catabolic pathway.[22] MSAD catalyzes the decarboxylation of **3** to **4**, where **4** can be shuttled into the Krebs cycle. MSAD is a homotrimer where each monomer is composed of 129 amino acids.[26] In this way, it is similar to *cis*-CaaD. MSAD contains the amino-terminal proline and the two arginine residues (Arg70 and Arg 73). Like CaaD and *cis*-CaaD, but unlike many other decarboxylases, MSAD uses no metals or cofactors to catalyze the decarboxylation reaction. A ^{15}N NMR titration of MSAD vs. pH showed that the pK_a of Pro1 is 9.2, similar to that of CaaD.[39] MSAD catalyzes the hydration of **7**, although slowly (as assessed by k_{cat}/K_m).[39] The purpose of this hydration activity is unknown. It could be a promiscuous activity carried over from a progenitor enzyme or it could be functional. One possibility is that the hydration activity is a rescue function. It would allow MSAD to remove covalent adducts formed on Pro1 by reaction with the aldehyde moieties in the substrate or product. Based on all of these observations and the crystal structure, a mechanism for the decarboxylation of **3** was proposed (Scheme 1.8).[40]

Scheme 1.8: Proposed Mechanism for Decarboxylation by MSAD



The crystal structure, along with mutagenesis studies, showed that Asp37 is critical for both the decarboxylation and the hydration activities.[40] Asp37 might act with Pro1 to coordinate an active site water molecule. This may help facilitate the decarboxylation of **3** to form **4** and could help stabilize the enolate intermediate. Asp37 might also be responsible for maintaining the pK_a value of Pro1. For the hydration activity, Asp37 could deprotonate the water molecule allowing it to add (in a Michael addition) to the acetylene moiety of **7** (Scheme 1.9). The difference between these two proposed functions of Asp37 could stem from a difference in binding of **3** and **7**, as **7** is two carbons longer than **3** and has limited flexibility.

Scheme 1.9: Proposed Hydration of 2-oxo-3-Pentynoate (7) by MSAD



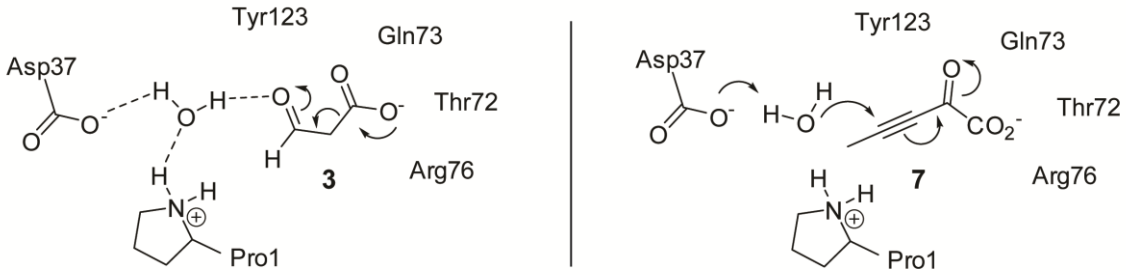
One major difference between the active site architecture of MSAD and the dehalogenases (*cis*-CaaD or CaaD) is the presence of a large hydrophobic wall in MSAD. The hydrophobic wall is located at the back, left side of the active site, opposite the two

arginines, which are located at the back, right side of the active site. In the dehalogenase enzymes, this region of the active site is occupied by the water activating residues (e.g., Glu52 in CaaD and Glu114 and Tyr103 in *cis*-CaaD). The hydrophobic wall is made up of 5 residues: Tyr39, Trp114, Phe116, Phe124, and Leu128. The role of this large hydrophobic wall is unclear, but it is conserved across other MSADs. The substrate for MSAD, **3,** exists as a mixture of the free aldehyde and hydrate. The equilibrium of the mixture is about 73% hydrate and 27% free aldehyde and is maintained at a nearly diffusion-controlled rate.[41] It is not known if the hydrate binds to the active site or MSAD only catalyzes the free aldehyde, as shown in Scheme 1.8. One role for the hydrophobic wall might be to dehydrate the hydrate, thereby generating the free aldehyde at the active site for decarboxylation. Another interesting feature about MSAD is that it does not show saturation kinetics with **3** (up to 20 mM). It is unclear saturation is not able to be achieved with **3**, but the hydrate could play a role.

A MSAD homologue, designated FG41 MSAD, was cloned from the *Coryneform bacterium* strain FG41. FG41 MSAD is a homotrimer where each monomer is 136 amino acids. The homologue was annotated as a decarboxylase due to the high sequence identity (~37%) with MSAD. In fact, FG41 MSAD decarboxylates **3** at a rate comparable to that of the *P. pavonaceae* MSAD.[42] However, the FG41 MSAD has a significantly lower hydration activity using **7**.[42] Interestingly, FG41 MSAD has one of the critical arginine residues (Arg76), while the other one is replaced by Gln73.

Once the crystal structure was solved, two other residues, Thr72 and Tyr123 were identified as possible substitutes for the missing arginine residue.[42] Subsequent mutagenesis studies confirmed the importance of these residues in maintaining the decarboxylation activity. Roles for these residues were tentatively assigned, as shown in Scheme 1.10.

Scheme 1.10: Decarboxylation and Hydration Reactions of FG41 MSAD



Thr72 and Tyr123, along with Gln73, are proposed to form multiple hydrogen bond interactions with **3** to help facilitate decarboxylation.[42] When Gln73 is changed to Arg73 in FG41 MSAD, the decarboxylation activity doesn't improve significantly. However, the hydration activity of this mutant using **7** increases 9-fold over that of wild-type FG41 MSAD (as assessed by k_{cat}/K_m).[42] There are at least three other characterized MSAD homologues. Some of these homologues, such as YrdN from *B. subtilis* strain 168 (53% sequence identity to MSAD), only contain a single arginine residue at position 76 or the equivalent position. This observation suggests that this second arginine residue might not be necessary for decarboxylation in other MSAD homologues, but the significance of this observation remains unclear. This set of MSADs illustrates how a single reaction can be catalyzed using different groups in a similar reaction strategy.

1.8 Cg10062 FROM *CORYNEBACTERIUM GLUTAMICUM*

A BLASTP search using the *cis*-CaaD amino acid sequence as the query identified a gene labeled *cg10062* from *Corynebacterium glutamicum.* Unlike the enzymes of the 1,3-dichloropropene catabolic pathway, the *cg10062* gene has no clear genomic context. A sequence alignment with Cg10062 and *cis*-CaaD shows 53% sequence similarity and 34% sequence identity.[43] Interestingly, Cg10062 has the same

crucial residues as *cis*-CaaD: Pro1, His28, Arg73, Arg75, Try103, and Glu114. A crystal structure of the homotrimeric Cg10062 shows an active site architecture that is nearly superimposable on that of *cis*-CaaD.[43] A pH rate profile suggested that the pK_a of Pro1 is about 9.2, similar to those of *cis*-CaaD and CaaD.[44] Cg10062 catalyzes the hydration of **7**, although poorly (with a similar k_{cat}/K_m to that of *cis*-CaaD, 50 M^{-1} s^{-1} versus 10 M^{-1} s^{-1}, respectively)[43] Cg10062 is also covalently modified at Pro1 by **9**. These observations are consistent with a high pK_a value of Pro1 and that Cg10062 is, in fact, a hydratase.[43] Due to the high sequence similarities with *cis*-CaaD, Cg10062 was placed in the *cis*-CaaD subfamily.

However, an examination of Cg10062's activity with *cis*-3-chloro- (**2**) and *cis*-3-bromoacrylate (**10**), it was found that Cg10062 is a poor *cis*-CaaD.[43] It shows about a 1,000-fold reduction in specificity (k_{cat}/K_m) with **10** than *cis*-CaaD, primarily due to the high K_m value (>50 mM versus the 160 μM for *cis*-CaaD).[43,45] More intriguing, Cg10062 dehalogenates *trans*-3-chloroacrylate (**1**), albeit poorly. However, *cis*-CaaD and all known *cis*-CaaD mutants do not process **1**. This is intriguing because the active sites appear nearly identical.

Another pair of compounds provided more evidence that more significant differences exist between the *cis*-CaaD and Cg10062 than the active site crystal structure and sequences might suggest. The enantiomeric epioxide compounds (*R*)-oxirane-2-carboxylate and (*S*)-oxirane-2-carboxylate were studied as potential covalent inhibitors of *cis*-CaaD and Cg10062. *cis*-CaaD is irreversibly inactivated by only the (*R*)-oxirane-2-carboxylate.[44] However, Cg10062 is inactivated by both epoxide compounds with a preference for the (*R*) isomer.[44] These observations parallel the observations made for the 3-haloacrylates. The mechanism by which these compounds inactivate these enzymes is not entirely clear. Cg10062's lack of specificity with any substrate examined thus far is

intriguing and has implications for the exquisite substrate specificity exhibited by *cis*-CaaD.

One possible source of the specificity differences between *cis*-CaaD and Cg10062 was identified while comparing the crystal structures of the modified and unmodified structures of both enzymes. A loop near the active site is observed in different positions in *cis*-CaaD and Cg10062 (Figure 1.5A and 1.5B, respectively).[45] In the non-liganded structure of *cis*-CaaD, the loop is in an "open" conformation (Figure 1.5A, light green). In the covalently modified structure (dark green), the loop appears in a "closed" conformation. However, in Cg10062, the loop is observed in the "open" conformation in both the non-liganded and modified structures (Figure 1.5B).

These observations led to a hypothesis and subsequent investigation into the loop residues. Two of these residues in *cis*-CaaD (Thr32 and Thr34) were individually mutated to their counterparts in Cg10062 (Ala32 and Ala34). These individual mutations do have a significant effect on the K_m values using **10**.[45] The K_m values raise approximately 20-fold and 4-fold, respectively. However, changing Gln35 in *cis*-CaaD to Pro35, dropped the observed K_m value ~10-fold (suggesting tighter binding). Also, changing Ala32 and Ala34 in Cg10062 to Thr residues does not result in a similar improvement in the K_m values. This led to the conclusion that if the loop is playing a role in specificity, there are probably multiple factors leading to the observed changes in K_m values and no one single residue is responsible.[45]

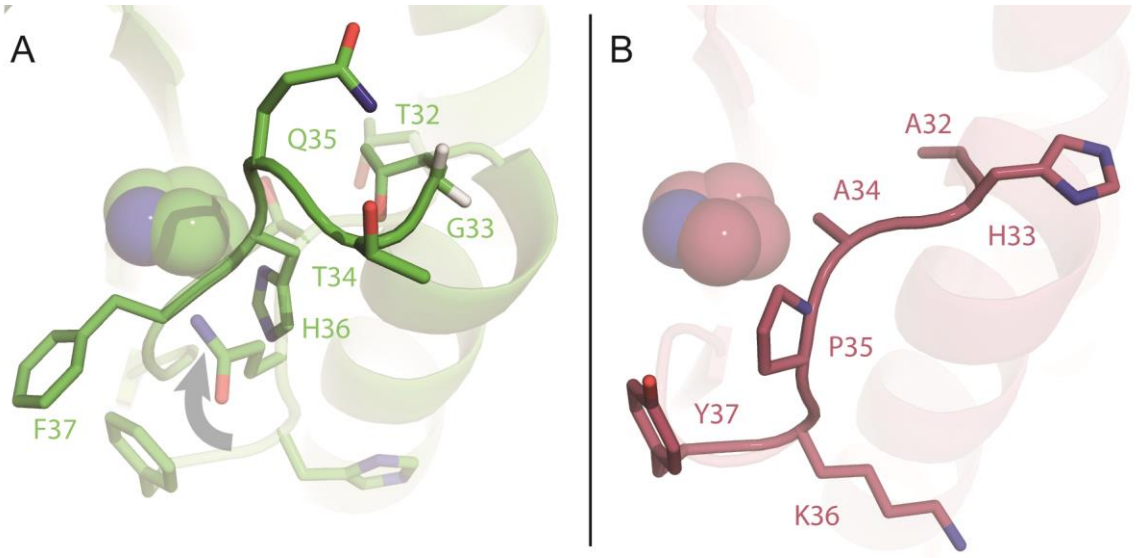


Figure 1.5: Loop Residues in *cis*-CaaD and Cg10062

A) Shown in dark green are the residues that make up the loop of *cis*-CaaD in the modified crystal structure. In the light green are the loop residues in the non-liganded structure. The gray arrow indicates the anticipated direction of movement from the "open" to the "closed" form of the enzyme (PDB codes: 3MF7 (dark green) and 3MF8 (light green). B) Shown in maroon are the residues of the loop in Cg10062 (PDB code: 3N4G). Cg10062 does not show this same movement upon modification of the proline residue. Shown in the spheres are the proline residues of both *cis*-CaaD and Cg10062.

In the course of the investigation of the roles of the loop residues, the allene substrate, **11**, was examined with Cg10062. It was found to be the best known substrate for Cg10062 (as assessed by k_{cat}/K_m). The k_{cat}/K_m value is comparable to that of *cis*-CaaD (5.1×10^3 versus 8.7×10^3 M^{-1} s^{-1} for Cg10062 and *cis*-CaaD, respectively)[45] More

importantly, the K_m values are the same, within experimental error (690 ± 30 μM for *cis*-CaaD and 780 ± 130 μM for Cg10062).[45] However, *cis*-CaaD processes **11** though a covalent intermediate using Pro1. This was confirmed by trapping the substrate on the enzyme with $NaCNBH_3$. When the same experiment was carried out with **11** and Cg10062, the enzyme is not inactivated and no covalently modified enzyme could be detected.[45] Cg10062 processes approximately 80 mM of **11** in the presence of either sodium borohydride or sodium cyanoborohydride. On this basis, it was concluded that Cg10062 does not use covalent catalysis like *cis*-CaaD, but rather does a direct hydration on **11**.

Subsequently, the T34A mutant of *cis*-CaaD was examined to see if it processed **11** through a covalent intermediate like wild-type *cis*-CaaD.[45] It was found that the T34A mutant of *cis*-CaaD goes through a covalent intermediate about 1 in every 5 turnovers of the enzyme. In other words, about 80% of the time, the T34A mutant carries out a direct hydration in contrast to the almost exclusive formation of the covalent intermediate in wild-type *cis*-CaaD. This observation suggests that the T34A mutant of *cis*-CaaD is more Cg10062-like in the mechanism it uses to process **11**.

Another unique feature of **11** is its ability to report the stereochemistry of the proton addition atom at C-2.[38,45]. *cis*-CaaD and Cg10062 were shown to add a proton stereospecifically at C2 on opposite faces of **11**.[45] The T34A mutant of *cis*-CaaD retains the same stereochemistry as wild-type *cis*-CaaD, even though it catalyzes the reaction by a different mechanism.[45] The implications of these results remain unclear, but again they highlight differences between the two enzymes (*cis*-CaaD and Cg10062), which were expected to be nearly identical (based on sequence and structural features).

1.9 SUMMARY

Enzymes are essential for life. Moreover, understanding how enzymes function is essential for advancing human health. Not only are enzymes the targets for many pharmaceuticals, but the ability to harness the power of enzymes has other benefits including bioremediation and synthesis of pharmaceuticals and other important chemicals. This process begins with an understanding of the exact features that define the specificity of an enzyme and how it catalyzes its reaction.

If we revisit the "lock and key" model of an enzyme proposed by Emil Fischer, it becomes obvious that this model is a drastic oversimplification. The rigidity suggested by the "lock and key" model is unrealistic. However, the "lock and key" model does provoke questions about the evolution of the "lock". By classifying enzymes into superfamilies and characterizing them, one gains a better understanding of how each individual enzyme functions and what is similar and different about them. In doing so, one gains a better understanding about the evolution of enzymes. The tautomerase superfamily is a model system to understand enzyme function and evolution. Within the TSF, the enzymes of the 1,3-dicholopropene catabolic pathway and their homologues show various catalytic strategies on a variety of different substrates while using very similar architectures. They provide a good model system to study divergent evolution and the features that define specificity and catalysis.

Specifically, the features that dictate the specificities of CaaD and *cis*-CaaD are unclear. The work presented in Chapter 2 and 3 focus on our mechanistic understanding of the dehalogenation reaction catalyzed by CaaD. This sets the stage for future work which may be able to elucidate the distinguishing features that define specificities in the two dehalogenases.

A *cis*-CaaD homologue, Cg10062 has shown little specificity with several substrates tested. Also, its lack of significant genomic context raises questions about its role in the organism. In Chapter 4, a substrate with high specificity is identified for Cg10062. Upon further investigation with this substrate and other acetylene compounds, it was determined that Cg10062 catalyzes a previously unknown reaction. These results begin to provide answers to some of the questions about Cg10062.

1.10 REFERENCES

(1) Shanmugam, S. *Enzyme Technology*; I.K. International Publishing House Pvt. Limited: New Delhi, 2009.

(2) Salque, M.; Bogucki, P. I.; Pyzel, J.; Sobkowiak-Tabaka, I.; Grygiel, R.; Szmyt, M.; Evershed, R. P. (2013) Earliest evidence for cheese making in the sixth millennium bc in northern Europe. *Nature. 493*, 522-525.

(3) Schulenburg, C.; Miller, B. G. (2014) Enzyme recruitment and its role in metabolic expansion. *Biochemistry. 53*, 836-845.

(4) Darwin, C. *Origin of Species: By Means of Nature Selection*; Cambridge University Press: Cambridge, 1876.

(5) Peregrin-Alvarez, J.; Sanford, C.; Parkinson, J. (2009) The conservation and evolutionary modularity of metabolism. *Genome Biol. 10*, R63.61-R63.17.

(6) Dellus-Gur, E.; Toth-Petroczy, A.; Elias, M.; Tawfik, D. S. (2013) What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J. Mol. Biol. 425*, 2609-2621.

(7) Neidhart, D. J.; Kenyon, G. L.; Gerlt, J. A.; Petsko, G. A. (1990) Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous. *Nature. 347*, 692-694.

(8) Petsko, G. A.; Kenyon, G. L.; Gerlt, J. A.; Ringe, D.; Kozarich, J. W. (1993) On the origin of enzymatic species. *Trends Biochem. Sci. 18*, 372-376.

(9) Khersonsky, O.; Tawfik, D. S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem. 79*, 471-505.

(10) Jensen, R. A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol. 30*, 409-425.

(11) Horowitz, N. H. (1945) On the evolution of biochemical syntheses. *Proc. Natl. Acad. Sci. USA. 31*, 153-157.

(12) Fondi, M.; Emiliani, G.; Fani, R. (2009) Origin and evolution of operons and metabolic pathways. *Res. Microbiol. 160*, 502-512.

(13) Babbitt, P. C.; Gerlt, J. A. (1997) Understanding enzyme superfamilies. *J. Biol. Chem. 272*, 30591-30594.

(14) Gerlt, J. A.; Babbitt, P. C. (1998) Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol. 2*, 607-612.

(15) Gerlt, J. A.; Babbitt, P. C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem. 70*, 209-246.

(16) Gerlt, J. A.; Babbitt, P. C.; Rayment, I. (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Arch. Biochem. Biophys. 433*, 59-70.

(17) Glasner, M. E.; Gerlt, J. A.; Babbitt, P. C. (2006) Evolution of enzyme superfamilies. *Curr. Opin. Chem. Biol. 10*, 492-497.

(18) Whitman, C. P. (2002) The 4-oxalocrotonate tautomerase family of enzymes: how Nature makes new enzymes using a β−α−β structural motif. *Arch. Biochem. Biophys. 402*, 1-13.

(19) Poelarends, G. J.; Veetil, V. P.; Whitman, C. P. (2008) The chemical versatility of the β−α−β fold: catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell. Mol. Life Sci. 65*, 3606-3618.

(20) Zandvoort, E.; Baas, B. J.; Quax, W. J.; Poelarends, G. J. (2011) Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: enamine formation and aldolase activity. *ChemBioChem. 12*, 602-609.

(21) Zandvoort, E.; Geertsema, E. M.; Baas, B. J.; Quax, W. J.; Poelarends, G. J. (2012) An unexpected promiscuous activity of 4-oxalocrotonate tautomerase: the *cis-trans* isomerisation of nitrostyrene. *ChemBioChem. 13*, 1869-1873.

(22) Poelarends, G. J.; Wilkens, M.; Larkin, M. J.; Van Elsas, J. D.; Janssen, D. B. (1998) Degradation of 1,3-dichloropropene by *Pseudomonas cichorii* 170. *Appl. Environ. Microbiol. 64*, 2931-2936.

(23) Janssen, D. B. (2004) Evolving haloalkane dehalogenases. *Curr. Opin. Chem. Biol. 8*, 150-159.

(24) Poelarends, G. J.; Saunier, R.; Janssen, D. B. (2001) *trans*-3-Chloroacrylic acid dehalogenase from *Pseudomonas pavonaceae* 170 shares structural and mechanistic similarities with 4-oxalocrotonate tautomerase. *J. Bacteriol. 183*, 4269-4277.

(25) Poelarends, G. J.; Serrano, H.; Person, M. D.; Johnson, W. H., Jr.; Murzin, A. G.; Whitman, C. P. (2004) Cloning, expression, and characterization of a *cis*-3-chloroacrylic

acid dehalogenase: insights into the mechanistic, structural, and evolutionary relationship between isomer-specific 3-chloroacrylic acid dehalogenases. *Biochemistry. 43*, 759-772.

(26) Poelarends, G. J., Johnson, H. W., Jr., Murzin, A.G., and Whitman, C.P. (2003) Mechanistic characterization of a bacterial malonate semialdehyde decarboxylase: identification of a new activity in the tautomerase superfamily. *J. Biol. Chem. 278*, 48674-48683.

(27) Azurmendi, H. F.; Wang, S. C.; Massiah, M. A.; Poelarends, G. J.; Whitman, C. P.; Mildvan, A. S. (2004) The roles of active-site residues in the catalytic mechanism of *trans*-3-chloroacrylic acid dehalogenase: a kinetic, NMR, and mutational analysis. *Biochemistry. 43*, 4082-4091.

(28) Poelarends, G. J.; Johnson, W. H., Jr.; Serrano, H.; Whitman, C. P. (2007) Phenylpyruvate tautomerase activity of *trans*-3-chloroacrylic acid dehalogenase: evidence for an enol intermediate in the dehalogenase reaction. *Biochemistry. 46*, 9596-9604.

(29) Wang, S. C.; Person, M. D.; Johnson, W. H.; Whitman, C. P. (2003) Reactions of *trans*-3-chloroacrylic acid dehalogenase with acetylene substrates: consequences of and evidence for a hydration reaction. *Biochemistry. 42*, 8762-8773.

(30) De Jong, R. M.; Bazzacco, P.; Poelarends, G. J.; Johnson, W. H., Jr.; Kim, Y. J.; Burks, E. A.; Serrano, H.; Thunnissen, A.-M. W. H.; Whitman, C. P.; Dijkstra, B. W. (2007) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase. Structural basis for substrate specificity and inactivation by *(R)*-oxirane-2-carboxylate. *J. Biol. Chem. 282*, 2440-2449.

(31) Poelarends, G. J.; Serrano, H.; Johnson, W. H., Jr.; Whitman, C. P. (2004) Stereospecific alkylation of *cis*-3-chloroacrylic acid dehalogenase by *(R)*-oxirane-2-carboxylate: analysis and mechanistic implications. *Biochemistry. 43*, 7187-7196.

(32) Robertson, B. A.; Schroeder, G. K.; Jin, Z.; Johnson, K. A.; Whitman, C. P. (2009) Pre-steady-state kinetic analysis of *cis*-3-chloroacrylic acid dehalogenase: analysis and implications. *Biochemistry. 48*, 11737-11744.

(33) Johnson, K. A. (1992) Transient-state kinetic analysis of enzyme reaction pathways In *The Enzymes* (Sigman, D. S., Ed.) 3rd Ed. pp 1-61, Academic Press, San Diego

(34) Johnson, K. A. (1983) The pathway of ATP hydrolysis by dynein. Kinetics of a presteady state phosphate burst. *J. Biol. Chem. 258*, 13825-13832.

(35) Johnson, K. A.; Simpson, Z. B.; Blom, T. (2009) Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem. 387*, 20-29.

(36) Johnson, K. A. (2009) Fitting enzyme kinetic data with KinTek global kinetic explorer In *Methods Enzymol.* (Michael, L. J., Ludwig, B., Eds.) pp 601-626, Academic Press,

(37) Johnson, K. A.; Simpson, Z. B.; Blom, T. (2009) FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal. Biochem. 387*, 30-41.

(38) Schroeder, G. K.; Johnson, W. H., Jr.; Huddleston, J. P.; Serrano, H.; Johnson, K. A.; Whitman, C. P. (2012) Reaction of *cis*-3-chloroacrylic acid dehalogenase with an allene substrate, 2,3-butadienoate: hydration via an enamine. *J. Am. Chem. Soc. 134*, 293-304.

(39) Poelarends, G. J.; Serrano, H.; Johnson, W. H., Jr.; Hoffman, D. W.; Whitman, C. P. (2004) The hydratase activity of malonate semialdehyde decarboxylase: mechanistic and evolutionary implications. *J. Am. Chem. Soc. 126*, 15658-15659.

(40) Almrud, J. J.; Poelarends, G. J.; Johnson, W. H., Jr.; Serrano, H.; Hackert, M. L.; Whitman, C. P. (2005) Crystal structures of the wild-type, P1A mutant, and inactivated malonate semialdehyde decarboxylase: a structural basis for the decarboxylase and hydratase activities. *Biochemistry. 44*, 14818-14827.

(41) Huddleston, J. P.; Schroeder, G. K.; Johnson, K. A.; Whitman, C. P. (2012) A pre-steady state kinetic analysis of the αY60W mutant of *trans*-3-chloroacrylic acid dehalogenase: implications for the mechanism of the wild-type enzyme. *Biochemistry. 51*, 9420-9435.

(42) Guo, Y.; Serrano, H.; Poelarends, G. J.; Johnson, W. H., Jr.; Hackert, M. L.; Whitman, C. P. (2013) Kinetic, mutational, and structural analysis of malonate semialdehyde decarboxylase from *Coryneform bacterium* strain FG41: mechanistic implications for the decarboxylase and hydratase activities. *Biochemistry. 52*, 4830-4841.

(43) Poelarends, G. J.; Serrano, H.; Person, M. D.; Johnson, W. H.; Whitman, C. P. (2008) Characterization of Cg10062 from *Corynebacterium glutamicum*: implications for the evolution of *cis*-3-chloroacrylic acid dehalogenase activity in the tautomerase superfamily. *Biochemistry. 47*, 8139-8147.

(44) Robertson, B. A.; Johnson, W. H.; Lo, H. H., Jr.; Whitman, C. P. (2008) Inactivation of Cg10062, a *cis*-3-chloroacrylic acid dehalogenase homologue in *Corynebacterium glutamicum*, by *(R)*- and *(S)*-oxirane-2-carboxylate: analysis and implications. *Biochemistry. 47*, 8796-8803.

(45) Schroeder, G. K.; Huddleston, J. P.; Johnson, W. H., Jr.; Whitman, C. P. (2013) A mutational analysis of the active site loop residues in *cis*-3-chloroacrylic acid dehalogenase. *Biochemistry. 52*, 4204-4216.

Chapter 2: Design of a Fluorescent Mutant of *trans*-3-Chloroacrylic Acid Dehalogenase

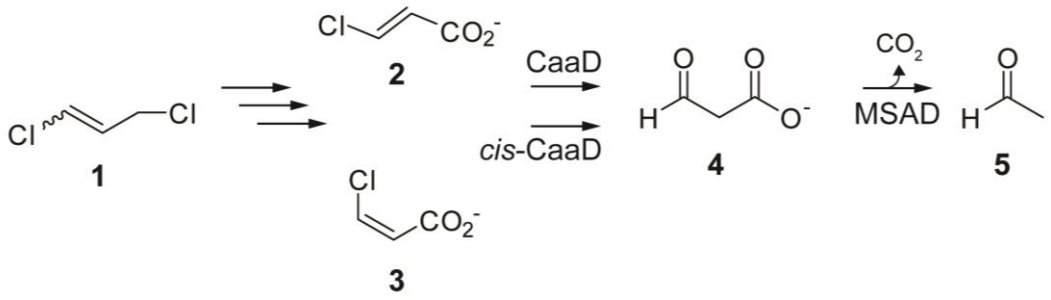
2.1 INTRODUCTION

A mixture of *cis*- and *trans*-1,3-dichloropropene (**1**, Scheme 2.1) constitutes the active ingredient in nematicides marketed as Telone II® and Shell D-D®.[1] The effectiveness of **1** in soil is limited by its rapid degradation due, in part, to the bacterial 1,3-dichloropropene catabolic pathway.[1-3] In three enzyme-catalyzed steps, the isomeric mixture is processed to the *trans*- and *cis*-isomers of 3-chloroacrylic acid (**2** and **3,** respectively).[2,4] Subsequently, these compounds are processed by the isomer-specific dehalogenases *trans*- and *cis*-3-chloroacrylic acid dehalogenase (CaaD and *cis*-CaaD, respectively) to malonate semialdehyde (**4**).[4-6] Malonate semialdehyde decarboxylase (MSAD) converts **4** to acetaldehyde (**5**), which is channeled into the Krebs cycle.[7] The three enzymes are members of the tautomerase superfamily, a group of structurally homologous enzymes characterized by a β–α–β building block and a catalytic amino-terminal proline (Pro-1).[8,9] These three enzymes provoke interesting mechanistic, structural, and evolutionary questions including how they became part of the same pathway. Hence, their properties are being intensely scrutinized.[9]

Adapted with permission from Huddleston et. al. *Biochemistry. 51 (46),* 9420-9435.

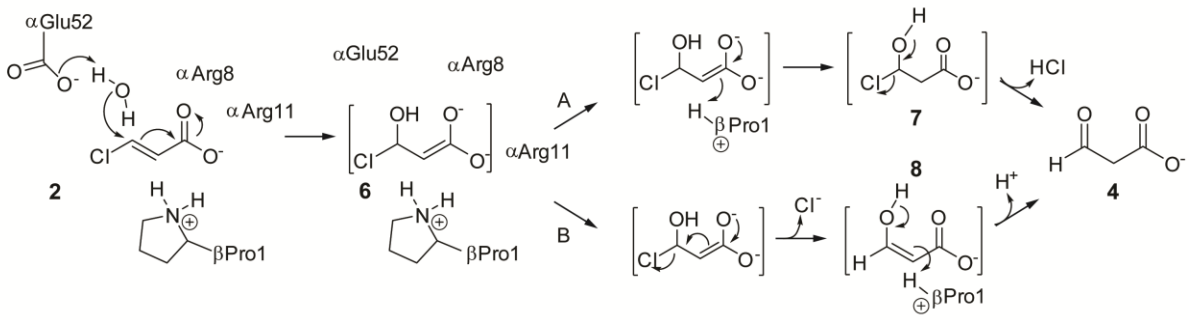
Dr. Schroeder helped with data analysis and experimental design. Dr. Johnson provided instrumentation, software, and data analysis expertise. Dr. Whitman provided helpful discussions, editing, and funding.

Scheme 2.1: 1,3-Dichloropropene Catabolic Pathway



Sequence analysis coupled with mechanistic and structural studies identified Pro-1, from the β-subunit, and Arg-8, Arg-11, and Glu-52, from the α–subunit, as key catalytic residues in the heterohexameric CaaD.[4,5,10,11] In the current working hypothesis for the mechanism (Scheme 2.2), αGlu-52 functions as a general base catalyst to activate a water molecule for attack at C-3 of **2**.[10,11] The two arginines (αArg-8 and αArg-11) interact with the C-1 carboxylate group to align the substrate in the active site and draw electron density away from C-3, which facilitates the addition of water to C-3 to produce the enediolate species, **6**. In one scenario (Scheme 2.2, route A), βPro-1 (pK_a ~9.2) protonates C-2 to complete the conjugate addition of water and form a chlorohydrin intermediate (**7**). The chemical or enzyme-catalyzed decay of **7** then yields **4**. In a second scenario (Scheme 2.2, route B), **6** rearranges to enol **8**, which is coupled with the elimination of the chloride (i.e., an α,β-elimination of HCl), and subsequent ketonization (chemical or enzyme-catalyzed) yields **4,** where the C-2 proton is again provided by βPro-1.

Scheme 2.2: Catalytic Mechanism of CaaD using *trans*-3-Chloroacrylic Acid (2)



Given the similarities in the overall active site organization, it was initially thought that the mechanisms of CaaD and *cis*-CaaD were largely the same, with differences in the primary (70 and 75 amino acids for the α- and β-subunits of CaaD vs 149 amino acids in *cis*-CaaD) and quaternary structures (heterohexamer vs trimer), and the corresponding isomer specificities.[6] However, in light of additional results, it became apparent that there are at least three major differences in the catalytic mechanisms. First, two extra residues, His-28 and Tyr-103, are involved in the dehalogenation of the *cis*-isomer.[12] The positions of these residues in a crystal structure suggest that His-28 aligns with the arginine pair and Tyr-103 assists a glutamate residue (Glu-114) in the activation of water. There are no obvious counterparts for these residues in CaaD. Second, a loop may be involved in *cis*-CaaD catalysis, where substrate binding induces loop closure, facilitating catalysis.[13] A similar event has not been discovered in the catalytic cycle of CaaD. Third, there is very recent evidence that *cis*-CaaD utilizes a covalent intermediate (via Pro-1) to process an alternate substrate, 2,3-butadienoate, and that some fraction of the reaction using *cis*-3-haloacrylates might also function by covalent catalysis.[14]

Potential roles for the loop and the covalent intermediate in the *cis*-CaaD mechanism were suggested in part from pre-steady-state kinetic experiments.[13,14] The experiments reported here were part of a larger effort to delineate similarities and difference between CaaD, *cis*-CaaD, and their mutants, and to establish a basis for their catalytic differences. In this chapter, we use *trans*-3-bromoacrylate (**9**) as the substrate because the release of bromide can be accurately quantified, as shown in our previous studies on *cis*-CaaD.[13] The stopped-flow fluorescence experiments on *cis*-CaaD relied on a signal provided by a single tryptophan residue (Trp-101) in the active site cavity.[13] CaaD does not have a significant native fluorophore: it only contains 2 tyrosine residues (αTyr-9 and αTyr-60), one of which (αTyr-60) is located near the active site.[10] Hence, a tryptophan residue was introduced to provide a stronger signal for stopped-flow fluorescence experiments. Here, we show how *in silico* docking studies guided our selection of a point mutation to introduce a tryptophan residue to provide a strong fluorescence signal to monitor enzyme states during catalysis. Mutation at this position had a minimal effect on the observable CaaD reaction, as assessed by the steady-state kinetic parameters, rapid quench experiments, and bromide inhibition studies. This mutant sets the stage for more in-depth kinetic studies to be carried out in an effort to gain more information about the features that distinguish CaaD from other family members.

2.2 MATERIALS AND METHODS

Materials. Chemicals, biochemicals, buffers, and solvents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO), Fisher Scientific Inc. (Pittsburgh, PA), Fluka Chemical Corp. (Milwaukee, WI) or EMD Chemicals, Inc (Gibbstown, NJ). The reagents used in the ion chromatography (IC) and rapid quench experiments were acquired from sources reported elsewhere.[13,14] The centrifugal filter devices (3,000 MW cutoff) were obtained from PALL Life Sciences (Ann Arbor, MI). Column resins were obtained from Sigma-Aldrich.

Bacterial Strains, Plasmids, and Growth Conditions. *Escherichia coli* strain BL21-Gold(DE3) was obtained from Stratagene (La Jolla, CA). The *E. coli* DH5α cells were obtained from Invitrogen (Carlsbad, CA). The construction of the pET-24a(+) vector (EMD Chemicals, Inc.) containing *caad1* and *caad2* (the α- and β-subunits of CaaD, respectively) is described elsewhere.[4,5,15] Cells were grown at 37 °C in Luria-Bertani (LB) media that contained kanamycin (Kn, 30 μg/mL).

General Methods. The PCR amplification of DNA sequences was conducted in a GeneAmp 2700 thermocycler (Applied Biosystems, Carlsbad, CA). Techniques for restriction enzyme digestion, ligation, transformation, and other standard molecular biology manipulations were based on methods described elsewhere.[16] DNA sequencing was performed by the DNA Core Facility in the Institute for Cellular and Molecular Biology (ICMB) at the University of Texas at Austin. Mass spectrometer (MS) data were collected on an LCQ electrospray ion-trap mass spectrometer (Thermo, San Jose,

CA) housed in the ICMB Protein and Metabolite Analysis Facility at the University of Texas. Steady-state kinetic assays were performed on an Agilent 8453 diode-array spectrophotometer at 22 °C.[5] Non-linear regression data analysis was performed using the program Grafit (Erithacus Software Ltd., Staines, U.K.). Protein concentrations were determined according to the method of Waddell.[17] Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on denaturing gels containing 15% or 20% polyacrylamide.[18]

2.2.1 Docking Studies

In order to identify and optimize positions for a tryptophan residue in the CaaD active site, docking studies were carried out using PyMOL with Autodock Vina.[19,20] To minimize differences in the active sites of the different CaaD crystal structures, 21 active sites in four reported crystal structures were compared (PDB codes: 1S0Y1, 3EJ3, 3EJ7, 3EJ9). [Each heterohexamer contains 3 active sites. One crystal structure (PDB code 3EJ9) contains a single heterohexamer, whereas those for PDB codes 1S0Y, 3EJ3, and 3EJ7 contain two heterohexamers. This gives a total of 21 active sites.] The crystal structures include those with covalent (malonic acid, 1S0Y) and non-covalent ligands (acetate and phosphate, 3EJ3), and those without any ligands (3EJ7 and 3EJ9). The side chains of αGlu-52 and αLeu-57 adopt different rotomers depending on the bound ligand limiting the available active site space. The active site with the covalently bound malonyl adduct was chosen as the receptor for docking studies (Figure 2.1A). The covalent adduct on the prolyl nitrogen of βPro-1 results from the reaction of CaaD with

3-bromo or 3-chloropropiolate, as described elsewhere.[5,10] The adduct was removed before docking experiments were performed (Figure 2.1B). A 10 Å × 15 Å × 10 Å box centered on the βPro-1 residue was selected as the origin of docking. The side chains of αGlu-52 and αLeu-57 were designated as flexible during docking routines.

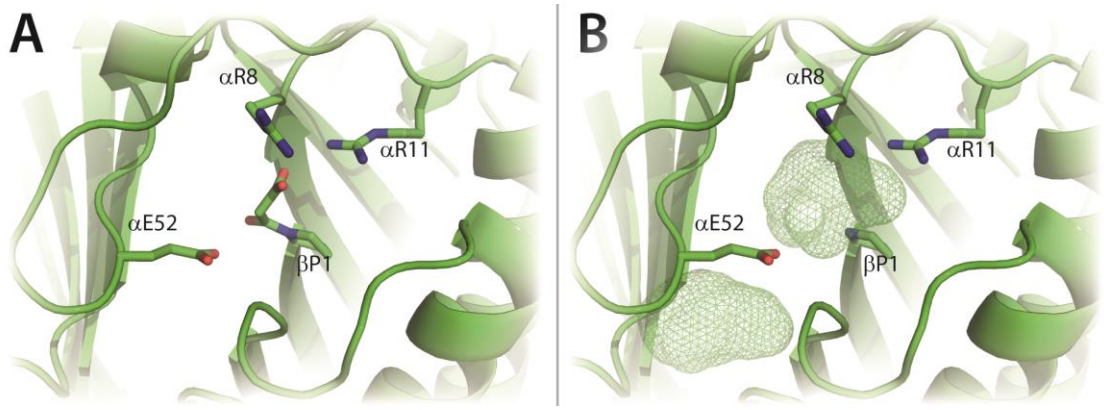


Figure 2.1: Active Site of CaaD

Two views of the active site of CaaD. A) The key catalytic residues of CaaD shown as sticks with βPro-1 covalently attached to malonate after reaction with 3-bromopropiolate.[10] B) The approximate active site volume as predicted in Pymol is shown in mesh after the malonate moiety has been removed from the active site. The mesh area below αE52 shows unoccupied space near the active site. The figure was prepared using PyMOL.[20]

2.2.2 Construction of the CaaD Mutants

Four CaaD mutants were constructed using a pET-24a(+) vector containing both *caad1* and *caad2* genes as the template. Mutations were introduced at the appropriate position using the QuikChange Site-Directed Mutagenesis Kit from Agilent (Santa Clara, CA) following the manufacturer's instructions. Briefly, the amplification mixtures (50 µL) contained 10X QuikChange buffer, 200 µM of each dNTP, 250 nM of each primer, 25-50 ng of template DNA, and 2.5 U of Pfu turbo Hot Start polymerase. Typically, the cycling parameters were 95°C for 2 min followed by 18 cycles of 95°C for 30 s, 55°C for 30 s and 68°C for 6 min with a final elongation step of 68°C for 10 min. The annealing temperature could be varied between 55°C and 75°C, which is approximately 5°C below the standard melting temperature of the primers. The extension time could also vary between 6 and 12 min. Oligonucleotide primers (coding and complementary) with the desired change to make the αY60W, αM7W, αL57W, and βI37W CaaD mutants were obtained from Sigma-Aldrich. The forward and reverse primers used to introduce the mutations were respectively: 5'-CTGCCGGACTGGGTGCCAG-3' and 5'-CTGGCACCCAGTCCGGCAG-3' (αY60W); 5'-CTTGCGACTGGCGCTATGGGAG-3' and 5'-CTCCCATAGCGCCAGTCGCAAG-3' (αM7W); 5'-CGAGCATTGGCCGGACTAC-3' and 5'-GTAGTCCGGCCAATGCTCG-3' (αL57W); and 5'-CGACCCCAAGTGGATCAATG-3' and 5'-CATTGATCCACTTGGGGTCG-3' (βI37W). After amplification , 1 U of the restriction enzyme, Dpn1, was added reaction mixture and allowed to incubate at 37°C for 2 hrs. Dpn1 digests methylated DNA

thereby removing most of the template from the reaction mixture. The DNA products were purified using the MiniElute PCR Purification Kit (Qiagen, Valencia, Ca.) and transformed into *E. coli* DH5α cells for plasmid preparation. Single colonies were used to inoculate 20 mL of LB/Kn media. The cultures were grown overnight at 37 °C and plasmids were isolated using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich).

2.2.3 Expression and Purification of CaaD and CaaD Mutants

CaaD was grown and expressed in *E. coli* BL21-Gold(DE3) cells and purified by a variation of published protocols.[5,15] An overnight starter culture (40 mL) was grown at 37 °C from a single colony and used to inoculate five 2 L Erlenmeyer flasks (each containing 500 mL of LB/Kn media). The cultures were incubated for ~2 h at 37 °C. Protein expression was induced with isopropyl β-D-thiogalactopyranoside (1 mM, final concentration), followed by a 4-5 h induction period at 37 °C. Cells were harvested by centrifugation (4,000 × g), to yield ~7 g of cells and stored at -80 °C. Cells were suspended in 10 mM Na_2HPO_4 buffer, pH 8 (Buffer A) and lysed by sonication. The resulting solution was centrifuged for 30 min (23,500 × g) and the supernatant was subjected to an additional 30 min of centrifugation (330,000 × g). Solid $(NH_4)_2SO_4$ (for a final concentration of 1.6 M) was slowly added (over 10 min) to the supernatant and the resulting solution was stirred on ice for 1 h, and then centrifuged for 15 min (23,500 × g). The clarified supernatant was filtered (0.2 μM pore) and loaded onto a Phenyl-Sepharose 6 Fast Flow column (0.5 × 10 cm, ~4 mL resin) pre-equilibrated with Buffer A containing 1.6 M $(NH_4)_2SO_4$ (Buffer B). The column was washed (~1 mL/min) with 5

mL of Buffer B followed by a linear salt gradient [1.6-0 M $(NH_4)_2SO_4$]. Protein typically elutes 45-60 min after initiation of the salt gradient. Protein purity in the individual fractions was evaluated by SDS-PAGE and activity assays. Fractions with the highest activity and purity were pooled and exchanged into Buffer C (50 mM Na_2CO_3 buffer, pH 10.5) using an Amicon stirred-cell concentrator (10,000 MW cutoff). Exchanged protein was then loaded onto a Q-Sepharose 6 Fast Flow column (~7 mL of resin) pre-equilibrated with Buffer C and washed (~1 mL/min) with Buffer C (5 mL), followed by elution using a linear salt gradient (0-0.1 M Na_2SO_4). Protein typically elutes 15-45 min after the start of the salt gradient. The appropriate fractions, assessed as described above, were pooled. The combined fractions were exchanged into 100 mM Na_2HPO_4 buffer, pH 8, using an Amicon stirred-cell concentrator (10,000 MW cutoff) and concentrated to ~16 mg/mL (~1 mM). Yields were typically ~75 mg of homogenous protein from 2.5 L of culture.

2.2.4 Mass Spectral Analysis and CaaD and CaaD Mutants

The monomer molecular masses of CaaD and the CaaD mutants were determined by electrospray ionization mass spectrometry (ESI-MS) using LCQ electrospray ion-trap mass spectrometer. Samples for ESI-MS analysis were made 1 mg/mL and prepared as reported.[5] The observed molecular mass (MH^+) for the α- and β-subunits of CaaD are 8343 Da and 7505 Da, respectively. The α-subunits of Y60W, M7W, and L57W-CaaD give molecular masses of 8367, 8398, 8416 Da, respectively. The β-subunit of I37W-

CaaD shows a molecular mass of 7578 Da. These masses agree with the calculated masses (within experimental error).

2.2.5 Steady-State Kinetics

The steady-state kinetic parameters for CaaD and the four CaaD mutants were measured using *trans*-3-bromoacrylate (**9**).[5] The assays were carried out at 22 °C in 100 mM Na_2HPO_4 buffer, pH 8.1, using a 2 μM enzyme solution (based on the α-β dimer molecular mass). Enzyme solutions were equilibrated for 1 h prior to use. The assay was initiated by the addition of **9** from a 20 mM stock solution (made up in 100 mM phosphate buffer, final pH 8.1). The decrease in absorbance at 224 nm, corresponding to the hydration of **9** (ε_{224}= 9100 M^{-1} cm^{-1}), was monitored over a 60 s time period, with readings recorded every 3 s. Initial rates were plotted vs. substrate concentration and fit to the Michaelis-Menten equation using Grafit to determine k_{cat} and K_m. Data for full time course experiments (300 s) were collected using five concentrations of **9** (20, 50, 80, 100, and 150 μM).

2.2.6 Stopped-Flow Experiments

The stopped-flow experiments were carried out on a SF 2004 series stopped-flow apparatus (KinTek Corp). Wild-type, αY60W- and αM7W-CaaD, were separately made up in 100 mM Na_2HPO_4 buffer, pH 8.1, to final concentrations of 20 μM (based on the α-β dimer molecular mass), and allowed to equilibrate at 22 °C for 1 h. Various concentrations of **9** (0-20,000 μM) were made up in 100 mM Na_2HPO_4 buffer, pH 8.1. The enzyme (10 μM after mixing) and substrate (0-10,000 μM after mixing) solutions

were then mixed in the stopped-flow apparatus at 22 ºC. The fluorescence was excited at 280 nm and emission was observed using a 340 nm band pass filter (Semrock, Rochester, NY). The slit width on the monochromator and the light filter were set at 0.6 mm. Traces represent an average of at least five runs at each substrate concentration. The time courses range between 1-300 s with 1000 data points collected for each trace. Stopped-flow fluorescence traces were fit initially to defined exponential functions by non-linear regression using the KinTek stopped-flow software. Data were fit either to a single exponential (Eq. 2.1) or double exponential function (Eq 2.2).

$$Y = A \cdot \exp(-\lambda t) + C \qquad \text{(Eq. 2.1)}$$

$$Y = A_1 \cdot \exp(-\lambda_1 t) + A_2 \cdot \exp(-\lambda_2 t) + C \qquad \text{(Eq. 2.2)}$$

The concentration dependence of the rate was fit to a hyperbola (Eq. 2.3) using non-linear regression (Grafit).

$$\lambda = \frac{k_{for}[S]}{K_d + [S]} + k_{rev} \qquad \text{(Eq. 2.3)}$$

Light transmittance (280 nm) was also monitored in the stopped-flow apparatus with various concentrations of **9** in buffer.

2.2.7 Rapid-Quench Experiments

The rapid-quench experiments were carried out at 22 °C on a RFQ-3 quench-flow apparatus (KinTek Corp). The experimental procedure followed previously published ones with the modifications noted below.[13] To correct for dilution errors introduced by sample preparation and the rapid quench apparatus, tartrate was included as an internal standard. Tartrate is well-resolved from other peaks (e.g. bromide, phosphate, and sulfate), readily quantified by ion chromatography (IC), and does not inhibit CaaD at concentrations of 4 mM (data not shown). Typically, ~8 mM tartrate (4 mM final concentration) was added to a concentrated enzyme solution (400-1000 μM) in 100 mM Na_2HPO_4 buffer, pH 8.1, and allowed to equilibrate for 1 h. For each experiment, one syringe was loaded with enzyme solution (400-1000 μM) containing 8 mM tartrate and a second syringe was loaded with freshly prepared **9** (from a 20 mM stock solution, pH 8.1). The reaction mixture (~15 μL enzyme and ~15 μL substrate) was quenched at intervals ranging from 3 ms to 1.5 s (depending on the k_{cat} of the enzyme) with 0.6 M H_2SO_4 (~90 μL). The individually quenched reaction mixtures (~110 μL) were then transferred to a PALL centrifugal filter device (3,000 MW cutoff), and centrifuged at 11,000 rpm for 30 min to remove enzyme. The effluent was saved (~80 μL) and used for bromide and tartrate quantification. An aliquot of the effluent (~25 μL) was diluted into 475 μL of 20 mM $NaHCO_3$ buffer, pH 9.0, to give a final volume of 500 μL, pH ~3.0. To quantify the amount of tartrate added to the concentrated enzyme solutions, an unreacted enzyme control sample (~100 μL) was filtered (as above) to remove enzyme and 10 μL aliquots of effluent were added to 490 μL of 20 mM $NaHCO_3$ buffer, pH 9.0,

to give a final volume of 500 µL. These control samples were made up in triplicate and the tartrate concentration was quantified by IC analysis. A dilution factor for each sample was determined by dividing the final tartrate concentration by the average initial tartrate concentration added to the enzyme solution. The bromide ion concentration was determined using IC on an ICS-1500 instrument as previously described with the following modifications.[13] The contents of the mixture elute over a 18 min period. Bromide ion was detected by suppressed conductivity with an applied current of 25 mA and quantified as described.[13] A linear plot of the peak area vs. the concentration of bromide gives a slope of 0.018 ± 0.002 µS × min^{-1} × ppm^{-1} (correlation coefficient ~ 0.995), where µS is microsiemens and ppm is parts per million. The calibration curve for tartrate was constructed using a stock solution of 100 mM sodium tartrate in 20 mM $NaHCO_3$ buffer, pH 9.0. Seven serial dilutions (0.08-50 mM) were made from the 100 mM sodium tartrate solution stock. A linear plot of the peak area vs. the concentration gave a slope of 0.0020 ± 0.0001 µS × min × mM^{-1} (correlation coefficient ~ 0.995).

Data were fit to the burst equation (Eq. 2.4) to estimate the rate (λ) and amplitude (*A*) of the reaction followed by steady-state turnover (k_{ss}):

$$Y = A \cdot (1 - \exp(-\lambda t)) + k_{ss} \cdot t \qquad \text{(Eq. 2.4)}$$

In this equation, the fitted line was constrained to begin at the origin (*Y*=0 at *t*=0).

2.3 RESULTS

2.3.1 *In silico* Docking Experiments

In order to provide a larger signal for stopped-flow fluorescence experiments on CaaD, we introduced a tryptophan residue at a position near the active site. We evaluated several potential sites based upon minimal impact on the reaction kinetics and enzyme structure. Inspection of the available crystal structures of CaaD (1S0Y, 3EJ3, 3EJ7, 3EJ9) and *cis*-CaaD (2FLZ) combined with a series of docking studies, identified six potential sites for the insertion of the tryptophan (αPhe-39, αPhe-50, αLeu-57, βIle-37, αMet-7, αTyr-60). Docking studies were carried out *in silico* with wild-type CaaD and six tryptophan mutants (αF39W, αF50W, αM7W, αY60W, αL57W, and βI37W generated by *in silico* using the PyMOL mutagenesis wizard) and **9** (Figure 2.2A).[20] With the exception of the αF39W and αL57W *in silico* mutants, (the tryptophan side chain of the αL57W mutant was held fixed during the docking studies) the consensus binding orientation for **9**, as assessed by the lowest predicted free energies of binding (Table 2.1), places the carboxylate group near the two arginine residues and the bromide moiety buried near the back of the active site, facing a potential halide binding pocket consisting primarily of αPhe-39 and αPhe-50. Interestingly, in this consensus binding mode, the side chains of αGlu-52 and αLeu-57 adopt rotomers that increase the size of the active site pocket (Figure 2.2B). This increase in the active site volume (shown in mesh, Figure 2.1B vs. Figure 2.2B) allows **9** to fit in the active site.

Enzyme	ΔG^a (kcal/mol)	ΔG^b (kcal/mol)
Wild-type	-4.8	-3.5
αY60W	-5.1	-3.6
αM7W	-4.8	-
αL57W	-4.0	-
βI37W	-4.8	-
αF39W	-4.9	-
αF50W	-5.0	-

[a]The ΔG value for binding as predicted by AutoDock Vina for the binding of **9** to CaaD and mutants (Figure 2.2A).[19]

[b]The ΔG for the binding of **9** to CaaD in which the αE52 points into the active site such that it can carry out the proposed chemistry (Figure 2.2C).

Table 2.1. Summary of the Docking Study Results using *trans*-3-Bromoacrylate (**9)**

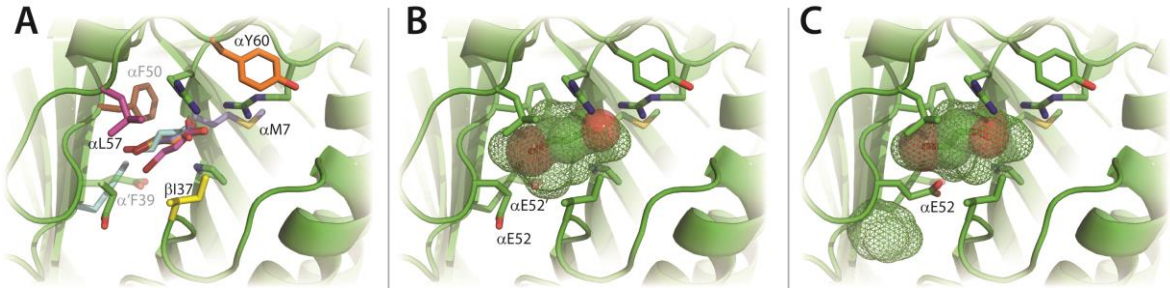


Figure 2.2: Substrate Docking Studies

A) Six docking poses of **9** in the CaaD active site having the lowest energies of binding (ΔG values can be found in Table 2.1 after each residue (shown in color) was independently replaced with a tryptophan (*in silico*). Four poses of **9** are superimposable, whereas the poses of **9** in the α'F39W (cyan) and αL57W (pink) enzymes are slightly different. All lowest energy poses move E52 (as found in cystal structure to solid sticks) as illustrated by the arrow. B) Docking of **9** (spheres) into wild-type CaaD.[19] Shown in mesh is the *in silico* active site pocket volume. The movement of αE52 to expand the active site volume is shown by the arrow. C) An alternative docking mode for **9** with αE52 in a position to perform the proposed chemistry with the newly predicted active site volume shown in mesh. The figure was prepared using PyMOL.[20]

At first glance, αPhe-39 and αPhe-50, seemed to be the most promising sites for a tryptophan substitution: αPhe-39 superimposes on Trp-101 in *cis*-CaaD (monitored in stopped-flow experiment on *cis*-CaaD,[13] and αPhe-50 aligns with a tryptophan in two other tautomerase superfamily members [i.e., a heterohexamer 4-oxalocrotonate tautomerase and a 4-OT homologue in the tomaymycin biosynthetic pathway designated TomN].[21,22] However, *in silico* mutagenesis of these residues indicates that the CaaD active site cannot accommodate the larger indole side chain of the tryptophan residue. In the αF39W *in silico* mutant, there are steric clashes with nearby residues and a significant volume reduction of the putative halide binding pocket. In the αF50W *in silico* mutant, all rotomers exhibit significant steric clashes including one with the β1 strand containing αArg-8. Moreover, docking studies using **2** and **9** show steric clashes with the halide moiety of each substrate. These results may partially explain the intractable kinetic data previously obtained with the αF39W mutant of CaaD using **9**.[23] Hence, the αF39W and αF50W positions were no longer considered as possible sites for tryptophan introduction.

The αY60W *in silico* mutant has a predicted energy of binding (with **9**) that is comparable to wild-type (–5.1 kcal/mol vs. –4.8 kcal/mol).[19] Moreover, docking conformations in which the side chain of αGlu-52 points *into* the active site (Figure 2.2C) such that it can carry out the proposed chemistry, (i.e. activation of a water molecule for attack at C-3), were observed for both the *in silico* αY60W mutant and wild-type CaaD (each at, -3.5 kcal/mol). A similar conformation for αGlu-52 is observed in the crystal structure of CaaD (1S0Y) containing the covalent malonyl adduct

on βPro-1, perhaps reflecting the position of αGlu-52 after the attack of water on **9**. Based upon our *in silico* analysis, four mutants, αM7W, αY60W, αL57W, and βI37W were constructed and their properties analyzed.

2.3.2 Steady-state Kinetic Parameters of CaaD and CaaD Mutants

Four CaaD mutants, αM7W, αY60W, αL57W, and βI37W, were constructed and purified. Mass spectral analysis shows nearly homogenous proteins (>95%) with the predicted subunit molecular masses (monomeric). The αL57W and βI37W mutants of CaaD showed no activity with **2** or **9**. Exploratory stopped-flow experiments carried out with the βI37W mutant and **9** suggest that βI37W-CaaD binds **9** very weakly with a K_d > 40 mM. Due to the similar locations of βI37 and αL57 in the active site, it appears that both mutants show difficulty with binding the substrate. One explanation for this observation is that the indole side chain of tryptophan is forming a "gate" that prevents **9** from entering the active site. No further experiments were carried out on these two mutants and their positions were no longer considered as possible sites for tryptophan introduction.

In contrast, the αM7W and αY60W mutants have steady-state kinetic parameters (with **9**) comparable to those of wild-type enzyme (Table 2.2). The K_m value for αM7W-CaaD was comparable to that of wild-type, whereas the k_{cat} decreased ~4-fold. As a result, there was a 4-fold decrease in k_{cat}/K_m compared with wild-type. The K_m value for

αY60W-CaaD decreased ~1.5-fold compared with wild-type, but the values of k_{cat} and k_{cat}/K_m were identical within experimental error.

Enzyme[a]	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
Wild-type	120 ± 10	2.5 ± 0.1	$1.3 \pm 0.1 \times 10^4$
αM7W	120 ± 10	0.6 ± 0.1	$0.3 \pm 0.1 \times 10^4$
αY60W	70 ± 3.5	2.5 ± 0.1	$1.2 \pm 0.1 \times 10^4$
βI37W	> 40,000[b]	-	-

[a]The kinetic parameters were measured in 100 mM Na_2HPO_4 buffer (pH 8.1) at 22 °C following a decrease in absorbance at 224 nm.[5] Errors are standard deviations.

[b]Determined in a stopped-flow experiment using **9**, as described in the text.

Table 2.2. Steady-State Kinetic Parameters for CaaD and Mutants using *trans*-3-Bromoacrylate (**9)**

Full progress curves (300 s) were also carried out with these two mutants following the disappearance of various concentrations of **9** at 224 nm. At all concentrations of **9**, the UV traces for αY60W-CaaD (shown in orange in Figure 2.3) mirror those of wild-type CaaD (shown in blue in Figure 2.3), reflecting the similar k_{cat} values. However, those corresponding to the αM7W mutant (shown in green in Figure 2.3) exhibit slower rates compared with wild-type, reflecting the reduced k_{cat} value.

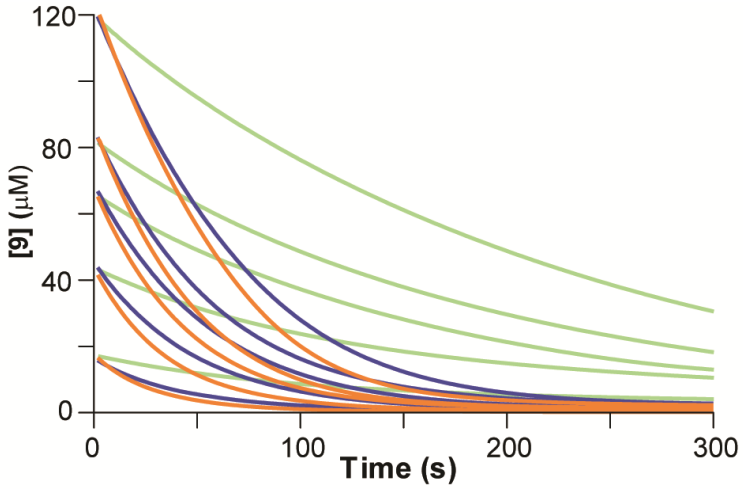


Figure 2.3: Full-Time Course Comparison of Wild-Type and Mutants of CaaD with *trans*-3-Bromoacrylate (**9**)

A) The UV traces following the disappearance of **9** (18, 42, 63, 83, and 125 μM) at 224 nm catalyzed by CaaD (blue), the αY60W mutant (orange), and the αM7W mutant (green). The full time courses for wild-type and the αY60W mutant are comparable. The full time course for the αM7W mutant is slower.

2.3.3 Pre-Steady-state Stopped-Flow Kinetic Experiments

The newly-created fluorescent CaaD mutants, αM7W and αY60W, were examined for their ability to report active site changes during the course of the enzymatic reaction. Changes in the enzyme fluorescence of both mutant proteins were monitored during the initial phases (12 s) and full time course (120 s) of the reaction with **9** (Figure 2.4). In the first second of the initial time course, the αY60W and αM7W mutants exhibited 2- and 3.5-fold decreases in fluorescence, respectively (Figures 2.4A and 2.4B). The full time course (120 s) shows that this decrease was followed by an increase in fluorescence, which eventually plateaus, as the reaction reaches completion (Figures 2.4C and 2.4D). Neither mutant shows significant photobleaching under the reaction conditions. Both mutants have similar initial reaction kinetics, which can be approximated by a single exponential, ignoring a possible lag at the beginning of the trace (The significance of this is discussed in Chapter 3). However, notable differences between the two mutants were observed in the full time course data (Figures 2.4C and 2.4D). The αY60W-CaaD mutant rapidly returns to a final fluorescence plateau, whereas the αM7W-CaaD showed a slower return to the final fluorescence value highlighting the slower catalytic turnover of this mutant (Table 2.2 and Figure 2.3).

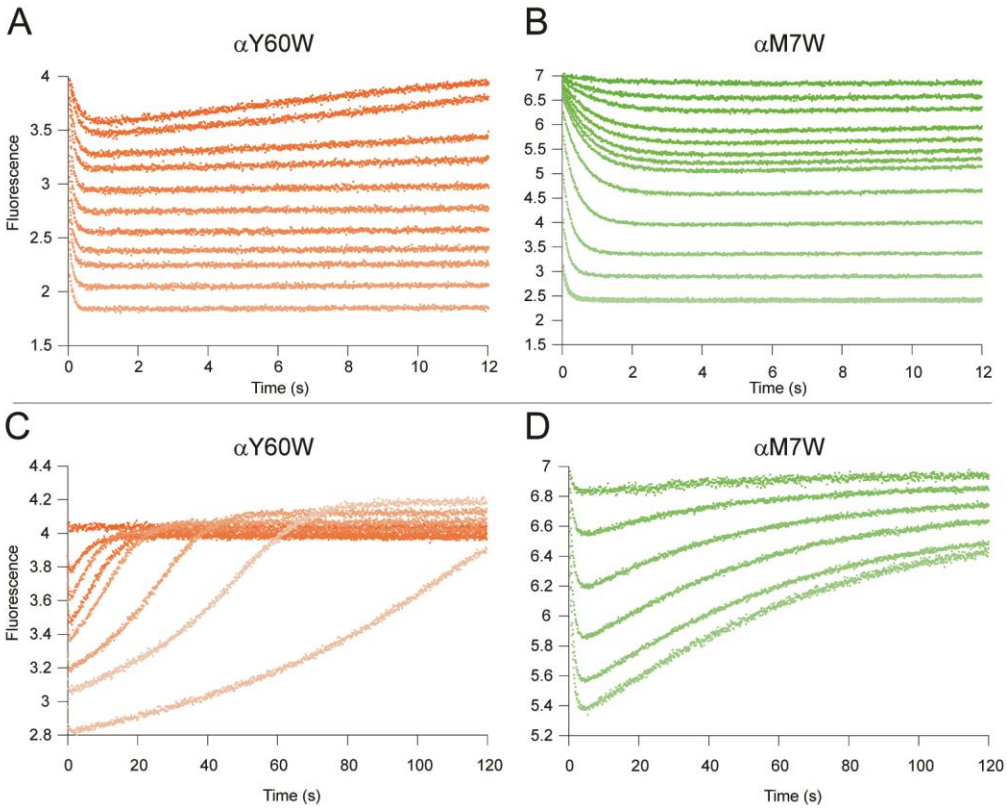


Figure 2.4: Stopped-Flow Fluorescence Traces of αY60W-CaaD and αM7W-CaaD

A) αY60W-CaaD for 12 s; B) αM7W-CaaD for 12 s; C) αY60W-CaaD for 120 s; and D) αM7W-CaaD for 120 s. The αY60W-CaaD mutant shows about a 2-fold decrease in fluorescence, whereas the αM7W-CaaD mutant shows a 3.5-fold decrease over the first 1 s of the reaction. Note the smaller overall amplitude observed with αY60W-CaaD.

Stopped-flow fluorescence measurements were also performed using wild-type CaaD, monitoring the native enzyme fluorescence at 340 nm (with excitation at 280 nm). The fluorescence in the wild-type enzyme is likely due to αY60 because of the location near the active site and the similarities observed for the wild-type (shown in blue in Figure 2.5) and αY60W CaaD mutant (shown in orange in Figure 2.5). Due to the poor quantum yield of tyrosine, the wild-type fluorescence data have a reduced signal-to-noise ratio. However, the data sets from both wild-type and the αY60W-CaaD mutant showed the same initial single exponential drop in fluorescence, and conventional analysis of both data sets (Table 2.3) yielded the same maximum rates (within error, shown in Figure 2.6B) when considering only the higher concentrations where the signal with wild-type enzyme was sufficient to afford reliable data fitting. Note that in both data sets, traces obtained at increasing substrate concentrations were offset due to the absorption of incident light by the substrate.

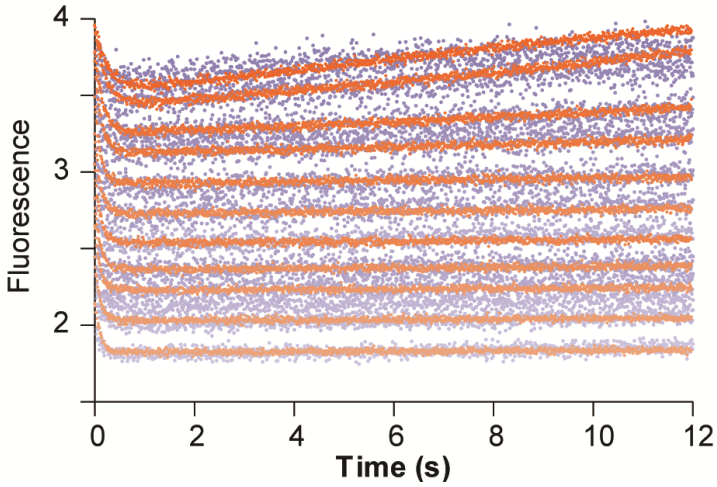


Figure 2.5: Stopped-Flow Enzyme Fluorescence of Wild-type and αY60W-CaaD mutant

Stopped-flow enzyme fluorescence (12 s) monitoring the reaction of wild-type CaaD (blue) and the αY60W-CaaD mutant (orange) with **9**. Both show single exponential decay in fluorescence in the first 1 s, but the αY60W CaaD mutant shows a significant improvement in the signal-to-noise ratio. Note, wild-type data set has been normalized to superimpose it on the αY60W-CaaD mutant data set.

Enzyme [a]	K_D [b] (μM)	k_{for} [b] (s^{-1})	k_{rev} [b] (s^{-1})
Wild-type	720	3	6
αY60W-CaaD	400	8	2.5
αM7W-CaaD	7500	12	1

[a]Rate constants are defined in Eq. 2.3. [b]Results from fit shown in Figure 2 (raw data shown in Figure 2.5). Fit for αM7W-CaaD not shown (raw data shown in Figure 2.4).

Table 2.3. Rates and Equilibrium Constants Obtained from Conventional Analysis of Stopped-flow Data using *trans*-3-bromoacrylate (**9**)

Data obtained with the αY60W-CaaD mutant gave a significantly larger signal-to-noise ratio at all concentrations, allowing more extensive analysis as shown in Figure 2.6. Initially, the first second of each trace was fit to a single exponential function (Figure 2.6A, Eq. 2.1). The concentration dependence of the rate (obtained from the single exponential) was fit to a hyperbola (Figure 2.6B, Eq. 3) suggesting a minimal two-step model with a rapid equilibrium step (K_d) followed by rate-limiting isomerization governed by forward and reverse rate constants (k_{for} and k_{rev}). Analysis of the concentration dependence of the rate defined an apparent K_d = 400 μM for initial substrate binding followed by a maximum rate of the fluorescence change of ~10 s^{-1}. A

well-defined intercept on the y-axis afforded resolution of a net forward rate ($k_{for} \sim 8\ s^{-1}$) and an apparent reverse rate ($k_{rev} \sim 3\ s^{-1}$). The numerical values above are summarized in Table 2.3. A more rigorous interpretation and explanation of these results are detailed in Chapter 3.

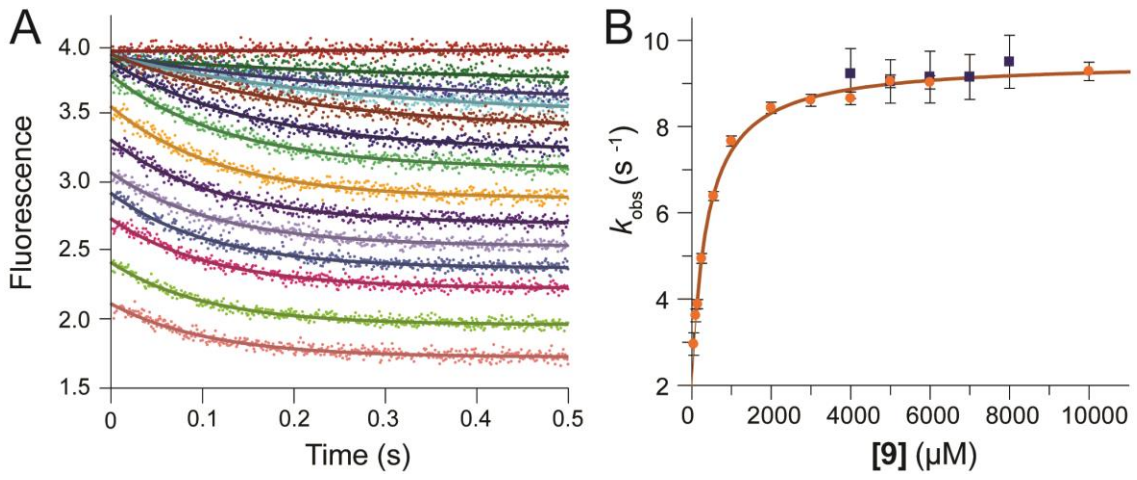


Figure 2.6: Conventional Fitting of Stopped-Flow Fluorescence Data with αY60W-CaaD

A) The stopped-flow fluorescence traces (0.5 s) showing reaction of αY60W-CaaD with **9** were fit to a single exponential (Eq 2.1) shown in solid lines. The concentrations of **9** are: 0, 50, 100, 150, 250, 550, 1000, 2000, 3000, 4000, 5000, 6000, 8000, 10000 μM. B) The observed rate constant (k_{obs}) from the 1 s stopped-flow data for the wild-type (■) and the αY60W (●) CaaD plotted versus the concentration of **9**. Error bars for the k_{obs} values are representative of the scatter in the data. Within error, the rates at high concentrations are identical. At lower concentrations, the signal to noise is to low too accurately resolve the rate. The solid line is a fit of the data to the hyperbolic equation (Eq 2.3).

2.3.4 Rapid-Quench Experiments with CaaD and αY60W-CaaD

In light of the similarities in the kinetic parameters for wild-type and αY60W CaaD, rapid quench experiments were carried out with both enzymes, to provide a more rigorous comparison. Experiments were conducted by mixing equal volumes of various concentrations of enzyme (CaaD: 400 μM; αY60W-CaaD: 250, 400, and 600 μM, after mixing) and an excess of **9** (10 mM, after mixing). At fixed intervals (3-1500 ms), the reactions were quenched with acid, and the amount of bromide ion quantified by IC and plotted versus time (Figures 2.7A and 2.7B). A pre-steady-state burst of product (bromide) formation was observed for both enzymes in all experiments, with a fast initial burst rate followed by a slower steady-state turnover rate. The observation of a pre-steady-state burst implies that a step after chemistry is at least partially rate limiting in both the CaaD- and αY60W-CaaD-catalyzed reactions.[24] Data were fit to the burst equation (Eq. 2.4) to yield the kinetic parameters summarized in Tables 2.4 and 2.5, defining the rate of the burst as approximately 85 s^{-1}, followed by steady-state turnover of 3 s^{-1}. The sum of these observations, along with the steady-state kinetic parameters and stopped-flow data presented above, indicated that the αY60W mutant of CaaD was a suitable candidate for subsequent in-depth analysis and use as a model for the wild-type CaaD reaction.

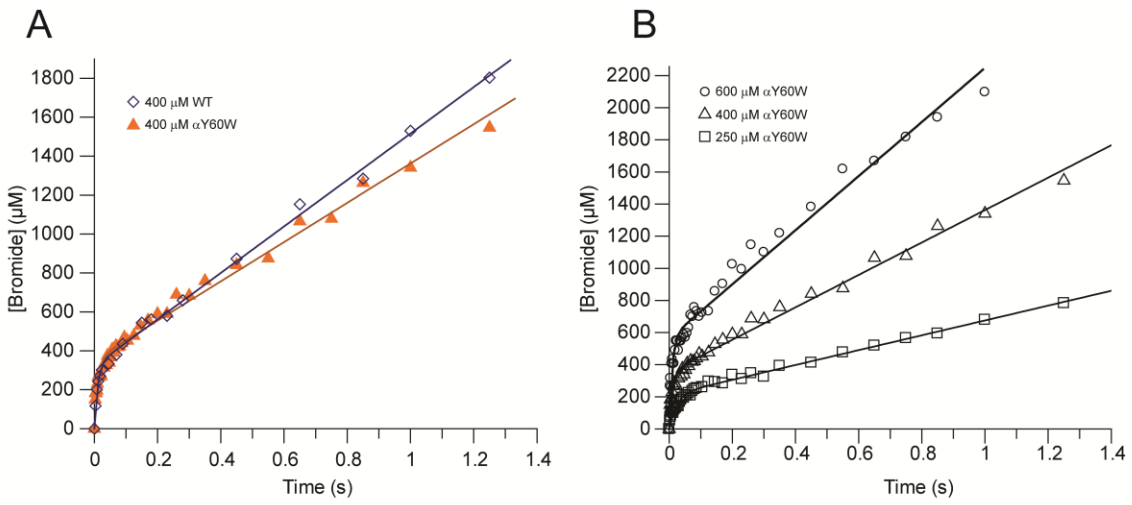


Figure 2.7: Pre-steady-state Burst of CaaD and αY60W-CaaD

Rapid quench experiments using CaaD and αY60W-CaaD fit to the burst equation. A) A comparison of the wild-type CaaD (◇) and the αY60W-CaaD (▲) showing the burst of bromide followed by steady-state turnover. The concentration of enzyme was 400 μM. B) Rapid quench experiment for three concentrations of αY60W CaaD: 250 μM (□), 400 μM (△), and 600 μM (○).

Enzyme[a] (400 μM)	k_{burst} (s^{-1})	k_{cat}[b] (s^{-1})	Burst Amplitude (*A*)
Wild-type	90 ± 20[c]	3 ± 0.5	0.8
αY60W	86 ± 10	2.5 ± 0.5	0.9

[a]The kinetic parameters were measured in 100 mM Na_2HPO_4 buffer (pH 8.1) at 22 °C with a saturating concentration of **9**. [b]Calculated from the burst equation.[24] [c]Errors are standard deviations.

Table 2.4. Burst Experiment Parameters Obtained using CaaD and the αY60W Mutant

Enzyme concentration (μM)	k_{burst} (s^{-1})	k_{cat}[b] (s^{-1})	Burst Amplitude (*A*)
200	43 ± 3[c]	1.8 ± 0.5	0.85
400	86 ± 20	2.5 ± 0.5	0.9
600	107 ± 15	2.8 ± 0.5	0.94

[a]The kinetic parameters were measured in 100 mM Na_2HPO_4 buffer (pH 8.1) at 22 °C with saturating concentration of **9** (10,000 μM). [b]Calculated from the burst equation.[24] [c]Errors are standard deviations.

Table 2.5. Burst Experiment Parameters Obtained at Different Concentrations of αY60W-CaaD

2.4 DISCUSSION

Enzymes have evolved ingenious catalytic strategies to remove halogens bonded to alkane, aromatic, and alkene substrates without the use of metal ions or other cofactors.[25,26] The haloalkane dehalogenases designated DhlA[26-29], 4-chlorobenzoyl-CoA dehalogenase[30-34], CaaD[2-5,10,11,15], and *cis*-CaaD[6,12-14] illustrate the reactions and their similarities and differences. The first two of these enzymes catalyze reactions that proceed by a well-characterized covalent ester intermediate[25,26], whereas *cis*-CaaD and CaaD do not.[5,6,9] Extensive mechanistic, structural, and kinetic studies have defined key features of all four enzyme-catalyzed reactions including the catalytic and binding residues, rate limiting steps, conformational changes, and the presence or absence of a halide-binding pocket.

Although *cis*-CaaD and CaaD have the same core catalytic residues (the amino-terminal proline, two arginines, and a glutamate), the *cis*-CaaD reaction appears to be more complex than that of CaaD because two additional active site groups (i.e., His-28 and Tyr-103) are required for activity.[12] There is no obvious explanation for the additional complexity, and the hydration of the *cis*-isomer is not expected to be energetically more difficult. As part of an effort to understand the basis for this difference and to define the fundamental reaction scheme, a kinetic analysis of both enzymes and their mutants is being pursued. A kinetic mechanism for *cis*-CaaD has been proposed.[13] Although upon further investigation, there is evidence that the kinetic model is more complex than initially thought.[14] A similar comprehensive analysis of the CaaD reaction was precluded by the absence of a strong fluorophore in the active site to provide

a signal to define kinetically significant enzyme states.[10] This prompted the construction of 5 fluorescent mutants of CaaD: αY60W, αM7W, αF39W, αL57W, and βI37W. Two of these mutants were inactive, αL57W, and βI37W. Two of these mutants, αM7W and αF39W, performed well as a fluorescent reporter, however αM7W showed significant kinetic deviations from the wild-type enzyme.

The αY60W mutant of CaaD was validated as an accurate fluorescent reporter of active site events and it yielded satisfactorily equivalent kinetic results to wild-type CaaD in all experiments. These experiments include: steady-state kinetics (Table 2.2 and Figure 2.3A), stopped-flow enzyme fluorescence (Table 2.3 and Figures 2.5 and 2.6B), rapid acid quench (Table 2.4 and Figure 2.7A) and steady-state bromide inhibition studies (not shown). The validation of this fluorescent mutant of CaaD allows for further experiments to be carried out which might provide insight into the distinguishing mechanistic features of the CaaD enzyme.

In Chapter 3, the Y60W CaaD data shown above (along with more data to be shown in Chapter 3) is globally fit to a single kinetic model defining six individual rate constants, an equilibrium constant for one product, and values for one of the two possible fluorescent factors of the enzyme observed during the course of the reaction.

2.5 REFERENCES

(1) Hartmans, S.; Jansen, M. W.; Van der Werf, M. J.; De Bont, J. A. M. (1991) Bacterial metabolism of 3-chloroacrylic acid. *J. Gen. Microbiol. 137*, 2025-2032.

(2) van Hylckama Vlieg, J. E.; Janssen, D. B. (1991) Bacterial degradation of 3-chloroacrylic acid and the characterization of *cis*- and *trans*-specific dehalogenases. *Biodegradation. 2*, 139-150.

(3) Poelarends, G. J.; Wilkens, M.; Larkin, M. J.; Van Elsas, J. D.; Janssen, D. B. (1998) Degradation of 1,3-dichloropropene by *Pseudomonas cichorii* 170. *Appl. Environ. Microbiol. 64*, 2931-2936.

(4) Poelarends, G. J.; Saunier, R.; Janssen, D. B. (2001) *trans*-3-chloroacrylic acid dehalogenase from *Pseudomonas pavonaceae* 170 shares structural and mechanistic similarities with 4-oxalocrotonate tautomerase. *J. Bacteriol. 183*, 4269-4277.

(5) Wang, S. C.; Person, M. D.; Johnson, W. H., Jr.; Whitman, C. P. (2003) Reactions of *trans*-3-chloroacrylic acid dehalogenase with acetylene substrates: consequences of and evidence for a hydration reaction. *Biochemistry. 42*, 8762-8773.

(6) Poelarends, G. J.; Serrano, H.; Person, M. D.; Johnson, W. H., Jr.; Murzin, A. G.; Whitman, C. P. (2004) Cloning, expression, and characterization of a *cis*-3-chloroacrylic acid dehalogenase: insights into the mechanistic, structural, and evolutionary relationship between isomer-specific 3-chloroacrylic acid dehalogenases. *Biochemistry. 43*, 759-772.

(7) Poelarends, G. J., Johnson, H. W., Jr., Murzin, A.G., and Whitman, C.P. (2003) Mechanistic characterization of a bacterial malonate semialdehyde decarboxylase: identification of a new activity in the tautomerase superfamily. *J. Biol. Chem. 278*, 48674-48683.

(8) Murzin, A. G. (1996) Structural classification of proteins: new superfamilies. *Curr Opin Struct Biol. 6*, 386-394.

(9) Poelarends, G. J.; Veetil, V. P.; Whitman, C. P. (2008) The chemical versatility of the β-α-β fold: catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell. Mol. Life Sci. 65*, 3606-3618.

(10) De Jong, R. M.; Brugman, W.; Poelarends, G. J.; Whitman, C. P.; Dijkstra, B. W. (2004) The x-ray structure of *trans*-3-chloroacrylic acid dehalogenase reveals a novel hydration mechanism in the tautomerase superfamily. *J. Biol. Chem. 279*, 11546-11552.

(11) Azurmendi, H. F.; Wang, S. C.; Massiah, M. A.; Poelarends, G. J.; Whitman, C. P.; Mildvan, A. S. (2004) The roles of active-site residues in the catalytic mechanism of *trans*-3-chloroacrylic acid dehalogenase: a kinetic, NMR, and mutational analysis. *Biochemistry. 43,* 4082-4091.

(12) De Jong, R. M.; Bazzacco, P.; Poelarends, G. J.; Johnson, W. H., Jr.; Kim, Y. J.; Burks, E. A.; Serrano, H.; Thunnissen, A.-M. W. H.; Whitman, C. P.; Dijkstra, B. W. (2007) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase. Structural basis for substrate specificity and inactivation by *(R)*-oxirane-2-carboxylate. *J. Biol. Chem. 282*, 2440-2449.

(13) Robertson, B. A.; Schroeder, G. K.; Jin, Z.; Johnson, K. A.; Whitman, C. P. (2009) Pre-steady-state kinetic analysis of *cis*-3-chloroacrylic acid dehalogenase: analysis and implications. *Biochemistry. 48*, 11737-11744.

(14) Schroeder, G. K.; Johnson, W. H.; Huddleston, J. P.; Serrano, H.; Johnson, K. A.; Whitman, C. P. (2012) Reaction of *cis*-3-chloroacrylic acid dehalogenase with an allene substrate, 2,3-butadienoate: hydration via an enamine. *J. Am. Chem. Soc. 134*, 293-304.

(15) Poelarends, G. J.; Johnson, W. H., Jr.; Serrano, H.; Whitman, C. P. (2007) Phenylpyruvate tautomerase activity of *trans*-3-chloroacrylic acid dehalogenase: evidence for an enol Intermediate in the dehalogenase reaction. *Biochemistry. 46*, 9596-9604.

(16) Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

(17) Waddell, W. J. (1956) A simple ultraviolet spectrophotometric method for the determination of protein. *J. Lab. Clin. Med. 48*, 311-314.

(18) Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature. 227,* 680-685.

(19) Trott, O.; Olson, A. J. (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem. 31*, 455-461.

(20) DeLano, W. L. (2002) *The PyMOL molecular graphics system*; DeLano Scientific, San Carlos, CA.

(21) Burks, E. A.; Fleming, C. D.; Mesecar, A. D.; Whitman, C. P.; Pegan, S. D. (2010) Kinetic and structural characterization of a heterohexamer 4-oxalocrotonate tautomerase

from *Chloroflexus aurantiacus* J-10-fl: Implications for functional and structural diversity in the tautomerase superfamily. *Biochemistry. 49*, 5016-5027.

(22) Burks, E. A.; Yan, W.; Johnson, W. H.; Li, W.; Schroeder, G. K.; Min, C.; Gerratana, B.; Zhang, Y.; Whitman, C. P. (2011) Kinetic, crystallographic, and mechanistic characterization of TomN: Elucidation of a function for a 4-oxalocrotonate tautomerase hmologue in the tomaymycin biosynthetic pathway. *Biochemistry. 50*, 7600-7611.

(23) Robertson, B. A. (2007) Evolution and divergence in the tautomerase superfamily: A pre-steady-state kinetic analysis of *cis*-3-chloroacrylic acid dehalogenase and an inhibition study of its homologue, Cg10062, in *Corynebacterium glutamicum* Dissertation, University of Texas at Austin

(24) Johnson, K. A. (1992) Transient-state kinetic analysis of enzyme reaction pathways In *The Enzymes* (Sigman, D. S., Ed.) 3rd Ed. pp 1-61 Academic Press, San Diego

(25) Copley, S. D. (1999) Microbial dehalogenases In *Compreheive Natural Products Chemistry* (Barton, D., Nakanishi, K., Eds.) pp 401-422 Elsevier, Amsterdam

(26) Janssen, D. B. (2004) Evolving haloalkane dehalogenases. *Curr. Opin. Chem. Biol. 8*, 150-159.

(27) Verschueren, K. H. G.; Kingma, J.; Rozeboom, H. J.; Kalk, K. H.; Janssen, D. B.; Dijkstra, B. W. (1993) Crystallographic and fluorescence studies of the interaction of haloalkane dehalogenase with halide ions. Studies with halide compounds reveal a halide binding site in the active site. *Biochemistry. 32*, 9031-9037.

(28) Schanstra, J. P.; Kingma, J.; Janssen, D. B. (1996) Specificity and kinetics of haloalkane dehalogenase. *J. Biol. Chem. 271*, 14747-14753.

(29) Schanstra, J. P.; Janssen, D. B. (1996) Kinetics of halide release of haloalkane dehalogenase: evidence for a slow conformational change. *Biochemistry. 35*, 5624-5632.

(30) Benning, M. M.; Taylor, K. L.; Liu, R.-Q.; Yang, G.; Xiang, H.; Wesenberg, G.; Dunaway-Mariano, D.; Holden, H. M. (1996) Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 Å resolution: An enzyme catalyst generated via adaptive mutation. *Biochemistry. 35*, 8103-8109.

(31) Zhang, W.; Wei, Y.; Luo, L.; Taylor, K. L.; Yang, G.; Dunaway-Mariano, D.; Benning, M. M.; Holden, H. M. (2001) Histidine 90 function in 4-chlorobenzoyl-coenzyme A dehalogenase catalysis. *Biochemistry. 40*, 13474-13482.

(32) Liu, R.-Q.; Liang, P.-H.; Scholten, J.; Dunaway-Mariano, D. (1995) Transient state kinetic analysis of the chemical intermediates formed in the enzymic dehalogenation of 4-chlorobenzoyl coenzyme A. *J. Am. Chem. Soc. 117*, 5003-5004.

(33) Xu, D.; Wei, Y.; Wu, J.; Dunaway-Mariano, D.; Guo, H.; Cui, Q.; Gao, J. (2004) QM/MM studies of the enzyme-catalyzed dechlorination of 4-chlorobenzoyl-CoA provide insight into reaction energetics. *J. Am. Chem. Soc. 126*, 13649-13658.

(34) Wu, J.; Xu, D.; Lu, X.; Wang, C.; Guo, H.; Dunaway-Mariano, D. (2006) Contributions of long-range electrostatic interactions to 4-chlorobenzoyl-CoA dehalogenase catalysis: a combined theoretical and experimental study. *Biochemistry. 45*, 102-112.

Chapter 3: The Kinetic Mechanism of the αY60W Mutant of *trans*-3-Chloroacrylic Acid Dehalogenase

3.1 INTRODUCTION

In Chapter 1 of this dissertation, one of the principals of enzymology was introduced (Figure 1.1). Emil Fischer proposed the lock-and-key model hypothesis to explain how enzymes work. In this model, Fischer hypothesized that the binding of a substrate to form an enzyme-substrate complex was followed by catalytic turnover to reform free enzyme and product.[1] This model was the first of several to explain the mechanisms of enzyme catalysis. Analysis of kinetic data provides one platform to achieve this goal.

In 1913, Maud Menten and Leonor Michaelis published work that highlighted the importance of analyzing kinetic data in order to gain information about an enzyme-catalyzed reaction.[2,3] Their work focused on the enzyme "invertase" (known today as β-D-fructofuranosidase), which converts sucrose to glucose and fructose (Scheme 3.1).[2,3]

Scheme 3.1: Reaction of "Invertase"

Although the kinetic work on "invertase" was started by Victor Henri, Michaelis and Menten recognized two deficiencies in Henri's methodology (i.e., failure to account

Adapted with permission from Huddleston et. al. *Biochemistry. 51 (46),* 9420-9435.
 Dr. Schroeder helped with data analysis and experimental design. Dr. Johnson provided instrumentation, software, and data analysis expertise. Dr. Whitman provided helpful discussions, editing, and funding.

for the interconversion of the α-β forms of glucose and control of pH) which prevented him from fully testing his theory about the formation of a catalytic enzyme-substrate complex.[2] By studying changes in initial velocity of substrate (sucrose) depletion as a function of increasing sucrose concentrations, Michaelis and Menten developed their now famous equation, the Michaelis-Menten equation (sometimes referred to as the Henri-Michaelis-Menten equation, Eq. 3.1). By 1925, the arguments were fully developed to support the validity of their experiments. The Michaelis-Menten equation is of the cornerstones of enzymology and remains so a century after its derivation.[2]

$$v = \frac{V_{max}[S]}{K_m + [S]} \qquad \text{Eq. 3.1}$$

Today, enzymologists are quite familiar with this equation and studying initial velocities of a reaction to gather information about the k_{cat} and K_m of enzymes. Although it is forgotten, Michaelis and Menten were the first to fit full time course kinetic data, not just the initial velocities.[2,3] They recognized that product inhibition played a significant role in the full time course kinetic data. By taking product inhibition into account, they were able to fit full time course data to an integrated form of the rate equation for their two-step model.[2] By fitting the data for several starting sucrose concentrations, they were able to obtain a fit to a family of curves and that yielded a single, reproducible "constant value".[2,3] In essence, this was the first global fit of data to a proposed model. Interestingly, the "constant value" Michaelis and Menten calculated was actually equivalent to k_{cat}/K_m, not the Michaelis constant or K_m as it is referred to today.[2]

Even though Michaelis and Menten realized the power in developing a "model" to account for all the observable data, this technique is not used nearly as frequently as the study of reactions using the initial velocity approach. This is most likely due to the

complexity of enzyme catalysis, which greatly increases the complexity of the mathematical functions necessary to describe the reactions. A common solution to this problem is to add assumptions to a model in order to simplify the functions.

In conventional data fitting, mathematical equations are used to model kinetic data.[4,5] These equations result from the solution of differential rate equations which give the time-dependent concentration for each species for a particular kinetic model.[4,6] For example, in the analysis of pre-steady-state kinetic data, the data are conventionally fit to an equation which is made up by a sum of exponentials (Eq. 3.2): where c is the endpoint, A_i is the amplitude, and λ is an observed rate (k_{obs}).[3,6]

$$y = c + \sum_{i=1}^{n} A_i \cdot e^{-\lambda \cdot t} \qquad \text{Eq. 3.2}$$

Scheme 3.2: One-step Reversible Model

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES$$

In order to understand the basis for Eq. 3.2, consider the simple reversible one-step binding of a substrate to an enzyme shown in Scheme 3.2. Integrating the differential rate equation for this simple model relies on the approximation that [S] is much larger than [E], results in a single exponential equation describing the concentration for each enzyme species over time.[6] The concentration of a species over time is a function of the observable rate (k_{obs}), which, in turn, is a function of the forward and reverse rates (k_1[S] and k_{-1}), and a complex amplitude value, which is a function of the equilibrium constant for the formation of the ES complex.[6] To fit kinetic data that follows this single exponential dependence, one assumes the model in Scheme 2.3 and

solves Eq 3.2 with n equal to 1, resulting in a single observable rate (λ) and amplitude (A_1). In general, each kinetically significant step of a model yields one exponential term.[3-6]

The components that make up the observable rates and amplitudes quickly grow in complexity. For example, for a two step reversible reaction (n = 2 in Eq 3.2) yielding two observable rates (λ_1 and λ_2), the observable rates are the roots of a complex quadratic equation (Eq. 3.3).[6]

$$\lambda_{1,2} = \frac{-(k_1[S] + k_{-1} + k_2 + k_{-2}) \pm \sqrt{(k_1[S] + k_{-1} + k_2 + k_{-2})^2 - 4 \times (k_1[S] \times (k_2 + k_{-2}) + k_{-1}k_2)}}{2}$$

Eq. 3.3

One can easily see how the complexity in the mathematical equations usually requires to assumptions to simplify them. To further complicate matters, fitting kinetic data to single or double exponential equations, yields only relative values for the observable rates and amplitudes, not absolute values of rate constants.[4]

As shown above, the observable rates and amplitudes are functions of multiple individual rate constants. In order to determine the actual values of the rate constants that make up the observed rates and amplitudes, the observed rates (k_{obs}) and amplitudes are subjected to a secondary graphical analysis.[4] For example, if we revisit the model in Scheme 3.2, the solution of the integrated rate equation suggests that $k_{obs} = k_1[S] + k_{-1}$.[6] This shows that the observed rates have a positive linear relationship with the substrate concentration. The plot of the observed rates as a function of [S], results in a straight line with a positive slope and the values for the rate constants, k_1 and k_{-1}, are the slope of the line and the y-intercept, respectively.[6]

This secondary plot is one most often described in the literature and used to make conclusions. Many times, data which deviate from the limited analysis provided by fitting to the sums of exponential equations are ignored or discarded.[4] Moreover, the amplitude values are completely ignored even though they contain important equilibrium information. Although fitting by conventional methods can be useful, making conclusions solely based on these results is an archaic and error-prone method for data analysis.

Today, the primary data can be fit directly using computer simulation because of the greatly improved computational power. This method uses numerical integration to directly solve the differential rate equations for a user-chosen model and starting concentrations.[3-5] As a result, the concentration of every species in the model can be calculated at any point in time. The results are presented on a computer screen as a solid line which can be optimized to the observed data using statistical regression calculations. This solid line is referred to as the simulated trace. Data fitting by simulation eliminates the requirements for assumptions and approximations usually found in conventional data analysis.[4] Also, the rate constants are determined directly from the primary data set, not through secondary means and both rate and amplitude information are included.[3-5]

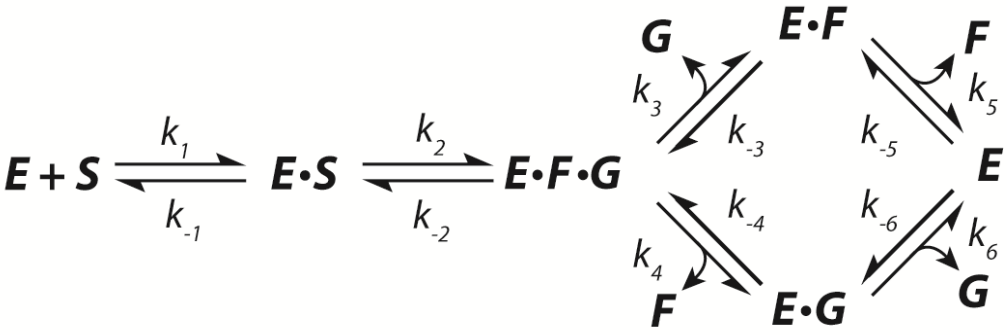
Nonetheless, there are two intrinsically linked limitations to this method of data fitting.[4] First, the user must enter a model (or the differential equation for the model) that accounts for all of the observed kinetic data. Second, that model cannot be more complex than is defined by the data.[4,5] Achieving this "Goldilocks" balance is the basis for fitting data by computer simulation. Because entering a model is the first step of fitting data and is absolutely required, many become disenchanted and return to the conventional methods of fitting.[4,5] Fortunately, conventional methods are usually a good start when coming up with an initial model to begin fitting by simulation.

The data set collected by Michaleis and Menten a century ago can be fit by computer simulation.[2,3] First a minimal model which completely accounts for all of the observed data must be entered. To simulate Michaelis and Menten's initial velocity steady state kinetic curves, only a simple model is required (Scheme 3.3).[2]

Scheme 3.3: Michaelis-Menten Initial Kinetic Model

This model is sufficient to define the values of k_{cat} and K_S (equivalent to K_m for their special case) for enzymatic turnover, which is all the rate information that can be defined by initial velocity steady state data.[2,3] In order to fit a more complex data set, like the full time course kinetic traces collected by Michaelis and Menten and fit using a mathematical equation, a more complete and realistic model is required (Scheme 3.4).[3] This model accounts for the k_{cat} and K_m and K_d values for each product.[3]

Scheme 3.4: Expanded Michaelis-Menten Model



In the case of the Michaelis and Menten, they were able to use some assumptions to simplify the mathematical rate equation to solve for a "constant value" at several different substrate concentrations and reaction times.[2,3] These assumptions were: rapid equilibrium of binding, irreversible chemistry, and a ternery complex, *EFG*, that doesn't

accumulate to any significant amount (i.e., it is a short-lived intermediate). Moreover, values for the dissociation constants of the substrate, sucrose, and the products glucose and fructose ($K_S = k_{-1}/k_1$, $K_F = k_{-5}/k_5$, and $K_G = k_{-6}/k_6$) were assumed from previous measurements.[2,3] Fitting their collected data by computer simulation under the assumptions of rapid equilibrium of binding for both the substrate and products, irreversible chemistry, and fast decomposition of the ternary complex (i.e., *EFG*), results in values defining the K_S (equivalent to K_d or K_m, under the conditions of their model), k_{cat} (k_2), and an average K_d for product inhibition.[3] The K_d for product inhibition is only an average because there is no data distinguishing the difference between binding of glucose or fructose to the enzyme from full time course data alone.[3] However, if the ratio of binding constants for glucose and fructose is considered, (K_F/K_G =1.53, as reported by Michaelis and Menten) the value of one (and hence both) can be obtained directly from fitting.[3] The amount of assumptions here could be reduced with more experiments. Data from transient state kinetic experiments, like stopped-flow, rapid chemical quench, or single turnover, could be incorporated resulting in a global fit with fewer assumptions.[3,4,6] As a result, the k_{on} values (or better defined limits) for substrate and products and/or the lifetime of the ternary complex, *EFG,* could be determined. Experiments providing this information would yield values for all 12 rate constants while completely eliminating any simplifying assumptions.[3]

Today, complex global fits, like the one published by Schroeder *et al.*, contain multiple data sets (including transient and steady-state data) with each separate experiment containing specific information about one or more individual rate constants of the defined model.[7] The model ultimately used for global fitting and the resulting values for each rate constant were introduced in Chapter 1 (Scheme 1.7). In this case, models derived from conventional analysis alone cannot account for all the data.[7]

It is important to note that even with a well-defined minimal model, a given data set may not contain data to sufficient define each individual rate constant of that model.[4] Either additional experiments must be conducted to address limitations of the existing dataset or some assumptions must be made. Unfortunately, restrictions in the enzyme system of interest or limitations of instrumentation may prevent the collection of data which would provide data defining a particular rate constant. As discussed above for the Michaelis-Menten data set, when fitting the steady-state full time course kinetic data to the model in Scheme 3.4, some assumptions were required to achieve well-constrained values for the fitted rate constants. Alternatively, additional experiments could better define values for the remaining rate constants. Therefore, assumptions or simplifications may be used to achieve well-defined parameters. The important distinction here is that these assumptions are not required, but they are most often required in conventional analysis.[4]

As a simple example of the use of assumptions when fitting by computer simulation, consider an enzyme that reversibly binds a ligand following the model outlined in Scheme 3.2. When following this simple reaction, the signal should become saturated with increasing ligand concentrations (this occurs when $k_1 \times [S] >> k_{-1}$). If the k_{on} of the ligand occurs in the dead time of the instrument, this rate constant can never be defined in the data. However, the equilibrium dissociation constant (K_d) will be defined by the experiment. If one assumes a "fast" k_{on} value, the value for k_{off} is easily defined in the dataset, as $k_{off} = K_d \times k_{on}$. Even though "fast" is a relative and qualitative term to use as an assumption to define a numerical value, the experiment suggests that the rate constant must be fast enough to occur in the dead time of the instrument. The dead time sets a lower limit on the value for k_{on}, but not an upper limit. However, the upper limit can be no higher than the rate of diffusion. Although, the experiment in this example can

never determine the actual value for k_{on}, limits on its value can be set. Other methods or experiments may be used to determine the value or narrow the range of those limits. In summary, if a dataset does not contain information about each individual rate constant, all is not lost because sometimes equilibrium information is defined in the data. Assumptions can still be used in conjunction with fitting by computational simulation, but are not required, to yield in well-constrained values for parameters. In general, one must clearly state all the assumptions or simplifications that were made to achieve the fit, along with the corresponding justification where appropriate.[4]

As mentioned previously, the basis of fitting by simulation is finding the minimal model that accounts for the observable data. Fitting data to an overly complex model results in rates that are not defined and have little or no justification for assumed values.[4] As the model has no restraints and is input by us, a problem arises as to how to estimate errors on fitted parameters and evaluate goodness of fit resulting from traditional regression calculations.[4,5,8]

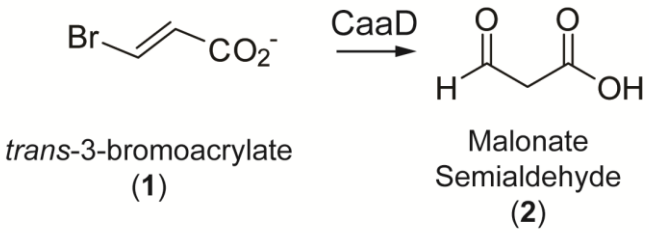
Traditionally, the general method of fitting data by computer simulation is based on minimizing the total chi-squared (χ^2) for the dataset.[4,8] The squared error is computed by squaring the difference between the simulated trace and an observed data point. Adding up all these squared error values and dividing by the standard deviation yields the chi^2.[4,5,8] Most fitting algorithms then set out to find the parameters which minimize the deviation between the simulated trace and the data as measured by chi^2 by nonlinear regression. The problem with this method of fitting is that it drastically underestimates the errors associated with the final values leading to a false sense of confidence in their accuracy, and it fails to identify any internal relationships among fitted parameters.[4,5,8] KinTek Global Kinetic Explorer (the software used in this work for global data fitting) uses this traditional method of minimizing the chi^2 to generate initial values for each

parameter. With these values, KinTek includes a feature called FitSpace to help overcome the problems of traditional fitting and achieves a more realistic estimates of the error associated with the individual rate constants.[5,8]

FitSpace is an exhaustive calculative method for determining uniqueness of fit using graphical plots. FitSpace is based on minimizing the chi-squared (χ^2) distribution of a fit.[8] The χ^2 value is similar to SSE, but is a more accurate statistical measurement of "goodness-of-fit", because it takes into account the standard deviation, or scatter, of individual experiments. If the standard deviation (also known as sigma) for a set of data is known or can be estimated, then the χ^2 can be calculated.[8] Initially, FitSpace uses traditional nonlinear regression fitting algorithms to find the optimal value each fitted parameter. FitSpace then monitors (calculates) the value of the global χ^2 (the summed χ^2 for each experiment with normalization) while each individual fitted parameter undergoes a systematic adjustment from its optimal value.[4,8] During this process of adjustment, all the other fitted parameters are allowed to be re-optimized in an effort to obtain a minimal value of χ^2. If by adjusting all other parameters a similarly low χ^2 can be obtained in the searched range of values, the parameter undergoing systematic adjustment is not well-constrained by the data. In other words, its optimal value is not unique because deviation from its optimum doesn't change the value of χ^2. However, if adjusting all other parameters cannot maintain the minimal χ^2 value, the fitted parameter is well-constrained by the data. This initial process is depicted as a 2-D plot showing how the χ^2 is affected as each parameter is systematically adjusted. This calculation is then carried out in a pair-wise fashion generating a 3-D contour plot of the χ^2 values for each pair of fitted parameters. This 3-D contour plot depicts the relationship between any two rate constants and the overall shapes of the curves are important in characterizing any dependencies.[8]

In summary, global data fitting is a powerful technique that yields a robust minimal model that accounts for all of the data from multiple experiments, complete with well-constrained estimates for all kinetically relevant parameters (along with reasonable error limits). Moreover, rather than basing data analysis on simplifying assumptions (conventional analysis), global fitting identifies which parameters of a given model are defined by the data and yields useful limits on their value. As such, this approach, particularly when coupled with transient state kinetic methods, is an invaluable tool for the elucidation of enzyme mechanisms. To this end, we employed this technique to investigate the dehalogenation of *trans*-3-chloroacrylate by CaaD (introduced in Chapter 2). The αY60W mutant of CaaD was designed as a fluorescent reporter for the CaaD-catalyzed dehalogenation of *trans*-3-bromoacrylate (**1**) reaction, and shown to be kinetically similar to the wild-type CaaD reaction (as reported in Chapter 2).

Scheme 3.5: Reaction of *trans*-3-Bromoacrylate (1) with CaaD



trans-3-bromoacrylate (**1**)

Malonate Semialdehyde (**2**)

In this chapter, all the kinetic data collected (stopped-flow fluorescence, rapid chemical quench flow data, and full progress curve) with the αY60W mutant of CaaD were globally fit by simulation to a single model. The global fit also includes data from other experiments introduced in this chapter (nine experiments in total), which helped minimize assumptions in the kinetic model. These experiments include inhibition studies

with the product, bromide, and the determination of the dissociation constants for bromide and malonate semialdehyde (**2**) to the free αY60W-CaaD (i.e., non-liganded). The global fit by simulation to a six-step model resulted in six well-constrained rate constants including rates defining chemistry, a proposed conformational change occurring after the chemistry, and a biased-random product release.

3.2 MATERIALS AND METHODS

Materials. Chemicals, biochemicals, buffers, and solvents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO), Fisher Scientific Inc. (Pittsburgh, PA), Fluka Chemical Corp. (Milwaukee, WI) or EMD Chemicals, Inc (Gibbstown, NJ). The centrifugal filter devices (3,000 MW cutoff) were obtained from PALL Life Sciences (Ann Arbor, MI).

General Methods. Steady state kinetic assays were performed on an Agilent 8453 diode-array spectrophotometer at 22 °C.[9] Non-linear regression data analysis was performed using the program Grafit (Erithacus Software Ltd., Staines, U.K.). Protein concentrations were determined according to the method of Waddell.[10] Global Fitting was performed using Kintek Global Kinetic Explorer Version 3.0 or higher.[5,8]

3.2.1 Inhibition of αY60W-CaaD by Bromide Ion

Product inhibition of CaaD by bromide ion was determined under the steady state conditions described in Chapter 2.2.5. Briefly, the assays were carried out at 22 °C in 100 mM Na_2HPO_4 buffer, pH 8.1 using a 2 μM enzyme solution. Aliquots of a sodium bromide solution (0.5 M in 100 mM Na_2HPO_4 buffer, pH 8.1) were added to 1 mL portions of enzyme solution to yield a total of six final inhibitor concentrations (0, 5000, 10000, 15000, 30000, 50000 μM). The assay was initiated by the addition of six different concentrations of **1** (35, 50, 75, 120, 160, and 200 μM), obtained from a 50 mM stock solution. The initial rate of the hydration of **1** was monitored over a 60 s time

period, recording readings every 3 s. To avoid, significant signal interference due to bromide absorbance at 224 nm, the kinetic traces were followed at 234 nm. The inhibition patterns were determined by plotting the initial rate of the reaction vs. the concentration of bromide at different concentrations of **1** in a Dixon plot.[11-13] Full time courses (220 s) at each inhibitor concentration were recorded under the same conditions using 300 μM of **1** (ε_{234} = 5200 M^{-1} cm^{-1}).

3.2.2 Binding of Bromide Ion to αY60W-CaaD

The binding of bromide ion to the αY60W-CaaD was observed in the stopped-flow apparatus (SD 2004 series), where the change in fluorescence at 340 nm was monitored. The assay was conducted under the conditions described in the Chapter 2, Section 2.6 using 10 μM enzyme solution. To test the binding of bromide, various concentrations of bromide (0-50,000 μM) were made up in 100 mM Na_2HPO_4 buffer, pH 8.1. The enzyme (10 μM after mixing) and substrate (0-50,000 μM after mixing) solutions were mixed in the stopped-flow apparatus at 22 ºC. At least five 1 s traces were collected and averaged for each concentration of bromide ion. Stopped-flow fluorescence traces were unchanged over 1 s. Therefore, the data for each trace fit to a horizontal line (slope =0) and plotted against the bromide ion concentration. The data in the resultant plot were fit to Eq. 3.4 using Grafit.

$$F = F_0 + \Delta F\left(\frac{[X]}{K_D + [X]}\right) \qquad \text{Eq. 3.4}$$

Stopped-flow traces were also fit by simulation using KinTek Global Kinetic Explorer and resulting values were used as constraints during global fitting.[5]

3.2.3 Binding of Malonate Semialdehyde (2) to αY60W-CaaD

The binding of malonate semialdehyde (**2)** to αY60W-CaaD was monitored in the stopped-flow apparatus (monitoring the change in fluorescence at 340 nm). The assay was conducted as described in the preceding section using a 10 μM enzyme solution. To generate **2**, *cis*-CaaD (20 μM) and propiolic acid (10,000 μM) were mixed in 100 mM Na_2HPO_4 buffer, pH 8.1 (W.H. Johnson, Jr., J.P. Huddleston, G.K. Schroeder, and C.P. Whitman, unpublished observations, 2012). After 4 min, the mixture was transferred to a PALL centrifugal filter device (3,000 MW cutoff), and centrifuged at 11,000 rpm for 10 min to remove enzyme. This procedure generates ~10,000 μM of **2** (assuming the reaction is quantitative). Malonate semialdehyde (**2)** is a mixture of the aldehyde (~25%) and hydrate (~75%).[14] Various concentrations of **2** (0-10,000 μM) were made up in 100 mM Na_2HPO_4 buffer, pH 8.1, immediately before use. The αY60W-CaaD (10 μM after mixing) and different concentrations of **2** (0-5,000 μM after mixing) were mixed (1 s) in the stopped-flow apparatus at 22 ºC. Stopped-flow traces were fit to a single exponential (Eq 3.2) using the KinTek stopped-flow software. The k_{obs} and constant values (C) were plotted vs. the concentrations of **2**. The plot of C vs. [**2**] was fit to the hyperbolic equation (Eq. 3.4) and k_{obs} vs. [**2**] was fit to the linear equation (Eq 3.5).[6]

$$k_{obs} = k_1[\mathbf{2}] + k_{-1} \qquad \text{Eq 3.5}$$

In addition to this conventional analysis, the data were fit by simulation and resulting values were used as constraints for global fitting.

3.2.4 Data and Global Fitting Analysis

Conventional data fitting of pre-steady state kinetic data (including concentration dependence) by non-linear regression using the program Grafit provided initial estimates for rate constants and a minimal number of steps in the overall reaction pathway. The steady state k_{cat}/K_m value was used as a lower limit for the rate of substrate binding and the k_{cat} value provided a lower limit for individual first order rate constants following substrate binding and proceeding through product release.[6,7] All of the data were then globally fit to a single kinetic model using KinTek Global Kinetic Explorer, as described elsewhere.[5-7,11] During the process of global optimization, the data collected over a series of concentrations within a given experiment were scaled using a correction factor (less than 5%) for each trace to correct for slight lamp drift between traces. The simulated traces also contain a correction factor to account for the change in the fluorescence intensity resulting from the reduced transmitted light (280 nm) due to absorbance by substrate (**1**, ε_{280} = 110 M^{-1} cm^{-1}) and product (**2**, ε_{280} ~ 27 M^{-1} cm^{-1}). The reduced light intensity (I) is provided by the logarithmic function, $I = I_{ref} 10^{-\varepsilon \cdot c \cdot (l/2)}$, where ε is the extinction coefficient at 280 nm, c is the concentration of the species, and l is the path length of the cell (0.5 cm). The average light intensity across the full fluorescence window was approximated by l/2 (0.25 cm).

3.3 RESULTS

3.3.1 Inhibition and Binding of Bromide to αY60W-CaaD

The mode of inhibition of bromide was investigated by monitoring the changes in the initial rate of the αY60W-CaaD-catalyzed reaction using **1** (35-200 μM) in the presence of various concentrations of bromide (0-50,000 μM). The inhibition pattern observed in the Dixon plot is consistent with competitive inhibition where the K_i value is ~10 mM (Figure 3.1).[12,13] Full time courses (300 s) were also monitored using the αY60W-CaaD mutant at a fixed concentration of **1** (300 μM) in the presence of various concentrations of bromide ion (Figure 3.2). A reduction in the rate of the reaction can only be observed at very high bromide concentration, consistent with the observed weak inhibition constant. Stopped-flow binding experiments were conducted using bromide ion and the αY60W-CaaD mutant (Figure 3.3, filled circles). At all bromide concentrations examined, the fluorescence change reached equilibrium within the dead time of the instrument (1.3 ms). Stopped-flow fluorescence remains constant over the 1 s collection. Therefore, the data for each trace were fit to a horizontal line yielding a single fluorescence value for each concentration. These single fluorescence values were plotted against the bromide ion concentration. The data in the resultant plot was fit to Eq. 3.4 using Grafit (Figure 3.4). Based on the saturation of the fluorescence signal, the αY60W-CaaD mutant binds bromide with a predicted K_d of 7.4 mM (similar to the K_i observed in the Dixon plot, Figure 3.1) with a ~14% increase in fluorescence.

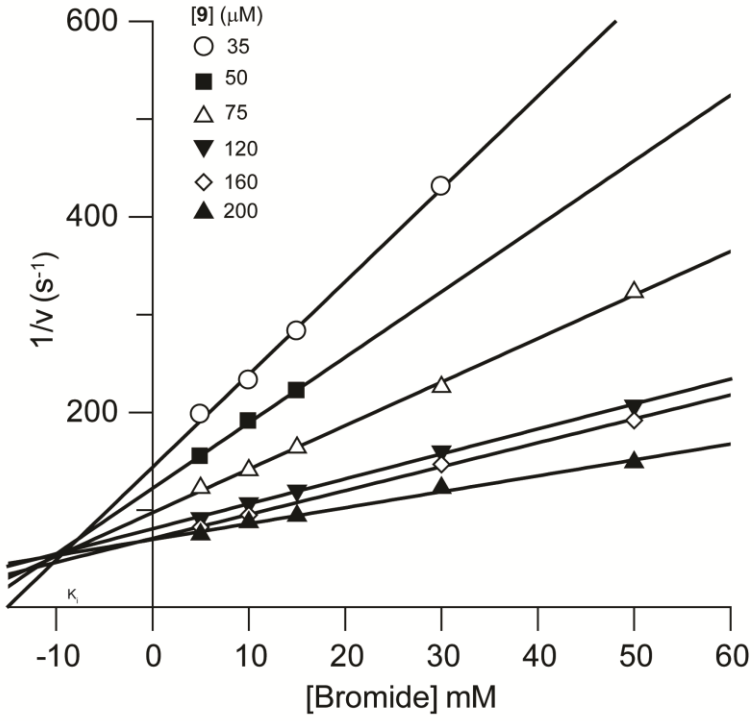


Figure 3.1: Dixon Plot of Bromide Inhibition of αY60W-CaaD

A plot of the initial reciprocal velocities (1/v) against bromide (5, 10, 15, 30 and 50 mM) at six concentrations of **1** (35, 50, 75, 120, 160, and 200 μM).[12] The K_I for inhibiton of αY60W-CaaD by bromide is determined by the intersection point of the lines resulting from a linear fit to the data (shown above as solid lines). The K_i of bromide (dashed lined extrapolation) was determined to be ~ 10 mM.

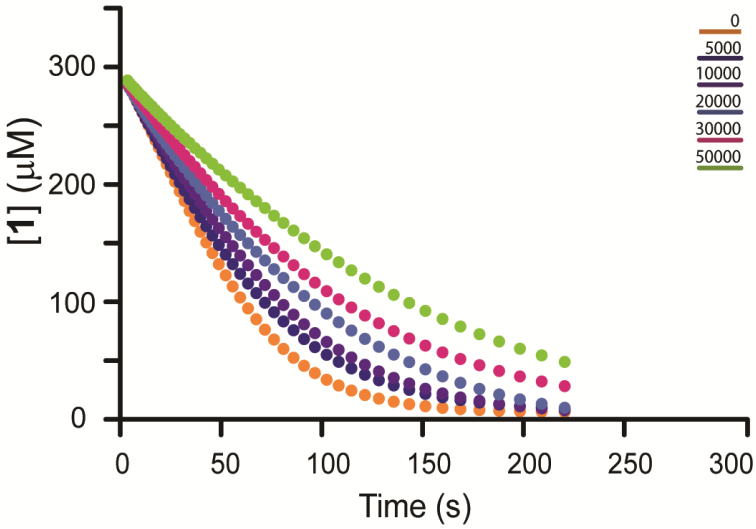


Figure 3.2: Bromide Inhibition of αY60W-CaaD.

Full progress curve kinetic traces observed at 300 μM of **1** with αY60W-CaaD in the presence of five concentrations of bromide (0, 5, 10, 20, 30, and 50 mM) are shown.

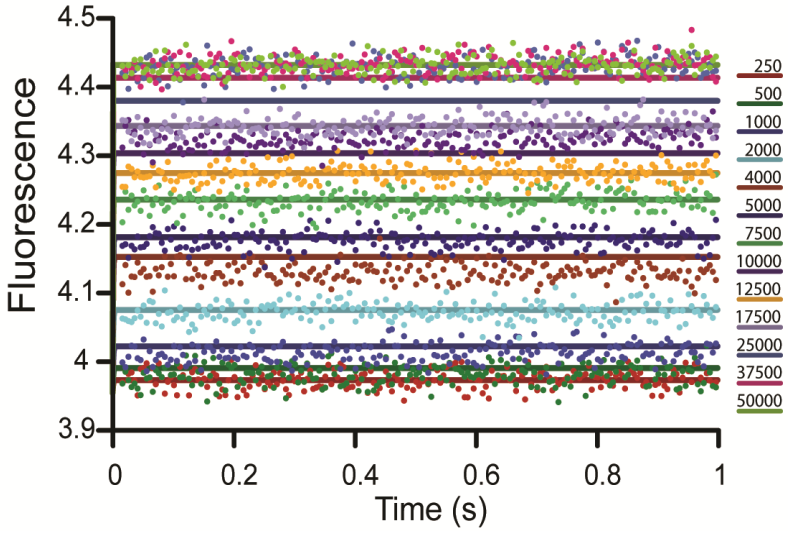


Figure 3.3: Bromide Binding to αY60W-CaaD.

A) The stopped-flow fluorescence traces (340 nm) of αY60W-CaaD with bromide (0-50000 μM) over 1 s. Each trace was fit to a line (slope = 0) and resulting constant values (C) were further analyzed as shown in Figure 3.4. Solid lines correspond to the fit by simulation. Results of the fit by simulation can be found in Table 3.1.

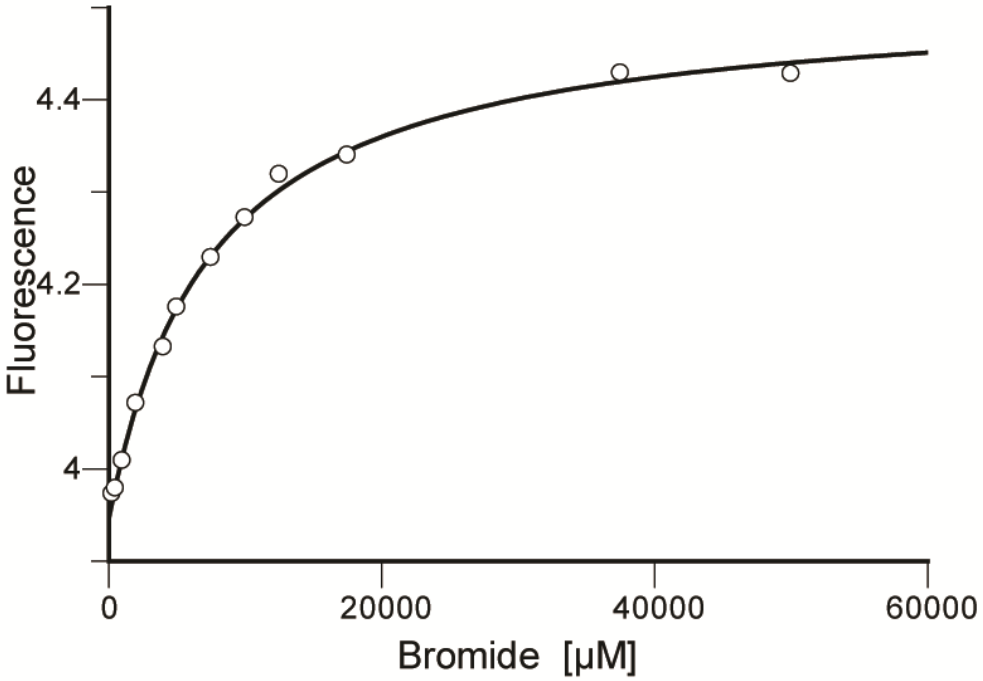


Figure 3.4: Plot of Constant Values (C) vs. Bromide Concentration with Fit to a Hyperbola

In addition to conventional fitting of data, the stopped-flow binding data were fit by simulation using KinTek Global Kinetic Explorer (Figure 3.3, solid lines).[5] Due to rapid equilibration between the αY60W-CaaD mutant and bromide ion (<1.3 ms), the k_{off} was held fixed at an arbitrarily fast rate (200 s^{-1}) where it didn't affect the simulated traces and fitting was based upon finding the optimal value for k_{on}. Accordingly, the data only define the equilibrium constant (K_d) for bromide binding to free enzyme, but not the individual on and off rate constants. Fitting by simulation also defined the fluorescence factor of 1.17 for enzyme-bound bromide, and a k_{on} affording calculation of a K_d of 12 mM. The values of k_{off} and the fluorescence factor for enzyme-bound bromide were used as constraints for subsequent global fitting of the reactions of **1** with αY60W-CaaD. A numerical summary of both fits (conventional and simulation) can be found in Table 3.1

Substrate	k_{on} (μM^{-1} s^{-1})	k_{off} (s^{-1})	K_d (μM)	F_0	ΔF	f/f_0 [a]
Bromide[b]	–	–	7400 ± 550	3.94 ± 0.01	0.56 ± 0.01	1.14 ± 0.02
	0.016 ± 0.002[c,d]	Fast[d]	12000[e] ± 250	3.95[f]	0.67 ± 0.12[c]	1.17 ± 0.03

[a]The value for f/f_0 was obtained using the following equation: $\frac{f}{f_0} = \frac{\Delta F + F_0}{F_0}$. Error ranges on these values were calculated using standard propagation of errors. [b]The first row is based upon the conventional analysis. The second row (shaded) is based from the individual fits to the data by simulation. [c]Error ranges were calculated from FitSpace analysis of the individual fit by simulation with a threshold of 5% deviation from the minimal SSE. [d]Due to the rapid equilibrium (<1.3 ms) between bromide ion and the αY60W mutant of CaaD, the data only defines the K_d for bromide binding to free *E*. This k_{on} affords the calculation of the defined K_d when the k_{off} is assumed to be 200 s^{-1}. [e]This value was calculated by dividing k_{off} / k_{on}. [f]These values were held fixed during the subsequent global fit by simulation.

Table 3.1. Summary of the Conventional Analysis and Individual Fit by Simulation of the Binding of Bromide Ion to αY60W-CaaD

3.3.2 Binding of Malonate Semialdehyde (2) to the αY60W-CaaD

Stopped-flow binding experiments were conducted with malonate semialdehyde, **2**, (0-5000 μM after mixing) and the αY60W-CaaD mutant (Figure 3.5, filled circles). Upon mixing αY60W-CaaD with **2**, a 16% increase in the fluorescence signal was observed (Figure 3.6), similar in magnitude to that observed for bromide. However, in contrast to the bromide binding (fast equilibration within the dead time), the binding of **2** occurs more slowly and requires ~100 ms to reach equilibrium. After approximately 0.5 s, the fluorescence signal decays, which is attributed to the instability of **2** and subsequent degradation to a chromophoric species, which absorbs strongly at 280 nm, thereby reducing the fluorescence signal (W.H. Johnson, Jr., G.K. Schroeder, J.P. Huddleston, and C.P. Whitman, unpublished observations, 2012). Data were fit using both conventional analysis (fitting to equations, Figure 3.5A) and simulation-based methods (fitting based on numerical integration of the rate equations, Figure 3.5B, solid lines). For the conventional fit, the initial phase of the time course (250 ms) for each trace was fit to a single exponential (Figure 3.5A, solid lines). A plot of the resulting first order rate constants versus the concentration of **2** is nearly linear and fitting to Eq 3.5 gave values of k_{on} = 0.025 $\mu M^{-1}s^{-1}$ and k_{off} = 28 s^{-1} (Figure 3.6A). These values provide an estimate of the K_d (k_{off} / k_{on}) for **2** of 1100 μM. Based on the end-point fluorescence titration fit to Eq. 3.1, the K_d for **2** was predicted to be 380 μM (Figure 3.6B). Fitting the data (250 ms) by simulation (Figure 3.5B, solid lines) gave a k_{on} of 0.028 $\mu M^{-1}s^{-1}$ and a k_{off} of 17 s^{-1} which affords the calculation of K_d ~ 600 μM with a fluorescence factor of 1.14 for

enzyme-bound to **2**. All values listed above are summarized in Table 3.2. The values for k_{on} and the fluorescence factor for enzyme bound to **2** were used as constraints for subsequent global fitting of the reactions of **1** with αY60W-CaaD, yielding refined estimates for k_{off} and K_d by accounting for the rate and amplitude of the reactions simultaneously.

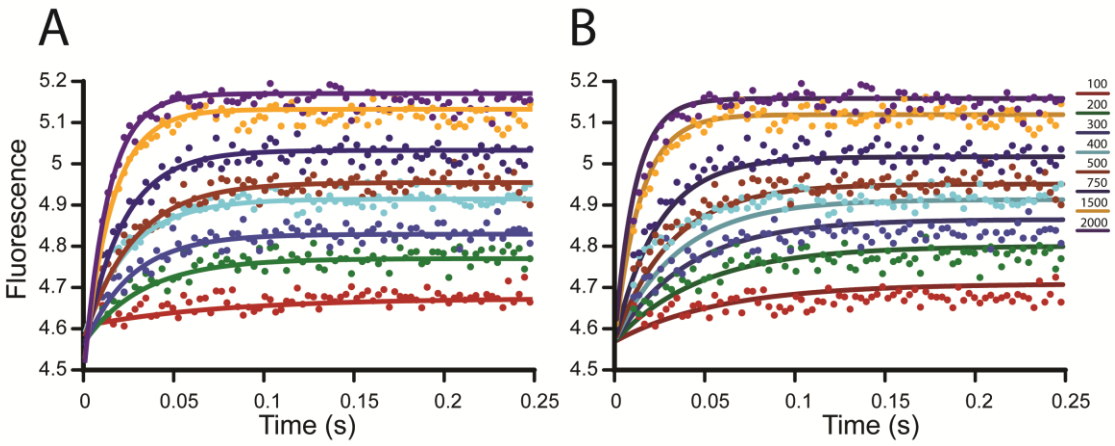


Figure 3.5: The Stopped-Flow Fluorescence Traces (250 ms) of αY60W-CaaD with Malonate Semialdehyde (**2**).

A) Solid lines represent a single exponential fit to the data. Resultant rate vs. concentration plots are shown in Figure 3.6. B) Solid lines represent the fit to the data by simulation. Rate constant values resulting from the fit by simulation are found in Table 3.2.

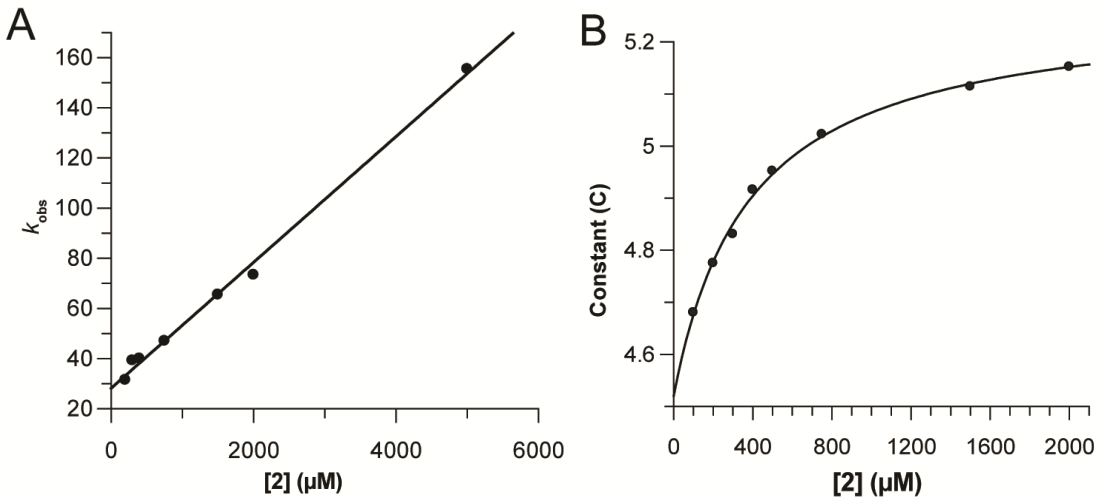


Figure 3.6: Conventional Analysis of the Binding of Malonate Semialdehyde (**2**) to the αY60W-CaaD Mutant.

A) Plot of the k_{obs} vs. [**2**] with fit to Eq 3.5 (solid line). B) Plot of constant values (C) vs. [**2**] with fit to Eq 3.4 (solid line). A summary of fitting results can be found in Table 3.2.

Substrate	k_{on} (μM^{-1} s^{-1})	k_{off} (s^{-1})	K_d (μM)	F_0	ΔF	f/f_0 [a]
2[b]	0.025 ± 0.001	28 ± 1.6	1120[c]/380[d]	4.52 ± 0.03	0.75 ± 0.02	1.16 ± 0.03
	0.028 ± 0.005[e]	17 ± 3[e]	600[f] ± 250	4.56[g]	0.64 ± 0.05[e]	1.14 ± 0.01

[a]The value for f/f_0 was obtained using the following equation: $\frac{f}{f_0} = \frac{\Delta F + F_0}{F_0}$. Error ranges on these values were calculated using standard propagation of errors. [b]The first row is based upon the conventional analysis. The second row (shaded) is based from the individual fits to the data by simulation. [c]K_d is calculated by dividing k_{off} / k_{on} using the values from the conventional analysis of the binding of **2** to αY60W-CaaD in a stopped-flow experiment (Figure 3.6A). [d]K_d is calculated from the conventional analysis of the saturation of the fluorescence signal in the stopped-flow experiment monitoring the binding of **2** to αY60W-CaaD (Figure 3.6B). [e]Error ranges were calculated from FitSpace analysis of the individual fit by simulation with a threshold of 5% deviation from the minimal SSE. [f]This value was calculated by dividing k_{off} / k_{on}. [g]This value was held fixed during the fit by simulation.

Table 3.2. Rates from the Conventional Analysis and Individual Fit by Simulation of the Binding of Malonate Semialdehyde (**2**) to the αY60W Mutant of CaaD

3.3.3 Global Fitting of the Data

Kinetic Model Development. A kinetic model was developed to allow global fitting of the full data set (KinTek Global Kinetic Explorer) in order to provide a more comprehensive test of the model. Conventional analysis (i.e. fitting to equations and plotting the observed rates versus concentration) was used to provide the initial kinetic model and estimates for some of the rate constants.[6] We initially fit the stopped-flow fluorescence data to a two-step model because the time dependence appeared to be adequately fit to a single exponential function, but showed a concentration dependence reaching a maximum rate of ~10 s^{-1}. However, the rate of the pre-steady state burst was approximately 100 s^{-1} (as shown in Chapter 2, Table 2.5), which defines the sum of the rates of chemistry and subsequent steps leading to release of product. These data imply that the chemistry step precedes the fluorescence change, and predicts that the fluorescence data should show a brief lag phase. Closer inspection of the stopped-flow fluorescence data revealed the data could be fit to a double exponential function (Figure 3.7A,B), but resolution of the lag phase was unreliable due to the large errors stemming from the low amplitude of the lag phase, and the inherent difficulties in fitting kinetic data to double exponential functions with arbitrary amplitude terms. In Figure 3.7C, we show a fit to the same data set based on computer simulation using the full kinetic model described below. The predicted lag phase is most prominent at the intermediate substrate concentrations (labeled with *). The simulated traces show the rate and the amplitude of the lag phase are consistent with the measured rates of the chemistry step, resulting in a more accurate representation of the lag phase kinetics based on fitting both experiments.

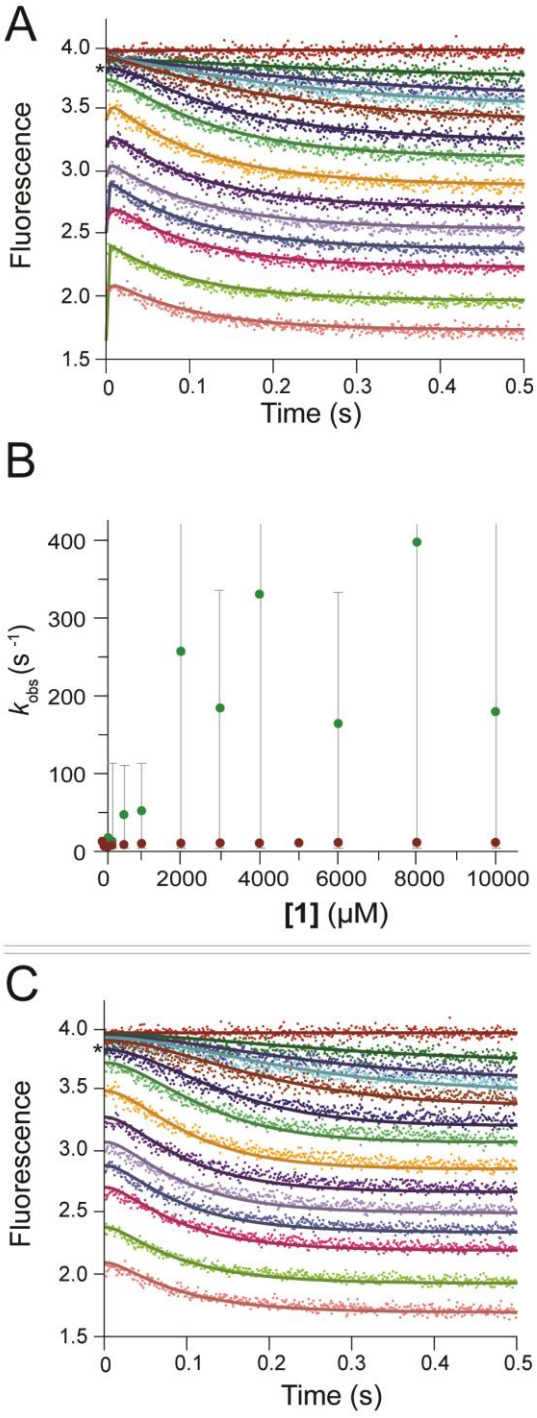
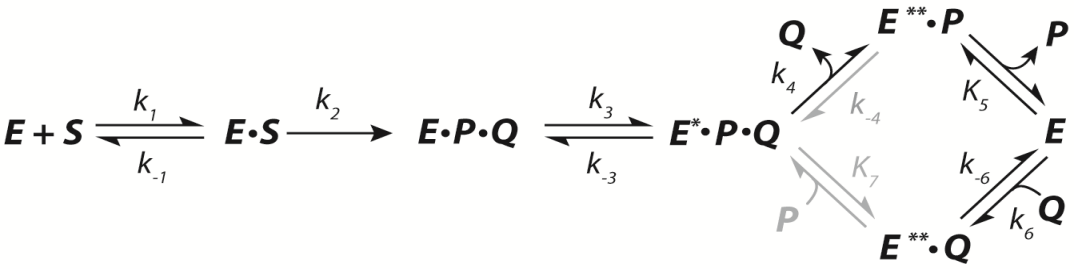

A
Fluorescence
Time (s)
B
k_{obs} (s^{-1})
[1] (μM)
C
Fluorescence
Time (s)

Figure 3.7: Fitting a Lag in the Stopped-Flow Fluorescence Data.

A) The stopped-flow fluorescence traces (0.5 s) of αY60W-CaaD with **1** were fit to a double exponential (Eq 3.1 with n = 2) yielding the solid lines shown. The concentrations of **1** are: 0, 50, 100, 150, 250, 550, 1000, 2000, 3000, 4000, 5000, 6000, 8000, 10000 μM. B) The observed rate constants (λ_1 and λ_2) from the fit to Eq 3.1 to the first 0.5 s of the stopped-flow data for the αY60W-CaaD are plotted versus the concentration of **1**. The faster rate constants are shown in green while the slower rate constants are shown in red. Shown in gray lines are the error bars on the first observable rate values. Note that some of these rates are off the *y*-axis and some of the error values exceed the rate constants values. Values for the slower rate constants and errors are similar to the rates observed from the single exponential fits shown in Chapter 2, Figure 2.6B. C) The stopped-flow fluorescence traces (0.5 s) of αY60W-CaaD with **1** fit by global simulation are shown in solid lines. The simulation lines are able to account for the lag phase significantly better than conventional methods. The concentrations are the same as listed above.

These results suggest a minimal model shown in Scheme 3.6, with a fluorescence change occurring after the chemistry step. Product release is proposed to follow a biased random pathway with kinetic preference for the upper branch of the pathway (see Discussion). Three distinct fluorescence states (*E*, E^* and E^{**}) were proposed based upon quantitative analysis of the kinetic and equilibrium titration data. Fitting of the data set using this model, and the features of the data requiring each element of the model will be discussed below.

Scheme 3.6: Minimal Kinetic Model Used for Global Fit



Global Fitting of Kinetic Data. The complete data set for the αY60W-CaaD mutant reacting with **1** was subjected to global data fitting using KinTek Global Kinetic Explorer[5] in order to rigorously test the model and provide a more comprehensive analysis of the lag phase in the stopped-flow fluorescence data relative to the observed rate of the pre-steady state burst. The global fit of the data according to Scheme 3.6 is shown in Figures 3.8A-I, where the smooth lines represent the global fit to the entire data set based upon Scheme 3.6 with the rate constants summarized in Table 3.3. The stopped-flow data at various concentrations of **1** (0-10,000 μM) are shown in Figures

3.8A-E. Figures 3.8A-B (see also Figure 3.7C) show the initial portion (100 ms) and first 1 s of the stopped-flow data, respectively, while Figure 3.8C shows the first 12 s of the reaction. Figures 3.8D-E (see also Figure 2.4C) show the change in fluorescence over the full-time course of the reaction (120 s and 300 s, respectively). Figure 3.8F (see also Figure 2.7B) shows the pre-steady state burst of bromide formation (using **1**) at various concentrations of αY60W-CaaD (250-600 μM). Figure 3.8G shows the time dependence for the binding of **2** to αY60W-CaaD (data from Figure 3.5A-B). This data set was not included explicitly in the global fit. However, the fit to this data by conventional analysis and simulation-based fitting of the single data set yielded a binding rate constant of ~0.025 s^{-1} (*vida supra*, Table 3.2 and Figure 3.6A); and this value was used for k_6 (Scheme 3.6) and held fixed during global fitting. Figure 3.8H (data from Figure 3.2) shows the dehalogenation of **1** by αY60W-CaaD in the presence of various concentrations of bromide (0-50,000 μM). Figure 3.8I (data from Figure 2.3) shows the full progress curves of the dehalogenation of **1**. A single fluorescence factor for enzyme bound product species ($E^{**}P$ and $E^{**}Q$ in Scheme 3.6) was defined based upon the conventional analysis and fits by simulation of the individual data sets; because these data were not included in the global analysis, the fluorescence factor was held constant at a value of 1.13 during global fitting, a value within the error ranges of each estimate (Table 3.1 and 3.2).

Figure 3.8

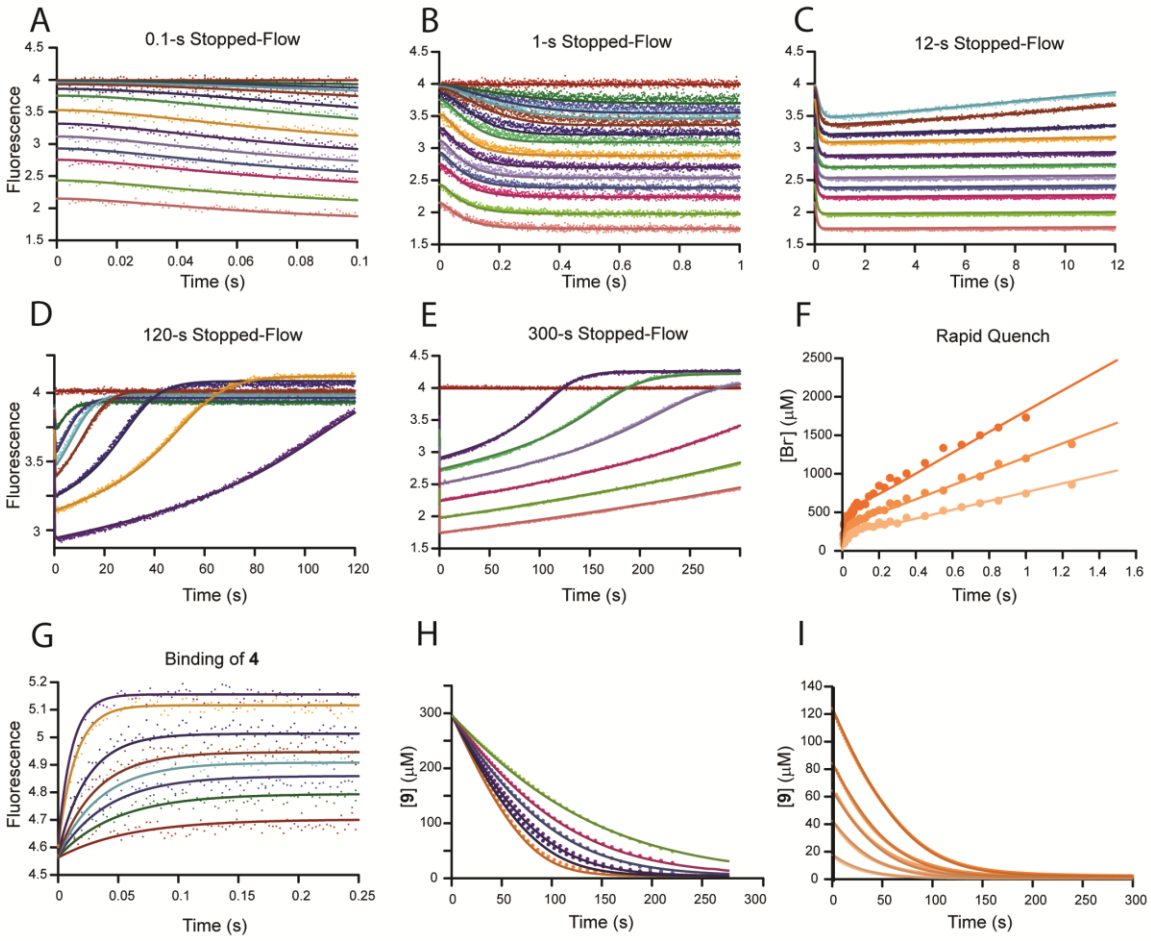

A
0.1-s Stopped-Flow
Fluorescence
Time (s)
B
1-s Stopped-Flow
Time (s)
C
12-s Stopped-Flow
Time (s)
D
120-s Stopped-Flow
Fluorescence
Time (s)
E
300-s Stopped-Flow
Time (s)
F
Rapid Quench
[Br⁻] (μM)
Time (s)
G
Binding of 4
Fluorescence
Time (s)
H
[9] (μM)
Time (s)
I
[9] (μM)
Time (s)

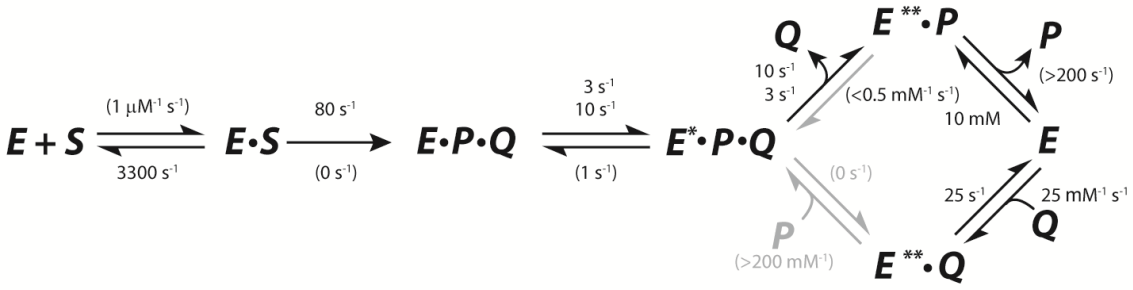
Figure 3.8: Global Fitting of Data Collected with αY60W-CaaD.

The αY60W-CaaD kinetic data with **1** and bromide (A-I) globally fit by simulation (solid lines) to the mechanism shown in Scheme 3.6 and (G) αY60W-CaaD kinetic data with **2** fit by simulation (solid lines) to give constraints on rates defined in Scheme 3.6 and Table 3.3.[5] (A-E) Stopped-flow fluorescence traces (0.1-300 s) for **1** (0-10000 μM) and αY60W-CaaD. Concentrations of **1** are: 0, 50, 100, 150, 250, 550, 1000, 2000, 3000, 4000, 5000, 6000, 8000, 10000 μM. Colors corresponding to the concentrations of **1** are maintained for graphs A-E. (F) Rapid-quench of αY60W-CaaD (200,400, and 600 μM) with **1**. (G) Stopped-flow fluorescence traces for **2** (100, 200, 300, 400, 500, 750, 1500, and 2000 μM) and αY60W-CaaD. (H) Full time progress curves following 300 μM of **1** at 224 nm with increasing concentrations of bromide (0, 5000, 10000, 20000, 30000, and 50000 μM). (I) Full time progress curves following the decrease of **1** (18, 42, 63, 83, and 125 μM) at 224 nm.

Data were not sufficient to define all of the rate constants in Scheme 3.6, so the less well-defined rate constants were assigned nominal values in order to fit the data to the full model. Because of the nature of the chemistry, we propose that the chemical reaction at the active site is essentially irreversible ($k_{-2} = 0$). The proposed conformational change occurring after chemistry (*vida infra*), was allowed to be reversible with the value of k_{-3} set at 1 s^{-1}, which represents a upper limit for k_{-3} based upon analysis by computer simulation; that is, in fitting the data we noted that any rate greater than 1 s^{-1} begins to affect the simulated traces, but rate constants below 1 s^{-1} have no effect, so any value less than or equal to 1 s^{-1} accounts for the data. The rate of bromide release from $E^{**}P$ appears to be rapid and therefore was held fixed at an arbitrarily fast rate (k_5 = 200 s^{-1}) during global fitting. Accordingly, the data only define the K_d for bromide, not the individual rate constants for binding and release.

Fitting with the constraints described above provides individual rate constants for each step in the model shown in Scheme 3.6 with error ranges shown in Table 3.3. Error ranges were provided by FitSpace[8] confidence contour analysis (Figure 3.9) and represent the upper and lower limits of a 3% threshold on the χ^2.[8] Global simulation of the data to the model shown in Scheme 3.6 with the constraints described above accounts for all aspects of the data and yields well-bounded rate constants. The final model with our best estimates for each rate constant is shown in Scheme 3.7. Numbers in parentheses represent rate constant values that were assigned nominal values and held fixed during fitting. Stacked values, 3 over 10 and 10 over 3, represent ambiguity in assigning these two rates to individual steps in the pathway, as discussed in detail below.

Scheme 3.7: Kinetic Model of CaaD with Rate Constants



Rate [a]	Lower Limit[b]	Upper Limit[b]	% Range[c]	Best fit	Rate	Lower Limit[b]	Upper Limit[b]	% Range[c]	Best fit
f/f_0[d]	0.708 (0.16)[f]	0.715 (0.23)[f]	0.5% (17%)[f]	0.710 (0.2)[f]	k_3[e]	5.0 (2.4)[f]	13.0 (4.0)[f]	41% (20%)[f]	9.6 s^{-1} (3.4 s^{-1})[f]
					k_4	2.6 (6.5)[f]	4.0 (11.4)[f]	21% (27%)[f]	3.3 s^{-1} (9.2 s^{-1})[f]
k_{-1}[g]	2600	4700	32%	3300 s^{-1}	$K_{d,Br}$[h]	9100	20000	43%	12500 µM
k_2	65	100	22%	78 s^{-1}	k_{-6}[i]	20	33	26%	25 s^{-1}

[a]The data were fit globally to the mechanism shown in Scheme 3.6. Chemistry and the initial release of **2** were assumed to be irreversible with reverse rates fixed at zero during global fitting. [b]The upper and lower limits reflect a threshold of 3% deviation from the minimal χ^2 in the confidence contours. FitSpace error confidence contours are shown in Figure 3.8.[8] [c]The percentage range was calculated by dividing the mean of the range by the best fit value as (upper-lower) / (2*best fit). This reflects the allowable variation of each best fit value as a percentage. [d]The fluorescence scaling factor is shown as the fractional change in enzyme fluorescence (see text). [e]Formation of E^*PQ was modeled as a reversible step associated with a conformational change in the enzyme. The value of k_{-3} was held fixed at 1 s^{-1} during global fitting (see text). [f]Rates for the alternate fit are provided (See Discussion 3.4). [g]The value for k_1 (substrate binding) was assumed to be fast and held fixed at 1 μM^{-1} s^{-1}, so fitting to derive k_{-1} defined only $K_{d,S}$ (see text). [h]The value for k_5 (release of bromide ion) was assumed to be fast (Figure 3.3) and held fixed at 200 s^{-1} (see text). As bromide binding is in rapid equilibrium with free enzyme, the data only defines the K_d for bromide [i]The value for k_6 was defined by fitting the binding of **2** to free enzyme (Figure 3.5 and Table 3.2) and held fixed at 0.025 μM^{-1} s^{-1} during global fitting (see text). Errors for this value estimated by FitSpace results to the binding of **2** to free enzyme can be found in Figure 3.8 inset. The value f/f_0 for the product bound enzyme species (E^{**}) was determined by fitting the products, **2** and bromide, binding to free enzyme, individually, (Table 3.1 and 3.2) and held fixed at 1.13 during global fitting (see text).

Table 3.3. Rates Constants Derived from the Global Analysis of the αY60W-CaaD Reaction with *trans*-3-Bromoacrylate (**1**).

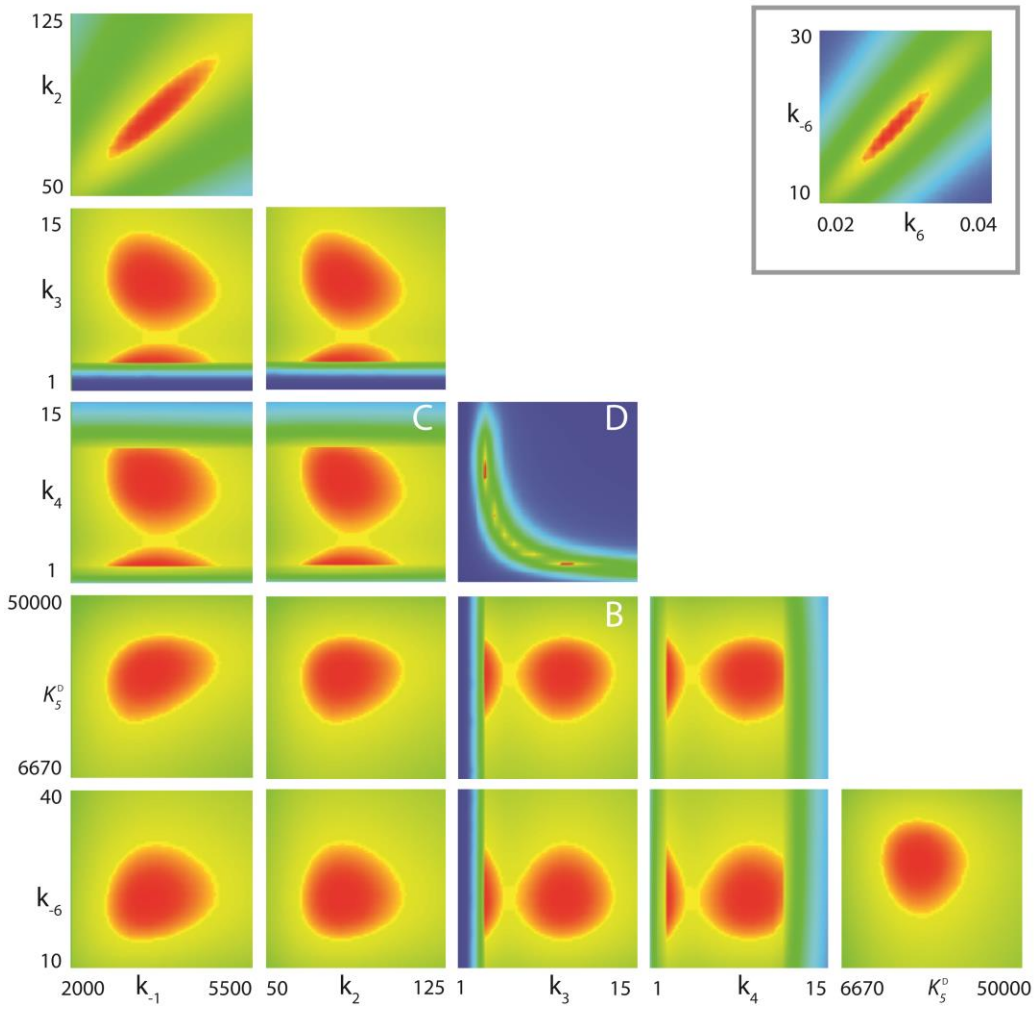


Figure 3.9. FitSpace Confidence Contours for the Global Fit.

FitSpace confidence contours for the global fit of the kinetic data shown in Figure 3.8A-I (αY60W-CaaD and **1**) to the model in Scheme 3.6.[8] The inset shows the FitSpace confidence contours for the fit by simulation of **2** binding to αY60W-CaaD (Figure 3.8G). The fluorescence factor of E^{**} was held fixed at 1.13. The letters B, C, and D (shown in white) refer to select results used in Figure 3.10.

3.4 DISCUSSION

Based on the kinetic similarities between the αY60W- and wild-type-CaaD enzymes, an in-depth analysis of the αY60W-CaaD data was pursued. Global fitting of the αY60W-CaaD data provided values for individual rate constants without many of the additional assumptions and deconvolution of complex rate functions required for conventional analysis.[5] Prior to global fitting (KinTek Global Kinetic Explorer), a minimal kinetic model was proposed based upon conventional analysis of the data.[5] The kinetic model (Scheme 3.6) used to account for the data was developed as follows. Observation of a pre-steady state burst in the rapid quench experiments implies that a step after chemistry is at least partly rate limiting. Fitting the burst data as shown in Chapter 2 (Section 2.3.4) defines the rate of the burst (≈ 85 s^{-1}) and the rate of steady state turnover, governed by k_{cat} (≈ 3 s^{-1}). Analysis of the pre-steady state stopped-flow data provided estimates for the rate constants governing a possible conformational change occurring at a rate slower than chemistry.[6,7] The stopped-flow fluorescence data could be fit to a double exponential decay with a fast phase (lag) occurring with a rate comparable to the rate of the pre-steady state burst of product formation. Although conventional analysis failed to provide reliable estimates for the rate constant governing the lag phase, simulation-based fitting supported the existence of a lag phase consistent with duration of the rate of chemistry. It is useful to note that the lag will be most noticeable when the rates governing the fast and slow phases are comparable, which should occur at a concentration of approximately 500 μM substrate according to the fitted parameters. This

is evident in the data, as marked by an asterisk (*) in Figures 3.7A and C. At higher concentrations of substrate, the lag phase becomes less distinct. This makes resolution of the lag by conventional fitting difficult, but the trend is included as part of the global fitting, which reinforces confidence in the results.

The slower phase of the fluorescence transients occurred at a maximum rate of ~10 s^{-1}, which defines the sum of all rate constants contributing to the final approach to the steady state,[6] including the rate constant that limits steady state turnover. Because chemistry is irreversible, the actual reverse of the fluorescence change occurs via the release of products to return the enzyme to the free E state. Accordingly, fitting the concentration dependence of the rate of the fluorescence change as shown in Chapter 2 (Section 2.3.3) gives $k_{for} = k_3 = 8\ s^{-1}$ and $k_{rev} = k_{cat} = k_4 = 3\ s^{-1}$.

Estimation of the equilibrium constant ($K_{d,S}$) for substrate binding was based on a rapid equilibrium assumption ($k_{-1} >> k_2$) in globally fitting the entire data set. In particular, the $K_{d,S}$ ($1/K_1$) was defined by the concentration dependence of the fluorescence transients and the need to account for k_{cat} and K_m values derived from the full time course kinetic traces. Although there was no observable lag in the burst data, suggesting a rapid equilibrium binding of the substrate prior to chemistry, this conclusion is inexact, but reasonable given the relatively weak apparent K_d (~ 400 μM). By assigning a second order rate constant of $k_1 = 1\ \mu M^{-1}\ s^{-1}$, where $k_{-1} >> k_2$, fitting the data to derive k_{-1} affords a simplistic calculation of the $K_{d,S} = k_{-1}/1\ \mu M^{-1}\ s^{-1}$. This value k_1 is significantly larger than the lower limit defined by k_{cat}/K_m ($0.01\ \mu M^{-1}\ s^{-1}$).

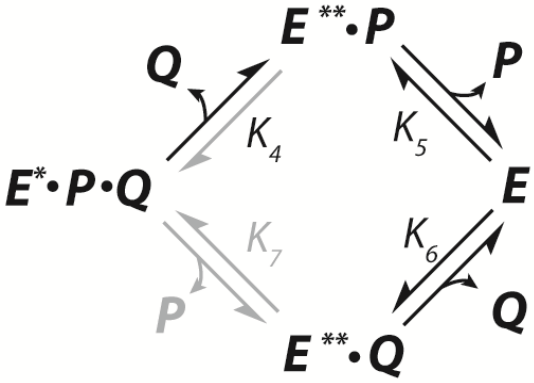
Because the rate of the fluorescence change is slower than chemistry, the formation of species $E^{*}PQ$ is proposed to occur *after* chemistry so that the binding and chemistry steps occur without a change in fluorescence. That is, the lag phase in the fluorescence transients can be accurately accounted for in the simulated traces by assigning the same fluorescence to all enzyme species through the chemistry step (i.e., *E*, *ES*, and *EPQ*). The maximum burst rate ($\approx 85\ s^{-1}$) provides an initial estimate of the rate for chemistry (k_2), but is actually the sum of the rate constants governing formation and decay of the *EPQ* complex.

Two product release steps are required to regenerate free enzyme, which is suggested to occur by a biased, random release pathway with kinetic preference for the upper branch of the pathway. This assignment was made as follows. As described above, bromide (*P*) is a competitive inhibitor ($K_i \sim 12$ mM). This might suggest an ordered product release pathway with loss of *Q* preceding the release of *P*.[15] However, malonate semialdehyde (*Q*, **2**) was shown to bind to free enzyme 20-fold more tightly than bromide (Table 3.2, $K_d \sim 600$ µM), perhaps suggesting a random pathway as shown in Scheme 3.6. Although a random pathway predicts a mixed inhibition pattern for both products, the apparent competitive inhibition pattern for *P* could be explained if the binding of *P* to $E^{**}\bullet Q$ is too weak to have an observable effect. However, the data suggest a kinetic preference for the upper pathway. This can be illustrated using thermodynamic box arguments based on the available product binding measurements.

The global fitting model shown in Scheme 3.6 incorporates two assumptions (shown in gray) regarding product release: 1) malonate semialdehyde (*Q* in Scheme 3.6)

release from $E^{*}PQ$ is largely irreversible, and 2) binding of P to $E^{**}Q$ is too weak or too slow to generate $E^{*}PQ$ (i.e., shows competitive inhibition, not mixed). The validity of these two assumptions can be demonstrated in the product release model shown in Scheme 3.8.

Scheme 3.8: Product Release Model



In Scheme 3.8, K_4 is the dissociation constant for release of Q from $E^{*}PQ$, K_5 is the dissociation constant for release of P from $E^{**}Q$, K_6 is the dissociation constant for release of Q from $E^{**}P$, and K_7 is the dissociation constant for release of P from $E^{*}PQ$. The product of all equilibrium constants contained in a thermodynamic cycle must be equal to unity. In other words, the product of constants on one side of the pathway must be equal to the product of the constants on the other side (Eq. 3.6):

$$K_4^D \bullet K_5^D = K_6^D \bullet K_7^D \qquad \text{Eq 3.6}$$

Rearrangement of Eq. 3.6 to make the binding of *Q* to $E^{**}P$ versus *E* equal to the binding of *P* to $E^{**}Q$ versus *E* gives Eq. 3.7:

$$\frac{K_4^D}{K_6^D} = \frac{K_7^D}{K_5^D} \qquad \text{Eq. 3.7}$$

Experimental data provide approximations of the values for K_6^D (~0.6 mM) and K_5^D (~12 mM^{-1}). Substituting these values in their respective positions and rearranging yields Eq. 3.8:

$$\frac{K_4^D}{0.6} = \frac{K_7^D}{12} \qquad \text{Eq. 3.8}$$

Rearranging Eq. 3.8 yields Eq. 3.9:

$$K_7^D = 20 \bullet K_4^D \qquad \text{Eq 3.9}$$

Eq. 3.8 and 3.9 demonstrate that the dissociation constants for K_4^D and K_7^D are proportional. As a result, Eq. 3.9 shows that *Q* binds to species $E^{**}P$ (K_4^D) approximately 20 times tighter than *P* binds to species $E^{**}Q$ (K_7^D). Therefore, the assumptions associated with the product release and rebinding in Scheme 3.6 are consistent with a competitive inhibition pattern for *P*, according to the upper (preferred) pathway.

Unfortunately, the product, *Q* (**2**) is too unstable to carry out more accurate product inhibition studies to probe product release in detail.

The preceding analysis also agrees with the docking results of **1** to CaaD, which indicate that substrate appears to bind with the halide oriented toward the back of the active site pocket (Figure 2.2A), suggesting that the physical release of bromide would occur second. In addition, an ordered product release model, with halide release as the final step, was proposed previously for the related enzyme *cis*-CaaD based on similar experimental observations.[7] Therefore, we prefer a biased-random product release pathway favoring the release of *Q* (**2**) first (slow) followed by bromide (fast), but accounting for the rebinding of *Q* to free enzyme as one leg of a random pathway. There is no direct evidence for the release of *P* from $E^{*}PQ$ or the binding of *P* to E^{**}•*Q*, so this reaction step is shown in gray in Scheme 3.6.

In Scheme 3.6, the formation of a new fluorescent state of the enzyme ($E^{*}PQ$) occurs after chemistry (k_2). Comparable rates of formation for $E^{*}PQ$ were observed for wild-type (Chapter 2, Figures 2.5 and 2.6B) and αY60W, suggesting that the slow formation for $E^{*}PQ$ is not an artifact caused by the αY60W mutation. The E^{**}•*P* and E^{**}•*Q* states accounted for an increase in the fluorescence endpoint (~13%) relative to that of the free enzyme when either product (i.e**., 2** or bromide) was bound to the enzyme (Tables 3.1 and 3.2).

Global fitting to the model shown in Scheme 3.6, constrained by the above information, yielded estimates for the individual rates constants for each step of the kinetic mechanism (summarized in Table 3.3). A more accurate error analysis for the

individual rate constants was obtained using FitSpace[8], which indicates that all of the rates are constrained as shown in Figure 3.9. Moreover, the kinetic parameters derived by global simulation provide reasonable estimates in agreement with the Michaelis-Menten parameters obtained from steady state initial velocity measurements. Using the numbers in Table 3.3 and Scheme 3.6, we calculate k_{cat} = 2.2 s^{-1} and K_m = 95 μM. The experimentally determined values for k_{cat} and K_m are 3.0 s^{-1} and 70 μM, respectively. Additionally, the calculated K_d values for bromide and **2** binding to free enzyme are in good agreement with those predicted from the individual fits by simulation and conventional analysis (Tables 3.1, 3.2, and 3.3).

Inspection of the FitSpace confidence contours (Figure 3.9) highlights various limitations of the data set and provides a more realistic estimation of errors on the rate constants (~20%, Table 3.3) compared to conventional data analysis.[6,8] For example, conventional analysis of the stopped-flow fluorescence data based upon fitting to exponential functions yields two rate constants for k_3 and k_4 (Scheme 3.6), which are ~8 s^{-1} and ~3 s^{-1}, respectively. Global fitting using these values as initial estimates for the rate constants k_3 and k_4, respectively, results in a fit shown in Figure 3.8 (with normal error estimates on all rate constants of <1%). However, FitSpace analysis revealed that an equivalent fit to the data can be obtained when the values of k_3 and k_4 are reversed (i.e. k_3 = 3 s^{-1} and k_4 = 10 s^{-1}), with a corresponding adjustment of the fluorescence scaling factor (f/f_0) to compensate for changes in the predicted transient concentration of E^*PQ. This is illustrated in Figure 3.10A, which depicts the parameter variation among k_3, k_4, and f/f_0 (the fluorescence scaling factor) required to maintain a low χ^2 value (within 3% of the

minimum). Figure 3.10A shows that as the value for k_3 increases from 3 s^{-1} to 12 s^{-1}, the value of k_4 decreases from 12 s^{-1} to 3 s^{-1}, following a hyperbolic relationship. Conversely, as the value for k_3 increases from 3 s^{-1} to 12 s^{-1}, the value for f/f_0 increases from 0.15 to 0.70 following an inverse hyperbolic relationship in which k_4 decreases. This suggests that the data can be fit equally well (maintaining an acceptably small χ^2) when k_3 and k_4 are defined as any appropriate combination of rates (i.e., following the hyperbolic relationship) from 3 s^{-1} to ~12 s^{-1} by proper adjustment of the fluorescence factor value (f/f_0). However, when comparing k_3 or k_4 to other rate constants, (e.g., k_3 vs. $K_{d,Br}$, Figure 3.10B or k_2 vs. k_4, Figure 3.10C) two sets of rates centered around ~3 s^{-1} and ~10 s^{-1} emerge as acceptable values. Therefore, all FitSpace confidence contours describing k_3 or k_4 in relation to other rate constants have two regions of minimum χ^2 (red spots) centered around 3 s^{-1} and 10 s^{-1}. In the direct comparison of k_3 vs. k_4, (Figure 3.10D) the hyperbolic relationship in the parameter variation is apparent, with two primary regions of a minimum χ^2 occurring when k_3 and k_4 are defined as 10 s^{-1} and 3 s^{-1}, or 3 s^{-1} and 10 s^{-1}, respectively (dashed lines in Figure 3.10D). Also, Figure 3.10D shows that when k_3 and k_4 divert from these rate constants (i.e., the hyperbolic relationship), the χ^2 increases rapidly, resulting in large areas of blue (45× the minimum χ^2).

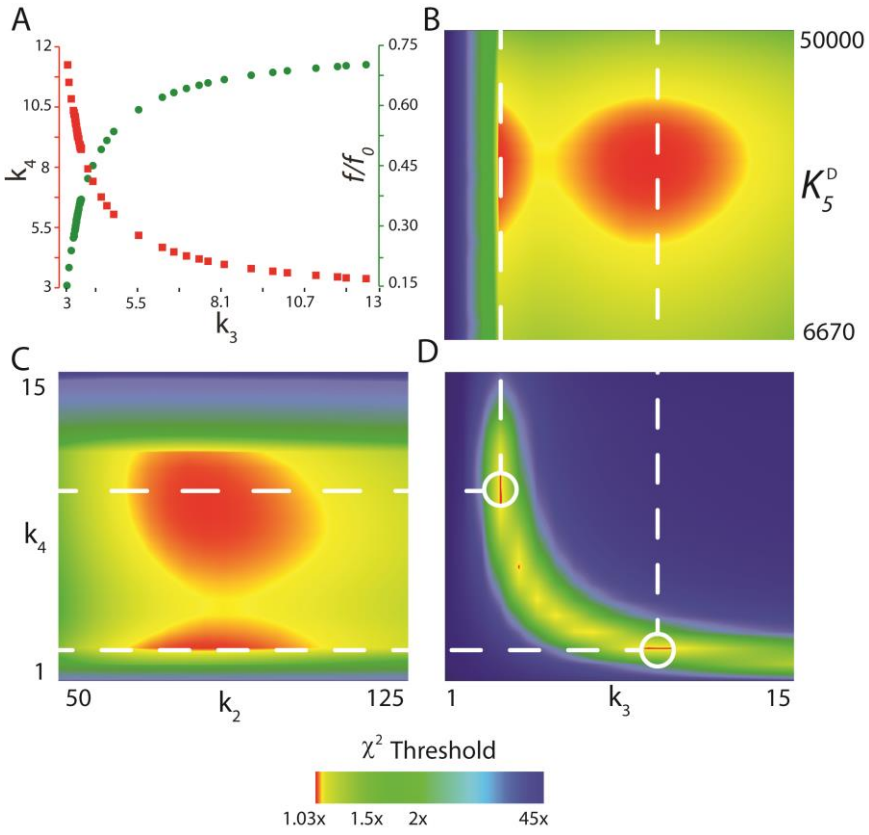


Figure 3.10: Selected Confidence Contour Plots.

Selected FitSpace confidence contours for the global fit to the kinetic data shown in Figure 3.7 (αY60W-CaaD and **1**). (A) Parameter variation for k_4 (■, left axis) and f/f_0 (●, right axis) plotted as a function of the value of k_3 required to maintain an SSE value within 3% of the minimum χ^2. FitSpace confidence contours for (B) k_3 vs. $K_{d,Br}$ (C) k_2 vs. k_4, and (D) k_3 vs. k_4. The dashed lines illustrate two sets of values for k_3 and k_4 that yield equivalent fits to the data (within 3% of the minimum χ^2 value). Graphs were generated using gnuplot 4.6.

The uncertainty in the order of these two rate constants is primarily due to the fact that the data lack explicit information to quantify the fluorescence factor for the species, E^*PQ. For example, a fast/slow sequence predicts a higher concentration of E^*PQ requiring a lower scaling factor to achieve the same fluorescence amplitude, compared to the slow/fast sequence of rate constants. This observation emphasizes the inability of kinetic analysis to provide the order of the two sequential rates in the absence of data to define the absolute amplitude of the intermediate species (E^*PQ in this case). Conventional methods of determining standard error drastically underestimate the error ranges for these rate constants because error analysis is based upon only exploring the area around a local minimum and the covariance analysis fails to reveal nonlinear relationships between fitted parameters.[8] The global fit to all the data with the confidence contour analysis reveals two acceptable fits to the data.

Although the data cannot define the order for k_3 and k_4, they do provide strict lower and upper limits for both rate constants. The lower limit for both k_3 and k_4 is 2.5 s^{-1} defined by k_{cat} and depicted in the FitSpace confidence contours by a rapid increase in χ^2 at rate constants below 2.5 s^{-1}. The upper limits for these rate constants are controlled by the range of allowable adjustments in the fluorescence factor values (f/f_0) (Figure 3.10A). The upper limit for k_3 is ~13 s^{-1} and is well constrained (Figure 3.10B). As k_3 increases, f/f_0 increases and begins to approach unity yielding a smooth, well defined upper limit (Figure 3.10B). Conversely, like the lower limits, the upper limit on k_4 is strictly defined. As seen in Figure 3.10A, as k_4 increases, f/f_0 decreases exponentially, reaching zero when

k_4 is ~12 s^{-1}. At all values above 12 s^{-1}, the χ^2 drastically increases, resulting in a strict upper limit for k_4 in the FitSpace confidence contours (Figure 3.10C). However, even at these limits, the values of k_3 and k_4 are within 5-fold of one another. This suggests that both the formation of E^*PQ and release of Q (**2**) are partially rate limiting.

A previous report indicated that product release is not rate limiting.[16] Horvat *et al*. based this conclusion on the observation that increasing solvent viscosity with trehalose did not decrease the values of k_{cat} or k_{cat}/K_m, based upon the simplistic analysis where viscosity effects probe the extent to which diffusion limits the observed rate.[16] At first glance, the observation of a burst of product formation appears at odds with this conclusion. However, the observation of a burst only implies that a step *after* chemistry, not *necessarily* product release *per se* (e.g., a conformational change), is rate limiting.[6] In the case of CaaD, product release appears to be only partially rate limiting, which may have produced the observed results of the viscosity experiments. Moreover, the release of *Q* from the E^*PQ state is accompanied by a change in enzyme state (to $E^{**}P$) as indicated by the different fluorescence level. Although the structural basis for the fluorescence change is not known with certainty, it is reasonable to propose that a change in enzyme structure precedes and limits the net rate of release of *Q*, defining k_{cat}. Therefore, the viscosity effect would have failed to reveal the rate-limiting product release, governed by a change in enzyme structure. It is reasonable to suppose that many slow product release steps are governed by rate-limiting conformational change steps that lead to rapid product release and weaken product rebinding. Such conformational change steps occurring before and after chemistry may provide a critical component of enzyme specificity.[17] Our

experimental results also provide a more accurate estimate of the rate enhancement for the CaaD reaction (k_2/k_{non} = 3.6 x 10^{13}), a value that is an order of magnitude greater than previously reported.[16] Our analysis also suggest caution in interpreting the implications of viscosity effects.

Our results raise the question about the nature of the partially rate-limiting conformational change steps following chemistry. One possible explanation is gleaned from the crystal structures and reproduced in the docking results (presented in Chapter 2, Section 2.3.1), which provide molecular details for the observed conformational change. In the crystal structure of covalently modified CaaD (1S0Y), αGlu52 is observed in this "closed" conformation, while in another structure of CaaD (3EJ3) without a ligand, αGlu52 is found in the "open" conformation.[18,19] In the docking poses with **1**, αGlu52 is reproduced in these two positions (Figures 3.11A and B, respectively).[20] It is reasonable that αGlu52 might have to rotate out of the active site after participating in the chemistry in order to facilitate product release. A possible conformational change, with measurable changes in fluorescence (E^*PQ), might rotate αGlu52 out of the active site so the products, **2** followed by bromide, can be released from the enzyme. This proposed conformational change could partially limit the rate of turnover. It is also reasonable to speculate that the efficiency of this enzymes can be optimized (perhaps by evolution) by accelerating the rate of the conformational change.

Mutation of a nearby asparagine residue (βAsn39to βAla39) had an unexpected, but significant effect on the kinetics of CaaD.[20] The βN39A mutant of CaaD shows

essentially no catalytic activity (k_{cat} < 0.01 s^{-1}) with **1**, despite being located more than 5 Å away from the amino-terminal proline.[20] One possibility for this observation is suggested by the docking study. Although βAsn39 is far from the amino-terminal proline, it is within hydrogen bonding distance to αGlu52. As suggested above, movement of αGlu52 might play a key role in both chemistry and product release. Regardless of the position of αGlu52, "open" or "closed", βAsn39 remains within hydrogen bonding distance of αGlu52 (dotted lines in Figures 3.11A and B, respectively). Based on the results of the kinetic from the βN39A mutant of CaaD, it appears that βAsn39 may be critical in guiding and positioning the αGlu52 residue for chemistry.[20] The breaking of the hydrogen bond formed in the "closed" conformation (Figure 3.11B) to allow for product release further supports the observed partially rate limiting conformational change with measurable changes in fluorescence.

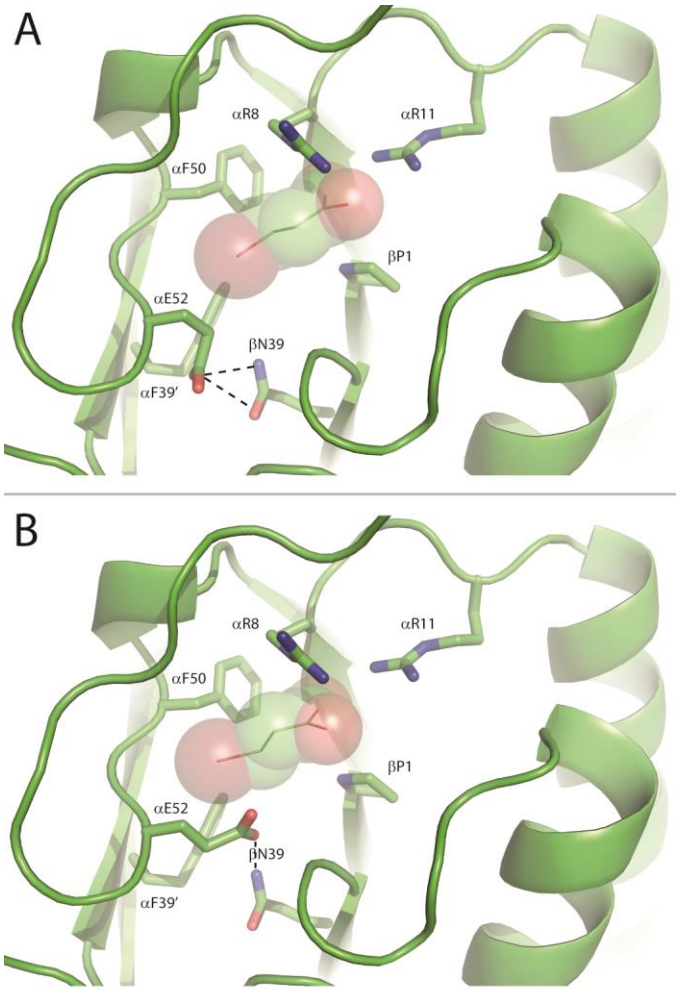


Figure 3.11: Results of the Docking Studies of CaaD with *trans*-3-bromoacrylate (**1**).

A) The "open" conformation pose where the carboxylate of αGlu52 points out of the active site. B) The "closed" conformation pose where the carboxylate group of αGlu52 is positioned to carry out the proposed chemistry. In both cases, the βAsn39 is within hydrogen bonding distance (shown in dotted lines) to the carboxylate moiety of αGlu52. The "open" conformation is predicted to be slightly favored conformation over the "closed" conformation with **1**.

The results reported here largely parallel those found for *cis*-CaaD. The kinetic data for the *cis*-CaaD reaction using *cis*-3-bromoacrylate are best accommodated by a five-step kinetic model[11] in which substrate binding is followed by a rapid isomerization to a new fluorescent state. Chemistry occurs next yielding an enzyme complex containing both products with no further change in fluorescence. Subsequent product release from the complex is ordered and rate-limiting: malonate semialdehyde is released first (and is rate-limiting), followed by the rapid release of bromide. The model suggested only two states for the enzyme, which could be associated with an open state and a closed state observed in the absence or presence of ligands, respectively. A comparison of the unliganded and liganded *cis*-CaaD structures provides a structural basis for this hypothesis.[11,21]

The similarities between CaaD and *cis*-CaaD include the ordered product release of malonate semialdehyde followed by bromide. Bromide appears to be a competitive inhibitor for both enzymes and the dissociation of malonate semialdehyde is at least partially rate determining for each enzyme. The only differences are the nature of the conformational change (movement of an active site group in CaaD after chemistry vs open and closed states in *cis*-CaaD) and in the product inhibition. Malonate semialdehyde binds more tightly to CaaD than bromide (~10-fold), whereas bromide binds more tightly to *cis*-CaaD than malonate semialdehyde (>10-fold). These differences do not explain the greater economy of the CaaD-catalyzed reaction (compared with that of the *cis*-CaaD reaction). This remains an active area of investigation, which can now be pursued based upon the foundation reported here.

3.5 REFERENCES

(1) Shanmugam, S. *Enzyme Technology*; I.K. International Publishing House Pvt. Limited: New Delhi, 2009.

(2) Johnson, K. A.; Goody, R. S. (2011) The original Michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry. 50,* 8264-8269.

(3) Johnson, K. A. (2013) A century of enzyme kinetic analysis, 1913 to 2013. *FEBS Lett. 587,* 2753-2766.

(4) Johnson, K. A. (2009) Fitting enzyme kinetic data with KinTek Global Kinetic Explorer In *Methods Enzymol.* (Johnson, L. M., Brand, L., Eds.) pp 601-626, Elsevier,

(5) Johnson, K. A.; Simpson, Z. B.; Blom, T. (2009) Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem. 387*, 20-29.

(6) Johnson, K. A. (1992) Transient-state kinetic analysis of enzyme reaction pathways In *The Enzymes* (Sigman, D. S., Ed.) 3rd Ed. pp 1-61, Academic Press, San Diego

(7) Schroeder, G. K.; Johnson, W. H.; Huddleston, J. P.; Serrano, H.; Johnson, K. A.; Whitman, C. P. (2012) Reaction of *cis*-3-chloroacrylic acid dehalogenase with an allene substrate, 2,3-butadienoate: hydration via an enamine. *J. Am. Chem. Soc. 134*, 293-304.

(8) Johnson, K. A.; Simpson, Z. B.; Blom, T. (2009) FitSpace Explorer: An algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal. Biochem. 387*, 30-41.

(9) Wang, S. C.; Person, M. D.; Johnson, W. H., Jr.; Whitman, C. P. (2003) Reactions of *trans*-3-chloroacrylic acid dehalogenase with acetylene substrates: consequences of and evidence for a hydration reaction. *Biochemistry. 42*, 8762-8773.

(10) Waddell, W. J. (1956) A simple ultraviolet spectrophotometric method for the determination of protein. *J. Lab. Clin. Med. 48*, 311-314.

(11) Robertson, B. A.; Schroeder, G. K.; Jin, Z.; Johnson, K. A.; Whitman, C. P. (2009) Pre-steady-state kinetic analysis of *cis*-3-chloroacrylic acid dehalogenase: analysis and implications. *Biochemistry. 48*, 11737-11744.

(12) Dixon, M. (1953) Determination of enzyme-inhibitor constants. *Biochem. J. 55*, 170-171.

(13) Cornish-Bowden, A. (1974) Simple graphical method for determining the inhibition constants of mixed, uncompetitive, and noncompetitive inhibitors. *Biochem. J. 137*, 143-144.

(14) Poelarends, G. J., Johnson, H. W., Jr., Murzin, A.G., and Whitman, C.P. (2003) Mechanistic characterization of a bacterial malonate semialdehyde decarboxylase: identification of a new activity in the tautomerase superfamily. *J. Biol. Chem. 278*, 48674-48683.

(15) Segel, I. H. *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems*; Wiley Classics Library: New York, 1993.

(16) Horvat, C. M.; Wolfenden, R. V. (2005) A persistent pesticide residue and the unusual catalytic proficiency of a dehalogenating enzyme. *Proc. Natl. Acad. Sci. USA. 102*, 16199-16202.

(17) Johnson, K. A. (2008) Role of induced fit in enzyme specificity: a molecular forward/reverse switch. *J. Biol. Chem. 283*, 26297-26301.

(18) De Jong, R. M.; Brugman, W.; Poelarends, G. J.; Whitman, C. P.; Dijkstra, B. W. (2004) The x-ray structure of *trans*-3-chloroacrylic acid dehalogenase reveals a novel hydration mechanism in the tautomerase superfamily. *J. Biol. Chem. 279*, 11546-11552.

(19) Pegan, S. D.; Serrano, H.; Whitman, C. P.; Mesecar, A. D. (2008) Structural and mechanistic analysis of *trans*-3-chloroacrylic acid dehalogenase activity. *Acta Crystallogr. Sect. D. Biol. Crystallogr. D64*, 1277-1282.

(20) Poelarends, G. J.; Serrano, H.; Huddleston, J. P.; Johnson, W. H.; Whitman, C. P. (2013) A mutational analysis of active site residues in *trans*-3-chloroacrylic acid dehalogenase. *FEBS Lett. 587*, 2842-2850.

(21) De Jong, R. M.; Bazzacco, P.; Poelarends, G. J.; Johnson, W. H., Jr.; Kim, Y. J.; Burks, E. A.; Serrano, H.; Thunnissen, A.-M. W. H.; Whitman, C. P.; Dijkstra, B. W. (2007) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase. Structural basis for substrate specificity and inactivation by *(R)*-oxirane-2-carboxylate. *J. Biol. Chem. 282*, 2440-2449.

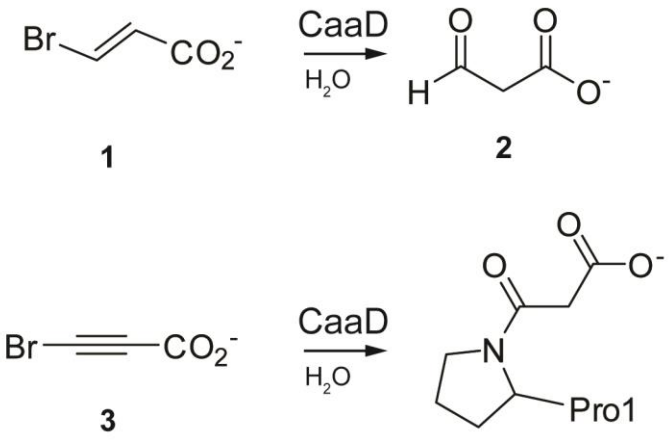
3.6 SUPPLEMENTAL

3.6.1 Introduction

Global fitting by simulation of the data generated by the αY60W mutant using *trans*-3-bromoacrylate (**1**) yielded a kinetic mechanism with six individual rate constants (Scheme 3.7). However, certain features of the model were not well defined. For example, the data doesn't explicitly define the fluorescence factor for the enzyme species designated as E^*PQ. This is mostly due to the similar magnitudes of the rate constants that govern its formation and decay. These rates are designated as k_3 and k_4, respectively, in Scheme 3.7. As a result, ambiguous values are obtained for the rates constants, k_3 and k_4, as discussed in Section 3.4, although limits on these values could be defined. This problem is inherent to this system.

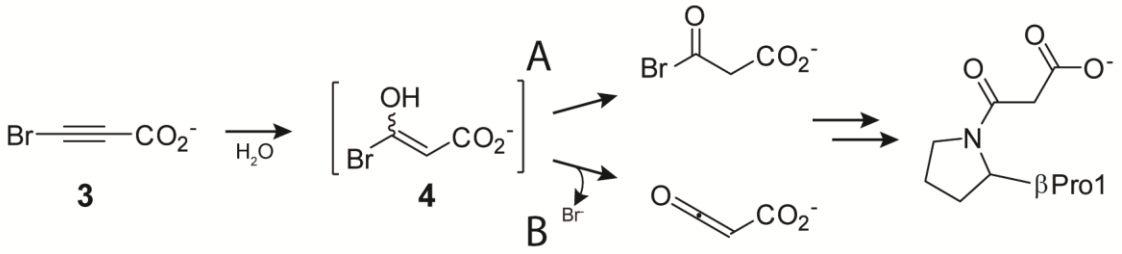
One potential strategy to circumvent this problem is to find an alternate system where the fluorescence factor for E^*PQ can be defined. Studying the reaction with other CaaD mutations or with alternate substrates could generate data that define the fluorescent factor. With the fluorescence factor in hand, the ambiguity in the rate constants for the αY60W mutant of CaaD with **1** can be resolved. One problem with alternate systems is that the fluorescent factor might not be the same.[1]

Scheme S3.1: Reactions of CaaD



As discussed in Chapter 1, Section 1.5 (Scheme S3.1), 3-bromopropiolate (**3**) is a mechanism-based inhibitor of CaaD that covalently labels βPro-1. Although the exact mechanism of action is unclear, it is proposed that **3** undergoes hydration to form an acyl bromide (Path A) or ketene (Path B). Either species can result in covalent modification of Pro-1 (Scheme S3.2). It was shown that CaaD is inactivated by **3** in less than 10 s when mixed in stoichiometric amounts.[2] Moreover, the inhibition of CaaD was shown to be protected by the addition of 2-oxo-3-pentynoate (2-OP). To increase our knowledge about the inactivation of CaaD by **3**, the reaction of the αY60W mutant of CaaD with **3** was followed using stopped-flow enzyme fluorescence. The resulting data were fit by both conventional methods and computer simulation.

Scheme S3.2: Proposed Mechanism of Inhibition by 3-Bromopropiolate (3)



3.6.2 Results

Following the same methodology outlined in Chapter 2, Section 2.2.6, various concentrations of **3** (80, 160, 240, 400, 1000, 2000, 4000, 10000 μM, initial concentrations) were mixed 1:1 with 20 μM (initial concentration) of the αY60W mutant of CaaD in 100 mM Na_2HPO_4 buffer, pH 8.1. The reaction was followed for 1 and 10 s. The 1 s data shows a decrease in the fluorescence signal followed by a slow rise to a fluorescent signal higher than free enzyme alone. After 10 s, the fluorescence signal remains unchanged at all concentrations. The resulting traces were initially fit to the double exponential function shown in Eq. 3.2 with n = 2. The resulting rates (λ_1 and λ_2) were plotted versus the concentration of **3**. These concentration-dependence plots were fit using conventional methods. The initial rate (λ_1) linearly increases with increasing concentrations of **3**. The second rate (λ_2) shows a hyperbolic dependence as defined by Eq. 2.3 (Figure S3.1). These results suggest a minimal irreversible two-step model with k_1 = 0.014 ± .0002 mM^{-1} s^{-1} and k_2 = 2.43 ± 0.2 s^{-1} (Scheme S3.3). There are two observable fluorescent species during the course of the reaction, designated here as E^*PR and $E'PR$. In this model, we designate *P* as bromide product and *R* as the reactive species that covalently labels the enzyme (Scheme S3.2). This is the model (Scheme S3.3) used for fitting of the data by computer simulation. The result of the fit by computer simulation is shown in Figure S3.2. The FitSpace results suggests that all the parameters, including the two fluorescent factors are well-constrained by the data set (Figure S3.3). The resultant rate constants, fluorescent factors and error ranges are summarized in Table S3.1 and Scheme S3.4.

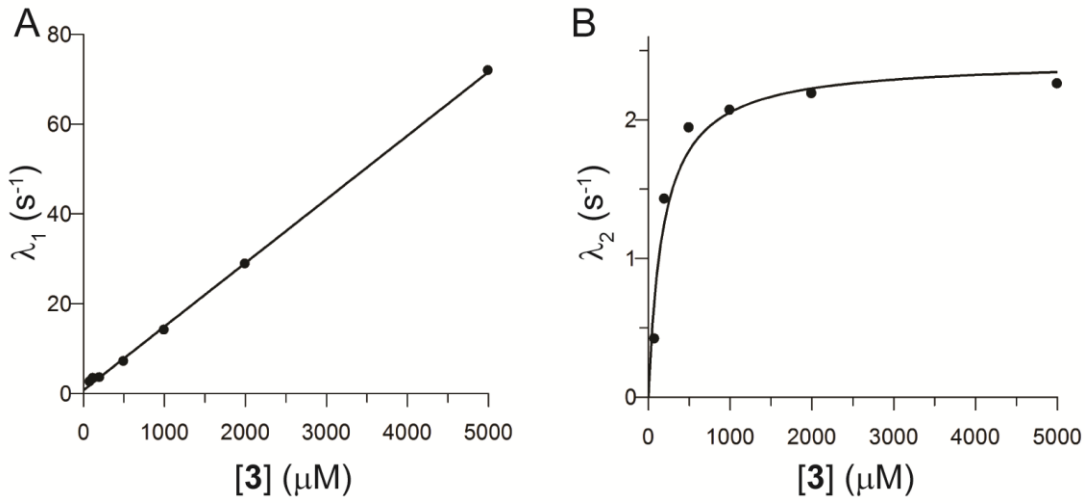


Figure S3.1: Conventional Analysis of Stopped Flow Fluorescence Transients Observed After Mixing 3-Bromopropiolate (**3**) with αY60W-CaaD.

A) Plot of λ_1 vs. **3** with fit to a line (Eq. 3.5). B) Plot of λ_2 vs. **3** fit to Eq 2.3 (solid line). The norminal values from these fits are described in the text.

Scheme S3.3: Kinetic Model Used for the Fit by Simulation.

$$E + S \xrightarrow{k_1} E^{*}\bullet P\bullet R \xrightarrow{k_2} E^{\prime}\bullet P\bullet R$$

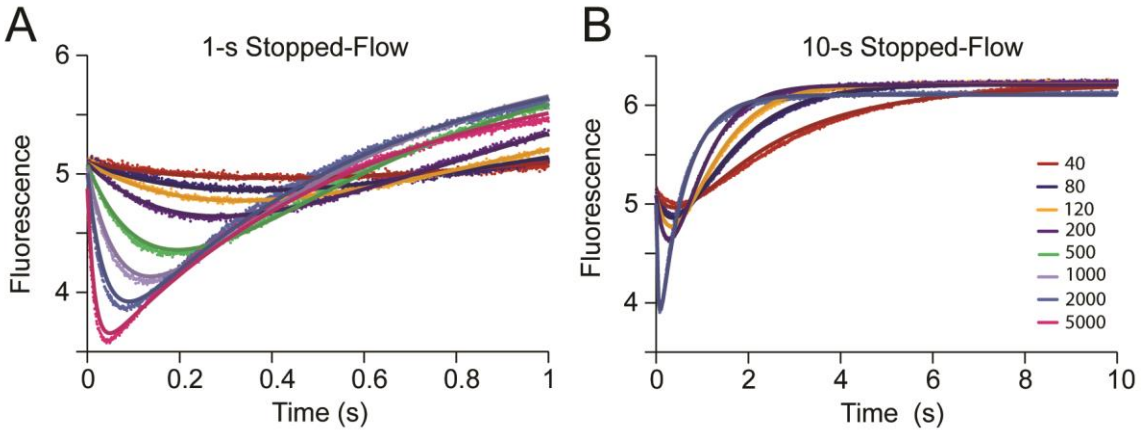


Figure S3.2: Global Data Fitting of 3-Bromopropiolate (**3**) with αY60W-CaaD.

The αY60W-CaaD kinetic data with **3** globally fit by simulation to the mechanism shown in Scheme S3.3) Stopped-flow fluorescence traces (1 s) for **3** (0-5000 μM) and αY60W-CaaD. B) Stopped-flow fluorescence traces (10 s) for **3** (0-2000 μM) and αY60W-CaaD. Concentrations of **3** are shown in the legend. Colors corresponding to the concentrations of **3** are maintained across both graphs. Solid lines are the fits to the data by simulation. A summary of the numerical values is shown in Table S3.1

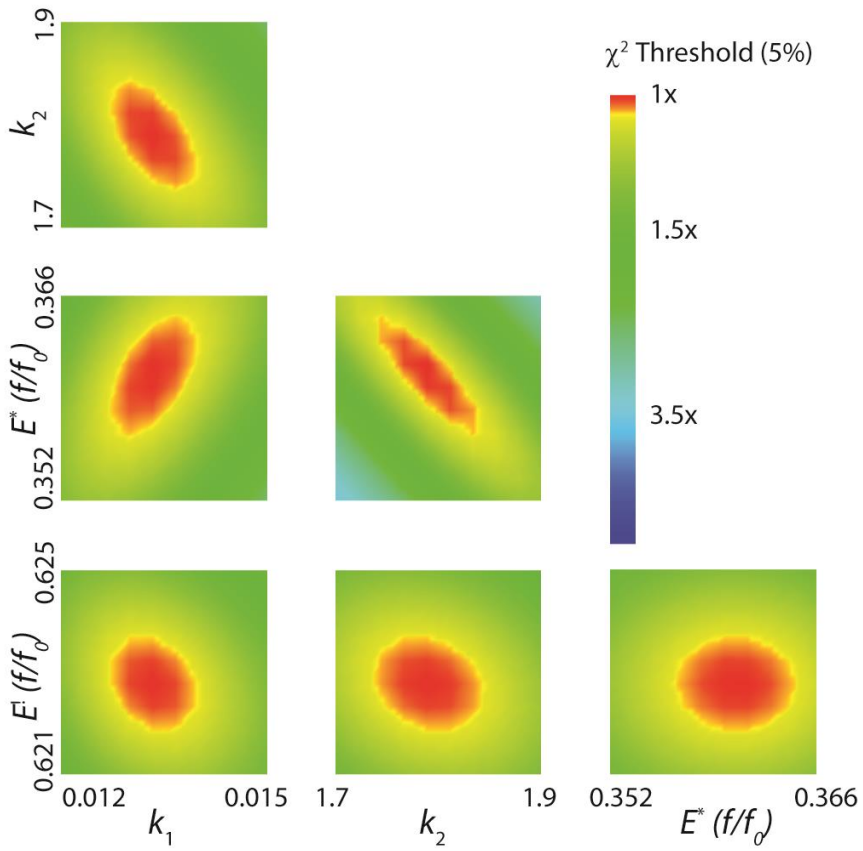


Figure S3.3: FitSpace Confidence Contours for the Global Fit for αY60W-CaaD and 3-Bromopropiolate (**3**)

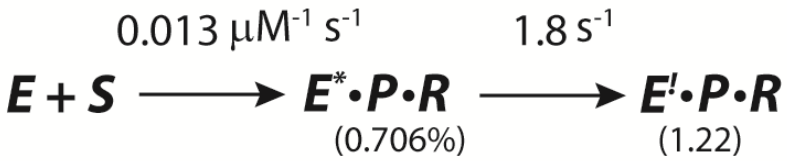
FitSpace confidence contours for the global fit of the kinetic data shown in Figure S3.2 (αY60W-CaaD and **3**) to the model in Scheme S3.3. Included in the FitSpace are the two observed fluorescence factors for enzyme species designated E^* and E'.

Rate [a]	Lower Limit[b]	Upper Limit[b]	% Range[c]	Best fit
k_1	0.0124	0.0136	4.6%	0.013 μM^{-1} s^{-1}
k_2	1.74	1.82	2.2%	1.78 s^{-1}
$E^*(f/f_0)$[d]	0.700	0.716	1.1%	0.706
$E^!(f/f_0)$[d]	1.220	1.224	0.17%	1.22

[a]The data are fit globally to the model shown in Scheme S3.3. [b]The upper and lower limits reflect a threshold of 5% deviation from the minimal χ^2 in the confidence contours. FitSpace error confidence contours are shown in Figure S3.3 [c]The percentage range was calculated by dividing the mean of the range by the best fit value as (upper-lower) / (2*best fit). This reflects the allowable variation of each best fit value as a percentage. [d]The fluorescence scaling factor is shown as the fractional change in enzyme fluorescence.

Table S3.1. Rate Constants and Fluorescence Factors Derived from the Fit by Simulation of the αY60W-CaaD with 3-Bromopropiolate (**3**).

Scheme S3.4: Final Kinetic Model of αY60W-CaaD using 3-Bromopropiolate (3)



If it is assumed that the observed fluorescent factor for enzyme species designated E^*PR in the reaction with **3**, is the same as E^*PQ in the reaction with *trans*-3-bromoacrylate (**1**), the rate ambiguities of k_3 and k_4 can be resolved. Given that the f/f_0 value for E^*PQ is 0.706 (the best fit value using **3**) and re-fitting the dataset with **1**, the best fit values for k_3 and k_4 are determined to be 9.05 s^{-1} and 3.36 s^{-1}, respectively. The resulting FitSpace is shown in Figure S3.4. This can be directly compared to the FitSpace shown in Figure 3.9. Visually, the fit by simulation is the same as shown in Figure 3.8. The numerical values and error ranges are shown in Table S3.2. Under the same assumptions outlined for fitting in Chapter 3, Section 3.3.3, all the fitted parameters are well-constrained.

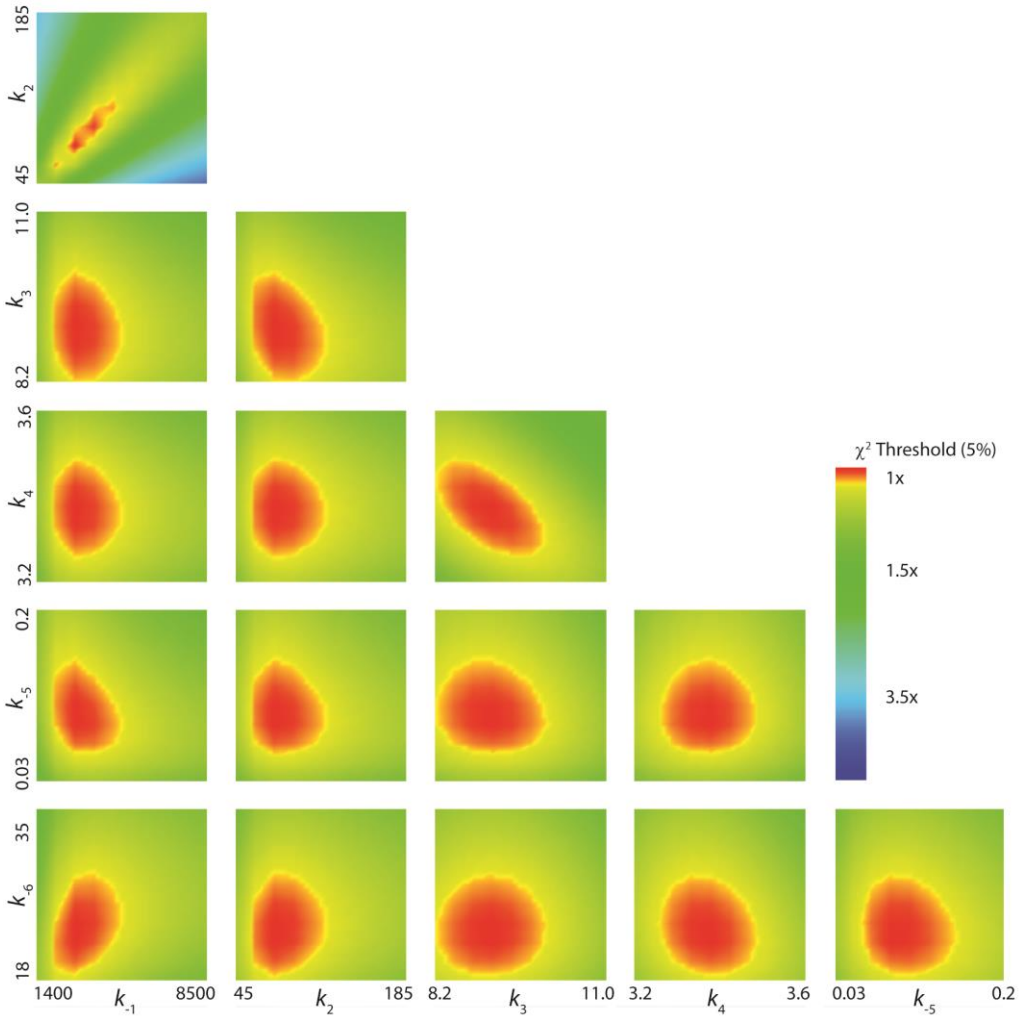


Figure S3.4: Refined FitSpace Confidence Contours for the Global Fit of αY60W-CaaD and *trans*-3-bromoacrylate (**1**)

FitSpace confidence contours for the global fit of the kinetic data shown in Figure 3.8A-I (αY60W-CaaD and **1**) to the model in Scheme 3.7 with the fluorescence factor for the enzyme species E^*PQ held fixed at 0.706, the value determined above (Table S3.1). Visually, the fit remains unchanged from Figure 3.8. A summary of the numerical rate constants is shown in Table S3.2

Rate [a]	Lower Limit[b]	Upper Limit[b]	% Range[c]	Best fit	Rate	Lower Limit[b]	Upper Limit[b]	% Range[c]	Best fit
k_{-1}[d]	2600	5000	36%	3340 s^{-1}	k_4	3.3	3.5	3%	3.37 s^{-1}
k_2	62	117	35%	78 s^{-1}	$K_{d,Br}$[f]	7200	22000	55%	13333 μM
k_3[e]	8.44	9.78	7%	9 s^{-1}	k_{-6}[g]	17	27	23%	22 s^{-1}

[a]The data are fit globally to the model shown in Scheme 3.7. Chemistry and the initial release of malonate semialdehyde were assumed to be irreversible with reverse rates fixed at zero during global fitting. [b]The upper and lower limits reflect a threshold of 5% deviation from the minimal χ^2 in the confidence contours. FitSpace error confidence contours are shown in Figure S3.4. [c]The percentage range was calculated by dividing the mean of the range by the best fit value as (upper-lower) / (2*best fit). This reflects the allowable variation of each best fit value as a percentage. [d]The value for k_1 (substrate binding) was assumed to be fast and held fixed at 1 μM^{-1} s^{-1}, so fitting to derive k_{-1} defined only $K_{d,S}$ [e]Formation of E^*PQ was modeled as a reversible step associated with a conformational change in the enzyme. The value of k_{-3} was held fixed at 1 s^{-1} during global fitting. [f]The value for k_5 (release of bromide ion) was assumed to be fast (Figure 3.3) and held fixed at 200 s^{-1}. As bromide binding is in rapid equilibrium with free enzyme, the data only define the K_d for bromide. [g]The value for k_6 was defined by fitting the binding of malonate to free enzyme (Figure 3.5 and Table 3.2) and held fixed at 0.025 μM^{-1} s^{-1} during global fitting. Errors for this value estimated by FitSpace results to the binding of malonate semialdehyde to free enzyme can be found in Figure 3.8 inset. The value f/f_0 for the product bound enzyme species (E^{**}) was determined by fitting the products, malonate semialdehyde and bromide, binding to free enzyme, individually, (Table 3.1 and 3.2) and held fixed at 1.13 during global fitting (see text). The value f/f_0 for the enzyme species (E^*) was held fixed at a value of 0.706 (see text, Table S3.1).

Table S3.2. Rates Constants Derived after Refinement of the Collected *trans*-3-Bromoacrylate (**1**) Data with the αY60W Mutant of CaaD.

The number of assumptions (i.e., fixing of values) can be further reduced by combining both datasets (**1** and **3**) into a single global fit. The model used for fitting is shown in Scheme S3.5. In this model we designate *S1* to be **1**, *S2* to be **3**, *P* to be bromide, *Q* to be **2**, and *R* to be the reactive species that covalently binds Pro1. There are three observed fluorescent species of the enzyme E^{*}, E^{**}, and $E^{!}$. The values shown in Scheme S3.5 in parenthesis were held fixed at the shown values during fitting. A forward only arrow assumes that the reverse arrow was held fixed at a value of "zero". This model results in well-constrained values for nine rate constants and all three fluorescent factors. The FitSpace and summary of numerical values are shown in Figure S3.5 and Table S3.3, respectively.

Scheme S3.5: Complete Model Used for Simultaneous Fitting of Both Datasets Collected with αY60W-CaaD.

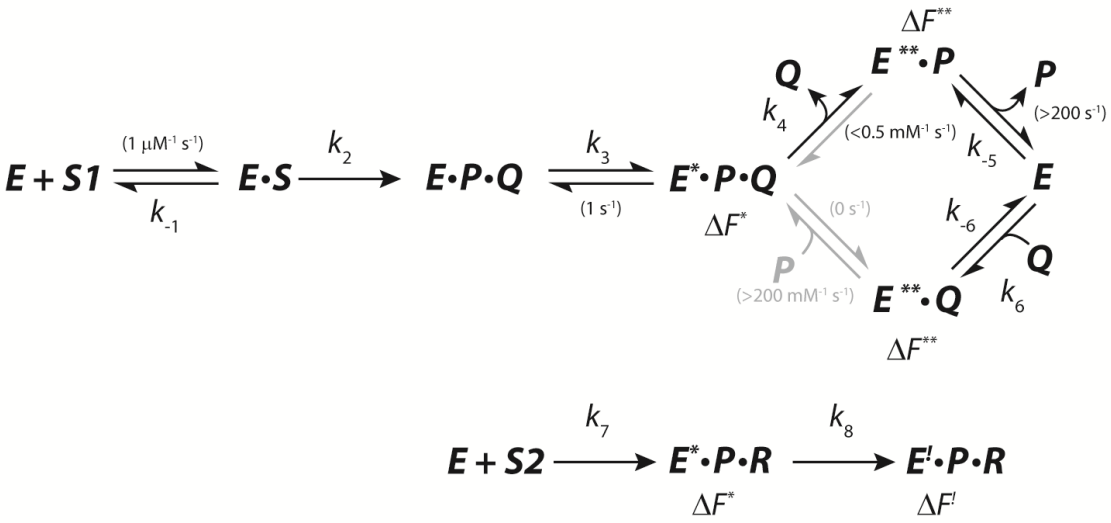


Figure S3.5

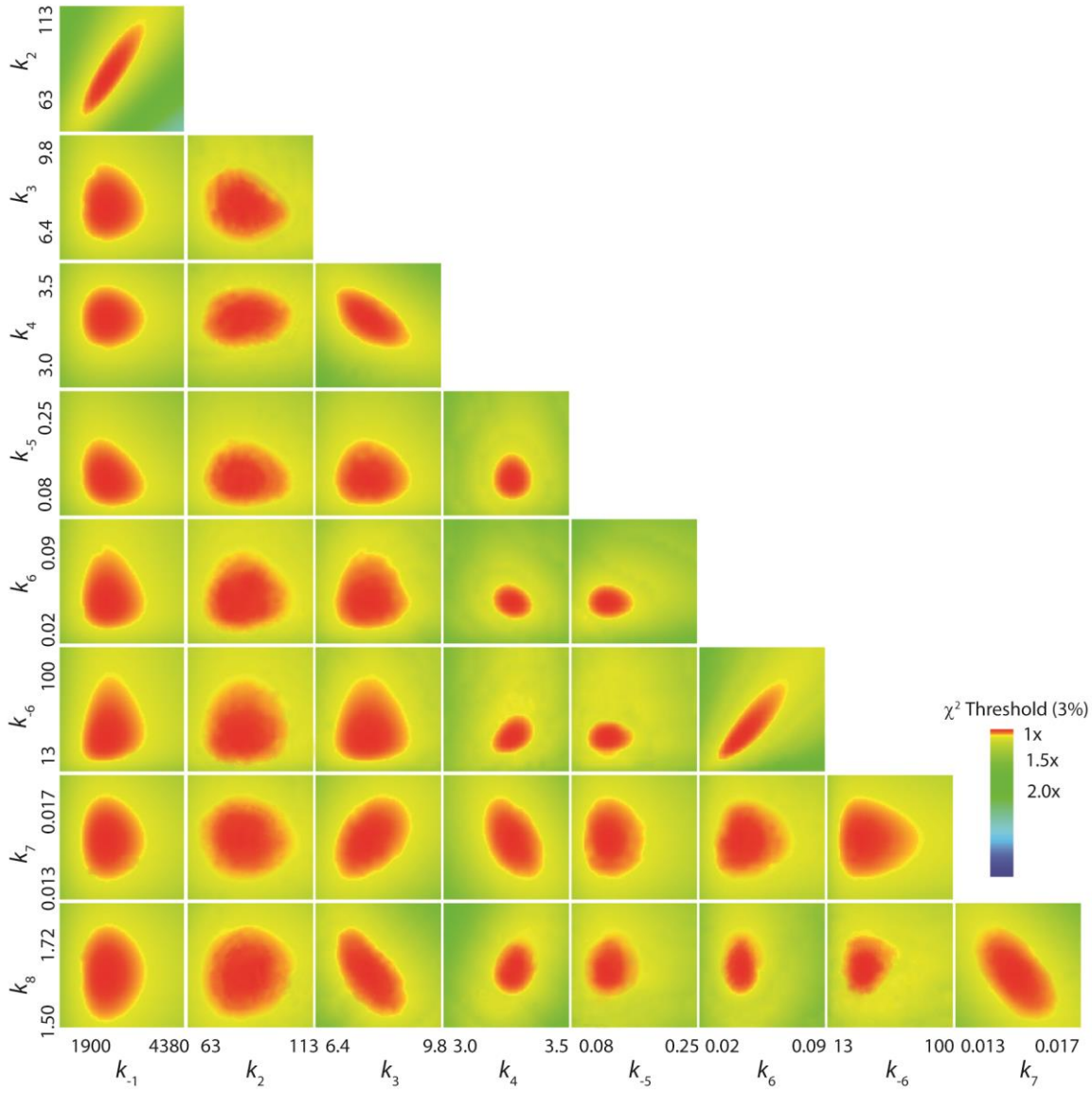

k_2 113 63
k_3 9.8 6.4
k_4 3.5 3.0
k_{-5} 0.25 0.08
k_6 0.09 0.02
k_{-6} 100 13
k_7 0.017 0.013
k_8 1.72 1.50
1900 4380 k_{-1}
63 113 k_2
6.4 9.8 k_3
3.0 3.5 k_4
0.08 0.25 k_{-5}
0.02 0.09 k_6
13 100 k_{-6}
0.013 0.017 k_7
χ^2 Threshold (3%)
1x
1.5x
2.0x

Figure S3.5: FitSpace Confidence Contours for the Simultaneous Global Fit of αY60W-CaaD with *trans*-3-Bromoacrylate (**1**) and 3-Bromopropiolate (**3**)

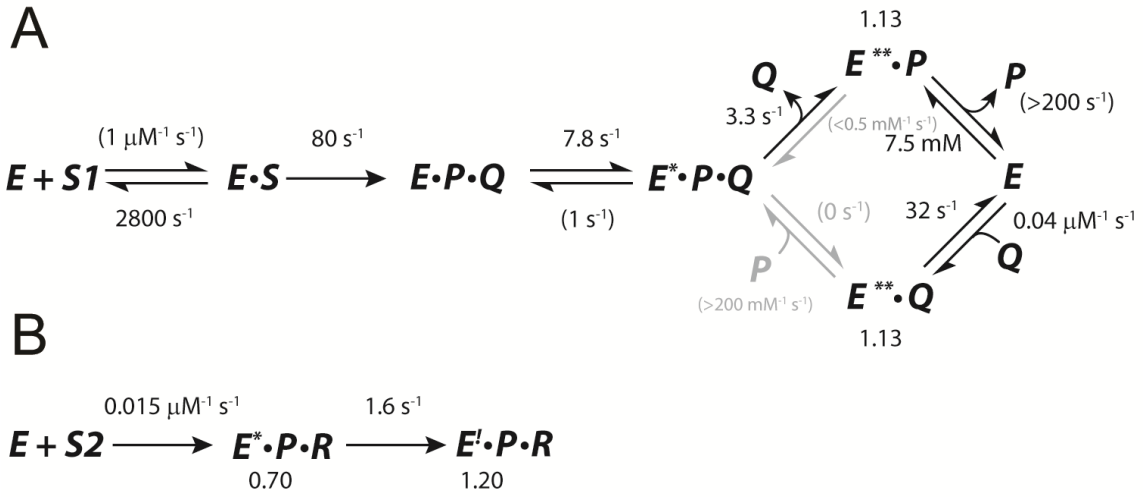
FitSpace confidence contours for the global fit of the kinetic data shown in Figure 3.8A-I (αY60W-CaaD and **1**) and Figure S3.1 (αY60W-CaaD and **3**) to the model in Scheme S3.5. Visually, the fits (not shown) only change slightly from Figure 3.8 and remain unchanged from the fits shown in Figure S3.2. A summary of the numerical rate constants are shown in Table S3.3

Rate [a]	Lower Limit[b]	Upper Limit[b]	% Range[c]	Best fit	Rate	Lower Limit[b]	Upper Limit[b]	% Range[c]	Best fit
k_{-1}[d]	2180	3830	30%	2765 s^{-1}	k_{-6}	19.9	61.6	66%	31.5 s^{-1}
k_2	68	102	22%	79 s^{-1}	k_7	0.014	0.016	6.6%	0.015 μM^{-1} s^{-1}
k_3[e]	7.1	8.9	12%	7.8 s^{-1}	k_8	1.54	1.7	5%	1.62 s^{-1}
k_4	3.13	3.53	6%	3.3 s^{-1}	$(f/f_0)^*$	0.683	0.708	1.7%	0.695
$K_{d,Br}$[f]	4850	10050	35%	7460 µM	$(f/f_0)^{**}$	1.12	1.15	1.3%	1.13
k_6[g]	0.032	0.066	40%	0.042 μM^{-1} s^{-1}	$(f/f_0)'$	1.200	1.203	0.1%	1.202

[a]The data are fit globally to the model shown in Scheme S3.5. The assumptions used in the fitting are discussed in the text. [b]The upper and lower limits reflect a threshold of 3% deviation from the minimal χ^2 in the confidence contours. FitSpace error confidence contours are shown in Figure S3.5. [c]The percentage range was calculated by dividing the mean of the range by the best fit value as (upper-lower) / (2*best fit). This reflects the allowable variation of each best fit value as a percentage. [d]The value for k_1 (substrate binding) was assumed to be fast and held fixed at 1 μM^{-1} s^{-1}, so fitting to derive k_{-1} defined only $K_{d,S}$ [e]Formation of E^*PQ was modeled as a reversible step associated with a conformational change in the enzyme. The value of k_{-3} was held fixed at 1 s^{-1} during global fitting. [f]The value for k_5 (release of bromide ion) was assumed to be fast (Figure 3.3) and held fixed at 200 s^{-1}. As bromide binding is in rapid equilibrium with free enzyme, the data only define the K_d for bromide.

Table S3.3. Rate Constants and Fluorescence Factors Resulting from a Comprehensive Global Fit of all Data.

Scheme S3.6: Final Complete Kinetic Model with Rate Constants and Fluorescence Factors for A) Caad with *trans*-3-bromoacrylate (1) and B) CaaD with 3-bromoproiolate (3).



3.6.3 Discussion

Previously, 3-bromopropiolate (**3**) was identified as a mechanism-based inhibitor of CaaD, which covalently labels the enzyme at βPro-1. The inactivation process was not well-defined. For example, at a concentration of 300 μM of **3**, the detectable activity of CaaD is less than 10% in 10 s. Due to the very fast inactivation, it is difficult to monitor the reaction using conventional methodology (e.g. UV spectrophotometry, NMR, etc). However, the reaction of **3** with the validated fluorescent mutant can be monitored using stopped-flow by following changes in enzyme fluorescence over the course of the reaction. The collected data were fit using both conventional methods and computational simulation. The stopped-flow data of αY60W-CaaD with **3** follows a two-step irreversible kinetic model. Fitting the data resulted in well-constrained rate constants for both steps as well as fluorescence factors for two enzyme species. The two fitted rate constants are equivalent to the classic inhibitor kinetic parameters for apparent rate of

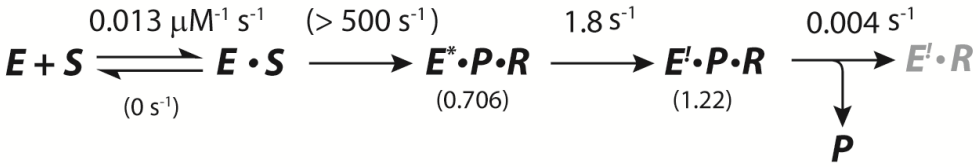
inactivation (k_{inact}) and inhibitor potency (k_{inact}/K_I). The values of k_{inact} and k_{inact}/K_I are 1.8 ± 0.04 s^{-1} and 0.013 ± 0.0006 μM^{-1} s^{-1}, respectively. With these values, an apparent K_I can be determined ($K_{I, app}$ = 140 ± 7 μM).

The result of this fit, specifically the fluorescence factor for E^*PR, can be used to re-fit the data of αY60W-CaaD with **1**. By assuming the fluorescence factor for enzyme species E^* is the same in both systems the ambiguity in the rate constants of k_3 and k_4 can be resolved. This results in well-constrained values for all the fitted parameters in Scheme S3.6. Moreover, combining the two sets of data results in a complete global fit with well-constrained values for nine rate constants and three fluorescents factors with a reasonable estimation of errors for each (Scheme S3.6 and Table S3.3).

These results are contingent on the assumption that E^* is the same in both systems (αY60W-CaaD with **1** and αY60W-CaaD with **3**). In the αY60W-CaaD with **1**, the proposed model has two florescence species, one occurring *after* chemistry and other occurring with initial product release. This conclusion is primarily based on the rate observed in the burst experiment (~80 s^{-1}) versus the observed rate in the stopped flow experiments (~10 s^{-1}). Further support for this conclusion comes from observation of a brief lag phase in the stopped-flow data which is comparable to that of the burst rate. The fitting results for the αY60W-CaaD data with **3** suggest that the fluorescence change occurs with chemistry or the rate of chemistry occurs so fast that it is not observed. Due to the reactivity of **3**, we expect the rate of chemistry to be as fast as or faster than that observed with **1**. Upon closer inspection of the stopped-flow data with **3**, there might be a very brief lag phase (<5 ms). For comparison, the lag phase in the αY60W-CaaD with **1** is around 20-30 ms. This suggests that the fluorescence change with **3** may also occur *after* chemistry, as was concluded with αY60W-CaaD and **1**. In fact, the αY60W-CaaD with **3** data can be fit to an expanded model (Scheme S3.7), with a lower limit on the rate

for chemistry (k_{chem} equal to ~500 s^{-1}), a value that is about 5-fold greater than that observed for αY60W-CaaD with **1**. Unfortunately, the lack of data in this region precludes further error analysis on this rate provided by during the fitting optimization routines. The lower limit of 500 s^{-1} is provided by visual inspection only. In Figure S3.6, the fit of the first 10 ms is shown. In the light, colored lines is the fit to Scheme S3.3. Shown in the solid, lines is the fit to Scheme S3.7 with the rate of chemistry set to 500 s^{-1}.

Scheme S3.7: Expanded Model to fit αY60W-CaaD with 3-Bromopropiolate (3)



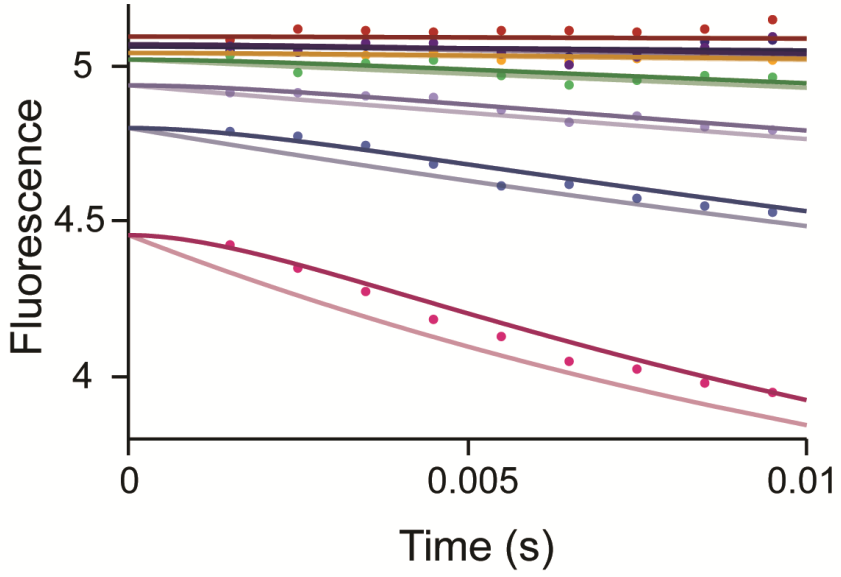


Figure S3.6: Fit of the First 10 ms of αY60W-CaaD with 3-Bromopropiolate (**3**)

Shown in filled circles are the data. The light colored lines are the result of the fit by simulation to Scheme S3.3. The solid colored lines are the fit by simulation to Scheme S3.7. The concentration series of **3** is the same as listed in Figure S3.2.

In the expanded model (Scheme S3.7), the product, *P*, bromide, is shown to be released at a very slow rate. There are no stopped-flow data to support the slow release of product. In fact, a model that allows *P* to be released from E^*PR to form $E'R + P$, fits the stopped-flow data equally as well. However, previous work showed that bromide release occurs very slowly with an observed rate of approximately $k_{obs} = 0.01 \pm 003$ s^{-1} (Figure S3.7).[2] This rate is at least 100-fold less than the observed rate of inactivation ($k_{inact} = 1.8$ s^{-1}) determined in the stopped-flow experiment. The "fast" rate for k_{inact} is supported by the enzyme activity versus time plots presented elsewhere.[2] The bromide release data can be included into the global fit with fit and fit by simulation by allowing

the enzyme to be inactivated "fast" to form species $E^{\prime}$ followed by a very slow release of bromide to form species $E^{\prime}R$ as shown in Scheme S3.7. By including the bromide release data in the global fit of αY60W-CaaD with **3** and allowing for some background bromide (about 20 μM), the release rate for bromide is determined to be 0.004 s^{-1}.

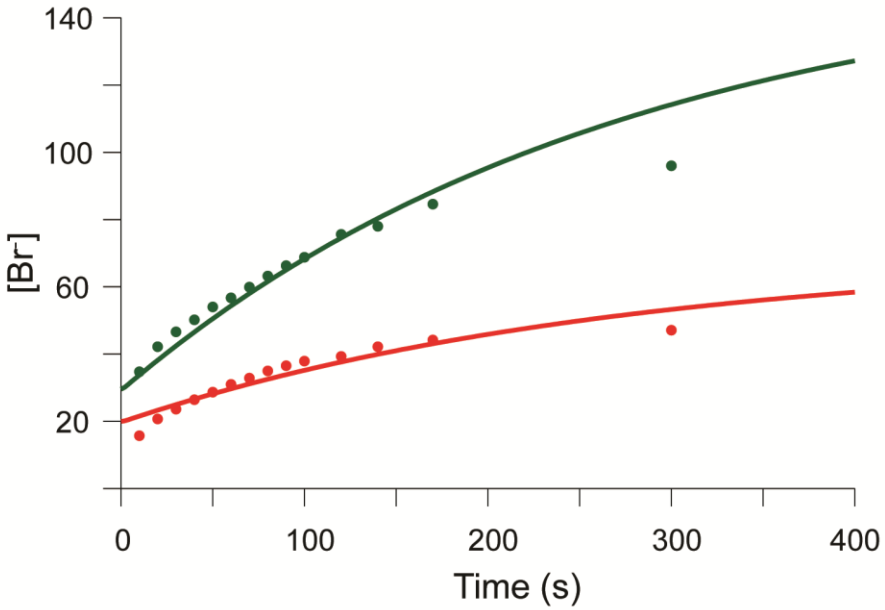


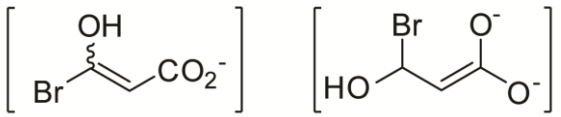
Figure S3.7: Fit of the Bromide Release Data with 3-Bromopropiolate (**3**) and wild-type CaaD

Shown in the filled circles are selected points from the data presented in Wang *et al*.[2] The lines are the result of the fit by simulation using Scheme S3.7. Including this data with the stopped-flow data and fitting using the expanded model does not change the fits from Figure S3.2.

At first glance this rate seems to contradict the rapid bromide release rate observed with αY60W-CaaD and **1**. However, in the model with **1**, we concluded that product release occurs in a biased random release with a preference for the upper pathway (Scheme S3.6A) where the release of **2** occurs first at a slow rate (k_4 = 3.3 s^{-1}), and is followed by rapid release of bromide. It was assumed that the bottom pathway (release of bromide followed by **2**) in Scheme S3.6A is not significant (shown in gray arrows). In the case of αY60W-CaaD and **3**, bromide release must occur "first" because the reactive intermediate is not released into solution. In effect, the enzyme is forced to follow the lower pathway of Scheme S3.6A. This forced release of bromide first is approximately 3-orders of magnitude slower than when **2** is released first (upper pathway of Scheme S3.6A) during the catalysis of **1** (3.3 s^{-1}) and, at least, 4-orders of magnitude slower than when bromide is released from the enzyme second. Interestingly, the upper limit of release of bromide first during the catalysis of **1** (lower pathway Scheme S3.6A) is about 0.01 s^{-1}, as predicted by inspection of the fit. These observations are consistent with the previous proposal that CaaD preferentially releases **2** followed by bromide in the dehalogenation of **1**.

As a final check, docking was carried out with the putative intermediates formed during the catalysis of **3** and **1**, designated as **4** and **5** (Scheme S3.8), respectively. These results for both intermediates suggest the bromide ion cannot be released from CaaD first without significant movement, because it is predicted to be buried at the back of the active site (Figure S3.8A and B). The 3-bromoenolate intermediate in Figure S3.8A is the *trans*-3-bromoenolate intermediate. The exact face of addition to **3** is unknown, however, the cis-3-bromoenolate intermediate docks very poorly into the active site of CaaD.

Scheme S3.8: Putative Intermediates, 4 and 5, Formed the During Catalysis of 3-bromoproiolate (3) and *trans*-3-bromoacrylate (1), respectively.



4 **5**

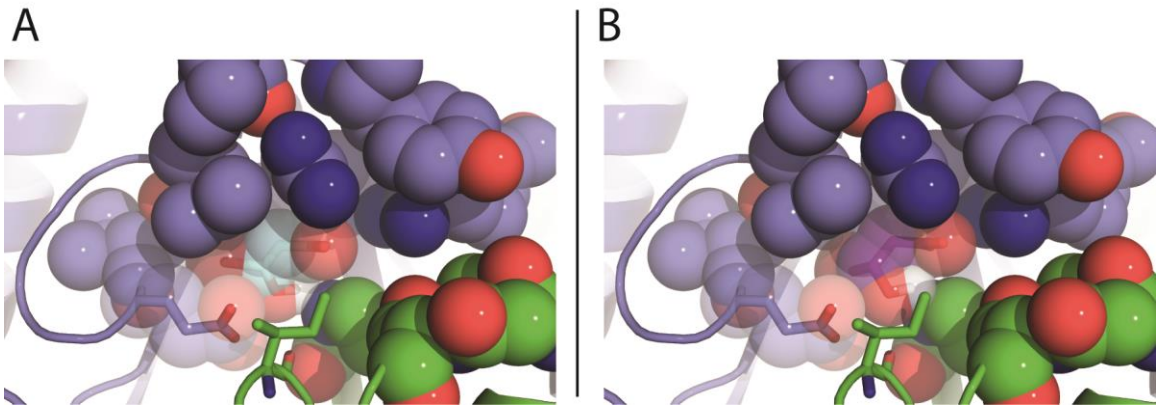


Figure S3.8: Docking of the Intermediates, **4** and **5**, into the Active Site of CaaD

A) Shown in light blue spheres is the docked 3-bromoenolate intermediate, **4**. The bromide moiety is shown in maroon spheres. B) Shown in purple spheres/sticks is the docked enolate intermediate formed during the catalysis of **1**. The bromide moiety is shown in maroon spheres. It is positioned almost directly behind C3 of **5**. Shown in blue and green spheres are the residues around the active site.

3.6.4 References

(1) Schroeder, G. K.; Huddleston, J. P.; Johnson, W. H., Jr.; Whitman, C. P. (2013) A mutational analysis of the active site loop residues in *cis*-3-chloroacrylic acid dehalogenase. *Biochemistry. 52*, 4204-4216.

(2) Wang, S. C.; Person, M. D.; Johnson, W. H., Jr.; Whitman, C. P. (2003) Reactions of *trans*-3-chloroacrylic acid dehalogenase with acetylene substrates: consequences of and evidence for a hydration reaction. *Biochemistry. 42*, 8762-8773.

Chapter 4: Reactions of Cg10062 with Acetylene and Allene Substrates: Evidence for a Hydration-dependent Decarboxylation

4.1 INTRODUCTION

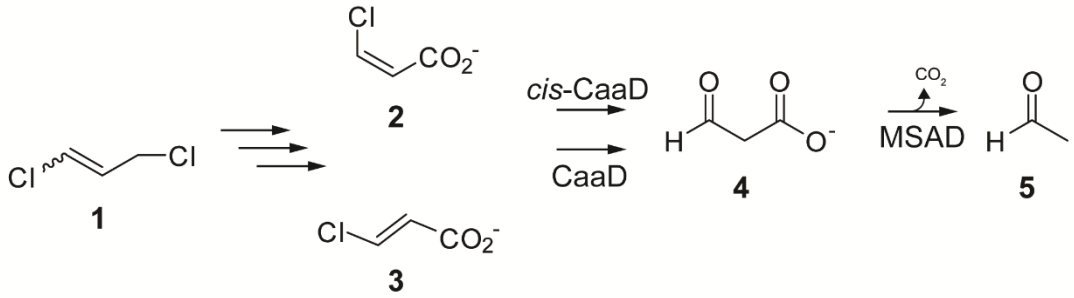
Cg10062 is a 149-amino acid protein from *Corynebacterium glutamicum* with unknown biological function(s).[1,2] It is a *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD) homologue (~34% pairwise sequence identity and 53% similarity), and is categorized in the *cis*-CaaD family, one of the five known families in the tautomerase superfamily.[3-5] Members of this superfamily share a characteristic β–α–β fold and a catalytic amino-terminal proline.

In contrast to Cg10062, *cis*-CaaD from *Pseudomonas pavonaceae* 170 carries out a known reaction in an established pathway.[3,6-8] The enzyme catalyzes the hydrolytic dehalogenation of *cis*-3-chloroacrylate (**2**, Scheme 4.1) to produce malonate semialdehyde (**4**).[3] A related enzyme, *trans*-3-chloroacrylic acid dehalogenase (CaaD), processes the *trans* isomer, **3**, to **4**.[9] A subsequent decarboxylation reaction catalyzed by

Adapted with permission from Huddleston et. al. *Biochemistry. 54,* (DOI: 10.1021/acs.biochem.5b00240). Copyright 2015 American Chemical Society. Dr. W.H. Johnson Jr. helped with ^{1}H NMR collection, data analysis, and experimental design. Dr. Schroeder provided helpful insight and data analysis. Dr. Whitman provided helpful discussions, editing, and funding.

malonate semialdehyde decarboxylase (MSAD) yields acetaldehyde (**5**), which is presumably shuttled to the Krebs cycle.[10] These reactions are two of the five steps that convert the isomeric mixture of the nematicide, 1,3-dichloropropene (**1**), to acetaldehyde.[6-8]

Scheme 4.1: The Enzyme-catalyzed Reactions of the 1,3-Dichloropropene Catabolic Pathway



Cg10062 and *cis*-CaaD share common features.[1,2,11,12] Both are homotrimers where each monomer is composed of 149 amino acids. Six residues critical for *cis*-CaaD activity (Pro-1, His-28, Arg-70, Arg-73, Tyr-103, and Glu-114) are also present in Cg10062.[1] Moreover, the structures and active site regions are largely superimposable including the position of the six key residues (Figure 4.1).[11,12]

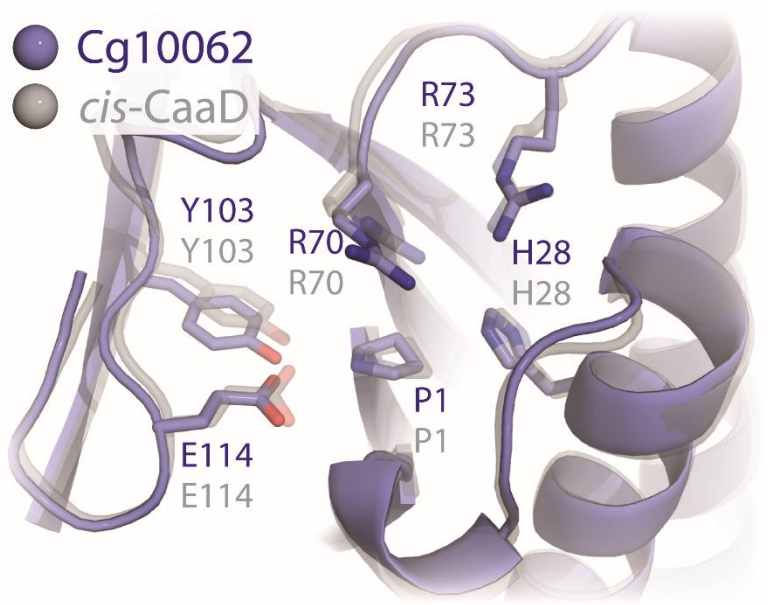


Figure 4.1: Overlay of Cg10062 and *cis*-CaaD Active Sites

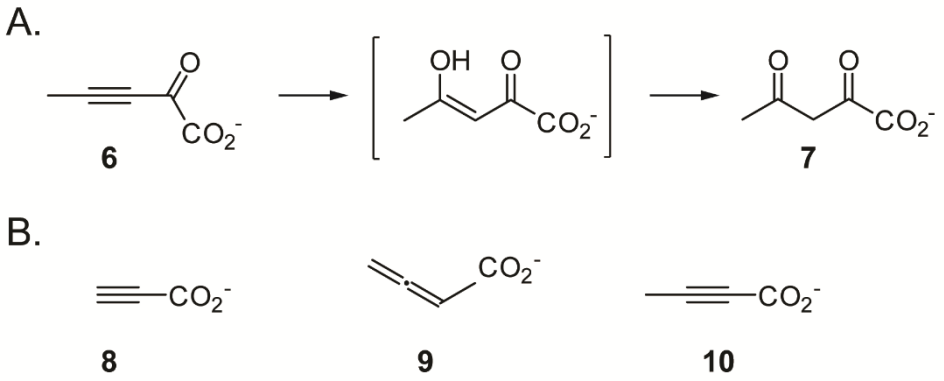
An overlay of the active sites of Cg10062 (blue, PDB code: 3N4G) and *cis*-CaaD (grey, PDB code: 2FLZ). Shown in sticks are the six key residues (Pro-1, His-28, Arg-70 Arg-73, Tyr-103, and Glue-114).

However, Cg10062 is not a very efficient *cis*-CaaD: there is a significant reduction in k_{cat}/K_m (~1000-fold using *cis*-3-bromoacrylate) and it processes both the *cis* and *trans* isomers of 3-chloroacrylate with a preference for the *cis* isomer.[1,2,11] In contrast, *cis*-CaaD and CaaD are highly specific for their respective isomers.[1,3,13]

The basis for these differences is not known. The substantial increase in K_m (~687-fold) observed for Cg10062 using (*cis*-3-bromoacrylate) suggests that it might originate in the elements that govern substrate specificity.[11,12] Two possibilities include the more spacious active site of Cg10062 or a 6-residue loop in *cis*-CaaD that is observed

closed down on the active site in the crystal structure of the enzyme that has been inactivated by an irreversible inhibitor.[11] A similar feature has not been observed in Cg10062. Both possibilities have been explored, but the results of these studies did not provide a definite explanation.[11]

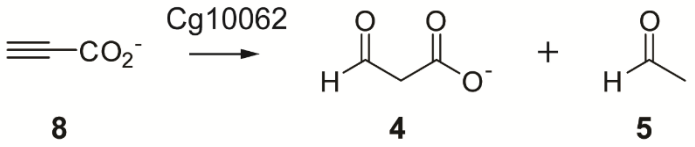
Scheme 4.2: A) The Enzyme-catalyzed Conversion of 2-oxo-3-Pentynoate (6) to Acetopyruvate (7). B) The Acetylene and Allene Substrates Used in this Work



In the course of our work on *cis*-CaaD, CaaD, and Cg10062, we desired a halide-free solution of **4**, which could only be generated with halide from **2** or **3** using the appropriate dehalogenase.[10] It is known that all three enzymes hydrate the acetylene compound, 2-oxo-3-pentynoate (**6**, Scheme 4.2A), to produce acetopyruvate (**7**) with varying degrees of efficiency (CaaD >> Cg10062 > *cis*-CaaD).[1,3,13] This observation prompted us to incubate *cis*-CaaD with propiolate (**8**, Scheme 4.2B), which was anticipated to produce **4** upon hydration. Indeed, **8** is a reasonably good substrate for *cis*-CaaD (k_{cat}/K_m = 6.4 × 10^3 $M^{-1}s^{-1}$), but an even better one for Cg10062. Interestingly, an analysis showed that **4** is only a minor product of the Cg10062-catalyzed reaction.

Instead, a mixture of **4** and **5** was obtained (25% and 75%, respectively), suggesting that Cg10062 functions as both a hydratase and decarboxylase (Scheme 4.3).

Scheme 4.3: The Cg10062-catalyzed Conversion of Propiolate (8) to a Mixture of Malonate Semialdehyde (4) and Acetaldehyde (5).



8 **4** **5**

These findings prompted a detailed investigation of this reaction with the wild type enzyme and active site mutants and confirmed that the hydration and decarboxylation reactions take place at the same active site. The reactions also appear to occur sequentially using the initial substrate (i.e., **8**), because the enzyme does not decarboxylate exogenously added **4**. The reactions of Cg10062 with 2,3-butadienoate (**9**) and 2-butynoate (**10**), as well as those of Cg10062 with **2** and **3**, were also examined. For **3**, **9**, and **10**, Cg10062 functions as a hydratase. With **2**, Cg10062 is a hydratase/decarboxylase, but it is not as efficient as with **8** (as assessed by the k_{cat}/K_m values). A mutational analysis of two water-activating residues (Glu-114 and Tyr-103) provided an interesting array of results. There was a reduction in wild type activity for some mutant/substrate combinations, as would be expected, but other combinations switched activities (hydratase to hydratase/decarboxylase or vice-versa). Moreover, the likely intermediates of the hydration and decarboxylation products can be trapped (using $NaCNBH_3$) covalently attached to Pro-1 for the E114D mutant with **9** and **10** and the

Y103F mutant with **10**. Three mechanisms are proposed to account for these observations. One mechanism involves a direct attack of water on the substrate followed by the decarboxylation of some of the hydration products. In a second mechanism, the substrate undergoes hydration, which is followed in some cases, by the formation of a Schiff base between the hydration product and Pro-1. Decarboxylation and hydrolysis of the Schiff base complete the reaction. A third mechanism involves the partitioning of the different substrates between covalent catalysis through Pro-1 and the direct attack of water. The route might be determined by the substrate's orientation in the active site with respect to Pro-1 and two water-activating residues (i.e., Glu-114 and Tyr-103). Finally, the possible biological relevance of the reaction of Cg10062 with **8**, the best substrate for Cg10062 identified to date, is discussed.

4.2 MATERIALS AND METHODS

Materials. Chemicals, biochemicals, buffers, and solvents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO), Fisher Scientific Inc. (Pittsburgh, PA), Fluka Chemical Corp. (Milwaukee, WI), or EMD Millipore, Inc. (Billerica, MA). Sodium phosphate buffer salts were at least 99.99% pure or greater, as indicated by the manufacturer. Propiolic acid (98%, **8**) and 2-butynoate (**10**) were purchased from Sigma-Aldrich. Both were purified further prior to use by distillation (**8**) or re-crystallization (**10**). 2,3-Butadienoate (**9**) was synthesized according to published methods.[14] The DEAE-Sepharose and Phenyl-Sepharose 6 Fast Flow resins used for protein purification were obtained from Sigma-Aldrich Chemical Co. (St Louis, MO). The Econo-Column® chromatography columns were obtained from BioRad (Hercules, CA). The Amicon stirred cell concentrators and the ultrafiltration membranes (10000 Da, MW cutoff) were purchased from EMD Millipore Inc. Malonate semialdehyde decarboxylase (MSAD) was purified as described previously.[10]

Bacterial Strains, Plasmids, and Growth Conditions. Escherichia coli strain BL21-Gold(DE3) was obtained from Agilent Technologies (Santa Clara, CA). The *E. coli* DH5α cells were obtained from Invitrogen (Carlsbad, CA). Cells were grown at 37 °C overnight in Luria-Bertani (LB) media that contained ampicillin (Ap, 100 μg/mL).

General Methods. The PCR amplification of DNA sequences was conducted in a GeneAmp 2700 thermocycler (Applied Biosystems, Carlsbad, CA). Techniques for restriction enzyme digestion, transformation, and other standard molecular biology manipulations were based on methods described elsewhere.[15] DNA sequencing was

performed in the DNA Core Facility in the Institute for Cellular and Molecular Biology (ICMB) at the University of Texas at Austin. Electrospray ionization mass spectrometry (ESI-MS) was carried out on an LCQ electrospray ion-trap mass spectrometer (Thermo, San Jose, CA). Peptide fragment masses were determined using Voyager-DE Pro matrix-assisted laser desorption/ionization (MALDI) mass spectrometer run in the linear mode. Both instruments are housed in the ICMB Protein and Metabolite Analysis Facility at the University of Texas. Steady-state kinetic assays were performed on an Agilent 8453 diode-array spectrophotometer at 22 °C. Non-linear regression data analysis was performed using the program Grafit (Erithacus Software Ltd., Staines, U.K.). Protein concentrations were determined by the Waddell method.[16] Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on denaturing gels containing 12% polyacrylamide.[17] Nuclear magnetic resonance (NMR) spectra were recorded on a Varian INOVA-500 or a Varian DirectDrive 600 MHz spectrometer (Palo Alto, CA). NMR signals were analyzed using the software program SpinWorks 3.1.6 (Copyright © 2009 Kirk Marat, University of Manitoba).

4.2.1 Construction of the Cg10062 Mutants and Expression, Purification of Wild-type and Cg10062 Mutants

The expression vectors for Cg10062 and the P1A, R70A, and E114Q mutants of Cg10062 were constructed as described elsewhere.[1,18] The remaining Cg10062 mutants were constructed using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) following the manufacturer's instructions. Oligonucleotide primers (coding and complementary) were obtained from Sigma-Aldrich. The forward and

reverse primers used to introduce the Y103F mutation were 5'-CCAATGAAGAAGTATGGGTATTTATAACGGAGATTCCTGGTTC-3' and 5'-GAACCAGGAATCTCCGTTATAAATACCCATACTTCTTCATTGG-3', respectively. The forward and reverse primers used to introduce the E114D mutation were 5'-GGTTCCAATATGACGGATTATGGCCGTCTCCTCAT-3' and 5'-ATGAGGAGACGGCCATAATCCGGTCATATTGGAACC-3', respectively. The codon used to introduce the mutation is underlined in each primer. The E114Q/Y103F double mutant of Cg10062 was constructed by the QuikChange method using the E114Q-pET3b plasmid as template DNA and the primers used for the Y103F mutant. Wild-type Cg10062 and the mutants were grown and expressed in *E. coli* BL21-Gold(DE3) cells and purified by previously published protocols.[1,2,18]

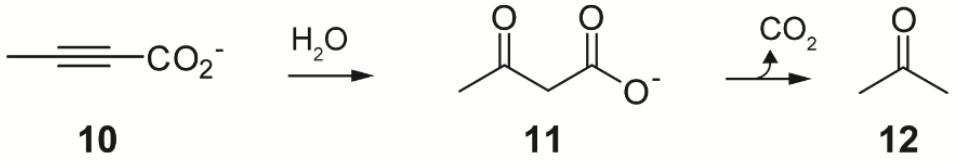
Mass Spectral Analysis of Cg10062 and the Cg10062 Mutants. The monomer molecular masses of Cg10062 and mutants were determined by ESI-MS. Samples for analysis were made 1 mg/mL, as described elsewhere.[13] The observed molecular masses (MH^+) for the wild-type enzyme, and the E114Q-, E114D-, Y103F-, and P1A-Cg10062 mutants are: 17100 Da, 17098 Da, 17085 Da, 17085 Da, and 17072 Da, respectively. These masses agree with the calculated masses (within experimental error).

4.2.3 Steady-State Kinetics

The assays were carried out at 22 °C in 100 mM Na_2HPO_4 buffer, pH 8.1, using various concentrations of enzyme (0.02–2 μM) depending on the k_{cat} of the enzyme with the substrate. A stock solution of **8** (~400 mg/mL, 5.71 M) was made up by pipetting (with a disposable glass pipet) ~0.34 g (~0.34 mL) of liquid **8** into 100 mM tribasic

phosphate buffer (0.5 mL). Subsequently, a 100 mM propiolate solution was made up in 100 mM phosphate buffer from this stock solution. The final pH was adjusted to 8.1. Diluted enzyme solutions were equilibrated for 1 h before the assays. Initial rates were plotted vs. substrate concentration and fit to the Michaelis-Menten equation using Grafit to determine k_{cat} and K_m. The reported errors are calculated from the standard errors of the fit to the data by nonlinear regression. If saturation kinetics were not observed, the data were fit to a straight line where the slope is equal to k_{cat}/K_m.[19] The reactions of **2**, **3**, and **8** were monitored by a coupled assay described elsewhere, where the reduction of acetaldehyde by alcohol dehydrogenase is coupled to the oxidation of NADH.[1,10,11] Cg10062 processes *cis*-3-bromoacrylate with comparable kinetic parameters and product ratios to those determined for **2** (data not shown). The hydration of **9** to acetoacetate (**11**) was followed by monitoring the decrease in absorbance at 224 nm (ε_{224} = 1125 μM^{-1} cm^{-1}), as described previously.[11] The hydration of **10** to acetoacetate (**11**, Scheme 4.4) was followed by monitoring the decrease in absorbance at 220 nm (ε_{220} = 624 μM^{-1} cm^{-1}). The decarboxylase activity of Cg10062 and the mutants with **2**, **3**, or **8** (yielding **5**), was determined using the coupled assay described above where MSAD was excluded from the mixture and by ^{1}H NMR spectroscopy. The decarboxylase activity of Cg10062 and mutants with **9** or **10** (yielding acetone, **12**) was determined by ^{1}H NMR spectroscopy.

Scheme 4.4: The Cg10062-catalyzed Conversion of 2-Butynoate (10) to a Mixture of Acetoacetate (11) and Acetone (12)



*Decarboxylation of **4** and **11** by Cg10062 and Mutants.* Malonate semialdehyde (**4,** ~ 10 mM) was generated using *cis*-CaaD and *cis*-3-chloroacrylate (**2**), as previously reported.[10,20] The decarboxylation of exogenous **4** by Cg10062 and the mutants was monitored by the coupled assay (above) without MSAD and by ^{1}H NMR spectroscopy.

4.2.4 Identification of the Products of the Cg10062- and Mutant-catalyzed Reactions by ^{1}H NMR Spectroscopy

The protocol for monitoring the reaction of Cg10062 with **2** or **3** by ^{1}H NMR spectroscopy is described elsewhere.[1] The same procedure was used with the following modifications for monitoring the reaction of Cg10062 with **8**, **9**, or **10**. Accordingly, an amount of **8**, **9**, or **10** (4 mg, ~95 mM for **8** and ~80 mM for **9** and **10**) was added to 100 mM Na_2HPO_4 buffer (pH 9.2) containing 30 μL of DMSO-d_6 (final total volume of 600 μL). The final pH of the solution is adjusted 8.0. Subsequently, a 7-μL aliquot of enzyme (wild-type or the E114Q, E114D, Y103F, Y103F/E114Q, and P1A mutant of Cg10062, from 18 mg/mL, 15 mg/mL, 17 mg/mL, 28 mg/mL, 16 mg/mL, or 17 mg/mL stock solutions, respectively) was added to the mixture and product formation was monitored every 3 min over a 3-90 min time interval depending on the enzymatic activity. DMSO-

d_6 was used as the lock signal and for the standardization of the chemical shifts (at 2.49 ppm). The chemical shifts for **4**, **5**, their hydrates, **11**, and **12** are reported elsewhere.[13,21] The approximate amount of product in the mixtures was determined by integration, as described previously.[1]

4.2.5 ESI-MS Analysis of $NaCNBH_3$-treated Mixtures Containing the Y103F or E114D Mutants of Cg10062 with 9 or 10.

Six reaction mixtures (300 μL) were made up containing 100 mM Na_2HPO_4 buffer (37 μL, pH 8.0), $NaCNBH_3$ (50 μL of a 1.6 M solution in 100 mM Na_2HPO_4, pH ~9.5) and the Y103F or E114D mutant of Cg10062 (20 μL) (3 reaction mixture each). After a 10-min incubation period, **9** or **10** (193 μL from a 14 mg/mL stock solution in 100 mM Na_3PO_4 buffer with the pH adjusted to 8.0 using a 50% NaOH solution) was added to the mixtures containing the Y103F- or E114D-Cg10062 (accounting for four of the six reaction mixtures). The remaining two mixtures (one with the Y103F mutant and one with the E114D mutant) served as controls where the substrate is replaced with 100 mM Na_2HPO_4 buffer, pH 8.0 (193 μL). The reaction mixtures incubated overnight at room temperature. Subsequently, the remaining hydratase activity in each sample (using **9** as substrate) was determined by the UV assay described previously.[11] The activities of three of the four reaction mixtures (**9** with E114D-Cg10062 and **10** with Y103F- and E114D-Cg10062) were reduced 10-100-fold from those observed in the control reaction mixtures. The incubation of **9** with Y103F-Cg10062 results in no reduction of activity from that observed in the control sample. The mixtures with reduced activity (i.e., **9** with E114D-Cg10062 and **10** with Y103F- and E114D-Cg10062) were then prepared for ESI-

MS analysis, as described elsewhere.[3,21] Four additional reaction mixtures were set up that contained the Y103F and E114D mutants of Cg10062 along with exogenous **11** or **12**. Following the protocol described above, an aliquot (230 μL) of **11** or **12** (from a 0.14 M stock solution, 14.95 mg/mL and 8.02 mg/mL, respectively) was added to 70 μL of pre-incubated solution containing enzyme (20 μL of the Y103F or E114D mutant of Cg10062) and $NaCNBH_3$ (50 μL). There was no reduction in the activity of these samples, as assessed by an assay using **9** (compared with that of control), so that ESI-MS analysis was not carried out on these samples.

4.2.6 Peptide Mapping and MALDI-MS Analysis.

The covalently modified Y103F and E114D mutants of Cg10062 (by their incubation with **9** and **10** and treatment with $NaCNBH_3$) were treated with a solution of *Staphylococcus aureus* endoproteinase Glu-C (protease V-8), following a previously published protocol.[3,21] After a 48-h incubation period, the resulting peptide mixture was analyzed by MALDI-MS, as described previously.[3,21]

4.3 RESULTS

4.3.1 Kinetic Parameters of Cg10062 with 8 and 10.

The steady-state kinetic parameters were measured for Cg10062 with the acetylene compounds, **2**, **8,** and **10**, and are summarized with those previously measured for Cg10062 with **3** and **9** in Table 4.1.[1,11,18] Cg10062 shows the highest k_{cat} and lowest K_m values with propiolate (**8**), yielding a k_{cat}/K_m value of 1.8×10^5 $M^{-1}s^{-1}$. This k_{cat}/K_m is 12,000-fold and 6000-fold higher than those measured respectively for **2** and 2-butynoate (**10**), and 35-fold higher than the one measured for 2,3-butadienoate (**9**). The k_{cat}/K_m values for Cg10062 with **2** and **10** are comparable. Cg10062 has the poorest activity and lowest efficiency with **3**. However, the ability of Cg10062 to process **3** is interesting in light of the inability of *cis*-CaaD to process **3** coupled with the many similarities between *cis*-CaaD and Cg10062.

Sub[a]	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
2	72,000 ± 8,500	1 ± 0.1	14 ± 2
3[b]	78,000 ± 36,000	0.06 ± 0.01	0.8 ± 0.4
8	33 ± 5	6 ± 0.2	$(1.8 \pm 0.2) \times 10^5$
9[c]	780 ± 120	4 ± 0.1	$(5.1 \pm 0.8) \times 10^3$
10[d]	-	-	30 ± 2

[a]The kinetic parameters for **8** and **10** were measured by the assay described in the text in 100 mM Na_2HPO_4 buffer (pH 8) at 22 °C. [b]Reported in refs 1 and 18. [c]Reported in ref 11. [d]Saturation kinetics could not be achieved with **10**. The initial rates were plotted versus [**10**] and fit to a straight line to provide a value for k_{cat}/K_m.[19]

Table 4.1. Steady-state Kinetic Parameters for the Cg10062-catalyzed Reactions

4.3.2 ^{1}H NMR Characterization of Products Resulting from Reactions Catalyzed by Cg10062.

^{1}H NMR spectroscopic analysis of the reaction of Cg10062 with **8** showed that **5** is the major product (~75%) of the reaction along with a small amount of **4** (~25%) (Figure 4.2A). The reaction is complete in less than 3 min, indicating that **5** does not result from the non-enzymatic decarboxylation of **4**. The same kinetic parameters are obtained with and without MSAD in the coupled assay, which is also consistent with **5** being the major product. The subsequent incubation of Cg10062 with exogenously

generated **4** (10-80 mM) did not produce **5** at a rate faster than that observed for the non-enzymatic rate, as assessed by UV and ^{1}H NMR spectroscopy. Based on these observations, the Cg10062 catalyzes the hydration-dependent decarboxylation of **8**.

These observations prompted a closer examination of the reactions of Cg10062 with **2** and **3** by ^{1}H NMR spectroscopy. Cg10062 converts **2** to both **4** (7%) and **5** (33%) after 48 min with 60% of **2** remaining (Figure 4.2B). (Two doublets corresponding to the protons of **2** appear at 6.11 and 6.19 ppm, which are outside the range shown in the figure.[3]) Although **3** is converted to **4** (2.9%) by Cg10062, **4** is not processed to **5** by the enzyme after 48 min (Figure 4.2C). A small amount of acetaldehyde, **5**, is observed (1.5%), but this is due to the non-enzymatic decarboxylation of **4** in the course of the 48-min incubation time. This was confirmed by the UV assay (data not shown). A large amount of unreacted substrate (95.6%) is also present. (Two doublets for the protons of **3** appear at 6.09 and 6.89 ppm, which are outside the range shown in the figure.[13]) The presence of **5** in the reaction of Cg10062 with **2** was noted previously, but was attributed to a non-enzymatic process or a promiscuous decarboxylase activity of Cg10062.[1]

With **9** and **10**, Cg10062 functions as a hydratase.[11] The reaction of Cg10062 with **9** is complete in less than 3 min and the ^{1}H NMR spectrum shows mostly **11**, which results from the hydration of **9** (Figure 4.2D). The reaction of **10** with Cg10062 yields a similar result, but over a longer time period: it is complete after 39 min and **11** is the major product observed (Figure 4.2E). There is a small amount of **12** present (~2.5% after 18 min), which exceeds the expected amount from the non-enzymatic

decarboxylation of **11**. Hence, Cg10062 functions primarily as a hydratase with this substrate, but has a weak decarboxylase activity.

Figure 4.2

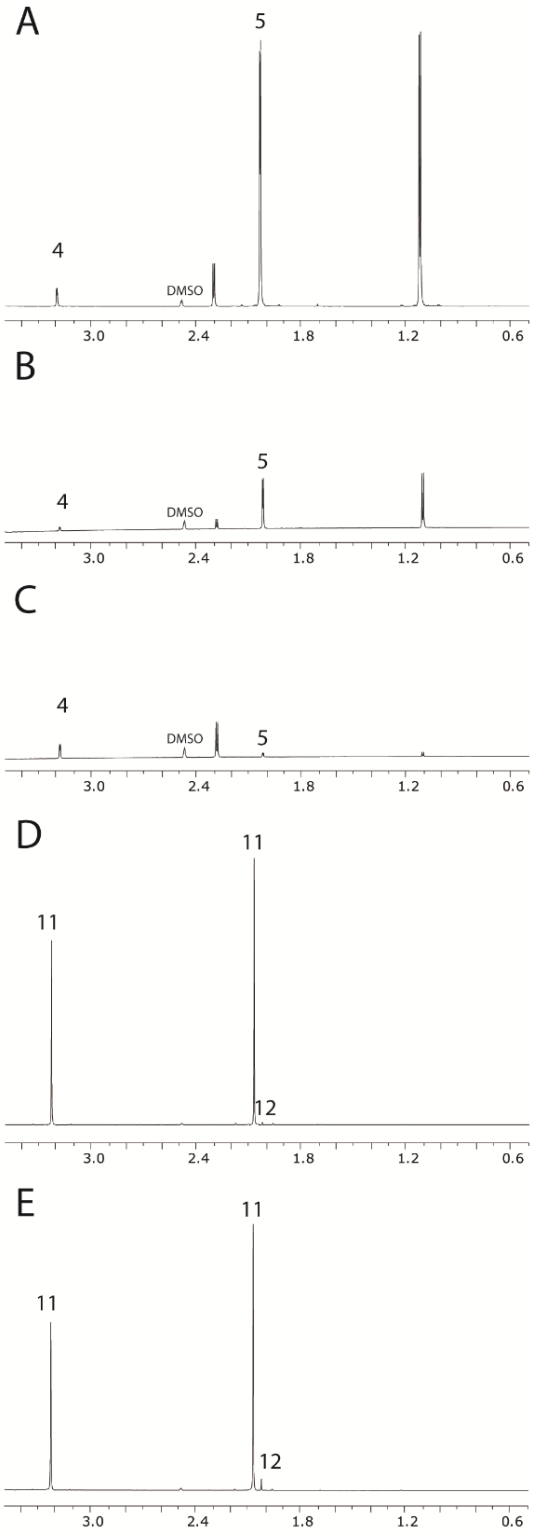

A
5
4
DMSO
3.0
2.4
1.8
1.2
0.6
B
5
4
DMSO
3.0
2.4
1.8
1.2
0.6
C
4
DMSO
5
3.0
2.4
1.8
1.2
0.6
D
11
11
12
3.0
2.4
1.8
1.2
0.6
E
11
11
12
3.0
2.4
1.8
1.2
0.6

Figure 4.2: ^{1}H NMR Spectroscopic Product Analysis of the Reaction of Cg10062 with **2**, **3**, **8**, **9**, or **10**.

A) Products of Cg10062 with **8** (~95 mM) after 3 min. B) Products of Cg10062 with **2** (~63 mM) after 48 min. The signals for the unreacted **2** are not shown. C) Products of Cg10062 with **3** (~63 mM) after 48 min. The signals for the unreacted **3** are not shown. The small amount of **5** present is consistent with non-enzymatic decarboxylation of **4**. D) Products of Cg10062 with **9** (~80 mM) after 3 min. The small amount of **12** present is consistent with the non-enzymatic decarboxylation of **11**. E) Products of Cg10062 with **10** (~80 mM) after 39 min. The amount of **12** present (2.5%) is more than expected for non-enzymatic decarboxylation. All reactions contained ~12.3 μM Cg10062 and 30 μL of DMSO-d_6. The reactions are individually scaled to show the products of the reactions. The scale can be assessed by the height of the DMSO peak (labeled in the spectra). The hydrates of **4** and **5** are also present: the signal at ~2.3 ppm corresponds to the methyl group of the hydrate of **4** and the signal at ~1.1 ppm corresponds to the methyl group of the hydrate of **5**.[13]

4.3.3 Incubation of the P1A, R70A, and Y103F/E114Q Mutants of Cg10062 with 2, 8, 9, and 10

In order to assess whether the reactions are taking place at the active site of Cg10062 (or another site), various active site mutants were examined with **2**, **8**, **9**, and **10**. The P1A mutant was examined with **2**, **8**, and **10**. The R70A mutant was examined with **2**, **8**, and **9**. The Y103F/E114Q mutant was examined with **8** and **9**. (The reactions of the P1A and R73A mutants of Cg10062 with **3** were previously reported.[1]) In all cases, the signals were greatly reduced in the ^{1}H NMR spectra (data not shown). These observations suggest that the reactions take place in the same active site.

4.3.4 Kinetic Parameters of E114Q-Cg10062 with 2, 8, and 9

The steady-state kinetic parameters for the E114Q mutant of Cg10062 with **2**, **8**, and **9** are summarized in Table 4.2. With **8**, the E114Q mutant of Cg10062 shows a 7.5-fold reduction in k_{cat} (vs. wild type) and an 11-fold reduction in K_m. As a result, there is a slight increase in the value of k_{cat}/K_m. With **9**, the k_{cat}/K_m value increases ~5-fold, which is due to an increase in the k_{cat}. Like **8**, the E114Q mutant of Cg10062 shows a reduction in both the k_{cat} and K_m values using **2**, and results in a slightly higher k_{cat}/K_m (vs. wild type). The steady-state kinetic parameters for the E114Q mutant using **3** or **10** could not be determined with the coupled assay because the reactions are too slow. Although the k_{cat}/K_m values for the E114Q mutant with **2** and **8** are comparable with the respective ones for wild type, the E114Q mutant does not carry out a decarboxylation reaction with the hydration product and MSAD must be present in the assay mixture in order to monitor the reactions by the UV assay.

Sub[a]	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
2	4,000 ± 800	0.4 ± 0.05	100 ± 25
8	3 ± 0.25	0.80 ± 0.02	$(2.7 \pm 0.2) \times 10^5$
9	660 ± 75	16 ± 1	$(2.4 \pm 0.3) \times 10^4$

[a]The kinetic parameters were measured using the assay described in the text in 100 mM Na_2HPO_4 buffer (pH 8) at 22 °C.

Table 4.2. Steady-state Kinetic Parameters for the E114Q-Cg10062-catalyzed Reactions

4.3.5 ^{1}H NMR Characterization of the E114Q-Cg10062-catalyzed Reaction with 2, 3, 8, 9 and 10

^{1}H NMR analysis of the five E114Q-Cg10062-catalyzed reactions uncovered some intriguing contrasts to the wild type reactions. With **8**, the E114Q-Cg10062 apparently functions as a hydratase because **4** is now the major product (8.5%) of the reaction after 18 min (Figure 4.3A). A small amount of **5** (0.5%) is observed in the NMR spectrum, but it is similar to that observed for the non-enzymatic decarboxylation of **4**. **(**This was confirmed by the UV assay using MSAD.) There is also a large amount of unreacted substrate present (91%), indicated by the singlet at ~2.9 ppm. A similar result is observed when **2** is incubated with E114Q-Cg10062 (Figure 4.3B). After 18 min, the mixture consists of unreacted substrate (80%), **4** (19%), and a small amount of **5** (1%).

The E114Q mutant of Cg10062 does process **3** to any significant extent after 90 min (Figure 4.3C).

With **9** and **10**, the ^{1}H NMR analysis suggests that the E114Q mutant of Cg10062 functions as a hydratase/decarboxylase. With **9**, the E114Q-Cg10062-catalyzed reaction is complete after 12 min, and yields mostly **11** (83%) with a significant amount of **12** present (~17%) (Figure 4.3D). Moreover, there is no further change in the amounts of **11** or **12** after the reaction mixtures incubate for an additional 15 min. This suggests that the E114Q mutant does not catalyze the decarboxylation of **11** when added exogenously, just as wild-type does not catalyze the decarboxylation of **4** (at a rate faster than non-enzymatic) when added exogenously. With **10**, the E114Q-Cg10062-catalyzed reaction produces **12** (2.2%, which is 73% of the total products observed) as the major product and **11** as the minor product (0.8%, which is 27% of the total products observed) after 90 min (Figure 4.3E). A large amount of unreacted substrate remains (97%), as indicated by the singlet at ~1.7 ppm. These results suggest that the E114Q mutant catalyzes a decarboxylation reaction (preceded by a hydration reaction), but only using **9** and **10** where the decarboxylation product is **12**.

Figure 4.3

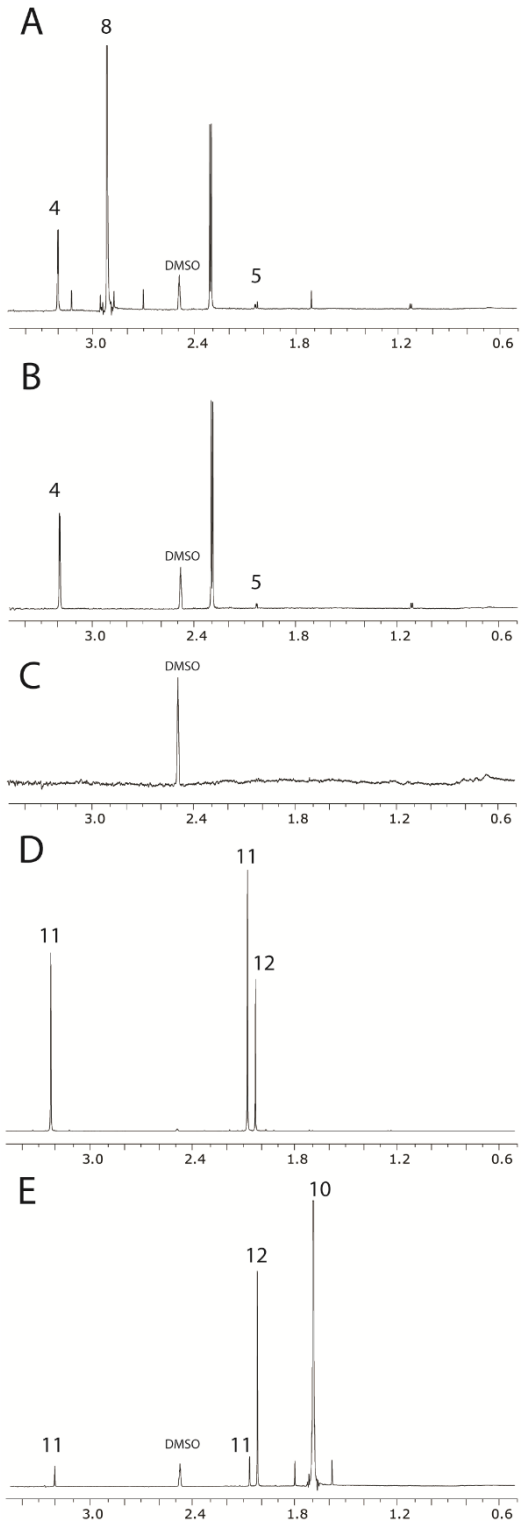

A
8
4
DMSO
5
3.0
2.4
1.8
1.2
0.6
B
4
DMSO
5
3.0
2.4
1.8
1.2
0.6
C
DMSO
3.0
2.4
1.8
1.2
0.6
D
11
11
12
3.0
2.4
1.8
1.2
0.6
E
10
12
11
DMSO
11
3.0
2.4
1.8
1.2
0.6

Figure 4.3: ^{1}H NMR Spectroscopic Product Analysis of the Reaction of E114Q-Cg10062 with **2**, **3**, **8**, **9**, or **10**.

A) Products of E114Q-Cg10062 with **8** (~95 mM) after 18 min. The small amount of **5** present is consistent with non-enzymatic decarboxylation of **4**. The reaction is not complete with a significant amount of **8** present (identified by the singlet at 2.90 ppm). B) Products of E114Q-Cg10062 with **2** (~63 mM) after 18 min. The small amount of **5** present is consistent with non-enzymatic decarboxylation of **4**. The reaction is not complete with a significant amount of **2** still present (NMR signals not shown). C) Products of E114Q-Cg10062 with **3** (~63 mM) after 90 min. There are no observable signals for **4** or **5**. D) Products of E114Q-Cg10062 with **9** (~80 mM) after 12 min. The signal at 2.03 ppm corresponds to the methyl group of **12**. The signals at 2.08 and 3.23 ppm correspond to the methyl and methylene groups respectively of **11**.[21] E) Products of E114Q-Cg10062 with **10** (~80 mM) after 90 min. All reactions contained ~12.3 μM E114Q-Cg10062 and 30 μL of DMSO-d_6. The reactions are individually scaled to show the product reaction. The scale can be assessed by the height of the DMSO peak, which is labeled in the spectra.

4.3.6 Kinetic Parameters of E114D-Cg10062 with 2, 8, 9, and 10

The steady state kinetic parameters for the E114D mutant of Cg10062 with **2**, **8**, **9**, and **10** are summarized in Table 4.3. Using **8**, there is a 6-fold reduction in k_{cat} and a 2.7-fold increase in K_m (vs. wild type), which results in a 16-fold decrease in the k_{cat}/K_m. The resulting k_{cat}/K_m (1.1×10^4 $M^{-1}s^{-1}$) is an order of magnitude less than that measured for wild type. With **2**, the E114D-Cg10062 exhibits poor activity. With **3**, the activity is so poor that it is not possible to measure kinetic parameters. The k_{cat}/K_m for the E114D mutant of Cg10062 using **9** increased slightly from that observed for wild type. The k_{cat}/K_m for the E114D mutant of Cg10062 with **10** is 14-fold higher than that for wild type. The k_{cat} is low (0.5 s^{-1}), but a K_m value can be measured unlike the wild type enzyme where saturation kinetics could not be achieved.

Sub[a]	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
2	40,100 ± 10,000	0.1 ± 0.02	2.5 ± 0.8
8	90 ± 10	1 ± 0.05	$(1.1 \pm 0.2) \times 10^4$
9	480 ± 90	3.2 ± 0.4	$(6.6 \pm 1.5) \times 10^3$
10	1400 ± 200	0.5 ± 0.03	$(4.1 \pm 0.5) \times 10^2$

[a]The kinetic parameters were measured by the assay described in the text in 100 mM Na_2HPO_4 buffer (pH 8) at 22 °C.

Table 4.3. Steady-state Kinetic Parameters for the E114D-Cg10062-catalyzed Reactions

4.3.7 ^{1}H NMR Characterization of the E114D-Cg10062-catalyzed Reaction with 8, 9 and 10

The reactions of the E114D mutant of Cg10062 with **8**, **9** and **10** were monitored by ^{1}H NMR spectroscopy. The major product of the reaction with **8** is **4** (22%), and not **5** (1%) after 3 min (Figure 4.4A). A significant amount of unreacted substrate (77%) remains (i.e., singlet at ~2.9 ppm). The E114D mutant of Cg10062 with **9** yields mostly **11** (97.2%) and a small amount of **12** (2.8%) after 3 min (Figure 4.4B). Similar observations are made for the reaction of **10** with the E114D mutant of Cg10062 (i.e., 91.5% of **11** and 2.5% of **12** and 8.2% unreacted substrate) after 3 min (Figure 4.4C). These results indicate that the E114D mutant of Cg10062 functions mostly as a hydratase with **9** and **10**, but also has a weak decarboxylase activity.

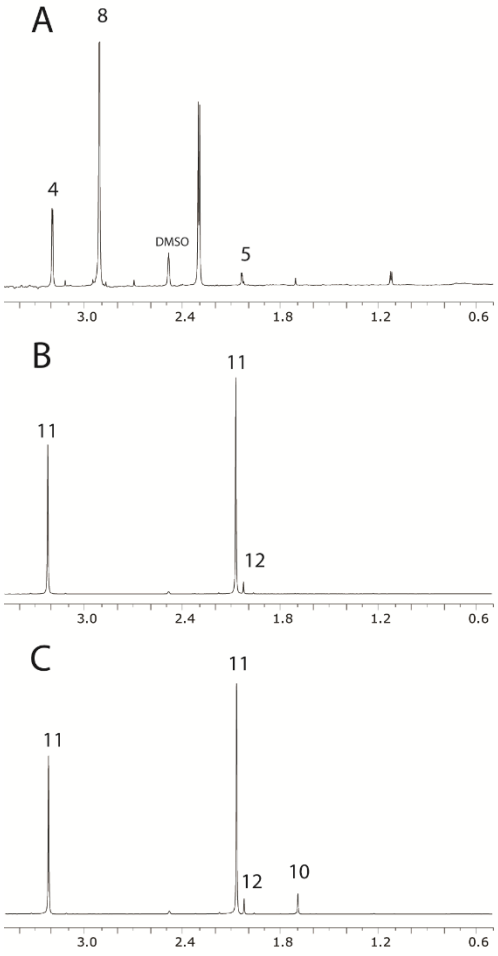


Figure 4.4: ^{1}H NMR Spectroscopic Product Analysis of the Reaction of E114D-Cg10062 with **8**, **9**, or **10**.

A) Products of E114D-Cg10062 with **8** (~95 mM) after 3 min. B) Products of E114D-Cg10062 with **9** (~80 mM) after 3 min. C) Products of E114D-Cg10062 with **10** (~80 mM) after 3 min. All reactions contained ~10.2 μM E114D-Cg10062 and 30 μL of DMSO-d_6. The reactions are individually scaled to show the product reaction. The scale can be assessed by the height of the DMSO peak, which is labeled in the spectra.

4.3.8 Kinetic Parameters of Y103F-Cg10062 with 2, 8, 9, and 10

The steady-state kinetic parameters for the reactions of **2, 8**, **9**, and **10** with the Y103F mutant of Cg10062 are summarized in Table 4.4. With **8**, there was a 12-fold decrease in k_{cat} and a 6-fold decrease in K_m (vs wild type), resulting in a slight decrease in the $k_{cat}/K_{m.}$ With **9**, the Y103F mutant of Cg10062 shows an 2-fold increase in the k_{cat}/K_m, which is due mostly to the 3.2-fold increase in k_{cat}. These observations parallel those made for the E114Q mutant of Cg10062. With **2**, there was a reduction in both k_{cat} and K_m (vs. wild-type) resulting in a slightly higher k_{cat}/K_m value (~4.3-fold) but lower than that for E114Q-Cg10062 (~1.7-fold). With **10**, there was a significant increase (~106-fold) in the value of k_{cat}/K_m. The Y103F mutant of Cg10062 showed very little activity with **3**, which precluded the measurement of kinetic parameters.

Sub[a]	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
2	4,700 ± 300	0.3 ± 0.04	60 ± 10
8	5 ± 2	0.5 ± 0.01	$(1.0 \pm 0.3) \times 10^5$
9	1200 ± 200	13 ± 2.5	$(1.1 \pm 0.3) \times 10^4$
10	700 ± 100	2.1 ± 0.2	$(3.2 \pm 0.8) \times 10^3$

[a]The kinetic parameters were measured by appropriate UV assay in 100 mM Na_2HPO_4 buffer (pH 8) at 22 °C.

Table 4.4. Steady-state Kinetic Parameters for the Y103F-Cg10062-catalyzed Reactions

4.3.9 ^{1}H NMR Characterization of the Y103F-Cg10062-catalyzed Reaction with 2, 3, 8, 9, and 10.

The reactions of the Y103F mutant of Cg10062 with **2**, **3**, **8**, **9** and **10** were monitored by ^{1}H NMR spectroscopy (Figure 4.5). The results with **8** and the Y103F mutant are comparable to those observed for the wild type, although slower. Accordingly, the reaction of **8** with the Y103F mutant yields **4** (0.7%) and **5** (2.1%), along with unreacted substrate (97.2%) after 27 min (Figure 4.5A), (as indicated by the singlet at ~2.9 ppm). The reaction of the Y103F mutant of Cg10062 with **2** yields **4** (1%) and **5** (8%) and unreacted substrate (91%) after 18 min (Figure 4.5B). The reaction with **3** is very slow although a small signal for **4** (~ 2.30 ppm) can be observed after 18 min (Figure 4.5C). The reaction of **9** with the Y103F mutant shows that **11** is the major product (99.5%) after 3 min (Figure 4.5D). A small amount of **12** (0.5%) is also present, which is slightly more than is expected for the non-enzymatic decarboxylation of **11**. The reaction of **10** with the Y103F mutant shows that the major product is **11** (84%), along with a significant amount of **12** (16%) after 6 min (Figure 4.5E). Hence, the Y103F mutant has hydratase/decarboxylase activity using **2**, **8**, and **10** as substrates. With **9**, it is mostly a hydratase.

Figure 4.5

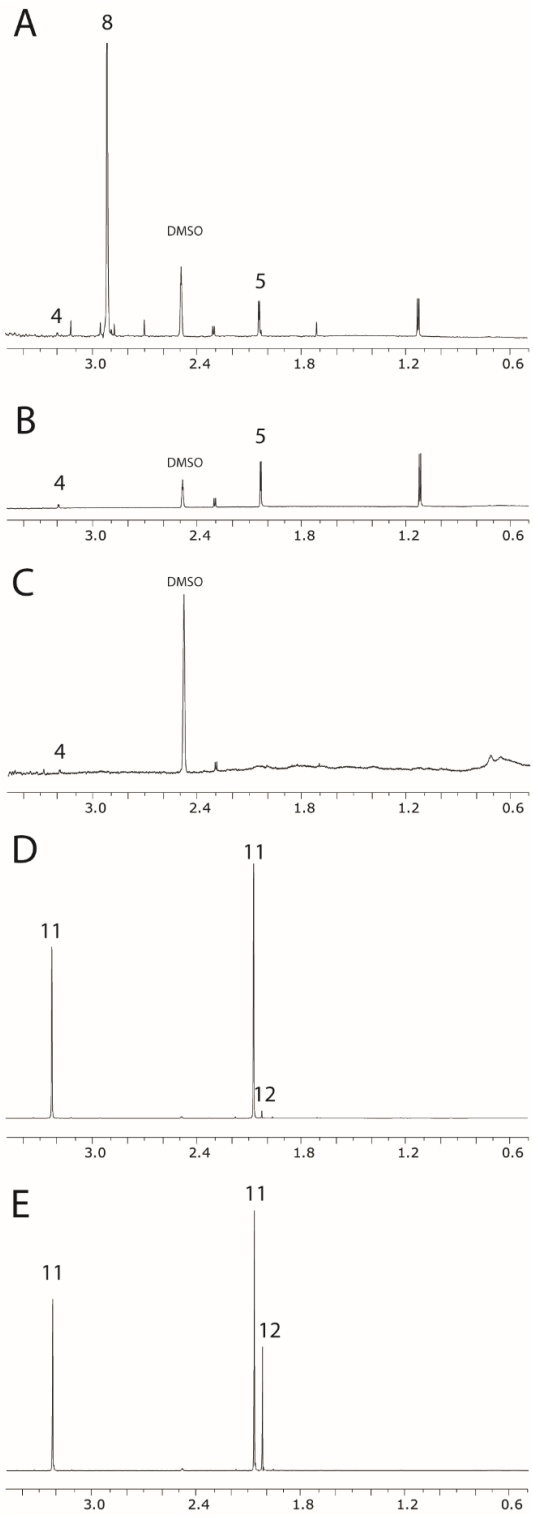

A
8
DMSO
5
4
3.0
2.4
1.8
1.2
0.6
B
5
DMSO
4
3.0
2.4
1.8
1.2
0.6
C
DMSO
4
3.0
2.4
1.8
1.2
0.6
D
11
11
12
3.0
2.4
1.8
1.2
0.6
E
11
11
12
3.0
2.4
1.8
1.2
0.6

Figure 4.5: ^{1}H NMR Spectroscopic Product Analysis of the Reaction of Y103F-Cg10062 with **2**, **3**, **8**, **9**, or **10**.

A) Products of Y103F-Cg10062 with **8** (~95 mM) after 27 min. The reaction is not complete with **8** still present (~2.9 ppm). B) Products of Y103F-Cg10062 with **2** (~63 mM) after 18 min. The reaction is not complete with **2** still present (NMR signals not shown). C) The reaction of Y103F-Cg10062 with **3** (~63 mM) after 18 min. There are no observable signals for products except for a trace of **4** in the baseline. D) Products of Y103F-Cg10062 with **9** (~80 mM) after 3 min. The small amount of **12** present is consistent with non-enzymatic decarboxylation of **11**. E) Products of Y103F-Cg10062 with **10** (~80 mM) after 6 min. All reactions contained ~19.1 μM Y103F Cg10062 and 30 μL of DMSO-d_6. The reactions are individually scaled to show the product reaction. The scale can be assessed by the height of the DMSO peak, which is labeled in the spectra.

4.3.10 Incubation of the E114D and Y103F mutants of Cg10062 with 9 and 10 in the Presence of $NaCNBH_3$.

Compounds **9** and **10** were incubated with the E114D and Y103F mutants overnight in the presence of $NaCNBH_3$ and the residual activity was determined using **9**. Three of the four mixtures had less than 10% remaining activity compared to the control (E114D and **9** or **10** and Y103F and **10**). The sample containing the Y103F mutant and **9** showed no loss in activity (vs the control). The ESI mass spectra of the three samples with reduced activities all showed two major signals: one at 17,083-17,085 Da and the other one at 17,125-17,126 Da (Figures 4.6A and 4.6B and Figure 4.7).[1,21]

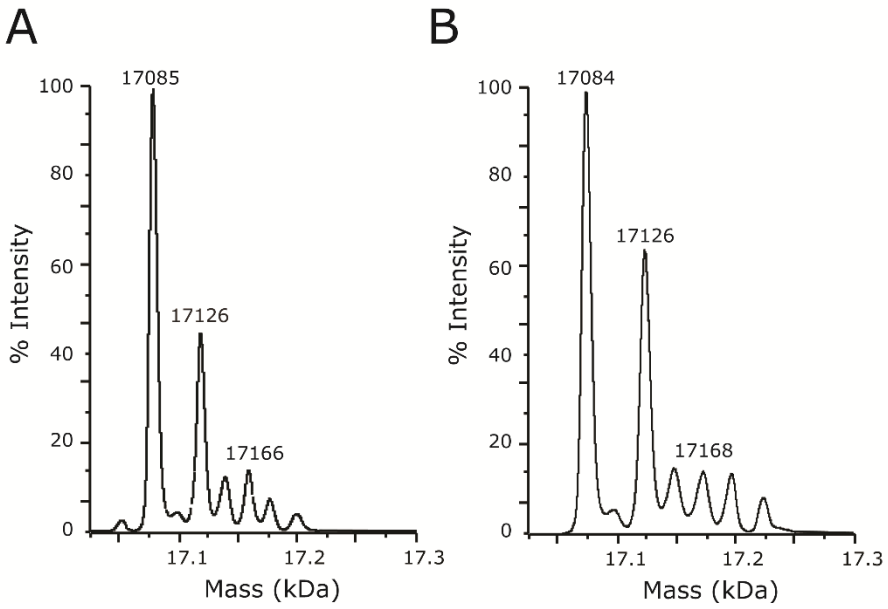


Figure 4.6: ESI-MS Spectra of the E114D Mutant of Cg10062 Incubated with **9** and **10** in the Presence of $NaCNBH_3$.

The signals at 17,085 Da and 17,084 Da (in A and B, respectively) correspond to the unlabeled E114D-Cg10062. The signals at 17,126 Da correspond to E114D-Cg10062 covalently modified by the reduced imine of **12**. The smaller signals at 17,166 Da and 17,168 Da (in A and B, respectively) correspond to E114D-Cg10062 covalently modified by the reduced imine of **11**.

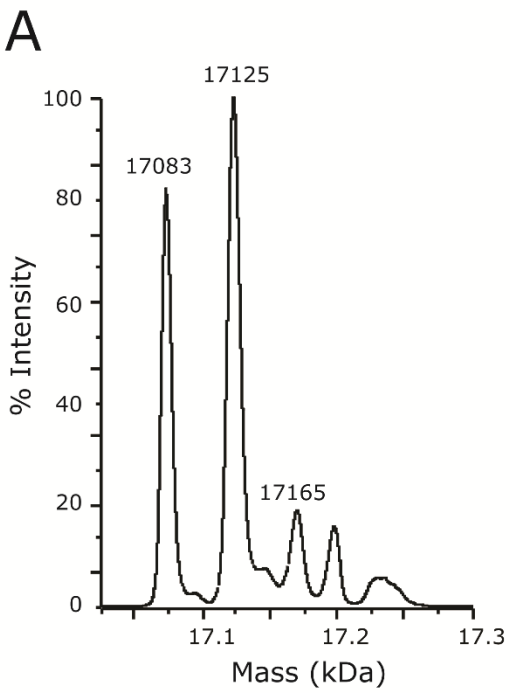


Figure 4.7: ESI-MS Spectra of the Y103F Mutant of Cg10062 Incubated with **10** in the Presence of $NaCNBH_3$.

The signal at 17,083 Da corresponds to the unlabeled Y103F-Cg10062. The signals at 17,125 Da and 17,165 Da correspond to Y103F-Cg10062 covalently modified by the reduced imine of **12** and **11**, respectively.

The first signal corresponds to the expected mass for the unlabeled Cg10062 (calc, 17,085 Da) and the other signal corresponds to the mass of the covalently modified enzyme. The difference in mass is 42 Da, which is consistent with the mass of the reduced imine of acetone (**12**), one of the expected products of the reactions. In all 3 spectra, there is also a smaller signal at 17,165-17,168 Da. The difference in mass from that of control is 81-84 Da, which is consistent with the mass of the reduced imine of acetoacetate (**11**), the other expected product of these reactions. (It is important to note that there is no loss in activity when the E114D or Y103F mutants are incubated with exogenous **11** or **12** in the presence of $NaCNBH_3$.)

4.3.11 Peptide Mapping and MALDI-MS Analysis

The location of the covalent modification was determined by proteolytic digestion using endoproteinase Glu-C (protease V-8), followed by MALDI-MS of the peptide fragments.[1,21] The spectra for the proteolytic digest of the E114D mutant with **9** or **10**, and the Y103F mutant with **10** are shown respectively in Figures 4.8A and 4.8B, and Figure 4.9. All three spectra show a signal corresponding to the unmodified peptide Pro-1-Glu-15 (1896.2-1896.4 Da) as well as second signal corresponding to the same peptide modified by an adduct with a mass 42 Da (1938.3-1938.5 Da). This mass difference is consistent with the addition of the reduced imine of acetone (**12**) to the fragment. An examination of the sequence indicates that Pro-1 is the only chemically reasonable site for imine formation and subsequent covalent modification by reduction with $NaCNBH_3$. A signal for the reduced imine of acetoacetate (**11**) was not detected. The spectra show signals for additional peptide species in the samples (four major peptide fragments for the

E114D mutant and three major peptide fragments for the Y103F mutant). However, none of these peptide fragments is modified by a covalent adduct.

Protease V-8 hydrolyzes peptides at the carboxylate side of glutamate and aspartate residues with a preference for the glutamyl residues.[22,23] Proteolysis is dependent on the buffer conditions as well as the residues adjacent to the site of hydrolysis. Interestingly, the E114D mutant shows a slightly different pattern than that for the Y103F mutant. The protease doesn't hydrolyze the protein at Asp-114 in the E114D mutant, but does hydrolyze at Glu-114 in the Y103F mutant. This observation provides additional confirmation for the identities of two enzymes.

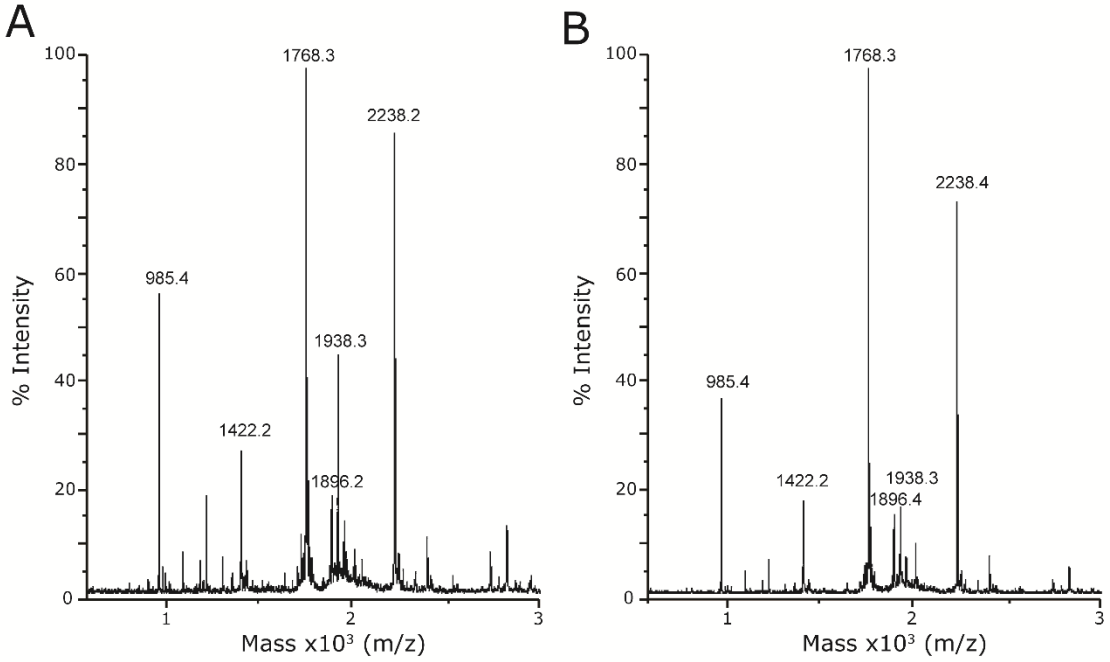


Figure 4.8: MALDI-MS of Spectra of Fragments from the Proteolytic Digest of the E114D Mutant of Cg10062 Incubated with **9** and **10** and Treated with NaCNBH3.

The signals for the six major peptide fragments are labeled and are assigned as follows: 985.4 Da, Leu-81 to Glu-88; 1422.2 Da, Tyr-115 to Glu-126; 1768.3 Da, Asn-61 to Glu-75; 1896.2/1896.4 Da, Pro-1 to Glu-15; 1938.3 Da, Pro-1 to Glu-15 (+ 42 Da); and 2238.2/2238.4 Da, Ile-107 to Glu-126. The mass difference of 42 Da on the Pro-1 to Glu-15 fragment is consistent with the covalent modification of the fragment by the reduced imine of **12**. Only the Pro-1 to Glu-15 fragment has a covalent adduct.

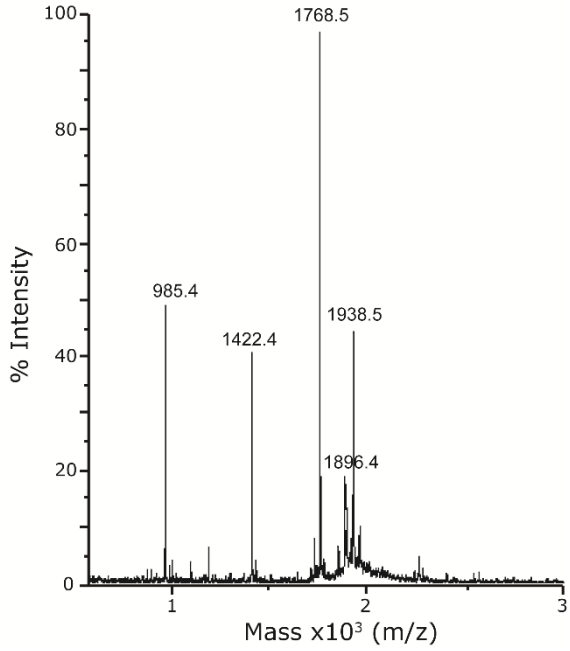


Figure 4.9: MALDI-MS of Spectra of Fragments from the Proteolytic Digest of the Y103F Mutant of Cg10062 Incubated with **10** and Treated with $NaCNBH_3$.

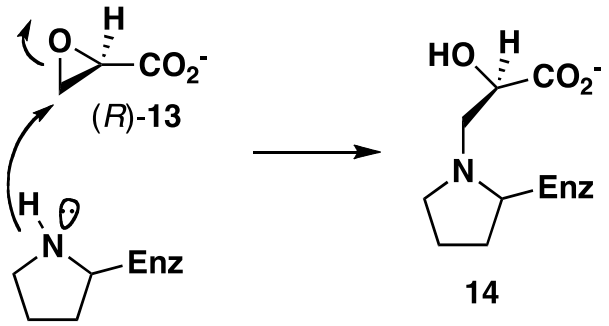
The signals for the five major peptide fragments are labeled and are assigned as follows: 985.4 Da, Leu-81 to Glu-88; 1422.4 Da, Tyr-115 to Glu-126; 1768.5 Da, Asn-61 to Glu-75; 1896.4 Da, Pro-1 to Glu-15; and 1938.5 Da, Pro-1 to Glu-15 (+ 42 Da). The mass difference of 42 Da on the Pro-1 to Glu-15 fragment is consistent with the covalent modification of the fragment by the reduced imine of **12**. Only the Pro-1 to Glu-15 fragment has a covalent adduct.

4.4 DISCUSSION

The mechanism of *cis*-CaaD has been studied extensively[3,11,12,21,24-27], and the results of these studies provide the context for the reactions of Cg10062 with **8**, **9**, and **10**. Sequence analysis first linked *cis*-CaaD to the tautomerase superfamily and identified four of the six residues (Pro-1, Arg-70, Arg-73, and Glu-114) now known to be required for activity.[3] Mutagenesis showed that three of these residues are critical for *cis*-CaaD activity: the P1A, R70A, and R73A mutants of *cis*-CaaD lacked any detectable activity. However, the E114Q mutant retained some activity (8-fold reduction in k_{cat}/K_m using **2**), which hinted that an additional residue(s) might be involved in the activation of a water molecule (later identified at Tyr-103). This initial characterization also showed that *cis*-CaaD catalyzed the hydration of the acetylene compound **6** to produce **7** with a 582-fold reduction in k_{cat}/K_m (compared with that measured for the CaaD-catalyzed reaction).

Analysis of a crystal structure of *cis*-CaaD covalently modified at Pro-1 by (*R*)-2-hydroxypropanoate (**14**, Scheme 4.5) identified a second water-activating residue (Tyr-103) and a sixth residue (His-28).[24,25] Covalent modification resulted from the incubation of *cis*-CaaD with the (*R*)-isomer of oxirane-2-carboxylate (**13**).[24] Mutagenesis confirmed the importance of these two residues for *cis*-CaaD activity. The structure also suggested how substrate might bind in the active site (Scheme 4.6).

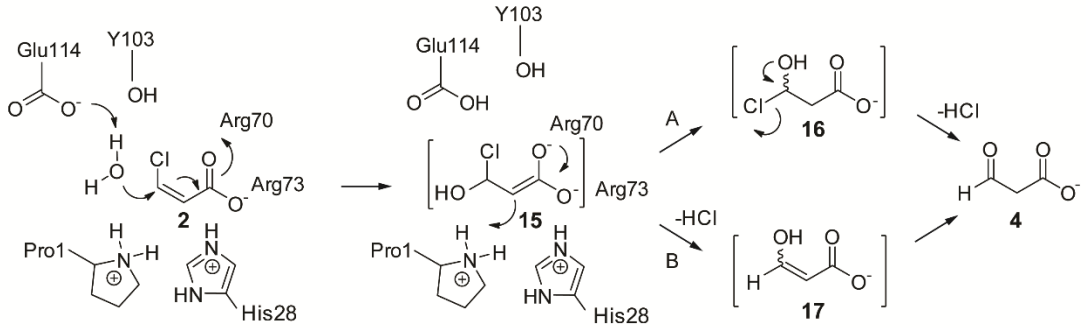
Scheme 4.5: The Irreversible Inactivation of *cis*-CaaD by *(R)*-Oxirane-2-carboxylate (13).



In the course of this study, a pH rate profile of the wild-type reaction identified a protonated group with a pK_a of ~9.3 that is required for activity.[24] The group is proposed to be Pro-1, which is consistent with the observed hydration of **6** to **7** involving a cationic Pro-1.[4]

Based on all of these observations, the mechanism shown in Scheme 4.6 was formulated for *cis*-CaaD (using **2**). In this mechanism, the carboxylate group interacts with the three positively charged residues (Arg-70, Arg-73, and His-28).[25,27] The interactions serve to bind and polarize the substrate. Polarization of the substrate generates a partial positive charge at C-3 and sets up the substrate for a Michael addition of water. Glu-114 and Tyr-103 activate the water molecule for attack at C-3 to produce the *aci*-carboxylate species **15**. This species undergoes one of the two fates shown in paths A and B. In path A, Pro-1 provides a proton at C-2 to produce **16**. Expulsion of HCl from **16** affords **4**. In path B, **15** undergoes an α,β-elimination reaction to produce **17**. Tautomerization of **17** yields **4**.

Scheme 4.6: Proposed Mechanism of *cis*-CaaD and *cis*-3-Chloroacrylate (2).



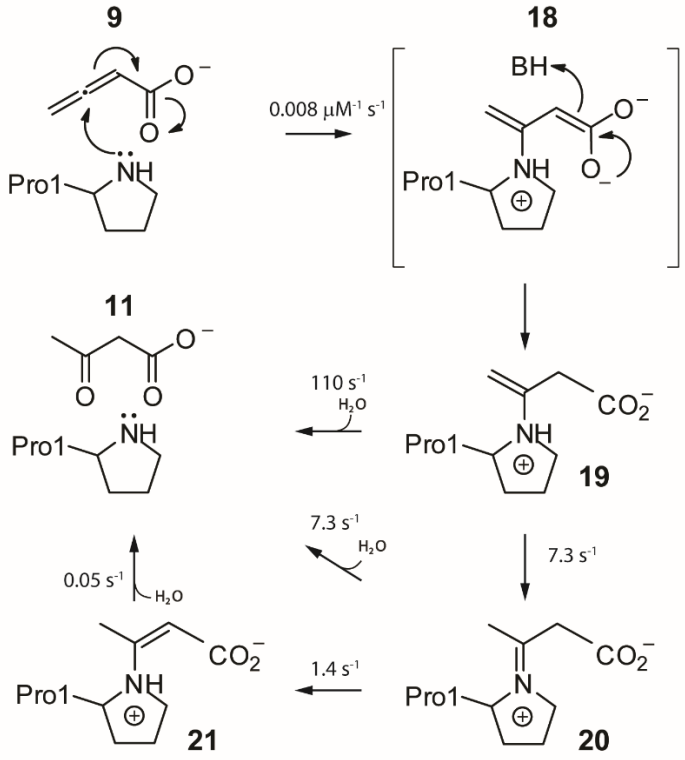
In addition to **6**, compounds **8**, **9**, and **10** were examined as substrates for *cis*-CaaD. Both compounds are reasonably good substrates for *cis*-CaaD with k_{cat}/K_m values of 6.4×10^3 $M^{-1}s^{-1}$ (**8**)[28] and 8.7×10^3 $M^{-1}s^{-1}$ (**9**).[11,21] Compound **10** is not processed to any appreciable extent.[28] For both of the substrates (**8** and **9**) with *cis*-CaaD, the decarboxylation products (**5** or **12**, respectively) are not observed (above non-enzymatic background) nor are the hydration products (**4** or **11**, respectively) decarboxylated by the enzyme.

It was anticipated that *cis*-CaaD would catalyze the direct addition of water to C-3 of **8** and **9**, paralleling the reaction mechanism for the enzyme with **6** (Scheme 4.2A). However, in the presence of **9** and $NaCNBH_4$, *cis*-CaaD was irreversibly inactivated by the covalent modification of the prolyl nitrogen of Pro-1.[21] This prompted a detailed investigation into the mechanism of *cis*-CaaD with **9**. As a result, *cis*-CaaD was shown to catalyze the hydration of **9** almost exclusively through a covalent intermediate (Scheme 4.7). The proposed mechanism for covalent catalysis is the result of observations in both pre-steady state kinetic data and intermediate trapping studies. In

the covalent mechanism, **9** becomes covalently attached to Pro-1 by nucleophilic (or electrophilic) addition at the

C-3 positon. This forms an unstable enolate intermediate (**18**), which *cis*-CaaD stereospecifically incorporates a proton at C-2 yielding an 3,4-enamine intermediate (**19**). Hydrolysis of **19** generates acetoacetate (**11)** and free enzyme. This is proposed to be the kinetically preferred pathway of catalysis. Alternatively, **19** can tautomerize to give the imine species (**20**). This is also subject to hydrolysis, or **20** can tautomerize to the theromodynamically favored 2,3 enamine (**21**). Hydrolysis of **21** yields **11**. This step is proposed to be the rate-limiting step in turnover. This is the first example of covalent catalysis in the tautomerase superfamily.

Scheme 4.7: Proposed Nucleophilic Mechanism of *cis*-CaaD and Allene (9).



A database search for *cis*-CaaD family members identified Cg10062.[1] The conserved six active site residues, the sequence similarity, and the nearly identical active site region suggested that the activity and mechanism would be comparable to those of *cis*-CaaD, even though the genomic context did not show the presence of any enzymes in the 1,3-dichloropropene catabolic pathway or homologues of these enzymes.[1] Characterization of Cg10062 showed that this assumption was not correct.

Our studies uncovered three striking differences. First, Cg10062 is a poor dehalogenase using **2**, and is a non-specific dehalogenase in that it will process **3** (although poorly).[1] Second, the enzyme is inactivated by both isomers of **13**, with the (*R*)-isomer being more potent.[2] This observation parallels the enzyme's ability to process

both **2** and **3**. Finally, Cg10062 catalyzes the direct addition of water to the allene (**9**), whereas *cis*-CaaD uses covalent catalysis to hydrate the compound (Scheme 4.7). A stereochemical analysis of the two enzymes (using **9**) shows that they incorporate a deuteron on opposite faces. The different stereochemical consequences could be a result of the different mechanisms, binding modes, proton sources, or some combination of the three.[21]

Frequently, two closely related enzymes do not have comparable levels of activity using the same substrate(s) or even the same activities.[29-31] This observation can be due to subtle changes in the active site that alter the positions of catalytic and/or binding residues. It can also result from elements outside the active site where the sequence similarities are not as high.[30,31] These elements might include a mobile loop or conformational dynamics.[11,30,31] Finally, the two enzymes might have different substrates or entirely different activities. The work presented here shows that Cg10062 has a different activity than *cis*-CaaD using the same substrate.

The reaction of *cis*-CaaD and **8** produces only **4**. In contrast, the reaction of **8** and Cg10062 produces a small amount of **4** (25%) and mostly **5** (75%). Hence, Cg10062 functions as a hydratase/decarboxylase, whereas *cis*-CaaD functions only as a hydratase. The unique feature of the Cg10062-catalyzed reaction is that the decarboxylation reaction *depends* on the hydration reaction. In other words, Cg10062 does not catalyze the decarboxylation of **4** when added exogenously. It is known that the exogenously added **4** binds at the active site with a K_d of about 25 mM. Moreover, although **4** is a reasonable intermediate in the transformation of **8** to **5**, it might not be generated as a discrete

intermediate. Understanding how Cg10062 catalyzes the decarboxylation reaction of this hydration-dependent reaction is critical for elucidating the enzyme's mechanism.

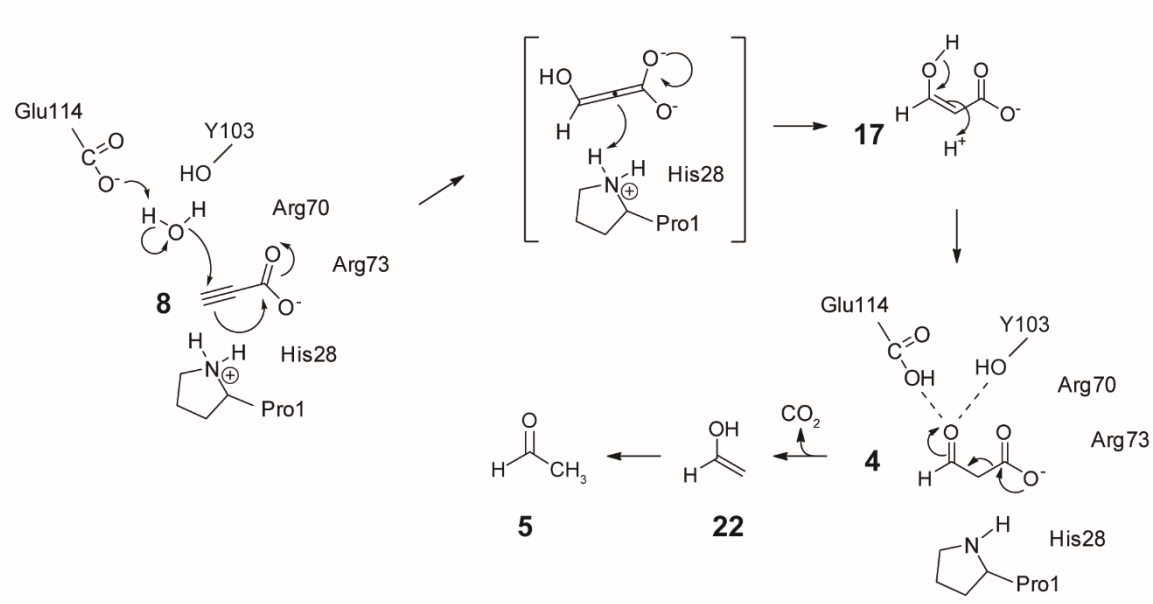
Several mechanisms have been reported for the decarboxylation of β-keto acids.[32-34] For a start, the hydration-dependent decarboxylation reaction reported here is coenzyme and metal ion independent (as is the hydration reaction), eliminating mechanisms that require a cofactor. Cofactor-independent β-ketoacid decarboxylases have been described that use a Schiff base mechanism (e.g., acetoacetate decarboxylase[32]) or an oxyanion hole (e.g., MSAD[10,20,33] and methylmalonyl CoA decarboxylase[34]). For Cg10062, the Schiff base would form between Pro-1 and substrate. These mechanisms might not be directly applicable to Cg10062 because the decarboxylation reaction is entirely dependent on the hydration reaction. This suggests that Cg10062 might proceed by an unprecedented mechanism.

In view of these considerations, three mechanisms can be envisioned for the hydratase/decarboxylase activity of Cg10062 where the interaction of the substrate carboxylate group with Arg-70, Arg-73, and His-28 most likely positions it in the active site. The first mechanism involves the Cg10062-catalyzed hydration of **8** to produce **4**, followed by the decarboxylation of **4** to yield **5** (Scheme 4.8). Water is activated by the combination of Glu-114 and Tyr-103 and **8** is polarized by the combination of Arg-70, Arg-73, and His-28 (analogous to the roles assigned to these residues in Scheme 4.6). Water is added to the C-3 of **8** resulting in an unstable allene intermediate that is immediately protonated at C-2. This results in **17**, which tautomerizes to **4**. Pro-1 (with an assumed pK_a value of 9.2) can provide the proton at C-2 in the initial hydration step

(to form **17**) or in the tautomerization step. A small amount of **4** (~25%) is released from the active site and the remaining **4** is decarboxylated to give **22** by an unknown mechanism. After decarboxylation, Glu-114 and Tyr-103 could be in a favorable protonation state to facilitate decarboxylation. **22** can be tautomerized to give **5**. An analogous mechanism can be envisioned for the hydration/decarboxylation of **2**.

This mechanism is at odds with the observation that exogenously added **4** is not decarboxylated by the enzyme and argues against it. However, if this mechanism is operative, one could argue that the hydration of **8** positions **4** in an orientation for decarboxylation. The positioning of **4** in the active site appears to be an important factor for its decarboxylation. The hydration of **3** also produces **4**, but **4** is not decarboxylated. Hence, the hydration of **8** (and **2**) could place **4** in a favorable state for decarboxylation, which is not achieved by the hydration of **3** or by exogenously added **4**. Another possibility is that **4** is not generated as a distinct entity in the hydration-dependent decarboxylation of **8** (and **2**).

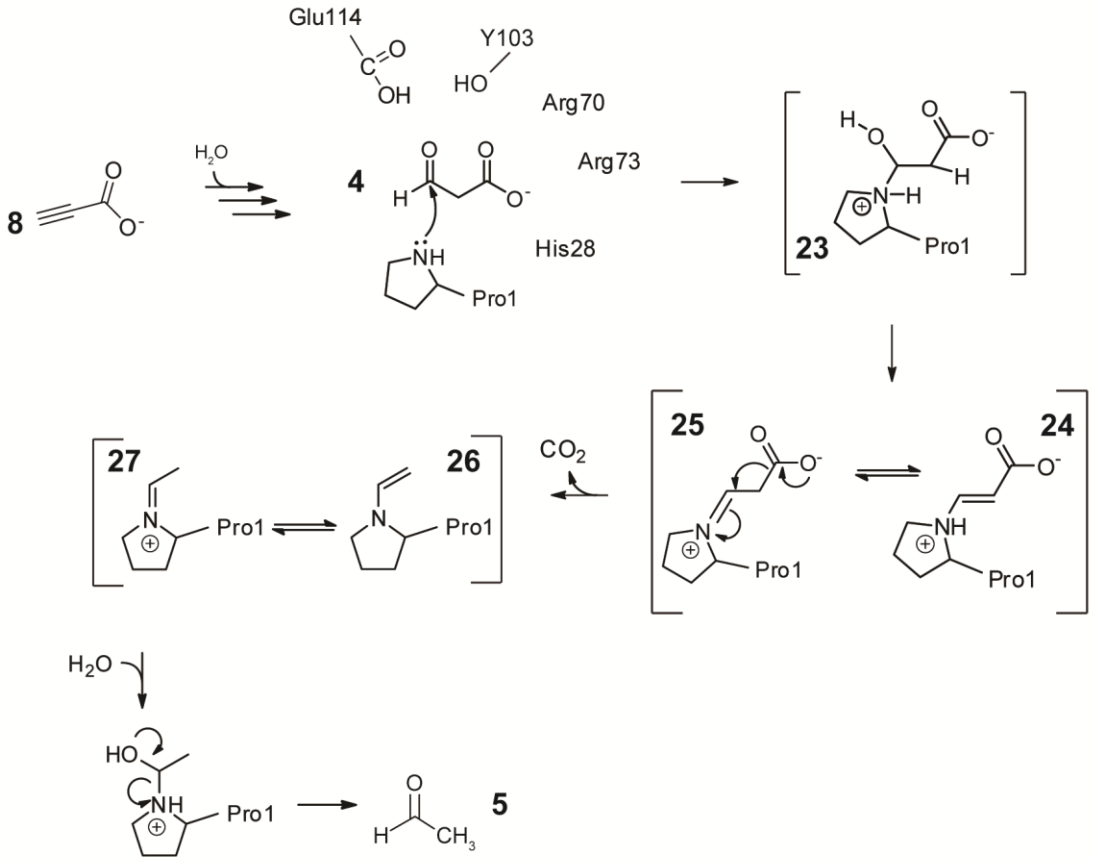
Scheme 4.8: The Cg10062-catalyzed Hydration of Propiolate (8) by Direct Attack of Water Followed by Decarboxylation



In a second mechanism, the hydration step is followed by the formation of a covalent intermediate between Pro-1 (now nucleophilic because the proton was removed in the hydration step) and C-3 of **4** (i.e., a Schiff base) (Scheme 4.9). The reactivity of the aldehyde could facilitate the process. This would form an unstable intermediate, **23**, which could undergo dehydration to give **24** (or **25**, depending on the proton elimination). Decarboxylation of the Schiff base (**25**) generates the enamine, **26**, which can tautomerize to **27**. Hydrolysis of **27** generates **5** and free enzyme. A small amount of **4** could escape from the active site and not form a Schiff base (accounting for the 25% of **4** in the final mixture). In this mechanism, the exogenously added **4** might not be decarboxylated because Pro-1 is not in the correct protonation state to form the Schiff

base intermediate. The hydration reaction must necessarily occur first in order to put Pro-1 into the correct state. The same sequence of events could process **2**.

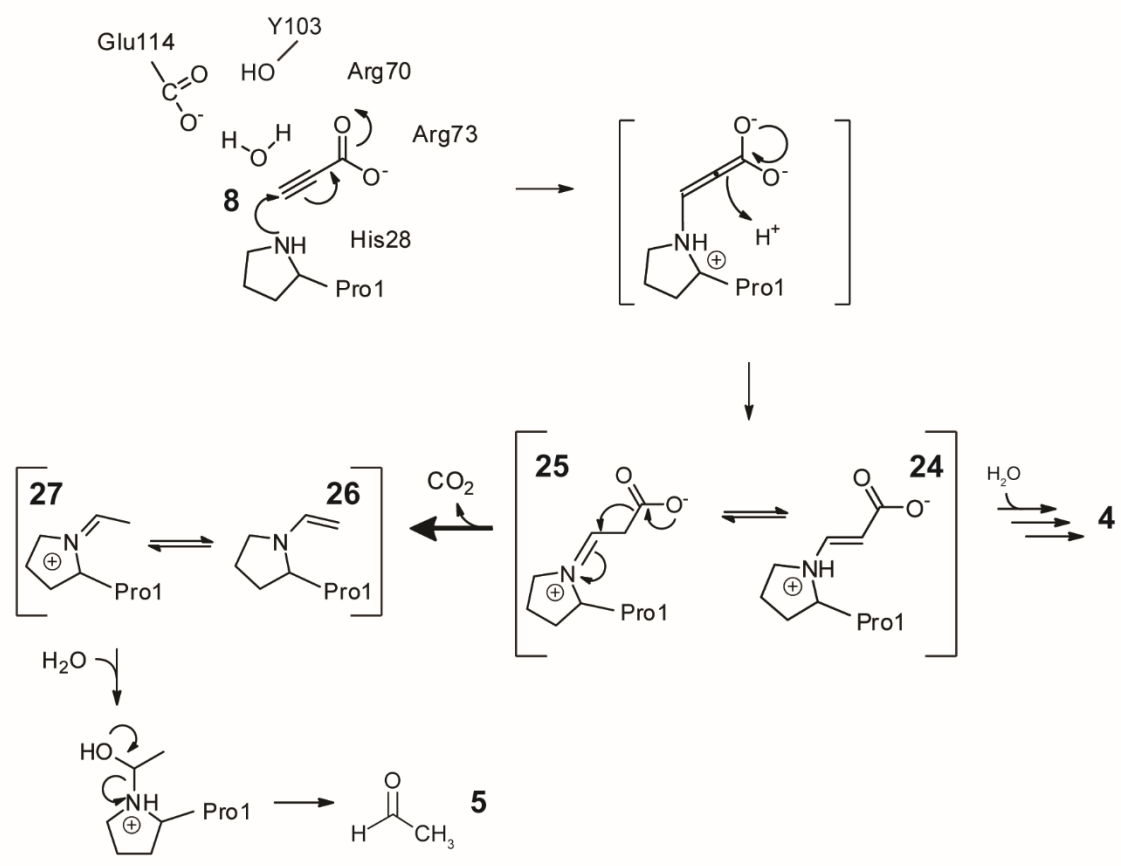
Scheme 4.9: The Cg10062-catalyzed Hydration of Propiolate (8) by Direct Attack of Water Followed by Decarboxylation via a Schiff Base.



The third mechanism involves covalent catalysis using Pro-1 (Scheme 4.10). Accordingly, Pro-1 adds to C-3 of **8** followed by a proton transfer to form **24**, which can rearrange to **25**. The occasional hydrolysis of **24** (or **25**) produces the small amount (25%) of **4**. Most of the time (75%), the decarboxylation of **25** forms **26**, which

tautomerizes to **27**. Hydrolysis of **27** yields **5** and free enzyme. An analogous mechanism is not likely for **2**.

Scheme 4.10: The Cg10062-catalyzed Hydration of Propiolate (8) by Covalent Catalysis Mechanism.



One potential drawback to this third mechanism is that the apparent pK_a of Pro-1 (~9.2), inferred again from the pH rate profile of *cis*-CaaD, suggests that the prolyl nitrogen is largely cationic and unable to function as a nucleophile. If this mechanism is operative, there could be two explanations for this discrepancy: the pH rate profile does

not reflect the actual pK_a of the prolyl nitrogen in Cg10062 or the reaction proceeds using the small amount of enzyme with the prolyl nitrogen in the uncharged form. (For the third mechanism, the small amount of enzyme in the uncharged form would readily react with the aldehyde moiety of exogenously added **4** and result in decarboxylation. This does not happen. A satisfactory explanation is not apparent.)

It is unknown how Cg10062 catalyzes the decarboxylation reaction in the hydration-dependent decarboxylation sequence in the first mechanism (Scheme 4.8). The Schiff base in the second and third mechanisms (Schemes 4.9 and 4.10) lends itself to decarboxylation (although it is not known which residues and other features of the active site are involved). Obviously, delineating the decarboxylation mechanism would narrow the mechanistic possibilities.

The reactions of **9** and **10** with Cg10062 produce only the hydration product (**11** for both substrates). Hence, for these substrates, Cg10062 functions as a hydratase. The most obvious difference between **8** and these substrates is length (3 vs. 4 carbons). These observations could suggest that the positioning of substrate in the active site plays an important role in the outcome (hydration vs hydration/decarboxylation). In addition, the hydration product, **11**, is a ketone with an extra methyl group (vs **4**). These features could further contribute to the reaction outcome. However, there still remains any number of explanations without knowing how Cg10062 catalyzes the decarboxylation part of the hydration-dependent decarboxylation of **8**.

In an effort to understand the Cg10062 reaction mechanism, the two water-activating residues (Glu-114 and Tyr-103) were examined by mutagenesis and their

reactions with **8**, **9**, and **10** characterized (Table 4.5, Figure 4.10). This analysis produced a range of results, which are mostly perplexing and not obviously explained. The easier results to rationalize are those that involve a reduction in wild type activity (e.g., Y103F with **8**). In these cases, the mutation could disrupt binding, the catalytic machinery, or both, to produce a less efficient enzyme.

Substrate	Wild Type	Y103F	E114Q	E114D
2	Hydratase/ Decarboxylase	Hydratase/ Decarboxylase	Hydratase	Hydratase
3	Hydratase	ND[a]	ND[a]	ND[a]
8	Hydratase/ Decarboxylase	Hydratase/ Decarboxylase	Hydratase	Hydratase
9	Hydratase	Hydratase	Hydratase/ Decarboxylase	Hydratase[b]
10	Hydratase[b]	Hydratase/ Decarboxylase	Hydratase/ Decarboxylase	Hydratase[b]

[a]Activity is not detected by UV or ^{1}H NMR spectroscopy. [b]Trace amounts of **12** (~0.3-2.8%) are present in the ^{1}H NMR spectra, but these enzymes function primarily as hydratases. Assay conditions are described in the text.

Table 4.5. Summary of Activities of Cg10062 and Mutants

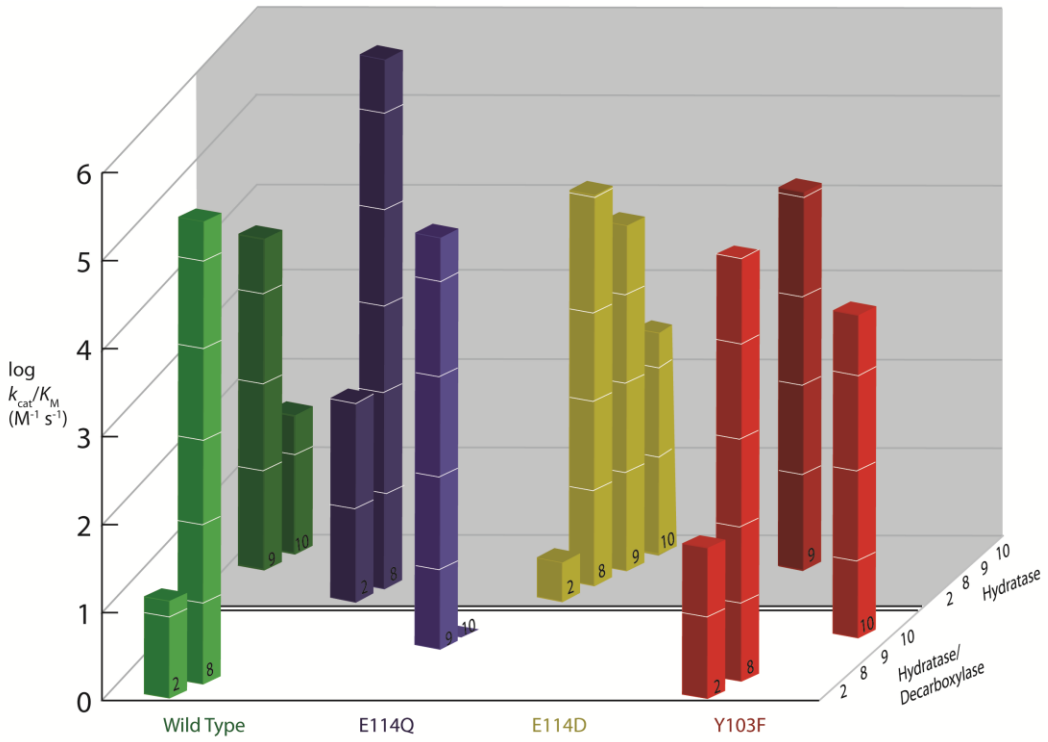


Figure 4.10: Visual Summary of Activities for the Wild Type, E114Q-, E114D-, and Y103F-Mutants of Cg10062 with **2**, **8**, **9**, and **10**.

Shown in shades of green, blue, yellow, and red columns (bars) are the activities for the four substrates (**2**, **8**, **9**, and **10**) with the wild-type, E114Q-, E114D-, and Y103F-mutants of Cg10062, respectively. Bars in front of the double line (i.e. in the white panel, lighter shades of color) represent enzyme and substrate combinations that showed hydratase/decarboxylase activity. Bars behind the double line (i.e. in the gray panel, darker shades of color) represent enzyme and substrate combinations that showed hydratase only activity. The height of the bars represents the log (k_{cat}/K_m) values for the four substrates.

It is more difficult to explain the switch in activities where the hydratase/decarboxylases lose decarboxylase activity or the hydratases gain decarboxylase activity. With **2** and **8**, the E114 mutants (E114Q and E114D) lose decarboxylase activity and function only as hydratases. With **2**, the E114D mutant is a poor hydratase, whereas with **8**, it has more significant activity. This is based on the ^{1}H NMR analysis, which shows that the mutant generates product at a faster rate using **8**. The same analysis shows that the E114Q mutant is a better hydratase with **2** than it is with **8** (but not as effective as the E114D mutant with **10**). The other switch in activity is the gain of function. With **9** and **10**, the E114Q mutant functions as a hydratase/decarboxylase as does the Y103F mutant with **10**. ^{1}H NMR analysis also suggests that the Y103F mutant (with **10**) and the E114Q mutant (with **9**) have significant hydratase/decarboxylase activity, (but not comparable to the hydratase/decarboxylase activity of the wild type with **8**). Several factors could be responsible for the gain of function ranging from a longer time in the active site to a more roomy, less constrained active site resulting in accessibility to C-3 of the substrate. Again, understanding how Cg10062 and the mutants catalyze the decarboxylation part of the hydration-dependent decarboxylation of **8** shortens the list of possibilities.

For two of the mutant enzymes (E114D and Y103F), it was possible to trap the equivalent of the hydration product (**11** from **9** or **10** processed by E114D and **11** from **10** processed by the Y103F mutant) and the decarboxylation product (**12**) as the reduced imine covalently attached to Pro-1 (using $NaCNBH_3$). It is not possible to trap the reduced imine by incubating the two mutants with exogenously added **11** or **12**. This

result is tantalizing, and seems to suggest that covalent catalysis is operative (Schemes 4.9 or 4.10). However, there are concerns. First, it is not possible to trap any imine when the same experiment is carried out with the E114Q mutant and **9**. This enzyme has a pronounced hydratase/decarboxylase activity (after 12 min, the mixture consists of 83% of **11** and 17% of **12**). Second, it is not clear why the imine of **12** is trapped when the E114D mutant is incubated with **9** and **10** because **12** is not present in very large amounts (~2.5%). These observations raise the possibility that the trapping results are experimental artifacts. Other experiments (not shown) with E114Q with **10** suggests that the presence of $NaCNBH_3$ alters the course of the reaction (i.e. more **11** is present than in the absence of $NaCNBH_3$, where **12** is observed to be the primary product).

Although **8** is clearly the "best" substrate identified for Cg10062 thus far (as suggested by the high k_{cat}/K_m of 1.8×10^5 M^{-1} s^{-1}), the biological relevance of the hydratase/decarboxylase activity with this substrate is not clear. Acetylene-bearing natural products are well known (e.g., the enediyne antibiotics) and there are various reports about the bacterial catabolism of acetylenes, but not in *C. glutamicum.*[35-37] The genomic context for Cg10062 is not informative because the adjacent genes have unknown functions or annotated functions that are not useful.[1] There are, however, two reports about the bacterial transformation of **8** that could have implications for the biological relevance. First, a strain of *Pseudomonas putida* that was isolated from rotting fruit was shown to use **8** as a sole carbon source.[36] ^{1}H NMR spectroscopy following the transformation of **8** in these *P. putida* cells showed the transformation of **8** to **4** and **5**. However, the enzyme(s) responsible for this transformation and whether the conversion

of **4** to **5** is an enzyme-catalyzed reaction (or not) were not reported. The identification of the enzyme responsible for this transformation and any genomic context could shed light on the role of Cg10062. A second report in the KEGG database outlines a pathway for the conversion of **8** to **4**, followed by the conversion of **4** to β-alanine.[37] The basis for this series of conversions is not entirely clear, but β-alanine can be processed to pantothenic acid in *C. glutamicum*, which is known for the overproduction of pantothenic acid.[38-40] (β-alanine can be routed to other pathways as well.[41]) This pathway doesn't account for the production of **5** in any obvious way and the connection between Cg10062 and the pathway is tenuous at best, but it is an intriguing possibility.

Finally, the hydratase/decarboxylase activity of wild type Cg10062 with the *cis*-3-haloacrylates has an intriguing evolutionary implication. The isomer-specific dehalogenases (*cis*-CaaD and CaaD) and MSAD are all tautomerase superfamily members that are found in the 1,3-dichloropropene catabolic pathway, where they catalyze successive reactions (Scheme 4.1).[4,6,7] *cis*-CaaD and MSAD have additional similarities: both are homotrimers with comparably sized monomers (129 amino acids for MSAD and 149 amino acids for *cis*-CaaD)[25,27,33] and both catalyze the hydration of **6** to yield **7** (Scheme 4.2).[3,10,13] For these reasons, the dehalogenases and MSAD might have diverged from a common ancestral enzyme that catalyzed both reactions.[1] Although Cg10062 might not be the progenitor, it might be representative of the progenitor.

4.5 REFERENCES

(1) Poelarends, G. J.; Serrano, H.; Person, M. D.; Johnson, W. H.; Whitman, C. P. (2008) Characterization of Cg10062 from *Corynebacterium glutamicum*: implications for the evolution of *cis*-3-chloroacrylic acid dehalogenase activity in the tautomerase superfamily. *Biochemistry. 47*, 8139-8147.

(2) Robertson, B. A.; Johnson, W. H.; Lo, H. H., Jr.; Whitman, C. P. (2008) Inactivation of Cg10062, a *cis*-3-chloroacrylic acid dehalogenase homologue in *Corynebacterium glutamicum*, by *(R)*- and *(S)*-oxirane-2-carboxylate: analysis and implications. *Biochemistry. 47*, 8796-8803.

(3) Poelarends, G. J.; Serrano, H.; Person, M. D.; Johnson, W. H., Jr.; Murzin, A. G.; Whitman, C. P. (2004) Cloning, expression, and characterization of a *cis*-3-chloroacrylic acid dehalogenase: insights into the mechanistic, structural, and evolutionary relationship between isomer-specific 3-chloroacrylic acid dehalogenases. *Biochemistry. 43*, 759-772.

(4) Poelarends, G. J.; Veetil, V. P.; Whitman, C. P. (2008) The chemical versatility of the β−α−β fold: catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell. Mol. Life Sci. 65*, 3606-3618.

(5) Murzin, A. G. (1996) Structural classification of proteins: new superfamilies. *Curr. Opin. Struct. Biol. 6*, 386-394.

(6) Poelarends, G. J.; Wilkens, M.; Larkin, M. J.; Van Elsas, J. D.; Janssen, D. B. (1998) Degradation of 1,3-dichloropropene by *Pseudomonas cichorii* 170. *Appl. Environ. Microbiol. 64*, 2931-2936.

(7) van Hylckama Vlieg, J. E.; Janssen, D. B. (1991) Bacterial degradation of 3-chloroacrylic acid and the characterization of *cis*- and *trans*-specific dehalogenases. *Biodegradation. 2*, 139-150.

(8) Hartmans, S.; Jansen, M. W.; Van der Werf, M. J.; De Bont, J. A. M. (1991) Bacterial metabolism of 3-chloroacrylic acid. *J. Gen. Microbiol. 137*, 2025-2032.

(9) Poelarends, G. J.; Saunier, R.; Janssen, D. B. (2001) *trans*-3-Chloroacrylic acid dehalogenase from *Pseudomonas pavonaceae* 170 shares structural and mechanistic similarities with 4-oxalocrotonate tautomerase. *J. Bacteriol. 183*, 4269-4277.

(10) Poelarends, G. J.; Johnson, H. W., Jr.; Murzin, A. G.; Whitman, C. P. (2003) Mechanistic characterization of a bacterial malonate semialdehyde decarboxylase: identification of a new activity in the tautomerase superfamily. *J. Biol. Chem. 278*, 48674-48683.

(11) Schroeder, G. K.; Huddleston, J. P.; Johnson, W. H., Jr.; Whitman, C. P. (2013) A mutational analysis of the active site loop residues in *cis*-3-chloroacrylic acid dehalogenase. *Biochemistry. 52*, 4204-4216.

(12) Robertson, B. A.; Schroeder, G. K.; Jin, Z.; Johnson, K. A.; Whitman, C. P. (2009) Pre-steady-state kinetic analysis of *cis*-3-chloroacrylic acid dehalogenase: analysis and implications. *Biochemistry. 48*, 11737-11744.

(13) Wang, S. C.; Person, M. D.; Johnson, W. H.; Whitman, C. P. (2003) Reactions of *trans*-3-chloroacrylic acid dehalogenase with acetylene substrates: consequences of and evidence for a hydration reaction. *Biochemistry. 42*, 8762-8773.

(14) Eglinton, G.; Jones, E. R. H.; Mansfield, G. H.; Whiting, M. C. (1954) Researches on acetylenic compounds. Part XLV. The alkaline isomerization of but-3-ynoic acid. *J. Chem. Soc.*, 3197-3200.

(15) Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

(16) Waddell, W. J. (1956) A simple ultraviolet spectrophotometric method for the determination of protein. *J. Lab. Clin. Med. 48*, 311-314.

(17) Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature. 227*, 680-685.

(18) Serrano, H. Characterization of the activites of *trans*-3-chloroacrylic acid dehalogenase and *cis*-3-chloroacrylic acid dehalogenase and malonate semialdehyde decarboxylase homologues: mechanism and evolutionary implications, University of Texas at Austin, 2009.

(19) Johnson, K. A. (1992) Transient-state kinetic analysis of enzyme reaction pathways In *The Enzymes* (Sigman, D. S., Ed.), 3rd Ed., pp 1-61, Academic Press, San Diego

(20) Guo, Y.; Serrano, H.; Poelarends, G. J.; Johnson, W. H., Jr.; Hackert, M. L.; Whitman, C. P. (2013) Kinetic, mutational, and structural analysis of malonate semialdehyde decarboxylase from *Coryneform bacterium* strain FG41: mechanistic implications for the decarboxylase and hydratase activities. *Biochemistry. 52*, 4830-4841.

(21) Schroeder, G. K.; Johnson, W. H., Jr.; Huddleston, J. P.; Serrano, H.; Johnson, K. A.; Whitman, C. P. (2012) Reaction of *cis*-3-chloroacrylic acid dehalogenase with an allene substrate, 2,3-butadienoate: hydration via an enamine. *J. Am. Chem. Soc. 134*, 293-304.

(22) Houmard, J.; Drapeau, G. R. (1972) Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. *Proc. Natl. Acad, Sci. U.S.A. 69*, 3506-3509.

(23) Sørensen, S. B.; Sørensen, T. L.; Breddam, K. (1991) Fragmentation of proteins by *S. aureus* strain V8 protease. Ammonium bicarbonate strongly inhibits the enzyme but does not improve the selectivity for glutamic acid. *FEBS Lett. 294*, 195-197.

(24) Poelarends, G. J.; Serrano, H.; Johnson, W. H., Jr.; Whitman, C. P. (2004) Stereospecific alkylation of *cis*-3-chloroacrylic acid dehalogenase by *(R)*-oxirane-2-carboxylate: analysis and mechanistic implications. *Biochemistry. 43*, 7187-7196.

(25) De Jong, R. M.; Bazzacco, P.; Poelarends, G. J.; Johnson, W. H., Jr.; Kim, Y. J.; Burks, E. A.; Serrano, H.; Thunnissen, A.-M. W. H.; Whitman, C. P.; Dijkstra, B. W. (2007) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase. Structural basis for substrate specificity and inactivation by *(R)*-oxirane-2-carboxylate. *J. Biol. Chem. 282*, 2440-2449.

(26) Sevastik, R.; Whitman, C. P.; Himo, F. (2009) Reaction mechanism of *cis*-3-chloroacrylic acid dehalogenase: a theoretical study. *Biochemistry. 48*, 9641-9649.

(27) Guo, Y.; Serrano, H.; Johnson, W. H.; Ernst, S.; Hackert, M. L.; Whitman, C. P. (2011) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase: implications for the catalytic and inactivation mechanisms. *Bioorg. Chem. 39*, 1-9.

(28) Huddleston, J. P.; Johnson Jr, W. H.; Schroeder, G. K.; Whitman, C. P. (2015) The accidental assignment of function in the tautomerase superfamily. *Perspectives in Science.* (in press).

(29) Lu, Z.; Dunaway-Mariano, D.; Allen, K. N. (2011) The x-ray crystallographic structure and specificity profile of HAD superfamily phosphohydrolase BT1666: comparison of paralogous functions in *B. thetaiotaomicron*. *Proteins. 79*, 3099-3107.

(30) Li, L.; Luo, M.; Ghanem, M.; Taylor, E. A.; Schramm, V. L. (2008) Second-sphere amino acids contribute to transition-state structure in bovine purine nucleoside phosphorylase. *Biochemistry. 47*, 2577-2583.

(31) Whittier, S. K.; Hengge, A. C.; Loria, J. P. (2013) Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. *Science (New York, N.Y.). 341*, 899-903.

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

(33) Almrud, J. J.; Poelarends, G. J.; Johnson, W. H., Jr.; Serrano, H.; Hackert, M. L.; Whitman, C. P. (2005) Crystal structures of the wild-type, P1A mutant, and inactivated malonate semialdehyde decarboxylase: a structural basis for the decarboxylase and hydratase activities. *Biochemistry. 44*, 14818-14827.

(34) Benning, M. M.; Haller, T.; Gerlt, J. A.; Holden, H. M. (2000) New reactions in the crotonase superfamily: structure of methylmalonyl CoA decarboxylase from *Escherichia coli*. *Biochemistry. 39*, 4630-4639.

(35) Yamada, E. W.; Jakoby, W. B. (1959) Enzymatic utilization of acetylenic compounds: II. Acetylenemonocarboxylic acid hydrase. *J. Biol. Chem. 234*, 941-945.

(36) Brecker, L.; Petschnigg, J.; Depiné, N.; Weber, H.; Ribbons, D. W. (2003) In situ proton NMR analysis of α-alkynoate biotransformations. *Eur. J. Biochem. 270*, 1393-1398.

(37) Kanehisa, M.; Goto, S.; Kawashima, S.; Nakaya, A. (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res. 30*, 42-46.

(38) Dusch, N.; Pühler, A.; Kalinowski, J. (1999) Expression of the *Corynebacterium glutamicum panD* gene encoding L-aspartate-α-decarboxylase leads to pantothenate overproduction in *Escherichia coli*. *Appl. Environ. Microbiol. 65*, 1530-1539.

(39) Hüser, A. T.; Chassagnole, C.; Lindley, N. D.; Merkamm, M.; Guyonvarch, A.; Elišáková, V.; Pátek, M.; Kalinowski, J.; Brune, I.; Pühler, A.; Tauch, A. (2005) Rational design of a *Corynebacterium glutamicum* pantothenate production strain and its characterization by metabolic flux analysis and genome-wide transcriptional profiling. *Appl. Environ. Microbiol. 71*, 3255-3268.

(40) Lange, C.; Rittmann, D.; Wendisch, V. F.; Bott, M.; Sahm, H. (2003) Global expression profiling and physiological characterization of *Corynebacterium glutamicum* grown in the presence of L-valine. *Appl. Environ. Microbiol. 69*, 2521-2532.

(41) Merkamm, M.; Chassagnole, C.; Lindley, N. D.; Guyonvarch, A. (2003) Ketopantoate reductase activity is only encoded by *ilvC* in *Corynebacterium glutamicum*. *J. Biotechnol. 104*, 253-260.

References

Almrud, J. J. *et .al.* (2005) Crystal structures of the wild-type, P1A mutant, and inactivated malonate semialdehyde decarboxylase: a structural basis for the decarboxylase and hydratase activities. *Biochemistry. 44*, 14818-14827.

Azurmendi, H. F. *et .al.* (2004) The roles of active-site residues in the catalytic mechanism of *trans*-3-chloroacrylic acid dehalogenase: a kinetic, NMR, and mutational analysis. *Biochemistry. 43*, 4082-4091.

Baas, B. J. *et .al.* (2011) Characterization of a newly identified mycobacterial tautomerase with promiscuous dehalogenase and hydratase activities reveals a functional link to a recently diverged *cis*-3-chloroacrylic acid dehalogenase. *Biochemistry. 50*, 2889-2899.

Babbitt, P. C. *et .al.* (1997) Understanding enzyme superfamilies. *J. Biol. Chem. 272*, 30591-30594.

Benning, M. M. *et .al.* (2000) New reactions in the crotonase superfamily: structure of methylmalonyl CoA decarboxylase from *Escherichia coli*. *Biochemistry. 39*, 4630-4639.

Benning, M. M. *et .al.* (1996) Structure of 4-chlorobenzoyl coenzyme A dehalogenase determined to 1.8 Å resolution: an enzyme catalyst generated via adaptive mutation. *Biochemistry. 35*, 8103-8109.

Brecker, L. *et .al.* (2003) In situ proton NMR analysis of α-alkynoate biotransformations. *Eur. J. Biochem. 270*, 1393-1398.

Burks, E. A. *et .al.* (2010) Kinetic and structural characterization of a heterohexamer 4-oxalocrotonate tautomerase from *Chloroflexus aurantiacus* J-10-fl: Implications for functional and structural diversity in the tautomerase superfamily. *Biochemistry. 49*, 5016-5027.

Burks, E. A. *et .al.* (2011) Kinetic, crystallographic, and mechanistic characterization of TomN: Elucidation of a function for a 4-oxalocrotonate tautomerase hmologue in the tomaymycin biosynthetic pathway. *Biochemistry. 50*, 7600-7611.

Copley, S. D. (1999) Microbial dehalogenases In *Compreheive Natural Products Chemistry* (Barton, D., Nakanishi, K., Eds.), pp 401-422, Elsevier, Amsterdam

Cornish-Bowden, A. (1974) Simple graphical method for determining the inhibition constants of mixed, uncompetitive, and noncompetitive inhibitors. *Biochem. J. 137*, 143-144.

Darwin, C. *Origin of Species: By Means of Nature Selection*; Cambridge University Press: Cambridge, 1876.

De Jong, R. M. *et .al.* (2007) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase. Structural basis for substrate specificity and inactivation by *(R)*-oxirane-2-carboxylate. *J. Biol. Chem. 282,* 2440-2449.

De Jong, R. M. *et .al.* (2004) The x-ray structure of *trans*-3-chloroacrylic acid dehalogenase reveals a novel hydration mechanism in the tautomerase superfamily. *J. Biol. Chem. 279,* 11546-11552.

DeLano, W. L. (2002) *The PyMOL molecular graphics system*; DeLano Scientific, San Carlos, CA.

Dellus-Gur, E. *et .al.* (2013) What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J. Mol. Biol. 425*, 2609-2621.

Dixon, M. (1953) Determination of enzyme-inhibitor constants. *Biochem. J. 55*, 170-171.

Dusch, N. *et .al.* (1999) Expression of the *Corynebacterium glutamicum panD* gene encoding L-aspartate-α-decarboxylase leads to pantothenate overproduction in *Escherichia coli*. *Appl. Environ. Microbiol. 65*, 1530-1539.

Eglinton, G. *et .al.* (1954) Researches on acetylenic compounds. Part XLV. The alkaline isomerization of but-3-ynoic acid. *J. Chem. Soc.*, 3197-3200.

Fondi, M. *et .al.* (2009) Origin and evolution of operons and metabolic pathways. *Res. Microbiol. 160*, 502-512.

Gerlt, J. A. *et .al.* (1998) Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol. 2,* 607-612.

Gerlt, J. A. *et .al.* (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem. 70*, 209-246.

Gerlt, J. A. *et .al.* (2005) Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Arch. Biochem. Biophys. 433*, 59-70.

Glasner, M. E. *et .al.* (2006) Evolution of enzyme superfamilies. *Curr. Opin. Chem. Biol. 10*, 492-497.

Guo, Y. *et .al.* (2011) Crystal structures of native and inactivated *cis*-3-chloroacrylic acid dehalogenase: implications for the catalytic and inactivation mechanisms. *Bioorg. Chem. 39*, 1-9.

Guo, Y. *et .al.* (2013) Kinetic, mutational, and structural analysis of malonate semialdehyde decarboxylase from *Coryneform bacterium* strain FG41: mechanistic implications for the decarboxylase and hydratase activities. *Biochemistry. 52,* 4830-4841.

Hartmans, S. *et .al.* (1991) Bacterial metabolism of 3-chloroacrylic acid. *J. Gen. Microbiol. 137,* 2025-2032.

Ho, M. C. *et .al.* (2009) The origin of the electrostatic perturbation in acetoacetate decarboxylase. *Nature. 459,* 393-397.

Horowitz, N. H. (1945) On the evolution of biochemical syntheses. *Proc. Natl. Acad. Sci. USA. 31,* 153-157.

Horvat, C. M. *et .al.* (2005) A persistent pesticide residue and the unusual catalytic proficiency of a dehalogenating enzyme. *Proc. Natl. Acad, Sci. U.S.A. 102*, 16199-16202.

Houmard, J. *et .al.* (1972) Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. *Proc. Natl. Acad, Sci. U.S.A. 69,* 3506-3509.

Huddleston, J. P. *et .al.* (2014) Identification and characterization of new family members in the tautomerase superfamily: Analysis and implications. *Arch. Biochem. Biophys. 564*, 189-196.

Huddleston, J. P. *et .al.* (2015) The accidental assignment of function in the tautomerase superfamily. *Perspectives in Science.* (in press).

Huddleston, J. P. *et .al.* (2012) A pre-steady state kinetic analysis of the αY60W mutant of *trans*-3-chloroacrylic acid dehalogenase: implications for the mechanism of the wild-type enzyme. *Biochemistry. 51*, 9420-9435.

Hüser, A. T. *et .al.* (2005) Rational design of a *Corynebacterium glutamicum* pantothenate production strain and its characterization by metabolic flux analysis and genome-wide transcriptional profiling. *Appl. Environ. Microbiol. 71,* 3255-3268.

Janssen, D. B. (2004) Evolving haloalkane dehalogenases. *Curr. Opin. Chem. Biol. 8*, 150-159.

Janssen, D. B. *et .al.* (2001) Microbial dehalogenation. *Curr. Opin. Biotechnol. 12*, 254-258.

Jensen, R. A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol. 30,* 409-425.

Johnson, K. A. (1983) The pathway of ATP hydrolysis by dynein. Kinetics of a presteady state phosphate burst. *J. Biol. Chem. 258*, 13825-13832.

Johnson, K. A. (1992) Transient-state kinetic analysis of enzyme reaction pathways In *The Enzymes* (Sigman, D. S., Ed.), 3rd Ed., pp 1-61, Academic Press, San Diego

Johnson, K. A. (2008) Role of induced fit in enzyme specificity: a molecular forward/reverse switch. *J. Biol. Chem. 283*, 26297-26301.

Johnson, K. A. (2009) Fitting enzyme kinetic data with KinTek global kinetic explorer In *Methods Enzymol.* (Michael, L. J., Ludwig, B., Eds.), pp 601-626, Academic Press,

Johnson, K. A. (2013) A century of enzyme kinetic analysis, 1913 to 2013. *FEBS Lett. 587*, 2753-2766.

Johnson, K. A. *et .al.* (2011) The original michaelis constant: translation of the 1913 Michaelis–Menten paper. *Biochemistry. 50*, 8264-8269.

Johnson, K. A. *et .al.* (2009) FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal. Biochem. 387*, 30-41.

Johnson, K. A. *et .al.* (2009) Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem. 387*, 20-29.

Kanehisa, M. *et .al.* (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res. 30*, 42-46.

Khersonsky, O. *et .al.* (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem. 79*, 471-505.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature. 227*, 680-685.

Lange, C. *et .al.* (2003) Global expression profiling and physiological characterization of *Corynebacterium glutamicum* grown in the presence of L-valine. *Appl. Environ. Microbiol. 69*, 2521-2532.

Li, L. *et .al.* (2008) Second-sphere amino acids contribute to transition-state structure in bovine purine nucleoside phosphorylase. *Biochemistry. 47*, 2577-2583.

Li, T. *et .al.* (2012) Decarboxylation mechanisms in biological system. *Bioorg. Chem. 43*, 2-14.

Liu, R.-Q. *et .al.* (1995) Transient state kinetic analysis of the chemical intermediates formed in the enzymic dehalogenation of 4-chlorobenzoyl coenzyme A. *J. Am. Chem. Soc. 117*, 5003-5004.

Lu, Z. *et .al.* (2011) The x-ray crystallographic structure and specificity profile of HAD superfamily phosphohydrolase BT1666: comparison of paralogous functions in *B. thetaiotaomicron*. *Proteins. 79*, 3099-3107.

Merkamm, M. *et .al.* (2003) Ketopantoate reductase activity is only encoded by *ilvC* in *Corynebacterium glutamicum*. *J. Biotechnol. 104*, 253-260.

Michaelis, L. *et .al.* (2011) The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry. 50*, 8264-8269.

Murzin, A. G. (1996) Structural classification of proteins: new superfamilies. *Curr. Opin. Struct. Biol. 6*, 386-394.

Neidhart, D. J. *et .al.* (1990) Mandelate racemase and muconate lactonizing enzyme are mechanistically distinct and structurally homologous. *Nature. 347*, 692-694.

Pegan, S. D. *et .al.* (2008) Structural and mechanistic analysis of *trans*-3-chloroacrylic acid dehalogenase activity. *Acta Crystallogr. Sect. D. Biol. Crystallogr. D64*, 1277-1282.

Peregrin-Alvarez, J. *et .al.* (2009) The conservation and evolutionary modularity of metabolism. *Genome Biol. 10*, R63.61-R63.17.

Petsko, G. A. *et .al.* (1993) On the origin of enzymatic species. *Trends Biochem. Sci. 18*, 372-376.

Poelarends, G. J. *et .al.* (2003) Mechanistic characterization of a bacterial malonate semialdehyde decarboxylase: identification of a new activity in the tautomerase superfamily. *J. Biol. Chem. 278*, 48674-48683.

Poelarends, G. J. *et .al.* (2007) Phenylpyruvate tautomerase activity of *trans*-3-chloroacrylic acid dehalogenase: evidence for an enol intermediate in the dehalogenase reaction. *Biochemistry. 46*, 9596-9604.

Poelarends, G. J. *et .al.* (2001) *trans*-3-Chloroacrylic acid dehalogenase from *Pseudomonas pavonaceae* 170 shares structural and mechanistic similarities with 4-oxalocrotonate tautomerase. *J. Bacteriol. 183*, 4269-4277.

Poelarends, G. J. *et .al.* (2013) A mutational analysis of active site residues in *trans*-3-chloroacrylic acid dehalogenase. *FEBS Lett. 587*, 2842-2850.

Poelarends, G. J. *et .al.* (2004) The hydratase activity of malonate semialdehyde decarboxylase: mechanistic and evolutionary implications. *J. Am. Chem. Soc. 126*, 15658-15659.

Poelarends, G. J. *et .al.* (2004) Stereospecific alkylation of *cis*-3-chloroacrylic acid dehalogenase by *(R)*-oxirane-2-carboxylate: analysis and mechanistic implications. *Biochemistry. 43*, 7187-7196.

Poelarends, G. J. *et .al.* (2004) Cloning, expression, and characterization of a *cis*-3-chloroacrylic acid dehalogenase: insights into the mechanistic, structural, and evolutionary relationship between isomer-specific 3-chloroacrylic acid dehalogenases. *Biochemistry. 43*, 759-772.

Poelarends, G. J. *et .al.* (2008) Characterization of Cg10062 from *Corynebacterium glutamicum*: implications for the evolution of *cis*-3-chloroacrylic acid dehalogenase activity in the tautomerase superfamily. *Biochemistry. 47*, 8139-8147.

Poelarends, G. J. *et .al.* (2008) The chemical versatility of the β–α–β fold: catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cell. Mol. Life Sci. 65*, 3606-3618.

Poelarends, G. J. *et .al.* (2004) Evolution of enzymic activity in the tautomerase superfamily: mechanistic and structural studies of the 1,3-dichloropropene catabolic enzymes. *Bioorg. Chem. 32*, 376-392.

Poelarends, G. J. *et .al.* (1998) Degradation of 1,3-dichloropropene by *Pseudomonas cichorii* 170. *Appl. Environ. Microbiol. 64*, 2931-2936.

Robertson, B. A. *et .al.* (2008) Inactivation of Cg10062, a *cis*-3-chloroacrylic acid dehalogenase homologue in *Corynebacterium glutamicum*, by *(R)*- and *(S)*-oxirane-2-carboxylate: analysis and implications. *Biochemistry. 47*, 8796-8803.

Robertson, B. A. *et .al.* (2009) Pre-steady-state kinetic analysis of *cis*-3-chloroacrylic acid dehalogenase: analysis and implications. *Biochemistry. 48*, 11737-11744.

Salque, M. *et .al.* (2013) Earliest evidence for cheese making in the sixth millennium bc in northern Europe. *Nature. 493*, 522-525.

Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.

Schanstra, J. P. *et .al.* (1996) Kinetics of halide release of haloalkane dehalogenase: evidence for a slow conformational change. *Biochemistry. 35*, 5624-5632.

Schanstra, J. P. *et .al.* (1996) Specificity and kinetics of haloalkane dehalogenase. *J. Biol. Chem. 271*, 14747-14753.

Schroeder, G. K. *et .al.* (2013) A mutational analysis of the active site loop residues in *cis*-3-chloroacrylic acid dehalogenase. *Biochemistry. 52*, 4204-4216.

Schroeder, G. K. *et .al.* (2012) Reaction of *cis*-3-chloroacrylic acid dehalogenase with an allene substrate, 2,3-butadienoate: hydration via an enamine. *J. Am. Chem. Soc. 134*, 293-304.

Schulenburg, C. *et .al.* (2014) Enzyme recruitment and its role in metabolic expansion. *Biochemistry. 53*, 836-845.

Segel, I. H. *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems*; Wiley Classics Library: New York, 1993.

Serrano, H. Characterization of the activites of *trans*-3-chloroacrylic acid dehalogenase and *cis*-3-chloroacrylic acid dehalogenase and malonate semialdehyde decarboxylase homologues: mechanism and evolutionary implications, University of Texas at Austin, 2009.

Sevastik, R. *et .al.* (2009) Reaction mechanism of *cis*-3-chloroacrylic acid dehalogenase: a theoretical study. *Biochemistry. 48*, 9641-9649.

Shanmugam, S. *Enzyme Technology*; I.K. International Publishing House Pvt. Limited: New Delhi, 2009.

Sørensen, S. B. *et .al.* (1991) Fragmentation of proteins by *S. aureus* strain V8 protease. Ammonium bicarbonate strongly inhibits the enzyme but does not improve the selectivity for glutamic acid. *FEBS Lett. 294*, 195-197.

Srivastava, D. *et .al.* (2010) The structure of the proline utilization a proline dehydrogenase domain inactivated by N-propargylglycine provides insight into conformational changes induced by substrate binding and flavin reduction. *Biochemistry. 49*, 560-569.

Trott, O. *et .al.* (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem. 31*, 455-461.

van Hylckama Vlieg, J. E. *et .al.* (1991) Bacterial degradation of 3-chloroacrylic acid and the characterization of *cis*- and *trans*-specific dehalogenases. *Biodegradation. 2*, 139-150.

Verschueren, K. H. G. *et .al.* (1993) Crystallographic and fluorescence studies of the interaction of haloalkane dehalogenase with halide ions. Studies with halide compounds reveal a halide binding site in the active site. *Biochemistry. 32*, 9031-9037.

Vogt, A. D. *et .al.* (2013) Conformational selection is a dominant mechanism of ligand binding. *Biochemistry. 52*, 5723-5729.

Waddell, W. J. (1956) A simple ultraviolet spectrophotometric method for the determination of protein. *J. Lab. Clin. Med. 48*, 311-314.

Wang, S. C. *et .al.* (2004) Reactions of 4-oxalocrotonate tautomerase and YwhB with 3-halopropiolates: analysis and implications. *Biochemistry. 43*, 748-758.

Wang, S. C. *et .al.* (2003) The 4-oxalocrotonate tautomerase- and YwhB-catalyzed hydration of 3*E*-haloacrylates: implications for the evolution of new enzymatic activities. *J. Am. Chem. Soc. 125*, 14282-14283.

Wang, S. C. *et .al.* (2003) Reactions of *trans*-3-chloroacrylic acid dehalogenase with acetylene substrates: consequences of and evidence for a hydration reaction. *Biochemistry. 42*, 8762-8773.

Whitman, C. P. (2002) The 4-oxalocrotonate tautomerase family of enzymes: how Nature makes new enzymes using a β−α−β structural motif. *Arch. Biochem. Biophys. 402*, 1-13.

Whittier, S. K. *et .al.* (2013) Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. *Science (New York, N.Y.). 341*, 899-903.

Wu, J. *et .al.* (2006) Contributions of long-range electrostatic interactions to 4-chlorobenzoyl-CoA dehalogenase catalysis: a combined theoretical and experimental study. *Biochemistry. 45*, 102-112.

Xu, D. *et .al.* (2004) QM/MM studies of the enzyme-catalyzed dechlorination of 4-chlorobenzoyl-CoA provide insight into reaction energetics. *J. Am. Chem. Soc. 126*, 13649-13658.

Yamada, E. W. *et .al.* (1959) Enzymatic utilization of acetylenic compounds: II. Acetylenemonocarboxylic acid hydrase. *J. Biol. Chem. 234*, 941-945.

Yang, J. *et .al.* (2005) Kinetic values for mechanism-based enzyme inhibition: assessing the bias introduced by the conventional experimental protocol. *Eur. J. Pharm. Sci. 26*, 334-340.

Zandvoort, E. *et .al.* (2011) Systematic screening for catalytic promiscuity in 4-oxalocrotonate tautomerase: enamine formation and aldolase activity. *ChemBioChem. 12*, 602-609.

Zandvoort, E. *et .al.* (2012) An unexpected promiscuous activity of 4-oxalocrotonate tautomerase: the *cis-trans* isomerisation of nitrostyrene. *ChemBioChem. 13*, 1869-1873.

Zhang, W. *et .al.* (2001) Histidine 90 function in 4-chlorobenzoyl-coenzyme A dehalogenase catalysis. *Biochemistry. 40*, 13474-13482.

Vita

Jamison Huddleston was born in Dallas, Texas. He was raised in Garland, Texas, where in graduated from South Garland High School in 2003. He attended the University of Texas in Austin, graduating with a B.S. in Biochemistry in December of 2007. He joined Dr. Christian Whitman's laboratory at the University of Texas in June of 2008 to begin his graduate work.

Contact via email: jphudd@hotmail.com

This dissertation was typed by the author