

STRUCTURAL ENZYMOLOGY
INVESTIGATING THE MECHANISM OF RICE BRANCHING ENZYME I,
RICE GRANULE BOUND STARCH SYNTHASE,
AND
CG10062, A MEMBER OF THE TAUTOMERASE SUPERFAMILY

By

Hadi Nayebi Gavvani

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry—Doctor of Philosophy

2020

ABSTRACT

STRUCTURAL ENZYMOLOGY INVESTIGATING THE MECHANISM OF RICE BRANCHING ENZYME I, RICE GRANULE BOUND STARCH SYNTHASE, AND CG10062, A MEMBER OF THE TAUTOMERASE SUPERFAMILY

By

Hadi Nayebi Gavvani

Enzymes are the factories of biological cells that drive most of the biochemical transformations necessary for living. These macromolecules evolved over billions of years to optimize their function to perform a vast range of chemical transformations. They represent some of the fastest and most selective catalysts known to us. In order to engineer enzymes and utilize their potential for new reactions, first we need to understand their native functions. Structural enzymology uses primarily x-ray crystallographic methods to observe enzymes at atomic scale.

Rice granule-bound starch synthase I (GBSSI) and rice branching enzyme I (BEI) are two key components of starch biosynthesis. Using structural enzymology and kinetic studies, an enhanced model of their biochemical function is generated. In the case of GBSSI, new structures with its native substrate were obtained and also for the first time its open conformer structure was solved. Developing kinetic experiments for rice BEI and investigating the structure of rBEI-maltododecaose complex revealed the binding site for the donor and acceptor chains.

Cg10062, a member of the tautomerase superfamily, transforms propiolate by hydration and decarboxylation. Intermediate-bound structures of various Cg10062 variants, identified many intermediates in the reaction. A mechanistic model is developed, and behavior of its mutants is clarified.

TABLE OF CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vi
CHAPTER ONE	1
Cg10062, a member of the tautomerase superfamily	1
1.1. Introduction.....	2
1.1.1. The tautomerase superfamily	4
1.1.2. Engineering tautomerase enzymes.....	16
1.1.3. Cg10062, a member of the TSF	18
1.2. Investigating Cg10062 hydratase/decarboxylase activity.....	25
1.2.1. X-ray crystal structures of native Cg10062	25
1.2.2. X-ray crystal structures of Cg10062-H28A soaked in propionate	32
1.2.3. X-ray crystal structures of Cg10062-R73A soaked in propionate.....	36
1.2.4. X-ray crystal structures of Cg10062-Y103F soaked in propionate	38
1.2.5. X-ray crystal structures of Cg10062-E114D soaked in propionate	38
1.2.6. Disordered C-terminus and the arginine residues in the active site.....	42
1.2.7. Comparison of ligand electron densities	44
1.2.8. Structures of native Cg10062 and its ligand-bound mutants	51
1.2.9. Catalytic cycle of Cg10062, hydration and decarboxylation.....	58
1.3. Materials and methods	61
1.3.1 Protein Crystallization	61
1.3.2 Ligand soaking experiments and solving the structures	68
REFERENCES.....	70
CHAPTER TWO	76
Investigating the Mechanism of Rice Branching Enzyme I and Structures of Rice Granule-Bound Starch Synthase I	76
2.1. Introduction.....	77
2.1.1. The structure of starch	78
2.1.2. Biosynthesis of starch	81
2.1.3. Application and relevance.....	85
2.1.4. Starch Synthases and GBSS.....	85
2.1.5. The Branching Enzymes	89
2.2. Structure of rice GBSSI enzyme.....	95
2.2.1. Structure of ADP-glucose bound Rice GBSSI (closed conformer).....	96
2.2.2. Structure of Rice GBSSI (open conformer).....	100
2.2.3. Structure of Rice GBSSI (open conformer) and ADP complex 1	102
2.2.4. Structure of Rice GBSSI (open conformer) and ADP complex 2	104
2.2.5. Structure of Rice GBSSI (open conformer) and UDP complex	106
2.3. Rice Branching Enzyme	108
2.3.1. Structure of rice branching enzyme and maltododecaose complex.....	108
2.3.2. Mutagenesis of rice branching enzyme.....	112

2.4. Materials and methods	124
2.4.1. Protein expression and purification	127
2.4.2. Crystallization of rice GBSSI	128
2.4.3. Protein expression and purification of rice branching enzyme.....	129
2.4.4. Chain length distribution assay for rice branching enzyme.....	130
REFERENCES.....	131

LIST OF TABLES

Table 1. 1 Major reaction types catalyzed by the members of the tautomerase superfamily.	6
Table 1. 2 Kinetic data for native and mutants of Cg10062 (data from Ms. Amaya Sirinimal in Dr. Karen Draths laboratory)	21
Table 1. 3 Steady-state kinetic parameters for the native Cg10062 and mutants. ⁴³	24
Table 1. 4 Reaction products ratio for native Cg10062 and mutants. ⁴³	24
Table 1. 5 Data collection and refinement statistics for Cg10062 Apo.	28
Table 1. 6 Data collection and refinement statistics for Cg10062-H28A soaked in propiolate. ..	34
Table 1. 7 Data collection and refinement statistics for Cg10062-R73A soaked in propiolate....	37
Table 1. 8 Data collection and refinement statistics for Cg10062-Y103F soaked in propiolate..	39
Table 1. 9 Data collection and refinement statistics for Cg10062-E114D soaked in propiolate..	41
Table 1. 10 Crystallization condition for native Cg10062 and mutants with successful data collection.....	62
Table 1. 11 Screening crystallization conditions for Cg10062 mutants.	63
Table 2. 1 Summary of available starch/glycogen synthase structures.	97
Table 2. 2 Data collection and refinement statistics for GBSSI-ADP-1-glucose complex.	99
Table 2. 3 Data collection and refinement statistics for GBSSI-ADP-1-glucose complex.	101
Table 2. 4 Data collection and refinement statistics for GBSSI-ADP-1-glucose complex.	103
Table 2. 5 Data collection and refinement statistics for GBSSI-ADP-1-glucose complex.	105
Table 2. 6 Data collection and refinement statistics for GBSSI-UDP complex.	107
Table 2. 7 Activities of rice branching enzyme and its mutants.....	118
Table 2. 8 General PCR cycle for mutagenesis	124
Table 2. 9 PCR primers for mutagenesis (Q5 style)	125
Table 2. 10 Crystallization conditions for GBSSI	129

LIST OF FIGURES

Figure 1. 1 Using Cg10062 to produce small organic molecules from methane and carbon dioxide.	3
Figure 1. 2 Sequence similarity network of the TSF superfamily ^{1,8} The 11395 sequences of the TSF superfamily are used to generate the network. These sequences are binned into 1323 nodes with > 50% pairwise sequence identity. Diamond-shaped nodes have one or more experimentally characterized proteins with a SwissProt annotation. square-shaped nodes have one or more structurally characterized nodes; triangular nodes have one or more proteins that are functionally and structurally characterized. Nodes containing the sequence of a founder protein are shown in bright yellow triangles.	5
Figure 1. 3 Three types of oligomers for members of the TSF. Side view (top row), top view (bottom row), monomer (yellow), isoform (blue) Pro-1 in the active site (green), (A) a homohexamer (PDB ID: 4X19) with 6 active sites. (B) a heterohexamer (PDB ID: 3MB2) with 3 active sites. (C) a trimer (PDB ID: 4LHP) with 3 active sites.	7
Figure 1. 4 Catalytic action of 4-OT. (A) 4-OT substrate in equilibrium with its two tautomers. (B) Proposed mechanism for 4-OT.	9
Figure 1. 5 Inactivated 4-OT by 2-oxo-3-pentynoate. Colors represent different chains in the hexamer.	10
Figure 1. 6 Modified mechanism for 4-OT from <i>Pseudomonas</i> sp.	10
Figure 1. 7 Catalytic mechanism of CaaD with trans-chloroacrylic acid as the substrate.	11
Figure 1. 8 Inactivated CaaD by 3-bromo-propionate.	12
Figure 1. 9 Proposed mechanism for cis-CaaD.	13
Figure 1. 10 The active site of CaaD (green) and cis-CaaD (magenta)	14
Figure 1. 11 Active site of MSAD, inactivated by 3-bromo-propionate.	15
Figure 1. 12 Aldolase reaction target for 4-OT directed evolution.	16
Figure 1. 13 non-Canonical analogs of proline. (2S,4R)-4-fluoroproline, (2S,4S)-4-fluoroproline, (2S)-3,4-dehydropyridine, and (4R)-1,3-thiazolidine-4-carboxylic acid.	17
Figure 1. 14 Target reaction for 4-OT enzymes with non-canonical proline.	17
Figure 1. 15 Cg10062 and cis-CaaD have identical active site residues. Cg10062 (yellow), cis-CaaD (magenta), CaaD (green), MSAD (purple)	19

Figure 1. 16 Hydration and decarboxylation of propiolate by Cg10062.21	Figure 1. 16 Hydration and decarboxylation of propiolate by Cg10062.....	21
Figure 1. 17 Hydration and decarboxylation substrates for Cg10062		21
Figure 1. 18 Proposed mechanism for Cg10062 hydration/decarboxylation reaction via Pro-1 as general acid catalyst.....		22
Figure 1. 19 Proposed mechanism for Cg10062 decarboxylation reaction via Schiff base intermediate.....		23
Figure 1. 20 Proposed mechanism for Cg10062 hydration/decarboxylation reaction via covalent intermediate.....		23
Figure 1. 21 Crystals of native Cg10062		26
Figure 1. 22 Structure of Cg10062 apo. (A) top view and (B) side view of the Cg10062 apo structure, single chain in the ASU (yellow), symmetry mates to complete the trimer (pink), active site Pro-1 (green). (C) A snapshot of Pro-1 fitted in the electron density. (contoured at rmsd 1Å) Tyr-103, His-28, and Glu-114 are present in this view. (D) overlay of 3N4G structure (green) to our native apo Cg10062 (pink).....		27
Figure 1. 23 Amino acid residues in the active site of the Cg10062 and their interactions. (A) Important amino acids in the active site, Tyr-103 from adjacent monomer (pink), (B) hydrogen bonding network between Pro-1 (green), Glu-114, Thr-2, and Gln-40 (yellow), Tyr-103 of adjacent monomer (pink), and a water molecule (red). (C) hydrogen bonding network between Pro-1 (green), Tyr-3, His-28, Arg-70, and Arg-73 (yellow), and two water molecule (red)		29
Figure 1. 24 Comparison of active site residues between Cg10062, cis-CaaD. Cg10062 (pink), cis-CaaD (green).....		30
Figure 1. 25 Plausible mechanism for the reaction of propiolate and Cg10062 via covalent intermediates.		31
Figure 1. 26 Crystal packing of Cg10062-H28A soaked in propiolate. (left) a one protein chain, b trimer, protein crystal is formed by the helical assembly of the trimer units in three dimensions. c overview of the protein crystal packing when only one chain is being searched during the molecular replacement. (right) a two protein chains, b a hexamer, protein crystal is formed by the assembly of the hexamer units in three dimensions. c overview of the protein crystal packing when two chains are being searched during the molecular replacement.		33
Figure 1. 27 X-ray crystal structure of Cg10062-H28A soaked in propiolate. Six chains of Cg10062-H28A, asymmetric unit (yellow), symmetry mates (pink), active site Pro-1 (green). (A) top view, (B) side view, (C) Pro-1 fitted in the electron density, the negative (red) density for H28, and positive (green) density for a ligand, (contoured at rmsd 1.5 Å) (D) active site of the H28A mutant highlighted by important residues.....		35

Figure 1. 28 X-ray crystal structure of Cg10062-R73A soaked in propiolate.(A) Trimer of Cg10062-R73A, single chain (yellow) is present in the ASU, symmetry mates (pink). (B) a positive electron density (green) connected to the electron density of Pro-1 (blue). (contoured at rmsd 1.5 Å)	36
Figure 1. 29 Asymmetric unit of Cg10062-E114D mutant crystal with twelve chains in four trimers.	40
Figure 1. 30 Different shape of electron density for twelve chains in E114D ASU. (A) – (L) are twelve chains of Cg10062-E114D that make up the ASU. 2FoFc (blue), FoFc (green). (contoured at rmsd 1.5 Å)	42
Figure 1. 31 Difference in C-termini of native and mutants. (A) Native (green), E114D (yellow), H28A, Y103F, R73A (shades of red). The C-terminus of native Cg10062 follows the trajectory of H28A, Y103F, and R73A. E114D deviates from the native structure. (B) Arg-117 has two conformations in all E114D structures.	43
Figure 1. 32 Arginine residues in the active site. native Cg10062 (yellow), E114D (red), R73A (green), H28A (light blue), Y103F (purple).	44
Figure 1. 33 Comparing the shape of the difference electron density map ($F_{obs}-F_{calc}$). (A) Y103F, (B) E114D, (C) H28A, and (D) R73A. (contoured at rmsd 1.5 Å)	45
Figure 1. 34 Covalent intermediates. 3-prolylacrylate 1, iminium tautomer of 3-prolylacrylate 2, 3-hydroxy-3-prolylpropionate 3.....	46
Figure 1. 35 Building 3-hydroxy-3-prolylpropionate in the active site of H28A mutants. (contoured at rmsd 1.5 Å)	47
Figure 1. 36 Building 3-hydroxy-3-prolylpropionate in the active site of R73A mutants. (contoured at rmsd 1.5 Å)	47
Figure 1. 37 trans-3-Prolylacrylate in the active site of Y103F mutants in the form of enamine and iminium. (contoured at rmsd 1.5 Å), $2F_{obs}-F_{calc}$ (blue, pink, purple). $F_{obs}-F_{calc}$ (green)... ..	48
Figure 1. 38 trans-3-prolylacrylate in the active site of E114D mutants and the appearance of new positive electron density. (contoured at rmsd 1.5 Å), $2F_{obs}-F_{calc}$ (blue). $F_{obs}-F_{calc}$ (green)	49
Figure 1. 39 3-prolylacrylate and malonate semialdehyde are both present in the active site of E114D mutants. (contoured at rmsd 1.5 Å), $2F_{obs}-F_{calc}$ (blue). $F_{obs}-F_{calc}$ (green)	50
Figure 1. 40 Ordered water molecules in the active site of native Cg10062.	51
Figure 1. 41 Overlay the active site residues for H28A and native Cg10062. The hydroxyl group of 3-hydroxy-3-prolylpropionate in the structure of H28A and W-1 in native Cg10062 are nearly overlapping. Native Cg10062 (yellow), H28A mutant (pink).	52

Figure 1. 42 Overlay of active site residues for R73A and native Cg10062. The hydroxyl group of 3-hydroxy-3-prolylpropionate in the structure of R73A and W1 water molecule in native Cg10062 structure nearly overlap. Native Cg10062 (yellow), R73A mutant (pink).....	53
Figure 1. 43 The difference between the active sites of native Cg10062, H28A and R73A mutants.	54
Figure 1. 44 3-Prolylacrylate in the active site of Y103F mutants. (A) enamine form overlaid on the native Cg10062, ordered water molecules in the native Cg10062 active site (red spheres), (B) iminium form overlaid on the native Cg10062, ordered water molecule in the Y103F active site (red sphere), ordered water molecules in the native Cg10062 active site (red crosses) (C) geometry of enamine form, (D) geometry of iminium form (dihedral angle is 2.6°), (E) enamine and iminium form of 3-prolylacrylate.	55
Figure 1. 45 The active site of E114D mutants overlaid on the structure of native Cg10062. The ordered water molecule in E114D active site (red sphere), ordered water molecules in native Cg10062 active site (red crosses).....	56
Figure 1. 46 The ordered water molecule in E114D active site. Ordered water molecule (W4) from native Cg10062 (red cross). E114D (magenta), Native Cg10062 (yellow).....	57
Figure 1. 47 Malonate semialdehyde in the active site of the E114D mutant.	58
Figure 1. 48 The proposed catalytic cycle of Cg10062. (1) substrate binding, (2) tautomerization of enamine/iminium, (3) hydration via activated water, (4) prolyl dissociation and product release (MSA), (5) decarboxylation of iminium intermediate, (6) hydration of decarboxylated intermediate, (7) prolyl dissociation and product release (acetaldehyde), (8) Variation of hydration in E114D mutants.	59
Figure 1. 49 Flat and multi-lattice crystals grown in 3% PEG 6000, 10mM Tris-SO ₄ , pH 4.0...	61
Figure 2. 1 Schematic diagram of starch granule structure. ^{16,17} (A) Whole granule of starch, consisting of alternating semi-crystalline and amorphous growth rings, (B) a stack of large and small blocklets, (C) crystalline and amorphous lamellae in a blocklet, (D) ordered double helices within crystalline lamellae and amylopectin branch points within an amorphous lamella. (modified, see references)	79
Figure 2. 2 Chemical structures of linear and branched glucose chains. (A) Chain of glucoses connected via α -1,4-glycosidic bond, (B) Branched glucose chain via α -1,6-glycosidic bond, (C) Helical conformation of single glucose chain, (D) Double helix formed by two adjacent glucose chains.....	80
Figure 2. 3 Biosynthesis of starch (or glycogen).	82
Figure 2. 4 Organization of ADPGPPase, SS, SBE, and DBE enzymes in plants. Stars represent all polyploidy events. Whole genome duplication, WGD (red), Whole genome triplication, WGT (green), Whole genome sextuplication, WGS (yellow). Total number of isoforms of the four-core enzyme families ³⁹	84

Figure 2. 5 Structural features in SS, GS, and GBSS. (A) Overlay of OsGBSSI, EcGS, RrGS, PaGS, HvSSI, CyGBSS, CpGBSSI, and AtSSIV. (B) rice GBSSI, two Rossmann folds (C-terminus in pink and N-terminus in green), C-terminus α -helix. (C) Conserved KXGGL active site sequence, UniProt classification: GLG and GYS for glycogen synthases, SSY for soluble starch synthases, and SSG for granule bound starch synthases.....	86
Figure 2. 6 Relative location of KXGGL motif in the active site of EcGS open and closed conformers (left) and surface oligosaccharide binding sites (right)	87
Figure 2. 7 An enzyme coupled assay for biochemical study.....	88
Figure 2. 8 Architecture of branching enzymes in GH13 family. (A) central (β/α) ₈ -barrel catalytic domain, (B) four structural motifs of branching enzymes, N-terminus β -sandwich domain common to bacterial glycogen branching enzymes I, CBM48 II, amylase catalytic domain III, C-terminus β -sandwich domain IV. (PDBID: 5gqy, 4lpc, 5clw, 3vu2 for cyanobacteria, E. coli, human, and rice respectively).....	90
Figure 2. 9 Surface binding sites in branching enzymes. Three different views of a branching enzyme and distribution of surface binding sites. EcBE (green surface), oligosaccharides (EcBE: yellow, CyBE: magenta, HsBE: blue).....	92
Figure 2. 10 Limitations of branching enzymes. (A) less than six glucose units to non-reducing end of donor chain, no hydrolysis, (B) at least six glucose units to non-reducing end of donor chain, successful hydrolysis, (C) less than six glucose units to the branch point on a branched donor chain. No hydrolysis. (D) at least six glucose units to the branch point on a branched donor chain, successful hydrolysis. (E) at least six glucose units to the branching point of an acceptor chain, successful transfer, (F) less than six glucose units to the branch point of an acceptor chain.	93
Figure 2. 11 A hierarchical classification of chain length distribution for 22 branching enzymes. Bacterial (aae ⁸⁸ , bacI ⁸⁹ , dge ⁹⁰ , dra ⁹⁰ , rmg ⁹¹ , smu ⁹² , syc ⁹³ , vvm ⁹⁴ , ehl ⁹⁵ , gse ⁹⁶ , cyt1 ⁹³ , cyt2 ⁹³ , cyt3 ⁹³ , ebd ⁹⁷), Plants (osa1 ⁸⁶ , pvu1 ⁹⁸ , pvu2 ⁹⁸ , zma1 ⁹⁹ , zma2 ⁹⁹ , zma3 ⁹⁹ , osa2 ⁸⁶ , dosa ⁸⁶). Left: Similarity of the chain length distribution for 22 branching enzyme CLD profiles (0, least similar. 1, identical). Right: five distinct clusters of CLD profiles obtained by hierarchical clustering method.....	94
Figure 2. 12 Sequence analysis of four rice branching enzyme isoforms. Conserved regions of I, II, III, IV common to all branching enzymes. CBM48 domain (blue),	95
Figure 2. 13 ADP-glucose bound Rice GBSSI, (A) ADP-glucose electron density. (B) hydrogen bond interactions between the enzyme and glucose unit. (C) hydrogen bond interactions between the enzyme and diphosphate. (D) hydrogen bond interactions between the enzyme and adenine unit.	98
Figure 2. 14 Overlay of open and closed conformers of rice GBSSI.	100
Figure 2. 15 ADP and GBSSI (open conformer) complex overlaid on closed form GBSSI and ADP-1-glucose complex. Closed form (green), open form (magenta).....	102

Figure 2. 16 Different binding mode for the ADP in the active site of GBSSI (open conformer). New binding mode (green)	104
Figure 2. 17 Binding mode for ADP and UDP are similar. ADP (green), UDP (purple).	106
Figure 2. 18 Surface binding sites in rice branching enzyme. (A) Cartoon representation of rBEI and three bound glucans. N-terminal domain (Gray), CBM48 domain (Slate Blue), Catalytic Core domain (Green), C-terminal domain (Yellow) and Glucans (Red and Magenta) depicted as space filling models. (B) Surface representation of rBEI (green) with bound glucans depicted as space filling models (C atoms, yellow, O atoms, red). Top, oriented as in A, bottom, rotated approximately 90° along the horizontal axis.	108
Figure 2. 19 Residues of rBEI interacting in site 1. (A) and site 3 (B). Glucans (Yellow), Residues (Green)	109
Figure 2. 20 Helical features of bound glucans to rBEI., rBEI bound glucans (Yellow), Reference glucans (Green) (A) Top, M12 from site 4 overlaid on one strand of a glycogen or amylopectin-like double helix. Bottom, the original double helix. (B) Top, M6 from site 1 overlaid on a model of an amylose single helix. Bottom, the amylose single helix model	110
Figure 2. 21 Detailed interactions between rBEI and M12 occupying site 4.	110
Figure 2. 22 Sequence alignment for some plant BEIs, BEII's, human, drosophila, yeast, cyanobacteria and Escherichia coli in site 4.	111
Figure 2. 23 Disordered loop (residues 468 and 474) adopts a new conformation upon site 4 M12 binding. Apo rBEI-chain A and B-3AMK (Magenta), M5 bound rBEI-chain A and B- 3VU2 (Slate Blue) and M12 bound rBEI (Green)	112
Figure 2. 24 Mutation map. Active site (red), region 1 (blue), CBM domain (green), region 2 (magenta)	113
Figure 2. 25 Relative activity of the wildtype enzyme and active site mutants.....	113
Figure 2. 26 Relative activity of the wildtype enzyme and region 1 mutants.	114
Figure 2. 27 Relative activity of the wildtype enzyme and CBM domain mutants.....	114
Figure 2. 28 Relative activity of the wildtype enzyme and region 2 mutants.	115
Figure 2. 29 Connecting the active site to the CBM48 domain.....	115
Figure 2. 30 Chain length distribution changes over time.	116
Figure 2. 31 Isoform defining loops of BEIs and BEIIs. (A) Sequence Alignment of the two loops that distinguish BEIs and BEIIs. (B) Location of the loops on M12-bound rBEI. rBEI	

(Green), Loop 143 and 541 (Magenta), M12 (stick model, C, yellow, all other atoms colored as above).	119
Figure 2. 32 Transfer Chain Specificity Assay. Fraction differences of transferred chains by wild type rBEI in 2 hrs. vs 1 min. (top panel), Fraction differences of transferred chains by wild type rBEI vs Y487A and D483A (middle panel), Fraction differences of transferred chains by wild type rBEI vs rBEII loop replacements (bottom panel).	121
Figure 2. 33 Overlay of M7 bound in Cyanothecce BE and M12-bound rBEI, rBEI (Green), glucans bound to rBEI (C, yellow, all other atoms as above), M7 bound to Cyanothecce BE (C, Pink), active site (Blue), Loop 143 (Orange), Loop 541 (Magenta).....	122
Figure 2. 34 SDS-page gel image for nickel column His-tag affinity purification of GBSSI (59 kDa). Lanes 1 through 13 are: insoluble fraction, run through, wash 1-3 (all washes are 20 mM imidazole in PBS), elution 1-8 (200 mM imidazole in PBS).....	127
Figure 2. 35 Size exclusion chromatograph for GBSSI.....	128
Figure 2. 36 Crystals of GBSSI. (A) before optimization, (B) after optimization	129

CHAPTER ONE

Cg10062, a member of the tautomerase superfamily

This chapter aims to comprehend the mechanism of the Cg10062 enzyme, a member of the tautomerase superfamily, using protein x-ray crystallographic methods. The first section provides a brief introduction to the tautomerase family and several well-studied proteins in this family to address the current understanding of the reaction mechanism controlled by these enzymes. Also, in this section, the Cg10062 enzyme will be introduced. The second section presents the experimental results obtained in this study. In this section, we try to address the questions related to the mechanism of Cg10062 by using the experimental results. Finally, the last section explains experimental methods and procedures for crystallization, data collection, and solving protein structures.

1.1. Introduction

Structural enzymology¹ utilizes the information gained from a protein crystal structure to elaborate on the catalytic activity of the enzyme. Generally, the topology of the active site, considering the chemistry of the reaction, allows for the construction of a model explaining the mechanism of the reaction. If the model could not explain the other experimental data such as activity and specificity, then it will be modified. Besides, by soaking the crystals of an enzyme in the solution of the substrate(s) or inhibitor(s) before x-ray data collection, intermediate steps are more likely to be observed as a result of the limited protein breathing. Studies show that protein breathing, a low-frequency conformational motion, is influential in the catalytic activity of the enzymes.² Since proteins packed in a crystal lattice have limited space and limited breathing activity, the intermediate steps of an enzymatic cycle might slow down, allowing the formation of intermediate species. Including the structure of mutants of an enzyme, preferably with a known biochemical outcome, is a powerful way to generate model-based predictions, test, and refine the

proposed mechanism.^{1,3-6} Cg10062 enzyme is a member of tautomerase superfamily and studies show that it catalyzes the hydration and decarboxylation of propiolate.⁷ This enzyme does not need any metal or cofactors to complete its catalytic cycle. Propiolate is the product of dehydrodimerization of CH₄ followed by carboxylation of the acetylene. Therefore, propiolate might be an intermediate for the production of small organic molecules from methane and carbon dioxide. Cg10062 enzyme catalyzes the hydration of the propiolate to form malonate semialdehyde. But the decarboxylation step removes a carbon and two oxygen atoms, which makes Cg10062 less useful. Understanding the mechanism of hydration and decarboxylation of Cg10062 will help us to engineer a variant of the enzyme that decarboxylation is prevented. Such an enzyme could be used in the process of producing small organic molecule with methane and carbon dioxide as the initial source (Figure 1.1).

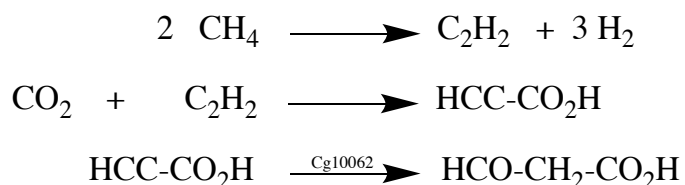


Figure 1. 1 Using Cg10062 to produce small organic molecules from methane and carbon dioxide.

1.1.1. The tautomerase superfamily

The tautomerase superfamily (TSF) is a group of structurally homologous proteins that share a β - α - β fold and a catalytic N-terminal proline residue. This superfamily has more than 11,000 non-redundant sequences from different domains of life,⁸ and they carry out a diverse range of enzymatic reactions. The complete range of these functions remains undiscovered. The β - α - β building block starts with a proline (Pro-1) and forms a β -strand followed by an α -helix and a 3_{10} helix, which continues with a second parallel β -strand⁹. There is a β -hairpin at the C-terminus of the protein. The β -hairpin is involved in the formation of the hexamer (or trimer).

From a primary sequence point of view (Figure 1.2), these enzymes are present in two sizes, proteins composed of one β - α - β fold and another group composed of two β - α - β folds. The proteins with one β - α - β unit, are active in the form of a hexamer (homo-hexamer or hetero-hexamer). And the proteins with two β - α - β building blocks form a trimer to catalyze the reaction. The N-terminus of one monomer combines with the C-terminus of an adjacent monomer to form the active site. For example, in a hexamer or a trimer, a conserved tyrosine residue from the C-terminus of an enzyme completes the active site of an adjacent monomer. This arrangement creates six active sites in the case of a hexamer and three active sites for a trimer form.⁹

Pro-1 in the tautomerase enzyme superfamily, can function as a general base or general acid catalyst. The catalytic role of Pro-1 depends on the pK_a value of the proline in the active site of the protein.¹⁰ Functionally, the tautomerase superfamily divides into five families. The families catalyze a similar reaction type and share the substrate. (see Table 1.1). A group of enzymes acts as a 4-Oxalocrotonate tautomerase (4-OT). 4-OT processes an enol-keto tautomerization of a pyruvoyl group, and an acidic Pro-1 residue catalyzes it. Generally, Pro-1 of these enzymes have low pK_a values.¹¹ The second family is 5-(carboxymethyl)-2-hydroxymuconate isomerase

(CHMI).¹² Macrophage migration inhibitory factor (MIF)^{13,14} has phenylpyruvate tautomerase activity.⁸ *cis*- and *trans*-3-Chloroacrylic acid dehalogenase (*cis*- and *trans*-CaaD)¹⁵ and malonate semialdehyde decarboxylase (MASD)¹⁶ perform mechanistically different reactions using their Pro-1. The pKa value of Pro-1 is higher in these families.

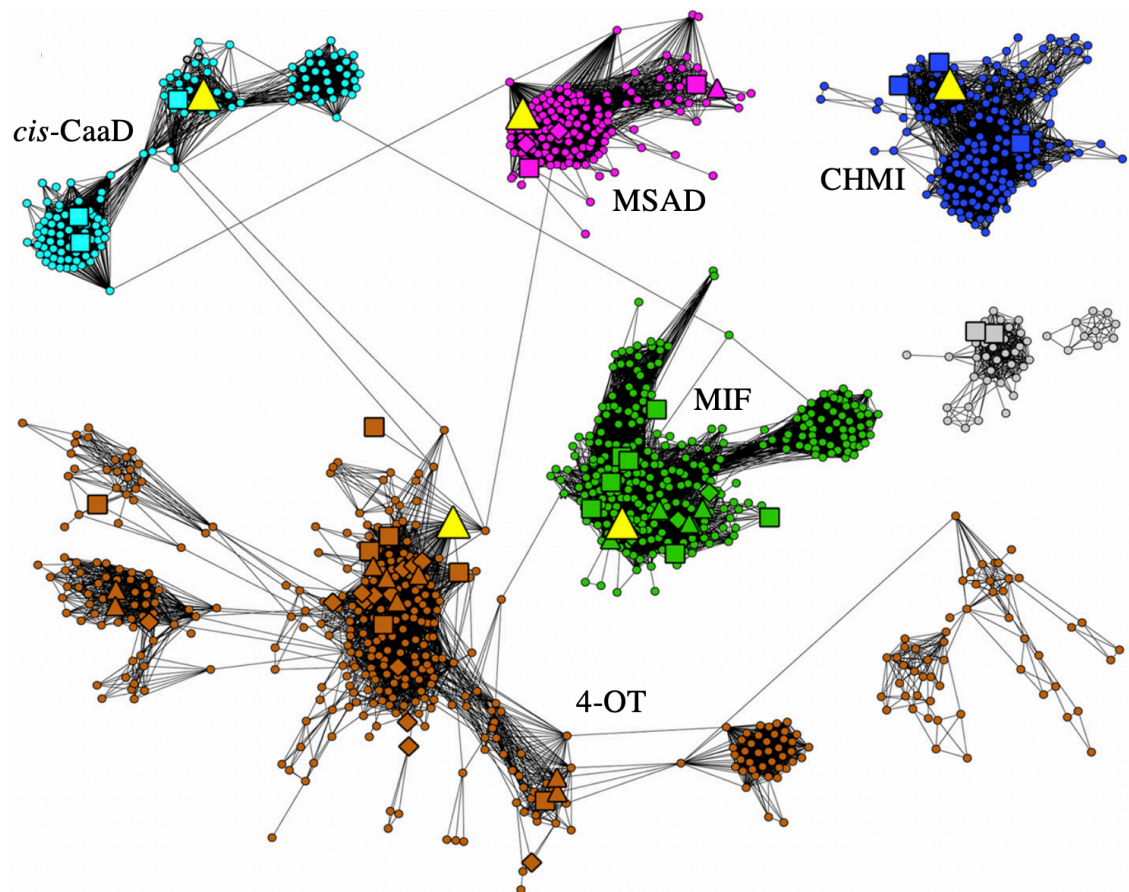
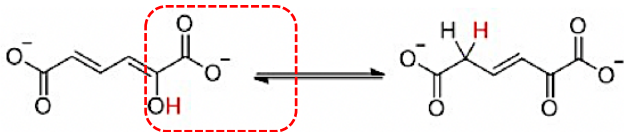
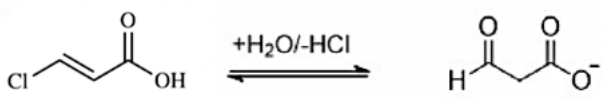
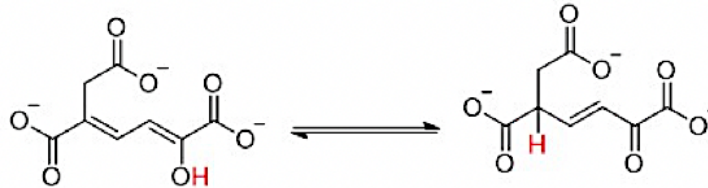
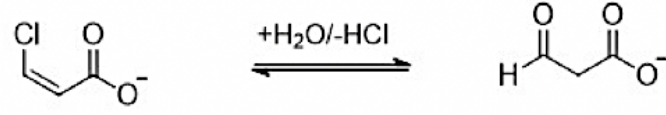


Figure 1. 2 Sequence similarity network of the TSF superfamily^{1.8} The 11395 sequences of the TSF superfamily are used to generate the network. These sequences are binned into 1323 nodes with > 50% pairwise sequence identity. Diamond-shaped nodes have one or more experimentally characterized proteins with a SwissProt annotation. square-shaped nodes have one or more structurally characterized nodes; triangular nodes have one or more proteins that are functionally and structurally characterized. Nodes containing the sequence of a founder protein are shown in bright yellow triangles.

Table 1. 1 Major reaction types catalyzed by the members of the tautomerase superfamily.

Family	Enzyme	Typical length	monomer	Oligomeric state	Reaction
4-OT *	4-OT	62		hexamer	
	CaaD	α -subunit 75 β -subunit 70		Heterohexamer	
CHMI *	CHMI	125		trimer	
MIF *	MIF	114		trimer	
cis-CaaD	cis-CaaD	149		Trimer	
MSAD	MSAD	129		trimer	

* The pyruvoyl moiety (in the red box), the common functional group for 4-OT, CHMI, and MIF.

From an evolutionary point of view, studies^{10,17} show that a gene duplication event gave rise to the generation of an isoform of the native enzyme, and subsequently, the formation of active enzymes in the form of heterohexamers. Over time one of the isoforms loses the Pro-1, the reactive residue, and results in losing half of the active sites in the heterohexamers. Later, a gene fusion event starts a new enzyme that is twice the size of the ancestor protein.⁸

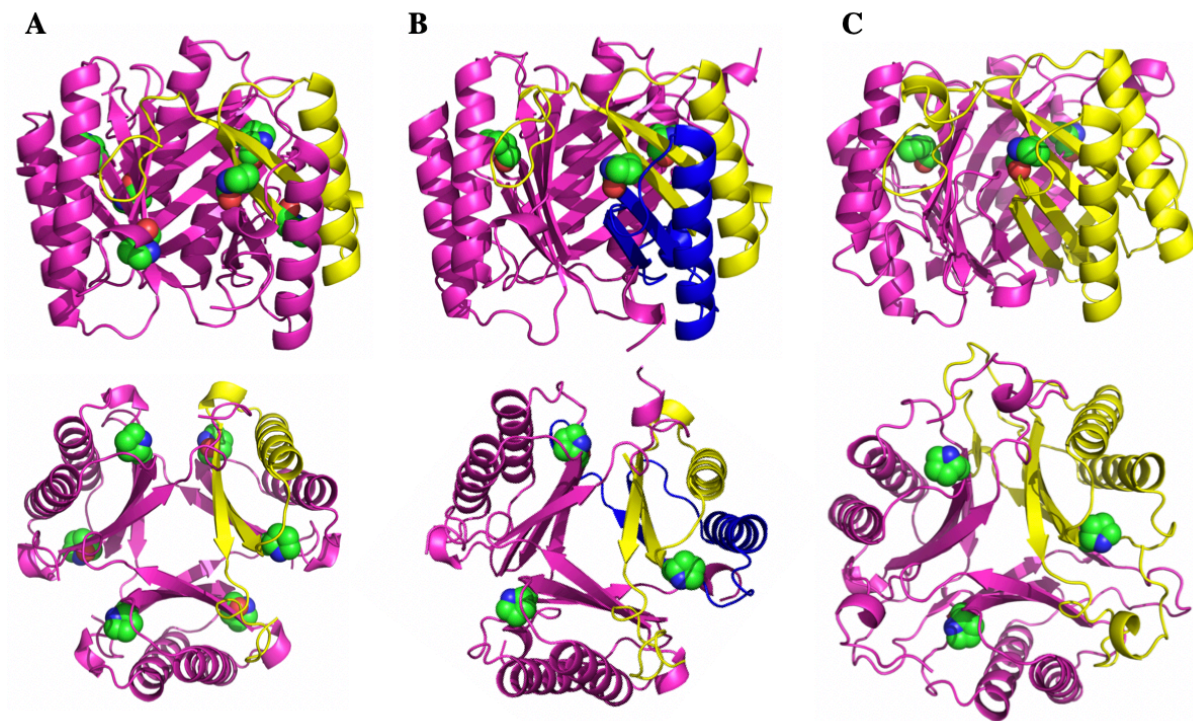


Figure 1. 3 Three types of oligomers for members of the TSF. Side view (top row), top view (bottom row), monomer (yellow), isoform (blue) Pro-1 in the active site (green), (A) a homohexamer (PDB ID: 4X19) with 6 active sites. (B) a heterohexamer (PDB ID: 3MB2) with 3 active sites. (C) a trimer (PDB ID: 4LHP) with 3 active sites.

Gene duplication, followed by a gene fusion event, resulted in four types of protein oligomers.⁸ A homohexamer has six active sites, while heterohexamers might have six or three active sites. The number of active sites for a heterohexamer depends on the extent of the evolution of the isoform. And, Trimers have three active sites. There are solved structures of homohexamer,

heterohexamers, and trimers, for more than fifty species. These structures are deposited in the Protein Data Bank.¹⁸ (see Figure 1.3)

Generally, a protein of this superfamily forms homo/hetero-hexamers if it contains only one β - α - β fold (4-OT).¹⁹ Homotrimers, on the other hand, consist of two β - α - β folds (CHMI, MIF, cis-CaaD, MSAD).^{10,19-21} TSF subfamilies are involved in the degradation of aromatic hydrocarbons (4-OT)²² and the degradation of aromatic amino acids (CHMI).^{12,23} MIF family processes phenylpyruvate tautomerization (MIF).^{14,24} The classification focuses on the primary reaction type catalyzed by the enzyme; however, the members of one family show relatively lower activities on the substrates of other families.²⁵

1.1.1.1. 4-Oxalocrotonate Tautomerase

4-Oxalocrotonate tautomerase (4-OT) is a member of the tautomerase superfamily that is made up of one β - α - β building block and is active in the form of a hexamer. 4-OT enzymes are involved in the degradation of aromatic hydrocarbons and enable the organism to use aromatic hydrocarbon as a source of carbon and energy.²⁶ Investigation of the active site for 4-OT enzymes reveals a hydrophobic characteristic, and as a result, the 1-Pro residue in 4-OT has a lower pK_a (~ 6.4). Pro-1 in 4-OT is unprotonated at cellular pH and functions as a general base.

Studies show²² that 2-hydroxymuconate **1** is the substrate for the 4-OT enzyme (from *Pseudomonas* sp.), and while the substrate is stable in crystalline form but it quickly tautomerase to 2-oxo-3-hexenedionate **2** in solution.²⁶ There exists another molecule, 2-oxo-4-hexenedionate **3**, that is in equilibrium with 2-hydroxymuconate (Figure 1.4). Based on NMR and crystallographic observation, done by the Whitman research group, a catalytic mechanism is proposed for the action of 4-OT.²⁶

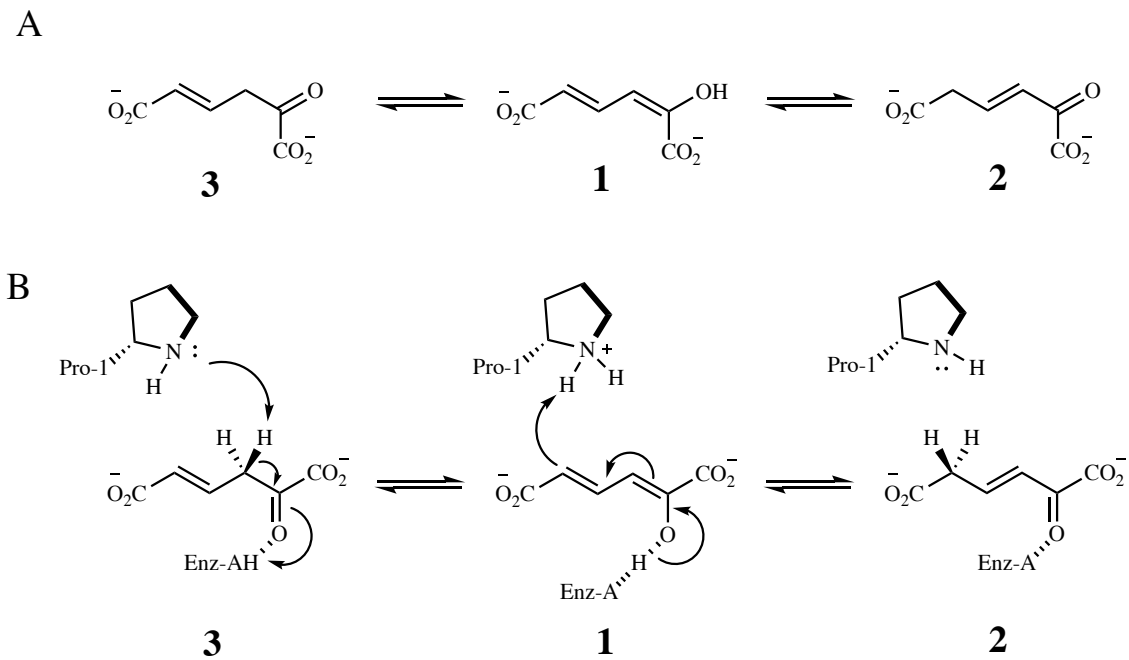


Figure 1. 4 Catalytic action of 4-OT. (A) 4-OT substrate in equilibrium with its two tautomers. (B) Proposed mechanism for 4-OT.

They identified Arg-39 as a possible candidate for general acid function and at the same time interacting with the C-1 carboxylate for binding. Arg-11 was another residue responsible for binding via the C-6 carboxylate group. The existence of two positively charged amino acids in proximity of Pro-1 and general hydrophobic characteristic of the active site pocket could explain the low pK_a value for the proline.²⁶ Later a ligand-bound structure of a 4-OT isozyme (*P. putida* mt-2) that shares a 73% sequence identity with the *Pseudomonas* sp. enzyme, was obtained. 2-oxo-3-pentynoate is an active site irreversible inhibitor for this enzyme and Pro-1 forms a covalent bond to C4 of 2-oxo-3-pentynoate.²⁷ (Figure 1.5)

Including new information and mutational studies, they proposed a modified mechanism. Arg-61 mutation to alanine did not affect the pK_a of Pro-1 suggesting that its role is not significant for the mechanism of the enzyme. Arg-11 was shown to be essential for binding and catalysis, while

the alanine mutation did not suggest a significant role in lowering the pK_a value of the Pro-1. Arg-39 has a significant role in catalysis and structural stability of the enzyme.²⁶ (Figure 1.6)

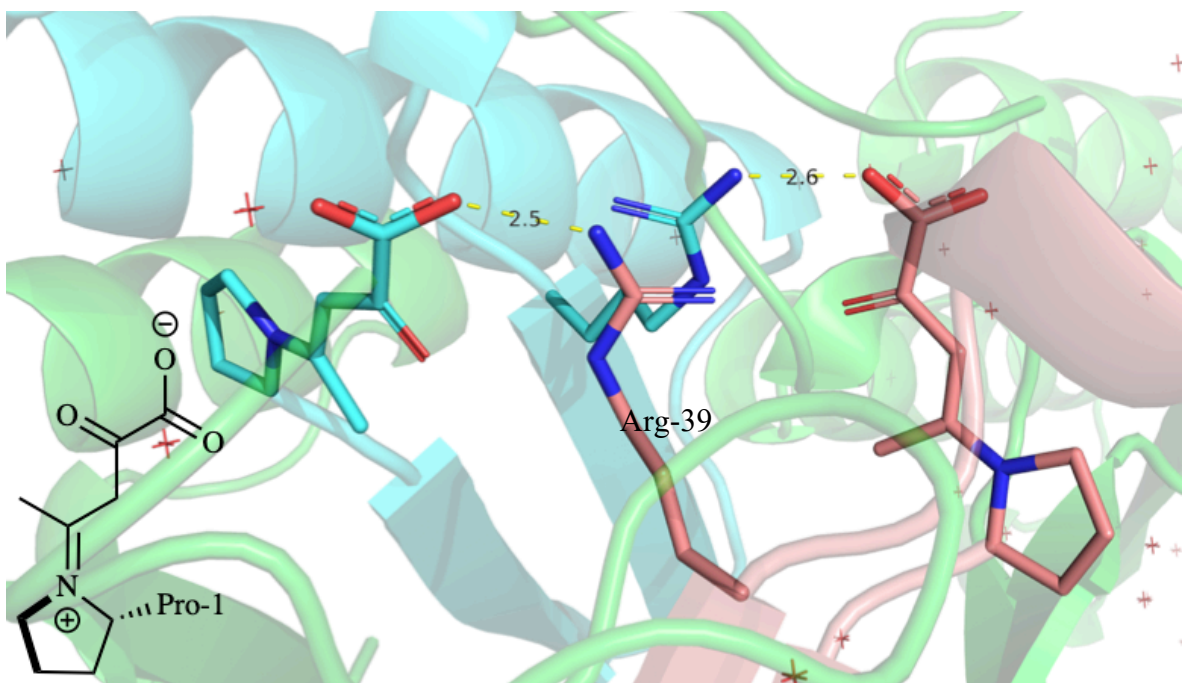


Figure 1. 5 Inactivated 4-OT by 2-oxo-3-pentynoate. Colors represent different chains in the hexamer.

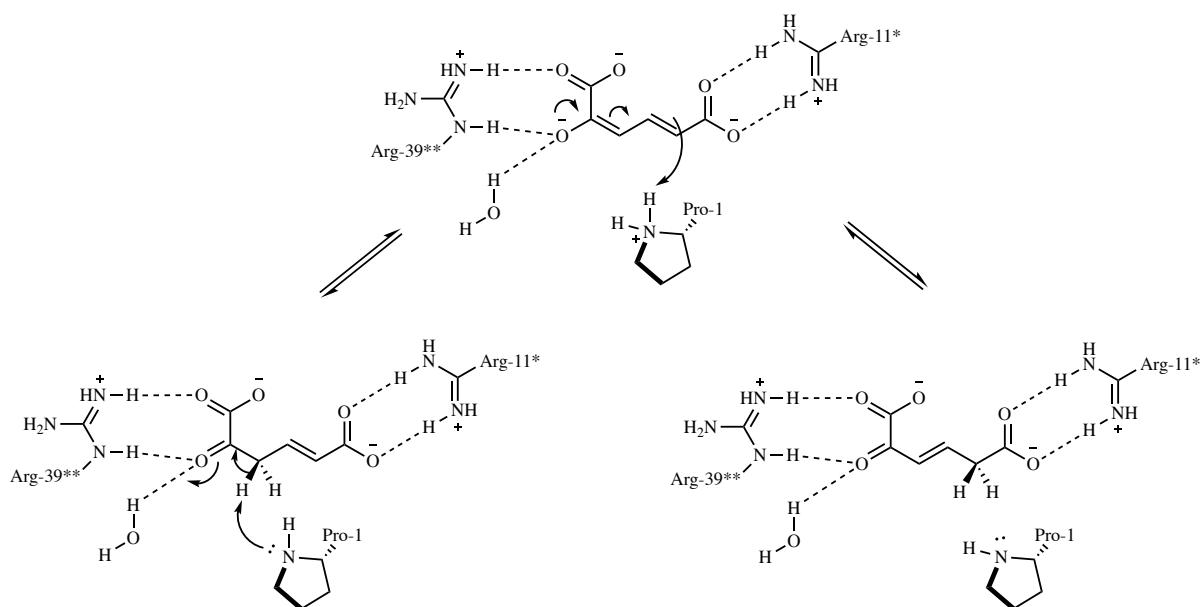


Figure 1. 6 Modified mechanism for 4-OT from *Pseudomonas* sp.

1.1.1.2. trans-3-Chloroacrylic acid dehalogenase

trans-3-Chloroacrylic acid dehalogenase (CaaD) and cis-3-Chloroacrylic acid dehalogenase (cis-CaaD) are both members of the tautomerase superfamily. CaaD has one β - α - β building block and is more homologous to the 4-OT family.^{28,29} However, CaaD forms a heterohexamer and only has three active sites suggesting that the CaaD family evolved from an ancestor 4-OT. After the gene duplication event, an isoform was evolved, and losing Pro-1 residue became possible. The benefit of this event is that the newly formed isoform is free to evolve, which helps the 4-OT to be able to expand its reaction scope. cis-CaaD, on the other hand, does not belong to the 4-OT family. They have two β - α - β building blocks and form a trimer. Both enzymes are involved in bacterial degradation of 1,3-dichloropropene.⁸ 1,3-dichloropropene is not a naturally occurring chemical, and it is the active ingredient of the pesticides used to kill nematodes. CaaD and cis-CaaD are involved in the catabolic pathway of 1,3-dichloropropene. These enzymes remove an HCl from cis/trans-3-chloroacrylic acid to produce melanic semialdehyde. Interestingly, the structural properties of TSF members allow them to evolve and start metabolizing unnatural substances such as 3-chloroacrylic acid. CaaD consists of two α and β subunits, arranged as a heterohexamer and β Pro-1, α Arg-8, α Arg-11, and α Glu-52 residues are the active site residues (Figure 1.8). cis-CaaD, however, has 149 amino acids and two β - α - β building blocks. cis-CaaD uses Pro-1, His-28, Arg-70, Arg-73, Tyr-109, and Glu-114 for its catalytic activity.³⁰

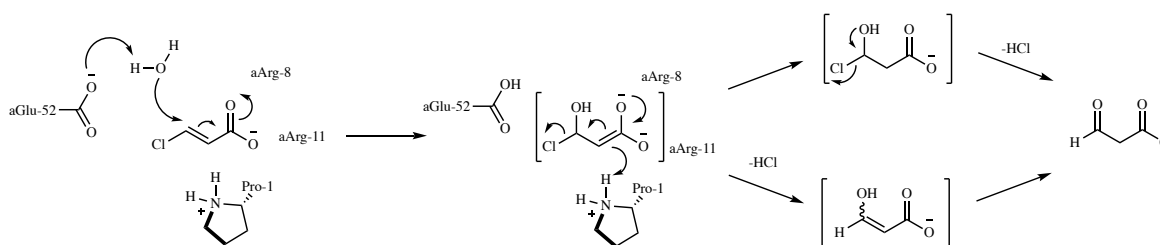


Figure 1. 7 Catalytic mechanism of CaaD with trans-chloroacrylic acid as the substrate.

Two arginine residues from CaaD are involved in binding and polarizing the substrate while α Glu-52 activates a water molecule for addition to C-3 of 3-chloroacrylic acid. The pK_a value of β Pro-1 is estimated to be 9.2, which is much higher than Pro-1 from 4-OT enzymes. At cellular pH, Pro-1 is protonated, and it transfers the proton to C-2, leading to the formation of chlorohydrin. Chlorohydrin is an unstable chemical, and it can break down chemically or by the enzyme (see Figure 1.7). In *cis*-CaaD, Tyr-103 is suggested to be involved in the water activation along with Glu-114, and His-28 might play a similar role as Arg-70 and Arg-73.³⁰

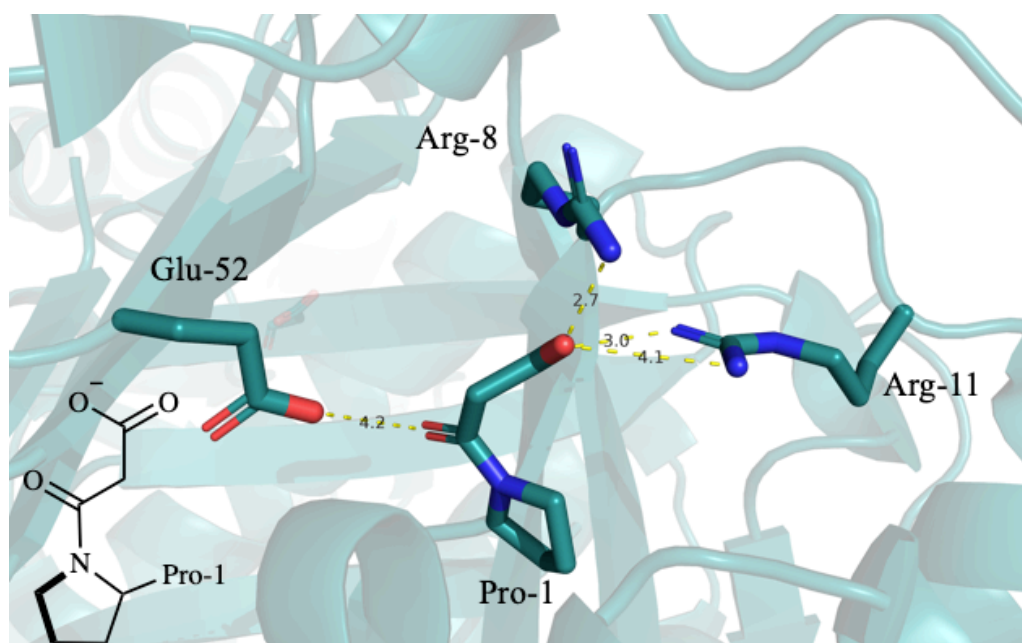


Figure 1. 8 Inactivated CaaD by 3-bromo-propiolate.

1.1.1.3. *cis*-3-Chloroacrylic acid dehalogenase

As mentioned in the previous section, *cis*-CaaD is one of the major families in the TSF with two β - α - β building blocks and active as a trimer. *cis*-CaaD catalyzes the hydrolytic dehalogenation of *cis*-3-chloroacrylic acid to produce malonate semialdehyde. Glu-114 and Tyr-103 are proposed to activate water for the attack on C-3 while Arg-70, Arg-73, and His-28 help the substrate to bind to the active site and polarize the α,β -conjugated carboxylate group.⁸

The Pro-1 residue is proposed to act as a general acid catalyst by providing a proton at C-2.⁸ An inactivated cis-CaaD crystal structure with (R)-oxirane-2-carboxylate shows that C-3 is in the perfect location for a prolyl nitrogen attack. Interactions of Arg-73 and His-28 with the carboxylate group of the ring-opened epoxide are suggesting their involvement in the binding. Arg-70 interacts directly or via a water network, with epoxide oxygen to facilitate the ring-opening.³¹

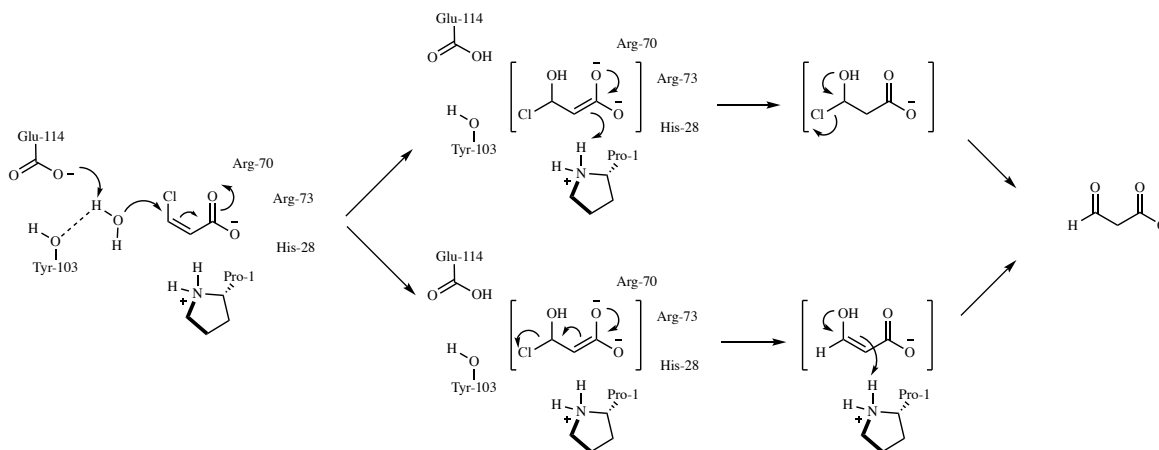


Figure 1. 9 Proposed mechanism for cis-CaaD.

Experiments show that Pro-1, Arg-70, and Arg-73 are all required for the inactivation. Glu-114 is not essential for the inactivation with (R)-oxirane-2-carboxylate. The hydroxyl group of the covalent adduct is interacting with new residues not observed before, such as hydrogen bonding with the backbone carbonyl group of Leu-38.³² The pK_a value of Pro-1 in cis-CaaD is estimated to be 9.3, similar to CaaD.

The mechanisms proposed for CaaD and cis-CaaD are similar (Figure 1.9), but mutational studies revealed differences between the two enzymes. $\alpha E52Q$ (involved in water activation similar to E114 from cis-CaaD) mutants completely lose their activity while the E114Q mutant from cis-CaaD is still active. Water activation in cis-CaaD is more complexed than CaaD, considering the importance of Y103F and more resilient to mutation of E114. The inhibition studies of CaaD and cis-CaaD identified the importance of Tyr-103 and His-28 in cis-CaaD. As a

result of these studies, water activation is more critical in CaaD, and substrate activation is more important in cis-CaaD catalysis. Both enantiomers of oxirane-2-carboxylate do not inactivate CaaD. α Arg-8 and α Arg-11 are both required for proton donation and facilitating the ring-opening. Since CaaD lacks the His-28, the carboxylate group might interact with α Arg-8 and α Arg-11 for binding. Interaction of α Arg-8 and α Arg-11 with the carboxylate group prevents them from protonating the epoxy oxygen.⁹ (Figure 1.10)

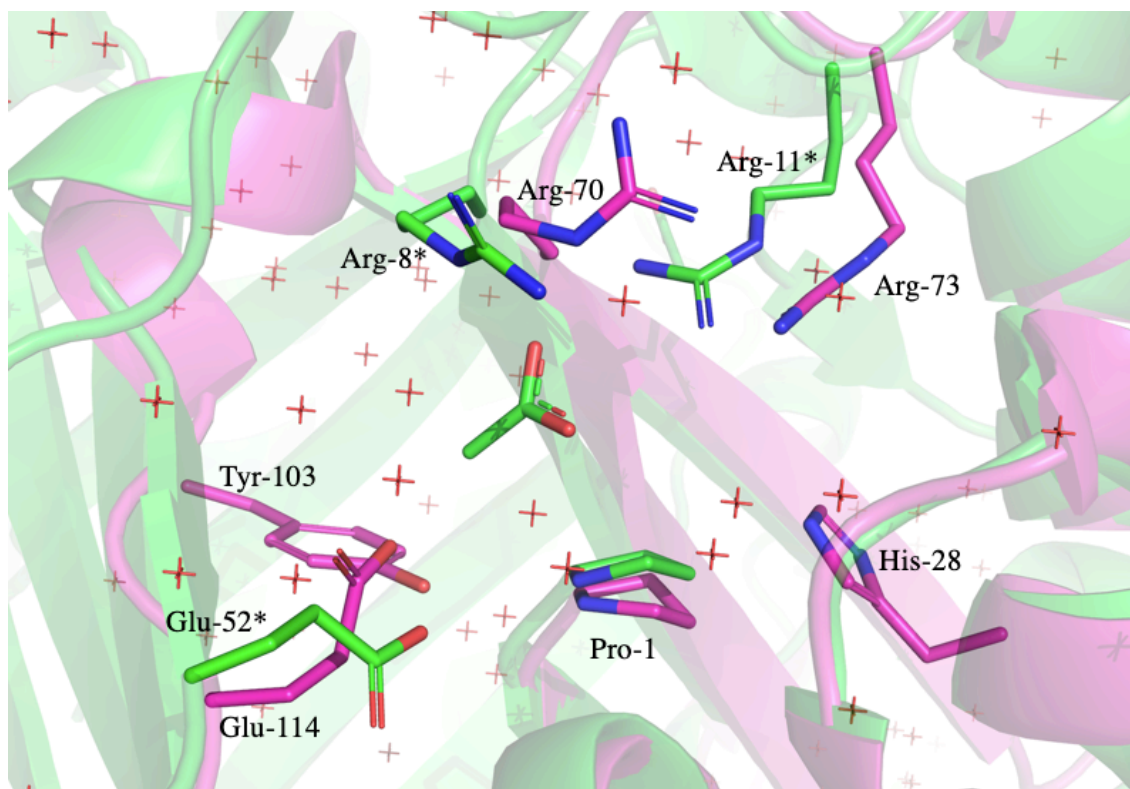


Figure 1. 10 The active site of CaaD (green) and cis-CaaD (magenta)

1.1.1.4. Malonate semialdehyde decarboxylase

Malonate semialdehyde decarboxylases, a subfamily of TSF, have two β - α - β building blocks and they are active in trimer form^{16,21,33–35} Pro-1 and Arg-75 are identified as the active site residues. The pK_a value of Pro-1 in MSAD is measured by titration using ^{15}N NMR spectroscopy and determined to be ~ 9.2 , suggesting a general acid catalysis role for Pro-1 in MSAD. A high pK_a

value suggests an electrostatic interaction, via hydrogen bonding, that polarizes the carbonyl oxygen of malonate semialdehyde.

Inhibition studies on MSAD generate inactivated enzyme with Pro-1 being covalently modified. Pro-1 can act as a nucleophile and form the covalent adduct, only if it is deprotonated by the initial hydration of the inhibitor. Structural studies of inactivated MSAD added Asp-37 and Arg-73 to the list of catalytic residues (Figure 1.11). The decarboxylation mechanism of MSAD is conducted by a water network between Pro-1 and Asp-37 to activate the carbonyl oxygen. Activated carbonyl oxygen creates an electron sink that facilitates decarboxylation.³³

The pK_a value for Pro-1 in the D37A mutant drops to ~ 6.4 , suggesting that Pro-1 interaction with Asp-37 via the water network turns Pro-1 to a general acid catalyst.⁹ Other than decarboxylation activity, MSAD has shown hydration activity on 2-oxo-3-pentynoate. The hydration mechanism of MSAD parallels that of CaaD. Asp-37 and α Glu-52 are equivalent and Arg-73/Arg-75 replaces α Arg-8/ α Arg-11.¹⁶

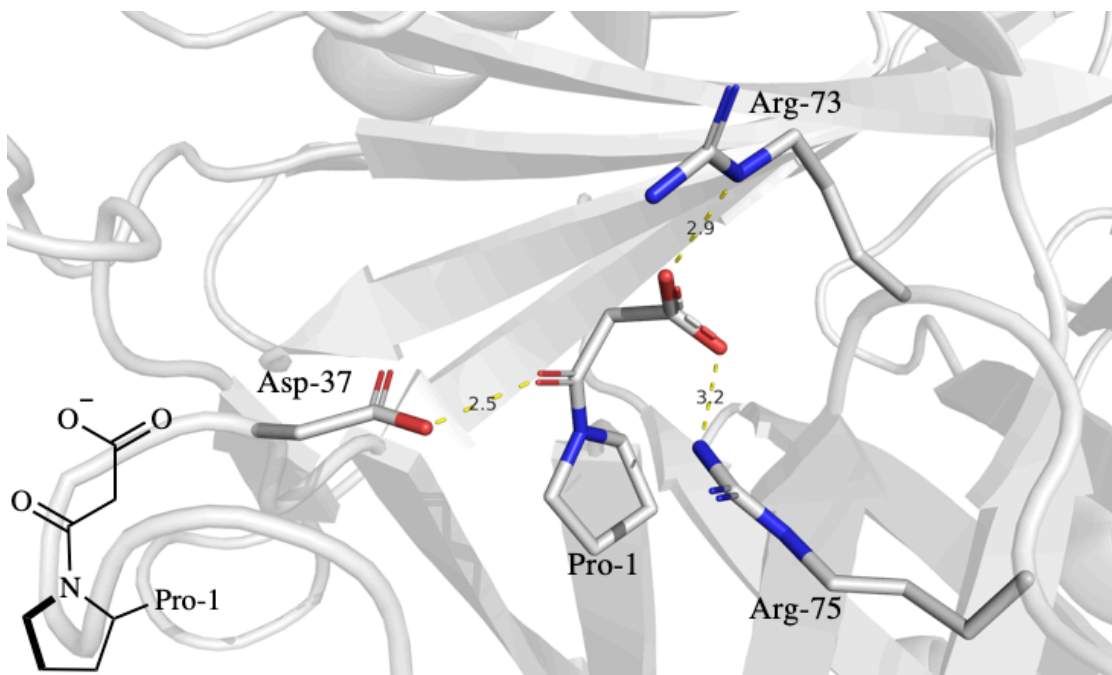


Figure 1. 11 Active site of MSAD, inactivated by 3-bromo-propionate.

1.1.2. Engineering tautomerase enzymes

The β - α - β building block is the basic structural unit of TSF enzymes. The gene duplication event created an isoform, which then evolved independently, and the gene fusion event combined the evolved isoforms with the ancestor unit to make use of the positive changes they accumulated for expanding their reaction scope. The small size of the β - α - β domain and the fact that they function without metals and cofactors, make them an excellent target for directed evolution studies³⁶⁻³⁸ and protein engineering.^{39,40}

Poelarends and coworkers⁴¹ screened 1040 single mutants of the 4-OT from *Pseudomonas putida* mt-2 for self-condensation of linear aliphatic aldehydes such as propanal via an aldolase reaction. (Figure 1.12)

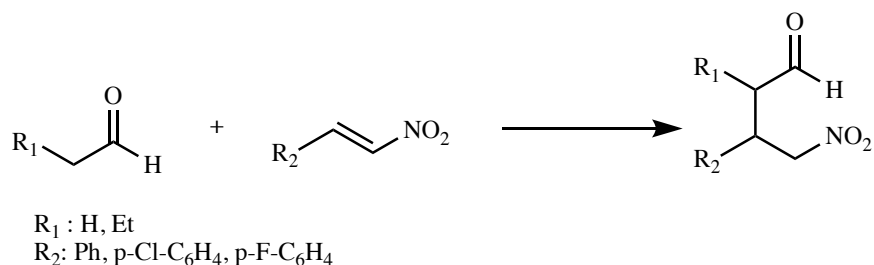


Figure 1. 12 Aldolase reaction target for 4-OT directed evolution.

Later they constructed a double-site (800 transformants) and a triple-site (3500 transformants) library. They were able to identify hot spots for mutations and discovered M45Y/F50V variant with significant aldolase activity in the self-condensation of propanal. The new variant has ~30-fold improvement over wild-type.^{36,41,42} Using the native substrate to examine the mutability space of 4-OT could be beneficial for understanding the mechanism. However, the high activity of the enzyme makes the comparison unreliable. Phenylpyruvate is not the native substrate for 4-OT, but it is ketonized by 4-OT slowly ($k_{\text{cat}} = 73 \text{ s}^{-1}$) to phenylpyruvate. Two regions in the mutability landscape for 4-OT, tautomerizing phenylpyruvate, show no or little effect on

activity. These regions are Ser-12 to Arg-29 and Gly-54 to Arg-62. The first region is between Arg-11 and Arg-39, the two amino acids that were identified by Whitman and colleagues to be required for tautomerization in 4-OT. Also, Gly-10 is found to be essential for the activity of the enzyme; however, the proximity of the residue to active site residue Arg-11 suggesting a structural tuning role.³⁶ Screening the mutability landscape led to the identification of hot spots that improves tautomerase activity (> 5-fold). Mutation sites enhancing the tautomerase activity are Ile-2, Gln-4, Leu-8, Ser-37, and Phe-50. Among these mutations, Gln-4 is ~11Å away from Pro-1, suggesting that distant mutation can improve the activity as well.³⁶

Recently Wiltschi and colleagues⁴⁰ engineered a 4-OT enzyme by substituting the Pro-1 with non-canonical analogs (Figure 1.13) such as (2S,4R)-4-fluoroproline **1**, (2S,4S)-4-fluoroproline **2**, (2S)-3,4-dehydroproline **3**, and (4R)-1,3-thiazolidine-4-carboxylic acid **4**.

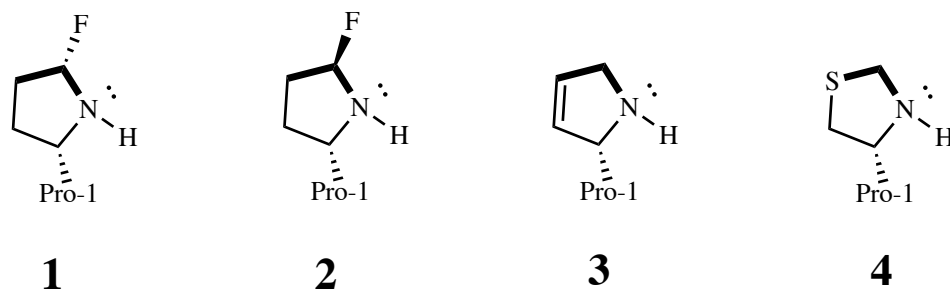


Figure 1. 13 non-Canonical analogs of proline. (2S,4R)-4-fluoroproline, (2S,4S)-4-fluoroproline, (2S)-3,4-dehydroproline, and (4R)-1,3-thiazolidine-4-carboxylic acid.

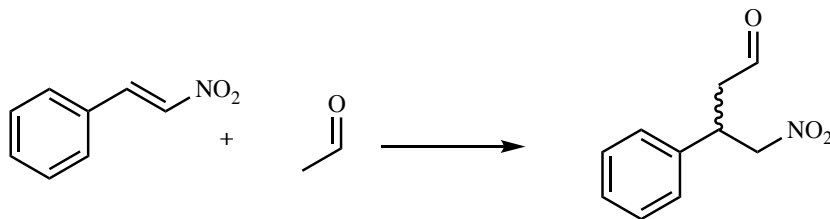


Figure 1. 14 Target reaction for 4-OT enzymes with non-canonical proline.

The target reaction for the modified enzyme is the Michael addition of acetaldehyde to nitrostyrene (Figure 1.14). Evaluating the activity of these analogues in the reaction showed a significant reduction of activity compared to the native enzyme. There exists another proline

residue in the native enzyme, Pro-34, which is replaced by the non-canonical analogs of proline as well. In order to eliminate the effect of Pro-34 substitution in the study, a mutant library at position 34 was constructed to identify the mutations with similar activity for Michael-type reaction as the native enzyme. P34E maintained similar activity and expression levels and is used for the incorporation of non-canonical analogs. However, the variants of 4-OT P34E were also inactive toward the Michael-type reaction. The crystal structure of 3,4-dehydroproline variant of 4-OT was obtained, and it has identical tertiary structure, and the active site residues show no deviation from the native enzyme. To investigate the reason for the loss of activity, theoretical pK_a of these analogs (9.7 for 4-fluoroproline, 10.7 for 3,4-dehydroproline, and 7.7 for 1,3-thiazolidine-4-carboxylic acid) are compared to that of free proline at 10.6. Since the pK_a of proline is lowered by approximately four units (~ 6.4) in the active site compared to the free amino acid, estimated pK_a values for 4-fluoroproline and 3,4-dehydroproline will be ~ 5.7 and ~ 6.7 , respectively. pK_a value for thiazolidine-4-carboxylic acid in the active site is estimated to be ~ 3.7 . Measuring the activity at different pH values did not improve the activity, showing that variability in pK_a values of analogs vs. proline is not responsible for the loss of activity.⁴⁰ Even though the study was not able to generate a successful variant for the Michael-type addition, it did demonstrate that the critical role played by Pro-1 in the activity of 4-OT makes it irreplaceable.

1.1.3. Cg10062, a member of the TSF

Since the first discovery of TSF in 1996, significant progress has been made in identifying and understanding its members. Protein Data Bank contains 178 structures from 49 members of the TSF in native or mutated forms and bound to the substrate or an inhibitor. Kinetic studies of mutated enzymes from different families of TSF combined with structural information enhanced

our understanding of the substrate variability and mechanistic diversity in the superfamily. Besides, the creation of sequence databases and improved bioinformatic tools vastly expands our ability to connect acquired knowledge from different members of the family. Bioinformatic will be able to identify previously uncharacterized members of the family and define the evolutionary connection between them.

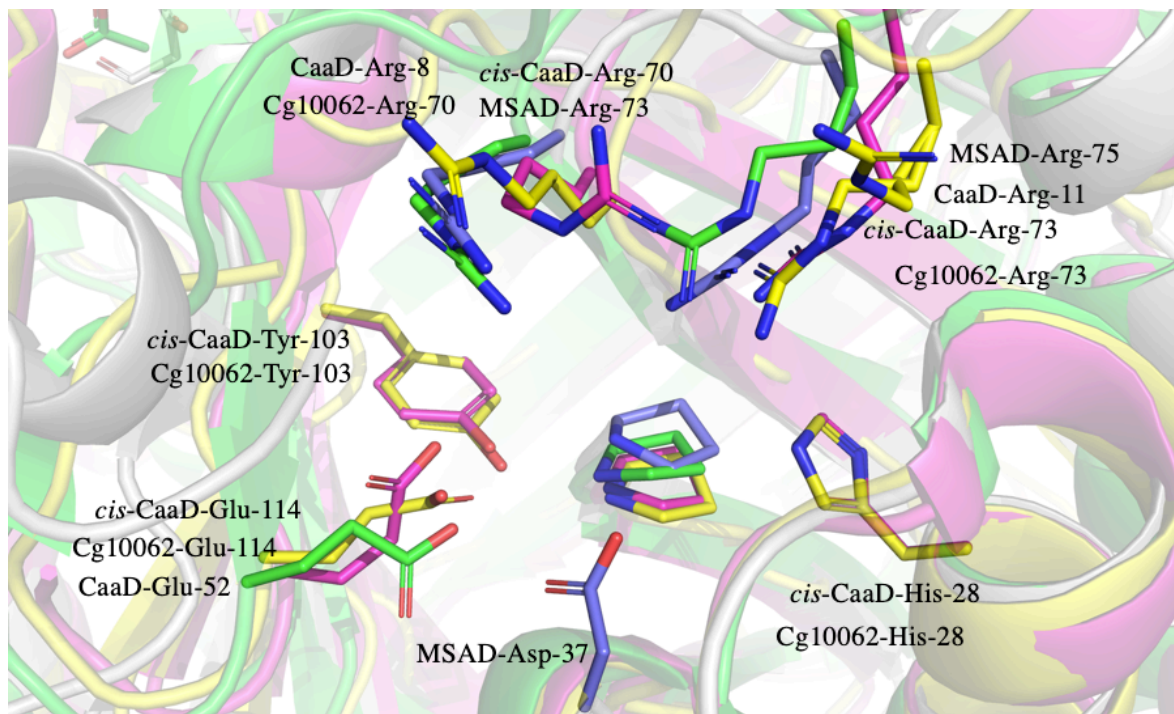


Figure 1.15 Cg10062 and cis-CaaD have identical active site residues. Cg10062 (yellow), cis-CaaD (magenta), CaaD (green), MSAD (purple)

Cg10062 is identified as a 149-amino acid homolog of cis-CaaD. It shares 24% sequence identity, and 53% sequence similarity with cis-CaaD and all the residues critical for cis-CaaD activity are present in the protein (Figure 1.15).⁷ Cg10062 converts 2-oxo-3-pentynoate to acetopyruvate and is inactivated by an acyl halide or a ketene, similar to cis-CaaD. However, it has much lower activity and lacks the stereospecificity.⁴³ Cg10062 acts on both isomers of 3-chloroacrylate slowly and with a preference for the cis-isomer. Cg10062 differs from cis-CaaD at position 69, where the His-69 from cis-CaaD is replaced by isoleucine in Cg10062. His-69 in cis-

CaaD is in the active site, and both His-28 and His-96 are interacting with the hydroxyl group of Tyr-3. Another significant difference is a nine-residue loop connecting the α -helix of the β - α - β unit to the second β -strand. RGLTGTQHF is the nine-residue loop from cis-CaaD, while Cg10062 has HELAHAPKY residues in the loop.⁹ This loop has residues facing the active site such as two threonines from the cis-CaaD loop, which are replaced by two alanine in the Cg10062. The cis-CaaD loop might be more flexible than the Cg10062 loop since it has two glycine residues compared to the Cg10062 loop that contains a proline residue.^{44,45}

The Cg10062 processes cis-3-bromoacrylate ~1000-fold slower than cis-CaaD and even dehalogenases trans-3-bromoacrylate unlike cis-CaaD, albeit slower than cis-isomer. This observation shows that the active site of the Cg10062 is not optimized for the dehalogenation reaction. The active site of Cg10062 is observed to be more spacious than the cis-CaaD active site.⁴⁵

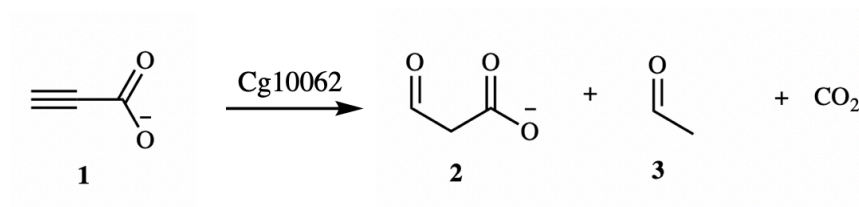


Figure 1. 16 Hydration and decarboxylation of propiolate by Cg10062.

Exploring the reactivity of CaaD, cis-CaaD, and Cg10062 toward various substrates helps us to understand the difference between these enzymes. Kinetic studies from Dr. Darths laboratory is summarized in table 1.2. In order to measure the ration of hydration and decarboxylation, an enzyme coupled assay is used to detect acetaldehyde. In the absence of MSAD, only acetaldehyde produced by decarboxylation is detected, but in the presence of MSAD, malonate semialdehyde is decarboxylated to produce more acetaldehyde. The difference between the acetaldehyde in the presence and the absence of MSAD gives the hydration product.

Table 1. 2 Kinetic data for native and mutants of Cg10062 (data from Ms. Amaya Sirinimal in Dr. Karen Draths laboratory)

Cg10062 variant	Specific Activity ($\mu\text{mol min}^{-1}\text{mg}^{-1}$)		Product ratio (%)	
	+ MSAD	-MSAD	2	3
native	8.3	6.9	19	81
E114N	2.9	0	100	0
E114Q	1.3	0	100	0
E114D	1.3	0	100	0
Y103A	0.14	0.08	43	57
Y103F	2.2	1.4	36	64
H28A	0.046	0.018	60	40
E114A	0.013	0.007	44	56
E114D-Y103F	0.74	0	100	0
R70A		inactive		
R70K		inactive		
R73A		inactive		
R73K		inactive		

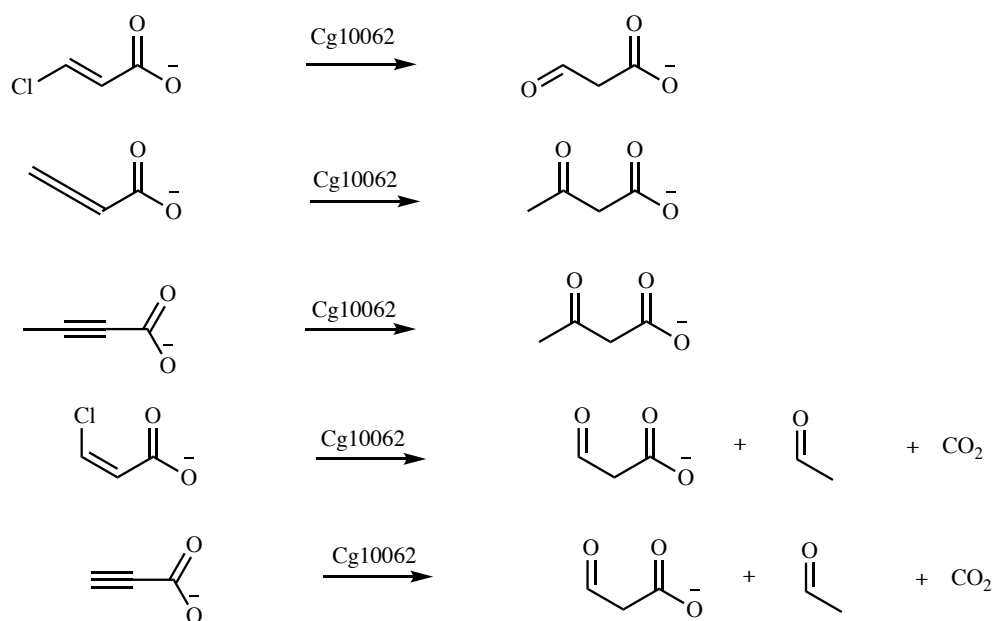


Figure 1. 17 Hydration and decarboxylation substrates for Cg10062

One of the first evidence of covalent intermediate in the mechanism of Cg10062 came from Whitman and coworkers. Incubating Cg10062-Y103F (and Cg10062-E114D) with 2-butyrate in

the presence of NaCNBH_3 . Analysis of ESI-MS spectra show an added mass that corresponds to the covalently modified imine intermediate. Three mechanisms are proposed by Whitman and coworkers for the hydration/decarboxylation activity of Cg10062.³⁰ (Figures 1.18, 1.19, 1.20)

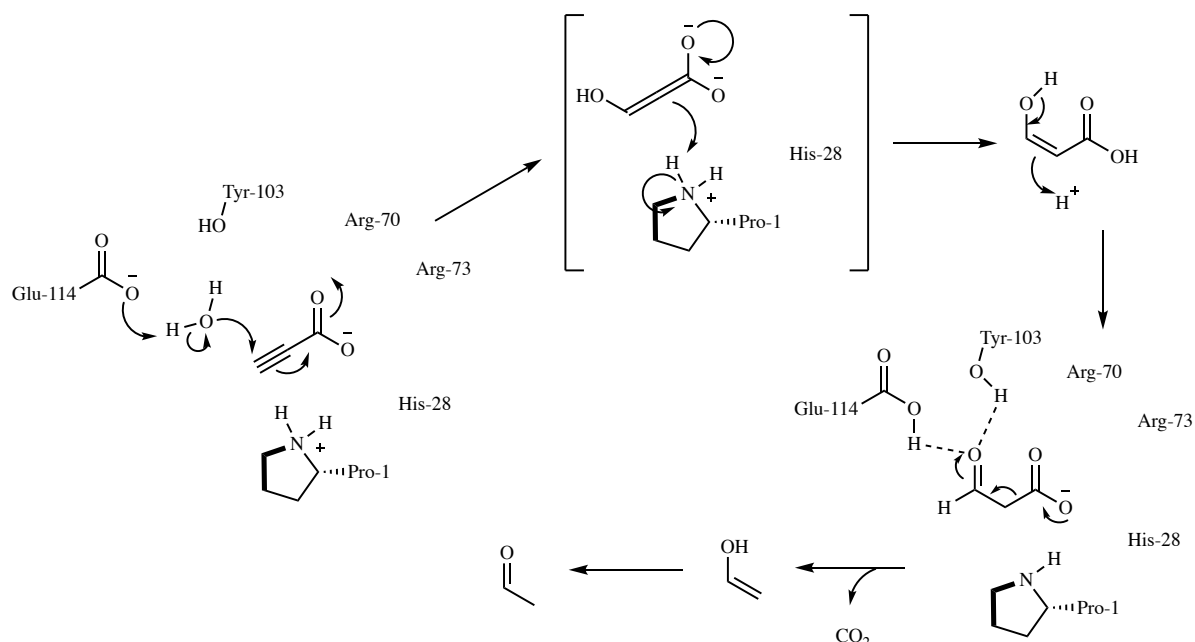


Figure 1. 18 Proposed mechanism for Cg10062 hydration/decarboxylation reaction via Pro-1 as general acid catalyst.

Evidence for the formation of iminium species in the reaction indicates the possibility of covalent intermediate. The steady-state kinetic parameters for Cg10062 are reported by Whitman⁴³ and colleagues, using acetylene compounds, *cis*-3-chloroacrylate, *trans*-3-chloroacrylate, propiolate, 2-butyrate, and 2,3-dibutadienoate as the substrate (see Table 1.3). They examined the products of the Cg10062 reactions with these substrates and the results are summarized in Table 1.4.

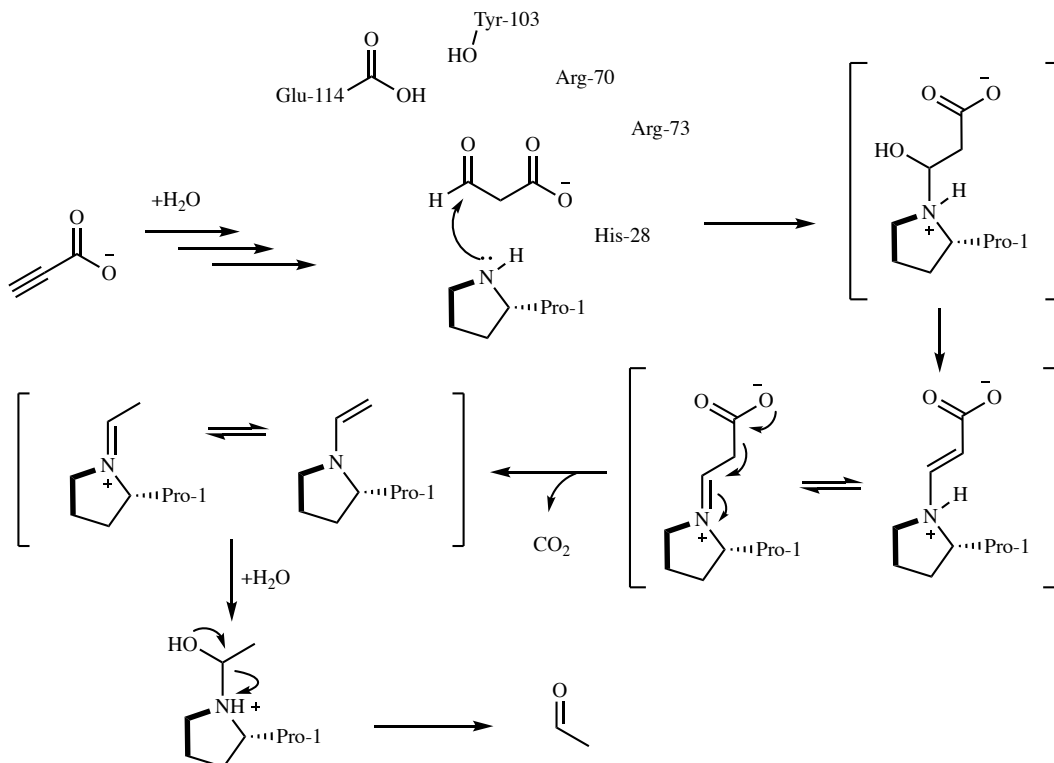


Figure 1. 19 Proposed mechanism for Cg10062 decarboxylation reaction via Schiff base intermediate.

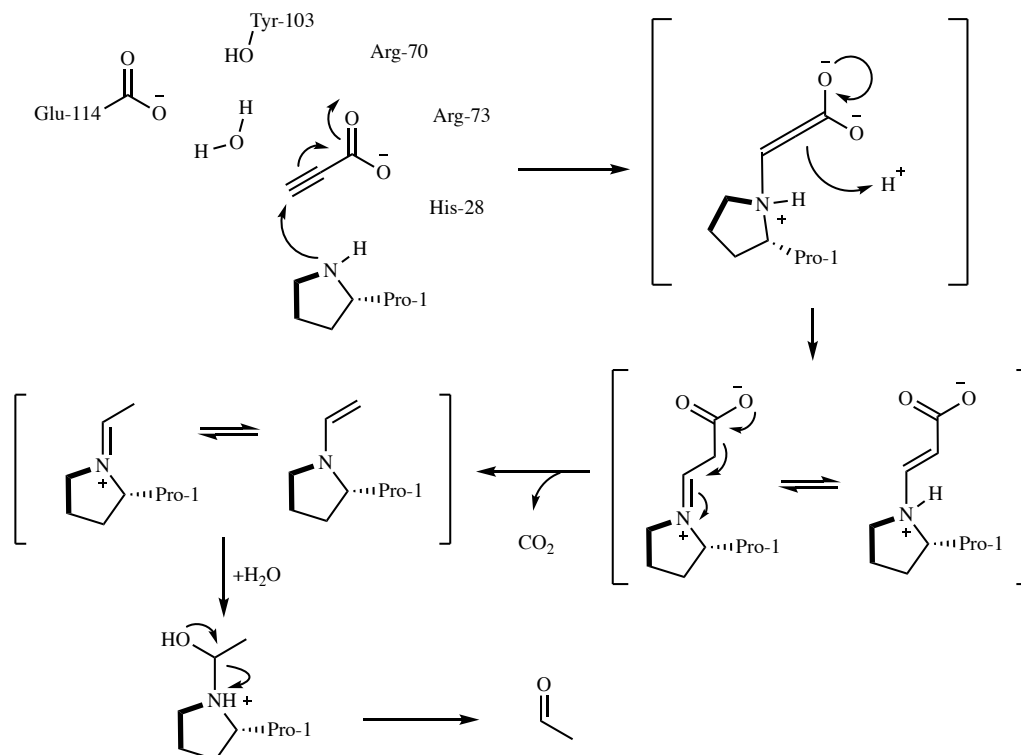


Figure 1. 20 Proposed mechanism for Cg10062 hydration/decarboxylation reaction via covalent intermediate

Table 1. 3 Steady-state kinetic parameters for the native Cg10062 and mutants.⁴³

Enzyme	substrate	K_m (μ M)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1} s^{-1}$)
Native	cis-3-chloroacrylate	72000 ± 8500	1 ± 0.1	14 ± 2
	trans-3-chloroacrylate	78000 ± 36000	0.06 ± 0.01	0.8 ± 0.4
	Propiolate	33 ± 5	6 ± 0.2	$(1.8 \pm 0.2) \times 10^5$
	2,3-dibutadienoate	780 ± 120	4 ± 0.3	$(5.1 \pm 0.8) \times 10^3$
	2-butyrate	-	-	30 ± 2
E114Q	cis-3-chloroacrylate	4000 ± 800	0.4 ± 0.05	100 ± 25
	Propiolate	3 ± 0.25	0.8 ± 0.02	$(2.7 \pm 0.2) \times 10^5$
	2,3-dibutadienoate	660 ± 75	16 ± 1	$(2.4 \pm 0.3) \times 10^4$
E114D	cis-3-chloroacrylate	40100 ± 10000	0.1 ± 0.02	2.5 ± 0.8
	Propiolate	90 ± 13	1 ± 0.05	$(1.1 \pm 0.2) \times 10^4$
	2,3-dibutadienoate	480 ± 86	3.2 ± 0.4	$(6.6 \pm 1.5) \times 10^3$
	2-butyrate	1365 ± 170	0.5 ± 0.03	$(4.1 \pm 0.5) \times 10^2$
Y103F	cis-3-chloroacrylate	4700 ± 320	0.3 ± 0.04	60 ± 10
	Propiolate	5 ± 1.8	0.5 ± 0.01	$(1.0 \pm 0.3) \times 10^5$
	2,3-dibutadienoate	1150 ± 175	13 ± 2.5	$(1.1 \pm 0.3) \times 10^4$
	2-butyrate	650 ± 140	2.1 ± 0.2	$(3.2 \pm 0.8) \times 10^3$

Table 1. 4 Reaction products ratio for native Cg10062 and mutants.⁴³

Enzyme	substrate	Hydration product (%)	Decarboxylation product (%)	Reaction completion time (min)
Native	cis-3-chloroacrylate	7	33	48
	trans-3-chloroacrylate	2.9	1.5	48
	Propiolate	25	75	< 3
	2,3-dibutadienoate	~ 100	< 1.5	< 3
	2-butyrate	~ 100	2.5	39
	Malonate semialdehyde		< 1.5	
	Malonate semialdehyde ^a		~ 1.5	
E114Q	cis-3-chloroacrylate	19	1	18
	trans-3-chloroacrylate	0	0	90
	Propiolate	8.5	0.5	18
	2,3-dibutadienoate	83	17	12
	2-butyrate	0.8	2.2	90
E114D	Propiolate	22	1	3
	2,3-dibutadienoate	97.2	2.8	3
	2-butyrate	91.5	2.5	3
Y103F	cis-3-chloroacrylate	1	8	18
	trans-3-chloroacrylate	~ 0	0	18
	Propiolate	0.7	2.1	27
	2,3-dibutadienoate	99.5	0.5	3
	2-butyrate	84	16	6

^a non-enzymatic decarboxylation

In our investigation of the mechanism of Cg10062, crystal structures of substrate-bound mutants are obtained. In the next section, these results are presented and elaborated to enhance our understanding of the Cg10062 reaction mechanism.

1.2. Investigating Cg10062 hydratase/decarboxylase activity

X-ray crystal structures of Cg10062 (native, E114D, Y013F, H28A, R73A) in the form of apo or ligand-bound is obtained. Details of the crystallization procedure and soaking conditions are described in the next section (material and methods).

1.2.1. X-ray crystal structures of native Cg10062

The x-ray crystal structure of native Cg10062 (Figure 1.21) was solved to 1.37Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.18 and 0.20, respectively (Table 1.5). The asymmetric unit contains one protein chain. All residues from Pro-1 to position 119 are fit into the electron density, but there was not any electron density for residues from 120 to the C-terminus.

The two α -helices in the C-terminus of the native Cg10062 were not built due to lack of electron density. This region was resolved in previously published structures of Cg10062 (PDB ID: 3N4G)⁴³. After examining the crystal packing and overlaying of the Cg10062 structure with folded C-terminus, we observed that in this crystal packing, a properly folded C-terminus region will clash into neighboring units. However, it is not clear that the crystal packing prevents the proper folding of the C-terminus, or a flexible C-terminus will lead to the observed crystal form. In the tertiary structure of the Cg10062 a pair of two-stranded β -sheets interacting in an anti-parallel form to create a four-stranded β -sheet. Also, two anti-parallel α -helices are interacting

with the four-stranded β -sheet. Three monomers are forming a barrel-like structure with hydrophobic residues located in the center of the trimer. There are many identical residues between Cg10062 and cis-CaaD, and the majority of them are the hydrophobic residues located in the center of the trimer.

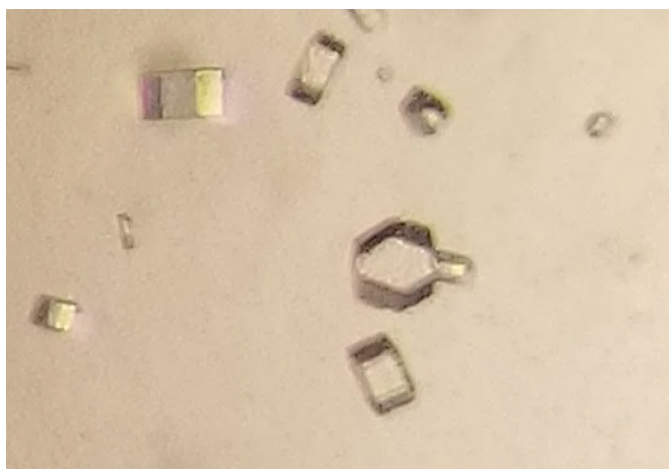


Figure 1. 21 Crystals of native Cg10062

A complete active site is formed by residues from N-terminus of a monomer and C-terminus of an adjacent monomer. The secondary structure of one monomer (Figure 1.22) consists of a β -strand starting from the 1-Pro residue (β 1). β 1 (residue 1 to 8) continues with a loop (residue 9 to 13) and an α -helix. α 1 (residue 14 to 32) is followed by a loop (residue 33 to 35) and a 3_{10} helix (residue 36 to 38). α 2 is then followed by a β -sheet parallel to β 1. β 2 (residue 39 to 46) is followed by a loop (residue 47 to 50), a β -strand (residue 51 to 53), and another loop (residue 54 to 62). The β 3 is a small β -sheet and is forming an anti-parallel interaction with β 2 of an adjacent monomer. Next, a β -strand (residue 63 to 70) forms an anti-parallel interaction with β 1, and then β 4 continues to form an α -helix after a loop (residue 71 to 74). α 3 (residue 75 to 93) is connected to a 3_{10} helix (residue 97 to 99) with a short loop (residue 94 to 96). Finally, β 5 (residue 100 to 108) forms a parallel interaction with β 4 and ends with a 3_{10} helix (residue 109 to 111) and an extended loop (residue 112 to 119).

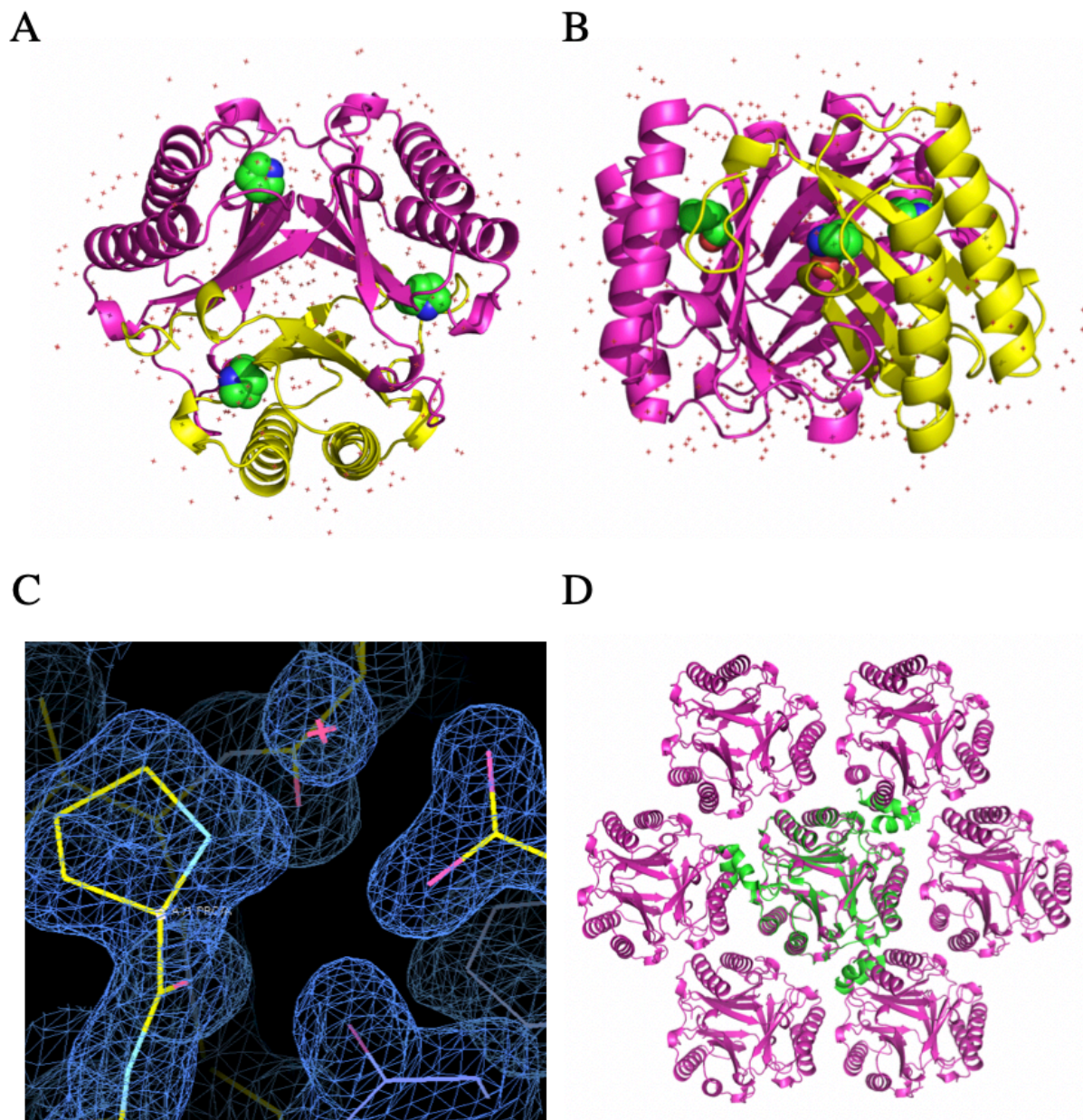


Figure 1. 22 Structure of Cg10062 apo. (A) top view and (B) side view of the Cg10062 apo structure, single chain in the ASU (yellow), symmetry mates to complete the trimer (pink), active site Pro-1 (green). (C) A snapshot of Pro-1 fitted in the electron density. (contoured at rmsd 1Å) Tyr-103, His-28, and Glu-114 are present in this view. (D) overlay of 3N4G structure (green) to our native apo Cg10062 (pink).

Table 1. 5 Data collection and refinement statistics for Cg10062 Apo.

Data statistics (last shell in parenthesis)	
Space group	P 3 2 1
No. of chains/ASU	1
Unit cell dimensions (Å, °)	a = 49.577, b = 49.577, c = 80.067 $\alpha = 90, \beta = 90, \gamma = 120$
Resolution range (Å) (outer shell)	42.93 - 1.37 (1.419 - 1.37)
Unique reflections (outer shell)	24563 (2401)
Completeness (%) overall (outer shell)	99.93 (99.42)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	24559 (2399)
Reflections used for R-free	2001 (193)
R-work	0.1808 (0.2401)
R-free	0.2037 (0.2621)
Number of non-hydrogen atoms/ solvent	1134/99
Protein residues	119
RMS (bonds) (Å)	0.007
RMS (angles) (Å)	1.21
Ramachandran plot (%) favored/allowed/outliers	99.15/0.00/0.85
Rotamer outliers (%)	0.00
Average B-factor/macromolecules/solvent	21.47/20.42/32.43

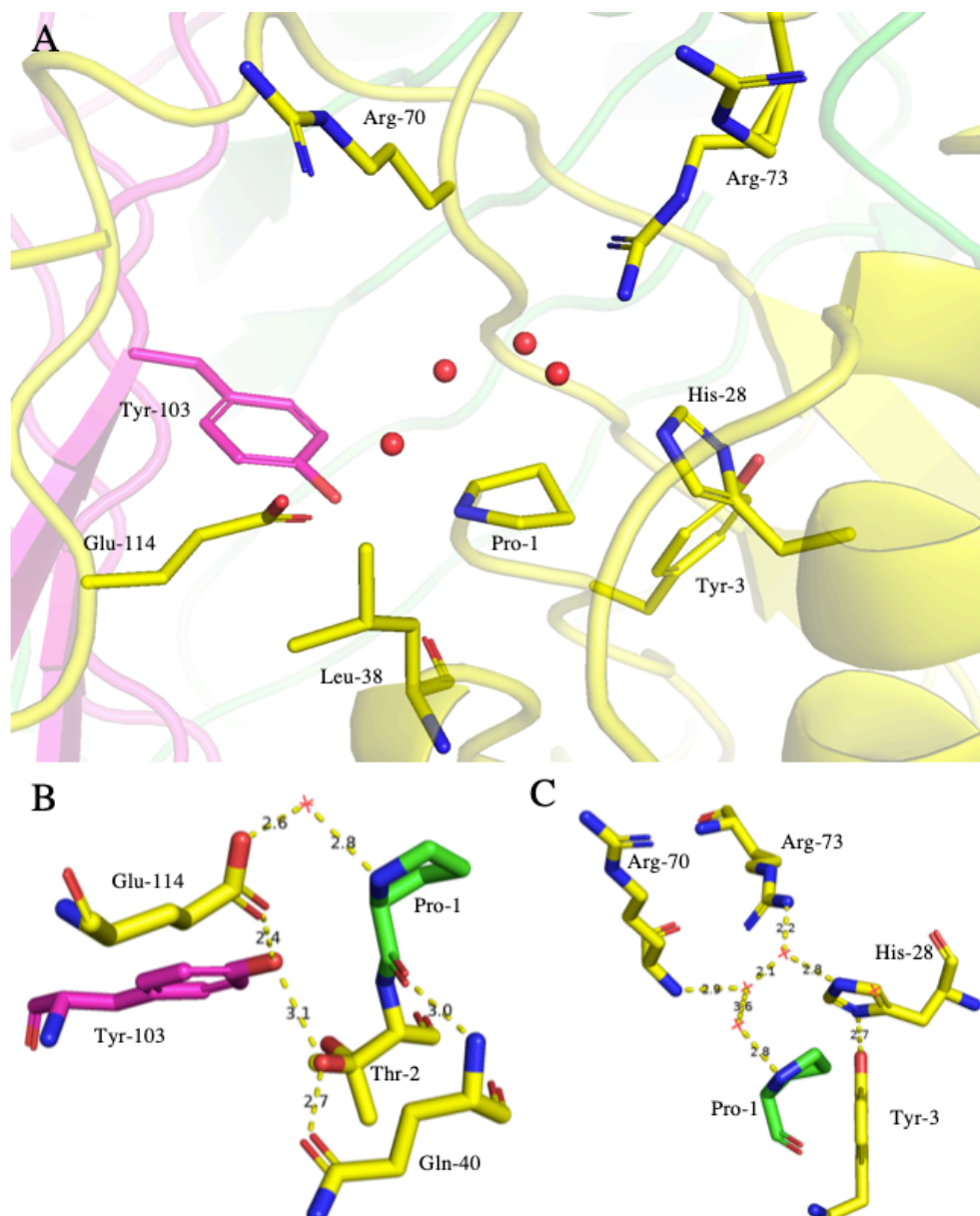


Figure 1. 23 Amino acid residues in the active site of the Cg10062 and their interactions. (A) Important amino acids in the active site, Tyr-103 from adjacent monomer (pink), (B) hydrogen bonding network between Pro-1 (green), Glu-114, Thr-2, and Gln-40 (yellow), Tyr-103 of adjacent monomer (pink), and a water molecule (red). (C) hydrogen bonding network between Pro-1 (green), Tyr-3, His-28, Arg-70, and Arg-73 (yellow), and two water molecule (red)

All functionally important residues from cis-CaaD are present in the active site of Cg10062.

Figure 1.23-A shows the important residues within 6 Å of the Pro-1. Among these residues, Tyr-103, and Trp-101, are present from a neighboring monomer. There exist four water molecules in the proximity of the Pro-1 and are involved in hydrogen-bond networks.

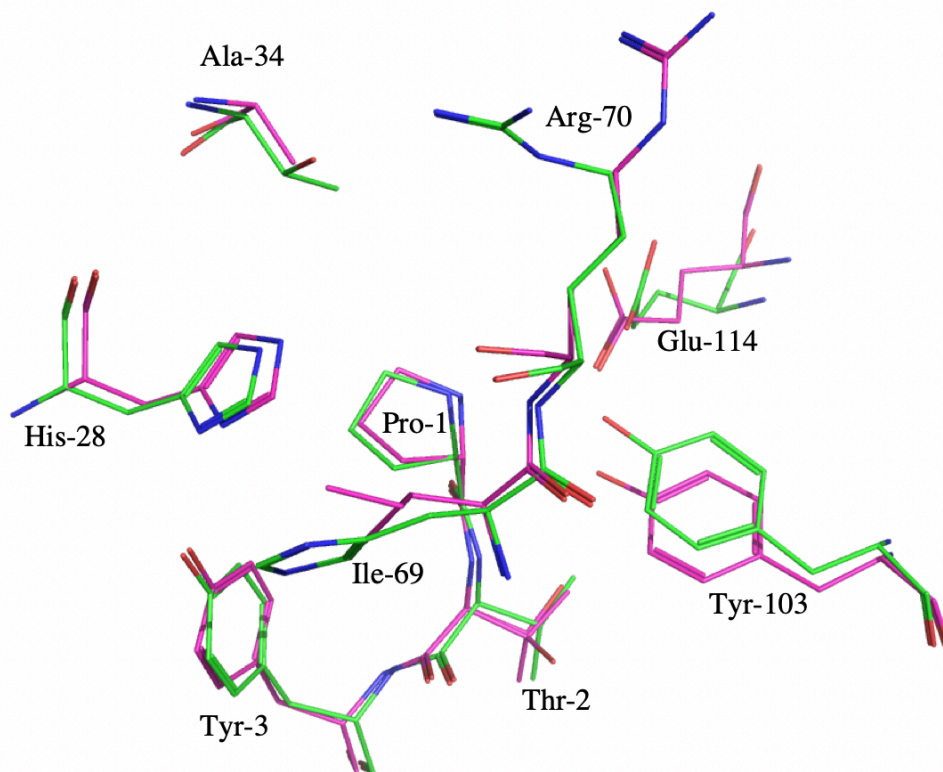


Figure 1. 24 Comparison of active site residues between Cg10062, cis-CaaD. Cg10062 (pink), cis-CaaD (green)

Two hydrogen-bond networks are present in the active site (Figure 1.23 B and C). In the first hydrogen-bond network, a water molecule is forming hydrogen bonds with Pro-1 and Glu-114. Glu-114 continues by forming a hydrogen bond with the hydroxyl group from Tyr-103. The backbone carbonyl from Pro-1 is also forming a hydrogen bond to the backbone amide of Gln-40. This bond is a typical hydrogen bond between β -strands. However, there is a hydrogen bond from Gln-40 to one of the conformers of Thr-2, which in turn forms a hydrogen bond to Tyr-103. Among these residues, only Glu-114 and Tyr-103 are known to be involved in the catalytic action of a cis-CaaD. Comparing the active site of Cg10062, CaaD, and cis-CaaD shows that the active site of Cg10062 is much more similar to cis-CaaD than that of CaaD. Tyr-103 is present in cis-CaaD as well, while Thr-2 is replaced by a valine residue. As a result, the hydrogen bonding between Thr-2 and Tyr-103 is not present in cis-CaaD.

A second hydrogen-bond network is present in the active site of the Cg10062 where the backbone amide of Arg-70 and guanidine moiety of Arg-73 are connected via two water molecules. The connection extends to His-28 via one of the water molecules. Pro-1 is also connected to these amino acids via the water molecule from the first hydrogen bonding network. His-28 also forms a hydrogen bond to Tyr-3. In *cis*-CaaD, His-69 is interacting with His-28; however, in the Cg10062 active site, an isoleucine residue is placed at position 69.

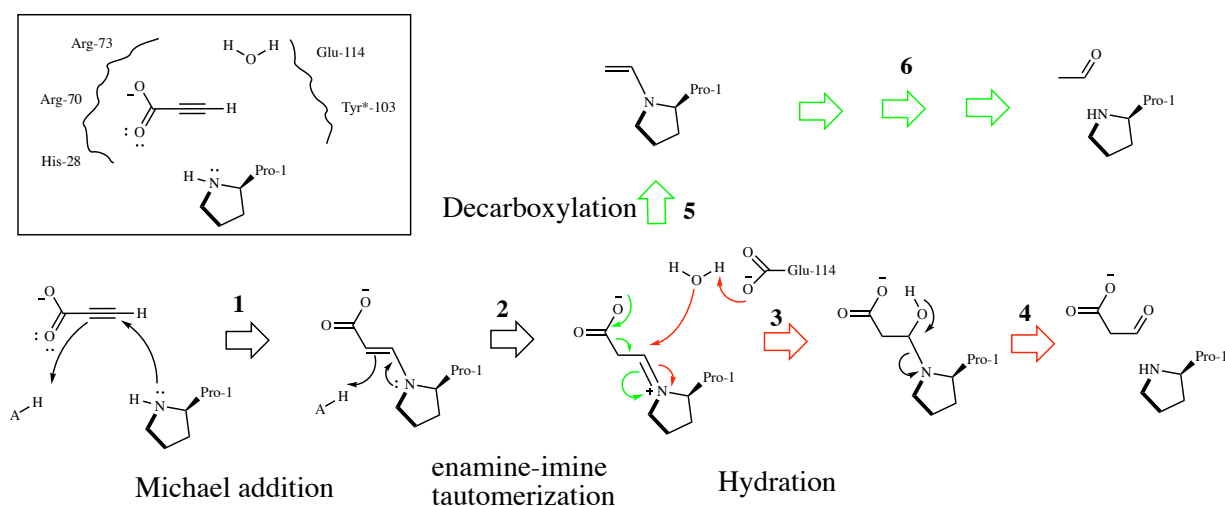


Figure 1. 25 Plausible mechanism for the reaction of propiolate and Cg10062 via covalent intermediates.

Overall, the active site of Cg10062 is quite similar to *cis*-CaaD (Figure 1.24) with a few differences, such as Thr-2 and Ile-69. These enzymes have different activities with identical substrates, and we will try to explain these differences. Soaking of native Cg10062 crystals in the propiolate solution was attempted, but the experiments were unsuccessful, and the crystals dissolved in the process. Therefore, we continued our investigation with the mutants of Cg10062. E114D, Y103F, H28A, and R73A mutants of Cg10062 were successfully crystallized and soaked in propiolate solutions to obtain substrate intermediate-bound structures. These structures show significant electron density connected to proline, and it is extended into the cavity of the active

site. As a result of this observation, we proposed a mechanism for the hydration/decarboxylation that involves the formation of a covalent intermediate.

In this mechanism (Figure 1.25), His-28, Arg-70, and Arg-73 are proposed to be involved in the binding of the substrate via electrostatic interactions. On the other side, Glu-114 and Tyr*-103 (the active site tyrosine belongs to the neighboring unit in the trimer) are involved in the water activation. The role of all these active site residues is similar to previous studies (discussed in the introduction section) on similar enzymes such as *cis*-CaaD. However, the Pro-1 plays an active role in forming a covalent intermediate, unlike other proposed mechanisms where it only acts as a general acid/base catalyst, or it is involved in the polarization of the π -bond. We propose that the decarboxylation step only occurs after the formation of iminium intermediate. The iminium acts as an electron sink to facilitate the decarboxylation process. However, in the absence of iminium after the hydration step, the substrate is immune to decarboxylation.

Based on this mechanism, the plausible intermediates and products for the reaction of propiolate with Cg10062 are propiolate, *cis*/*trans*-3-prolylacrylate, 1-prolylethylene, 3-hydroxy-3-prolylpropionate, 1-hydroxy-1-prolyl-ethylene, malonate semialdehyde, and acetaldehyde.

1.2.2. X-ray crystal structures of Cg10062-H28A soaked in propiolate

The x-ray crystal structure of Cg10062-H28A soaked in propiolate (Figure 1.27) was solved to 2.95 Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.30 and 0.24, respectively (Table 1.6). The crystal packing for this mutation is different than that of native Cg10062. The crystal of Cg10062-H28A soaked in propiolate has P4₁32 symmetry. First, the structure was solved by molecular replacement by searching for one copy of the protein in the asymmetric unit. Examining the crystal packing and relative distance between the protein chains

revealed a large space in the unit cell of the crystal. The crystal is formed by the trimer units (Figure 1.26). Further investigation showed a much weaker electron density in the empty regions of the unit cell. These electron densities are indicating that more than one protein chain is present in the asymmetric unit of the crystal, and the new chain is positioned in the hollow space of the crystal. However, the low intensity of the electron density is suggesting that the occupancy of the second chain is not 100%.

We decided to solve the crystal structure by searching for two chains in the asymmetric unit and lowering the occupancy of chain B in the refinement step.

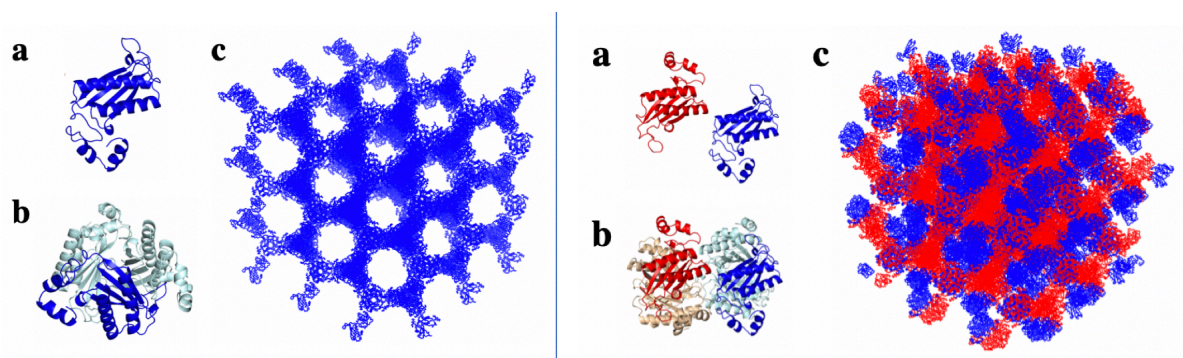


Figure 1. 26 Crystal packing of Cg10062-H28A soaked in propiolate. (left) a one protein chain, b trimer, protein crystal is formed by the helical assembly of the trimer units in three dimensions. c overview of the protein crystal packing when only one chain is being searched during the molecular replacement. (right) a two protein chains, b a hexamer, protein crystal is formed by the assembly of the hexamer units in three dimensions. c overview of the protein crystal packing when two chains are being searched during the molecular replacement.

The benefit of the crystal packing with big interspace is that soaking experiments are more likely to succeed. Another benefit of this packing is that the C-terminus of the protein is ordered. Residues from Pro-1 to G-145 are present in the structure of Cg10062-H28A.

The tertiary structure of Cg10062-H28A is similar to the native Cg10062 structure, except in the C-terminus region. In addition to the description of the structure from the native Cg10062 section, the extended loop continues to 125, and then there are two α -helices (α 4 from residue 125 to 131 and α 5 from residue 134 to 146) that are connected by two amino acids.

Table 1. 6 Data collection and refinement statistics for Cg10062-H28A soaked in propiolate.

Data statistics	
Space group	P4 ₁ 32
No. of chains/ASU	2
Unit cell dimensions (Å, °)	a = 146.04, b = 146.04, c = 146.04 $\alpha = 90, \beta = 90, \gamma = 90$
Resolution range (Å) (outer shell)	35.42 - 2.953 (3.059 - 2.953)
Unique reflections (outer shell)	11335 (1112)
Completeness (%) overall (outer shell)	96.45 (95.33)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	11297 (1083)
Reflections used for R-free	1127 (104)
R-work	0.2122 (0.2846)
R-free	0.3009 (0.4036)
Number of non-hydrogen atoms/solvent/ligands	2302/10/6
Protein residues	286
RMS (bonds) (Å)	0.011
RMS (angles) (Å)	1.47
Ramachandran plot (%) favored/allowed/outliers	90.07/8.87/1.06
Rotamer outliers (%)	0.00
Average B-factor/macromolecules/solvent	73.55/73.55/74.48

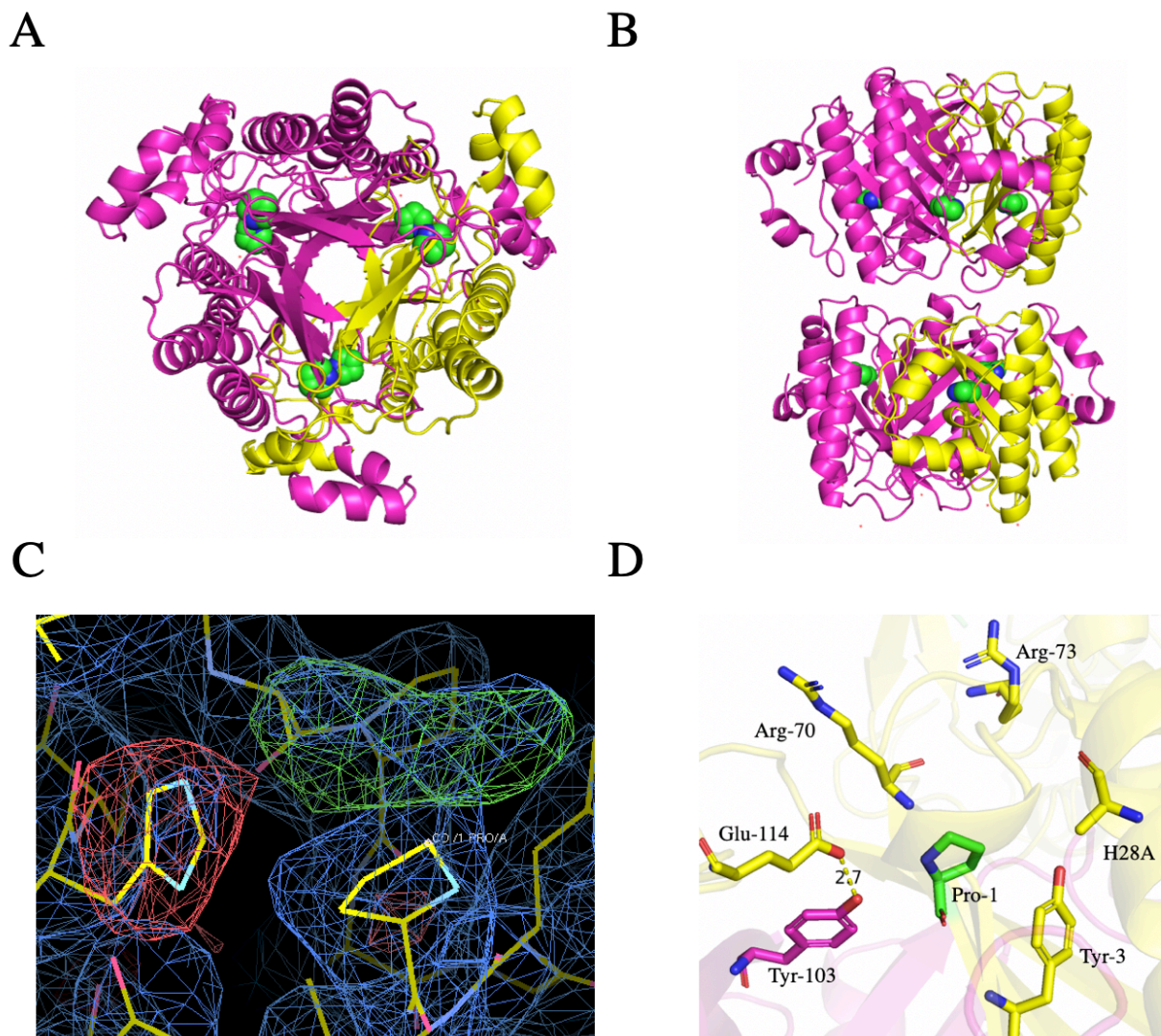


Figure 1. 27 X-ray crystal structure of Cg10062-H28A soaked in propiolate. Six chains of Cg10062-H28A, asymmetric unit (yellow), symmetry mates (pink), active site Pro-1 (green). (A) top view, (B) side view, (C) Pro-1 fitted in the electron density, the negative (red) density for H28, and positive (green) density for a ligand, (contoured at rmsd 1.5 Å) (D) active site of the H28A mutant highlighted by important residues.

After the molecular replacement step, the difference electron density map ($F_{\text{obs}} - F_{\text{calc}}$) presents a negative electron density for His-28 confirming the success of the mutation, and a positive electron density above Pro-1 is indicating that there exists an intermediate in the active site. Also, the difference electron density map ($2F_{\text{obs}} - F_{\text{calc}}$) extends from proline to the positive electron density, so it is more likely that intermediate is covalently bound to proline.

Since the resolution of the electron density map is ~ 3 Å, it is challenging to decide what molecule or intermediate is present in the active site. Later, the positive densities from all the successful soaking experiments are analyzed together to propose ligands or intermediates that are more likely to occupy the positive electron densities. Kinetic studies done in Dr. Draths laboratory show that the H28A mutation has residual activity. (Table 1.2)

1.2.3. X-ray crystal structures of Cg10062-R73A soaked in propiolate

The x-ray crystal structure of Cg10062-R73A (Figure 1.28) was solved to 3 Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.24 and 0.20, respectively (Table 1.7). The crystal packing for this mutation is different than that of native Cg10062.

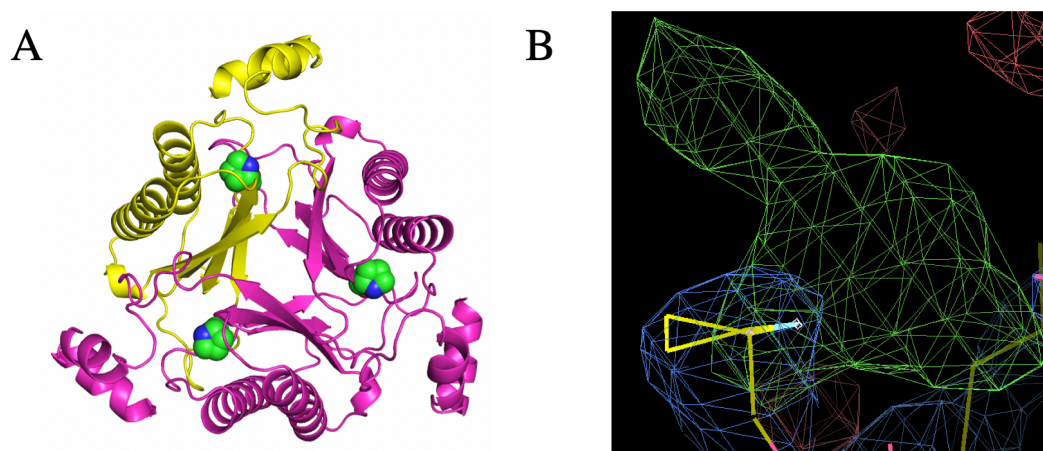


Figure 1. 28 X-ray crystal structure of Cg10062-R73A soaked in propiolate.(A) Trimer of Cg10062-R73A, single chain (yellow) is present in the ASU, symmetry mates (pink). (B) a positive electron density (green) connected to the electron density of Pro-1 (blue). (contoured at rmsd 1.5 Å)

Similar to H28A, the crystal of Cg10062-R73A has $P4_132$ symmetry. However, searching for a single chain was correct in the molecular replacement. There is no residual electron density for a second molecule. The crystal packing is similar to the one shown in Figure 1.26 (left).

Table 1. 7 Data collection and refinement statistics for Cg10062-R73A soaked in propiolate.

Data statistics	
Space group	P4 ₁ 32
No. of chains/ASU	1
Unit cell dimensions (Å, °)	a = 146.55, b = 146.55, c = 146.55 $\alpha = 90, \beta = 90, \gamma = 90$
Resolution range (Å) (outer shell)	35.54 - 2.998 (3.105 - 2.998)
Unique reflections (outer shell)	11178 (1086)
Completeness (%) overall (outer shell)	98.69 (99.91)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	11174 (1085)
Reflections used for R-free	1117 (107)
R-work	0.2035 (0.3566)
R-free	0.2558 (0.4129)
Number of non-hydrogen atoms/solvent/ligands	1181/10/6
Protein residues	144
RMS (bonds) (Å)	0.009
RMS (angles) (Å)	1.29
Ramachandran plot (%) favored/allowed/outliers	95.77/2.82/1.41
Rotamer outliers (%)	0.00
Average B-factor/macromolecules/solvent	78.39/78.18/120.30

Similar to H28A mutants, after the molecular replacement step, the difference electron density map ($F_{\text{obs}}-F_{\text{calc}}$) shows a positive electron density above Pro-1 indicating that there exists an intermediate in the active site. In both, H28A and R73A, the shape of the difference electron density maps ($F_{\text{obs}}-F_{\text{calc}}$) are similar. Cg10062-R73A mutant crystal structure is solved with all 146 residues and a complete C-terminus with two helices.

1.2.4. X-ray crystal structures of Cg10062-Y103F soaked in propionate

The x-ray crystal structure of Cg10062-Y103F was solved to 2.5 Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.22 and 0.21, respectively (Table 1.8). The crystal packing for this mutation is different than that of native Cg10062, but it is similar to the crystal packing of R73A mutant. Similar to H28A and R73A, the crystals of Cg10062-Y103F have $P4_132$ symmetry. Y103F crystal structure has one protein chain in the asymmetric unit and shows a positive electron density connected to Pro-1. The shape of the positive electron density suggesting the presence of a covalent intermediate. Similar to H28A and R73A, structure of Y103F mutant is complete with all 146 residues.

1.2.5. X-ray crystal structures of Cg10062-E114D soaked in propionate

Kinetic studies done in Dr. Draths laboratory and previously published studies⁴³ showed that the E114D mutation would significantly decrease the amount of decarboxylation product. Crystals of Cg10062-E114D mutant were soaked in propionate solutions, and a ligand-bound structure was obtained. The x-ray crystal structure of Cg10062-E114D (Figure 1.29) is solved to 2.20 Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.18 and 0.22, respectively

Table 1. 8 Data collection and refinement statistics for Cg10062-Y103F soaked in propiolate.

Data statistics	
Space group	P4 ₁ 32
No. of chains/ASU	1
Unit cell dimensions (Å, °)	a = 146.68, b = 146.68, c = 146.68 $\alpha = 90, \beta = 90, \gamma = 90$
Resolution range (Å) (outer shell)	33.65 - 2.497 (2.587 - 2.497)
Unique reflections (outer shell)	19303 (1895)
Completeness (%) overall (outer shell)	99.84 (100.00)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	19301 (1895)
Reflections used for R-free	1931 (190)
R-work	0.1974 (0.2819)
R-free	0.2313 (0.3255)
Number of non-hydrogen atoms/solvent/ligands	1272/20/5
Protein residues	146
RMS (bonds) (Å)	0.007
RMS (angles) (Å)	1.20
Ramachandran plot (%) favored/allowed/outliers	96.53/2.78/0.69
Rotamer outliers (%)	2.96
Average B-factor/macromolecules/solvent	60.40/60.18/91.22

(Table 1.9). The asymmetric unit contains twelve protein chains and the monomers in the asymmetric unit superimpose with rmsd values of ~ 0.134 Å on average. All residues from Pro-1 to position 121 are fit into the electron density but there was not any electron density for residues from 122 to the C-terminus. The two α -helices in the C-terminus of the protein are not included due to lack of electron density. The tertiary structure of the trimer of Cg10062-E114D is similar to the native Cg10062 and the average rmsd of (four trimers against native trimer) is 0.26 Å.

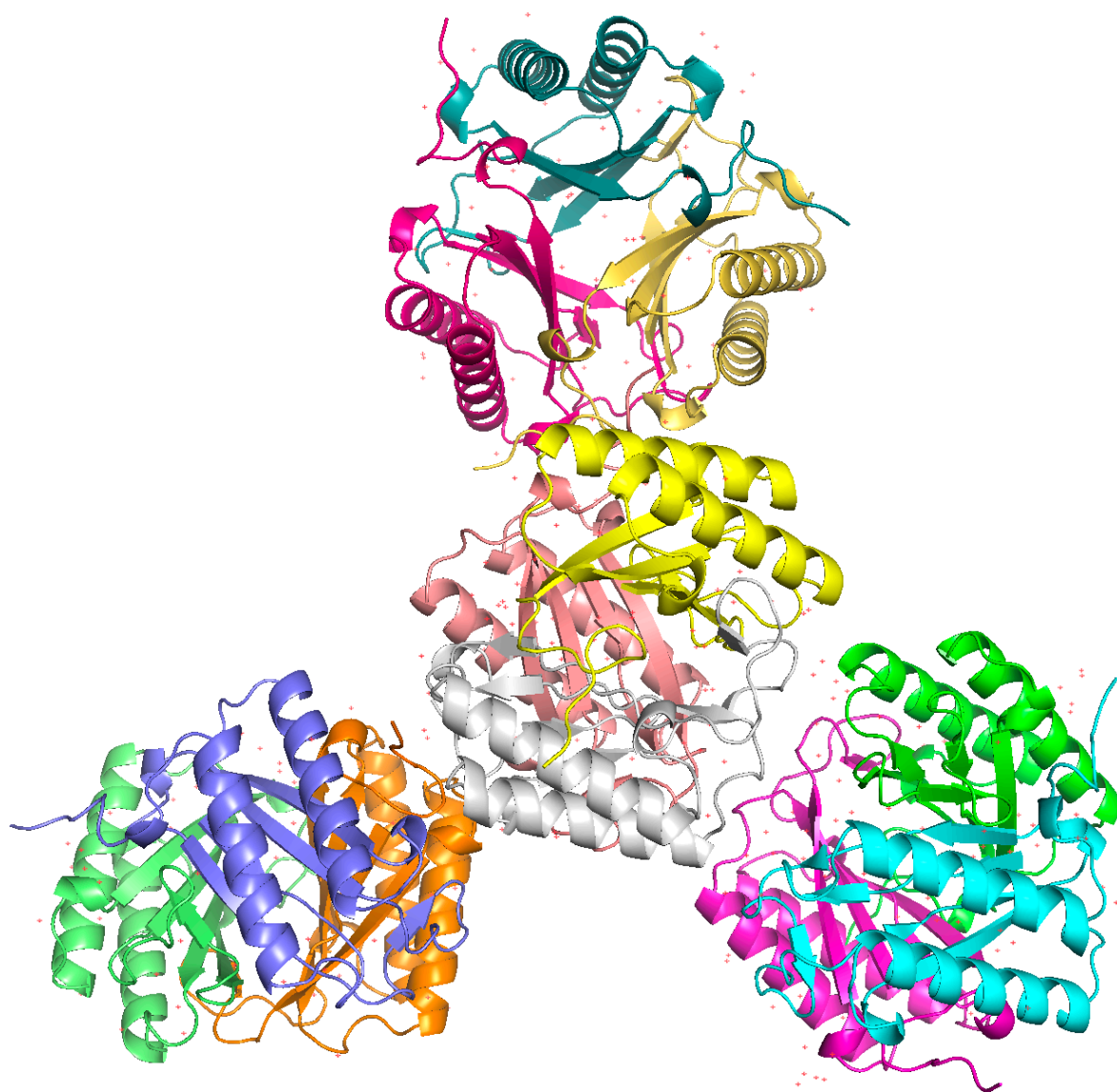


Figure 1. 29 Asymmetric unit of Cg10062-E114D mutant crystal with twelve chains in four trimers.

Table 1. 9 Data collection and refinement statistics for Cg10062-E114D soaked in propiolate.

Data statistics	
Space group	P 21 21 21
No. of chains/ASU	12
Unit cell dimensions (Å, °)	a = 107.319, b = 146.299, c = 146.41 $\alpha = 90, \beta = 90, \gamma = 90$
Resolution range (Å) (outer shell)	35.04 – 2.20 (2.28 – 2.20)
Unique reflections (outer shell)	116532 (11226)
Completeness (%) overall (outer shell)	99.47 (96.15)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	116462 (11167)
Reflections used for R-free	2004 (193)
R-work	0.1705 (0.2755)
R-free	0.2204 (0.2981)
Number of non-hydrogen atoms/solvent/ligands	12538/259/106
Protein residues	1452
RMS (bonds) (Å)	0.010
RMS (angles) (Å)	1.06
Ramachandran plot (%) favored/allowed/outliers	96.01/2.87/1.12
Rotamer outliers (%)	1.22
Average B-factor/macromolecules/solvent	49.12/49.03/53.75

All twelve chains in the ASU show a significant positive electron density ($F_{\text{obs}} - F_{\text{calc}}$) above the Pro-1 (Figure 1.30). The electron density is connected to the electron density of Pro-1, suggesting the presence of a covalent intermediate. This positive electron density is extended toward the space between E114D and Pro-1 and its shape suggesting the presence of an ordered water molecule between the two amino acids. In structures of the native Cg10062 and all other mutants, Glu114 occupies this space. However, mutating Glu-114 to aspartic acid creates the space for a water molecule to occupy.

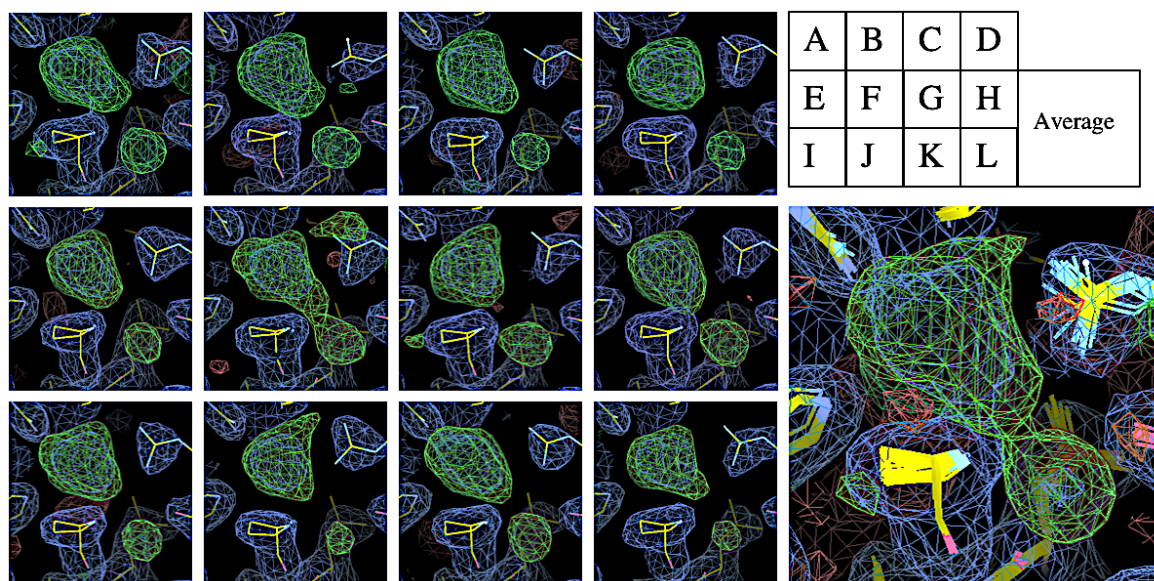


Figure 1. 30 Different shape of electron density for twelve chains in E114D ASU. (A) – (L) are twelve chains of Cg10062-E114D that make up the ASU. 2FoFc (blue), FoFc (green). (contoured at rmsd 1.5 Å)

1.2.6. Disordered C-terminus and the arginine residues in the active site

As mentioned earlier, the structures of native Cg10062 and E114D mutants lack the electron densities for the c-terminal residues (from 122 to 148). However, previously published structures of Cg10062 (PDBID: 3n4g, 3n4h, 3n4d), H28A, R73A, and Y103F structures show an ordered C-terminus. Overlaying structures of native, E114D, H28A, R73A, and Y103F reveals a new salt

bridge formation between E114D and Arg-117. In all structures of Cg10062 with an ordered C-terminus, and in one native structure with a disordered C-terminus, Arg-117 is facing the solvent, rotated away from residue 114. Mutating Glu-114 to aspartic acid allows for the formation of a salt bridge between Arg-117 and Asp-114 and as a result the amino acids following Asp-114 adapt a different trajectory (see Figure 1.31)

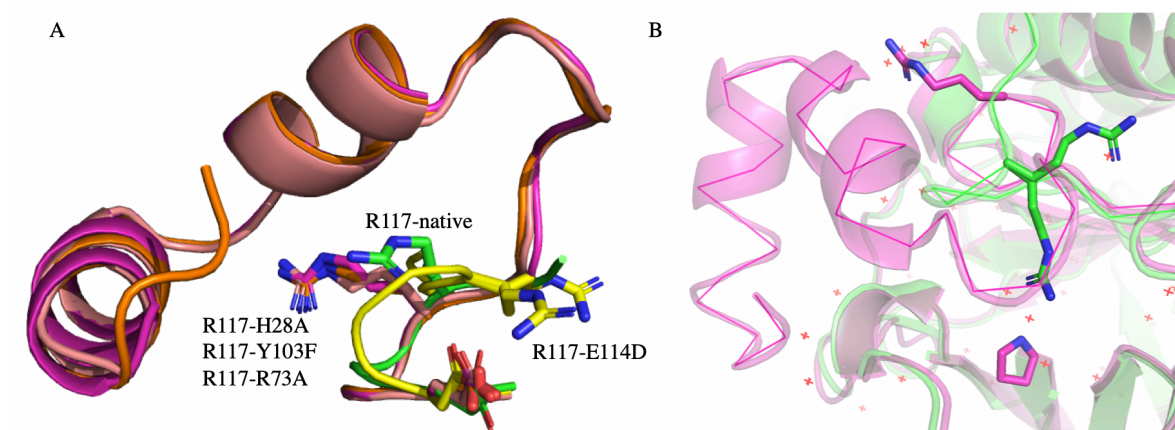


Figure 1. 31 Difference in C-termini of native and mutants. (A) Native (green), E114D (yellow), H28A, Y103F, R73A (shades of red). The C-terminus of native Cg10062 follows the trajectory of H28A, Y103F, and R73A. E114D deviates from the native structure. (B) Arg-117 has two conformations in all E114D structures.

Biochemically, the E114D mutant is kinetically slower than wildtype, but it does not give the decarboxylation product. The formation of the salt bridge allows Arg-117 to enter the active site but the extent of its effect on the catalytic process is unknown.

The active site of Cg10062 contains two arginine residues. The presence of these positively charged residues is critical for the binding of the negatively charged substrate molecules. In the case of the native Cg10062, Arg-73 is refined with two conformations, one facing the active site Pro-1, located adjacent to His-28, and the other one facing away from the active site. However, Arg-73 has only one conformation for H28A and Y103F mutants, and it is facing away from the active site. In the structure of the E114D mutant, Arg-73 is only facing the active site. The

trajectory of Arg-70 is similar for the structures of native Cg10062, H28A, and R73A mutants, and it has a unique trajectory for E114D mutants. Figure 1.32 shows the difference in the trajectory of various arginine residues in the active site of all these variants.

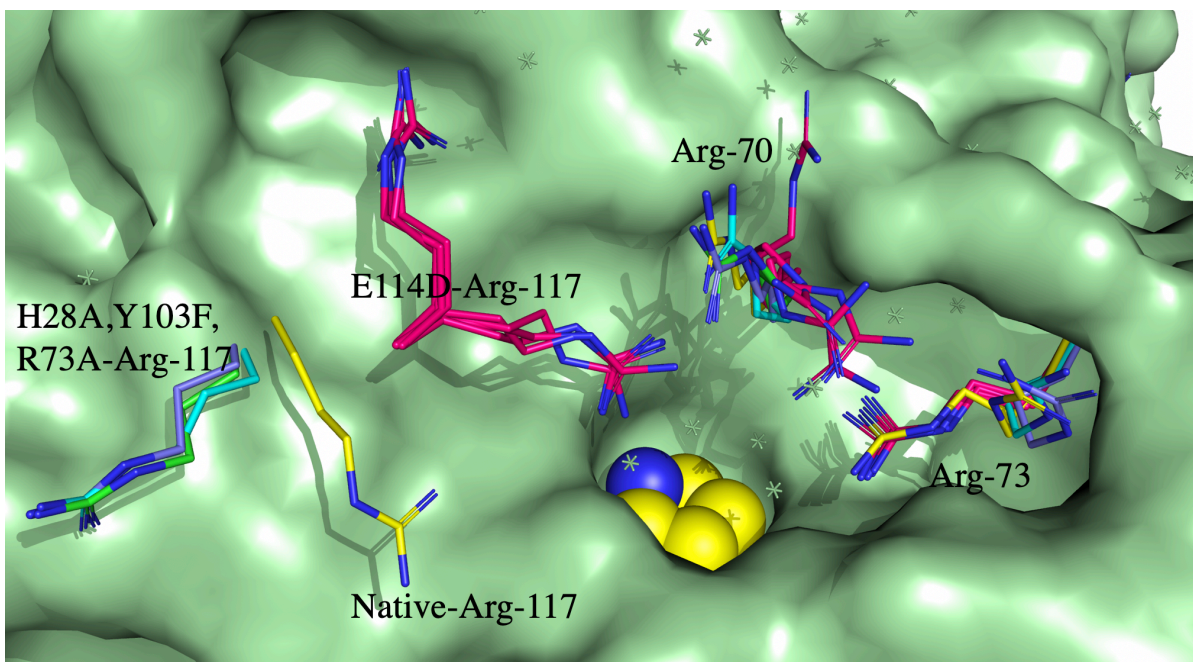


Figure 1. 32 Arginine residues in the active site. native Cg10062 (yellow), E114D (red), R73A (green), H28A (light blue), Y103F (purple).

1.2.7. Comparison of ligand electron densities

Resolutions of diffraction data for the ligand-bound mutants of E114D, Y103F, H28A, and R73A are 2.2, 2.5, 2.95, and 3.0 Å respectively. Careful observation of the shapes of these positive electron densities in the difference electron density map ($F_{\text{obs}} - F_{\text{calc}}$) suggests that similar intermediates might occupy them in H28A and R73A since in both cases it extends in both directions (assuming the nitrogen of the Pro-1 in the center, Figure 1.33). However, in the cases of E114D and Y103F, the majority of the density resides in only one side of the Pro-1 (ignoring the density for ordered water in E114D).

Matching the electron densities to the shape of possible intermediates (based on the proposed model for the mechanism), 3-hydroxy-prolylpropionate **3** is selected to complete the structures of H28A and R73A mutants (3-hydroxy-prolylpropionate is the intermediate after the hydration step, Figure 1.34). On the other hand, since E114D and Y103F structures do not show a significant electron density for a hydroxyl group, trans-3-prolylacrylate **1** was first built into the positive density of these mutants. Though the positive electron densities for all the ASU chains of E114D match the shape of trans-3-prolylacrylate, the positive electron density of the Y103F mutant the iminium tautomer **2**, because the torsion angles of this intermediate, where the double bond is between the prolyl nitrogen and C-1, is more consistent with the density observed.

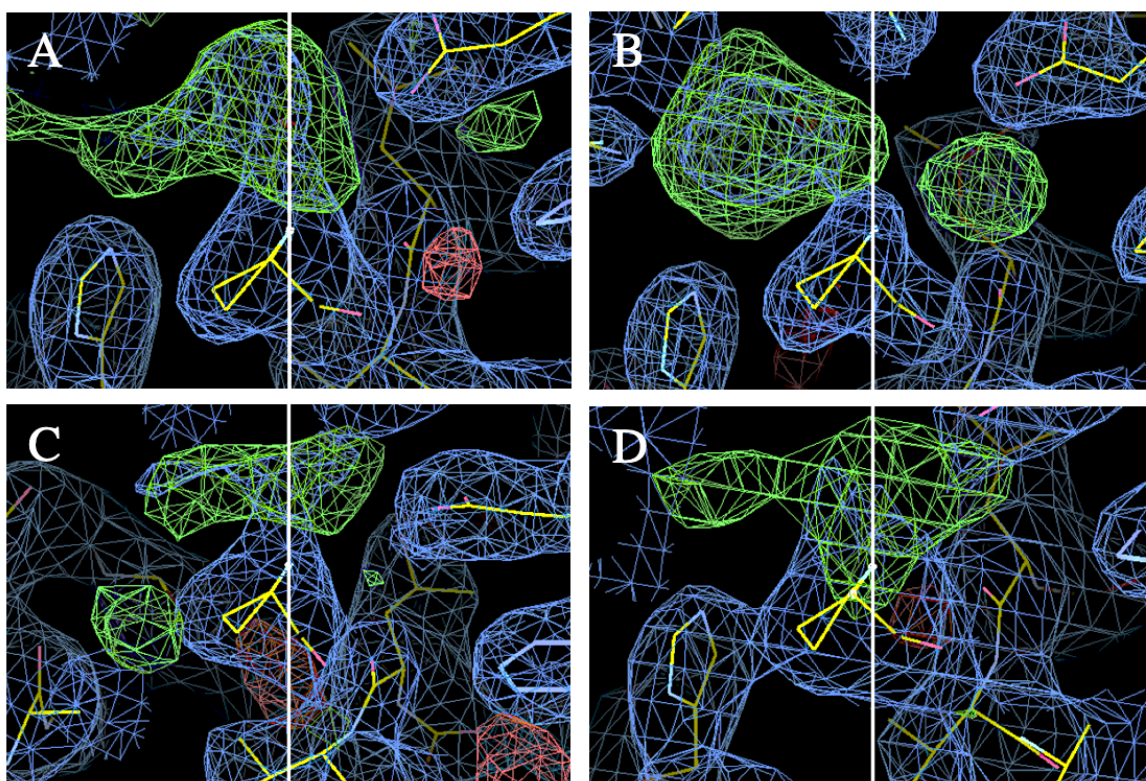


Figure 1. 33 Comparing the shape of the difference electron density map ($F_{\text{obs}} - F_{\text{calc}}$). (A) Y103F, (B) E114D, (C) H28A, and (D) R73A. (contoured at rmsd 1.5 Å)

Refining the structure of H28A with 3-hydroxy-3-prolylpropionate removes the positive electron density from the difference electron density map ($F_{\text{obs}} - F_{\text{calc}}$) (Figure 1.35). The resulting

carbon-nitrogen bond length is 1.41Å. Similarly, building 3-hydroxy-3-prolylpropionate (carbon-nitrogen bond length is 1.39Å) in the active site of the R73A removes the positive electron density from the difference electron density map ($F_{\text{obs}}-F_{\text{calc}}$) (Figure 1.36). Biochemical characterization of these mutants shows nearly 1000-fold reduced activity for H28A and no activity for R73A (data from Ms Amaya Sirinimal in Dr. Karen Draths laboratory). His-28 and Arg-73 are conserved residues that are expected to play significant roles in the binding of the substrate in the cis-CaaD family. The high concentration of the substrate in the substrate-soaking procedure (2mM) could lead to the formation of these intermediates. Also, though turnover may be super slow, that does not mean intermediates do not form reasonable fast, they just do not go to product very rapidly.

The x-ray crystal structure of Y103F mutants is refined by building both trans-3-prolylacrylate and its iminium tautomer. Refining the structure with the iminium form adds an ordered water molecule to the active site as well (Figure 1.37).

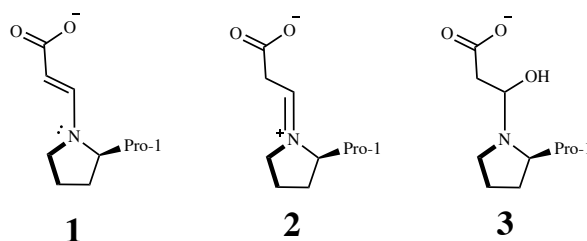


Figure 1. 34 Covalent intermediates. 3-prolylacrylate 1, iminium tautomer of 3-prolylacrylate 2, 3-hydroxy-3-prolylpropionate 3.

The enamine form contains a carbon-nitrogen single bond of length 1.39Å, and the carbon-nitrogen double bond in the iminium form is 1.31Å. Building trans-3-prolylacrylate for all twelve chains in the ASU of the E114D structure leaves a significant positive electron density in the difference electron density map ($F_{\text{obs}}-F_{\text{calc}}$) (Figure 1.38). The new positive electron density seems to indicate the presence of another ligand in the active site.

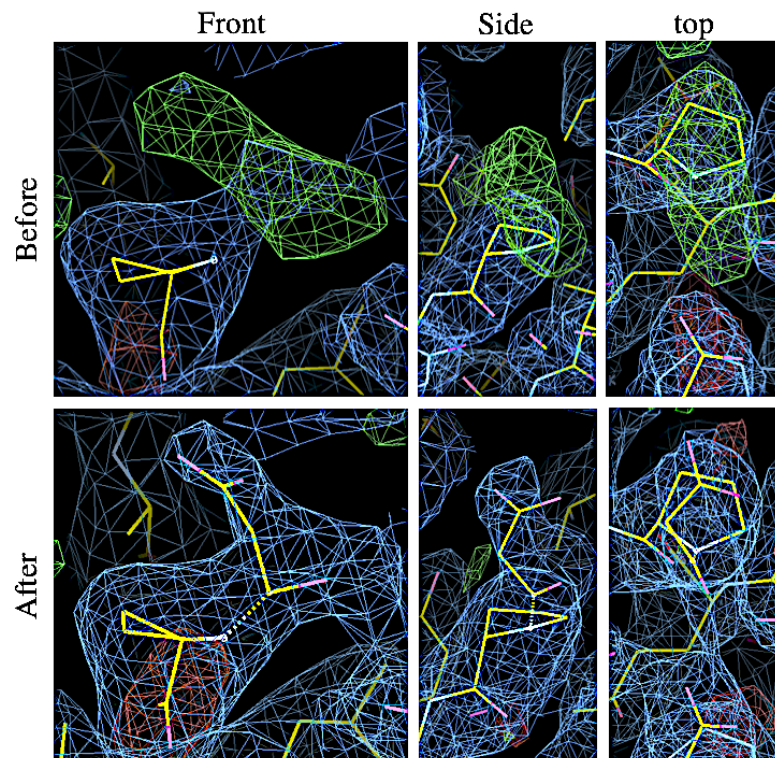


Figure 1. 35 Building 3-hydroxy-3-prolylpropionate in the active site of H28A mutants.
(contoured at rmsd 1.5 Å)

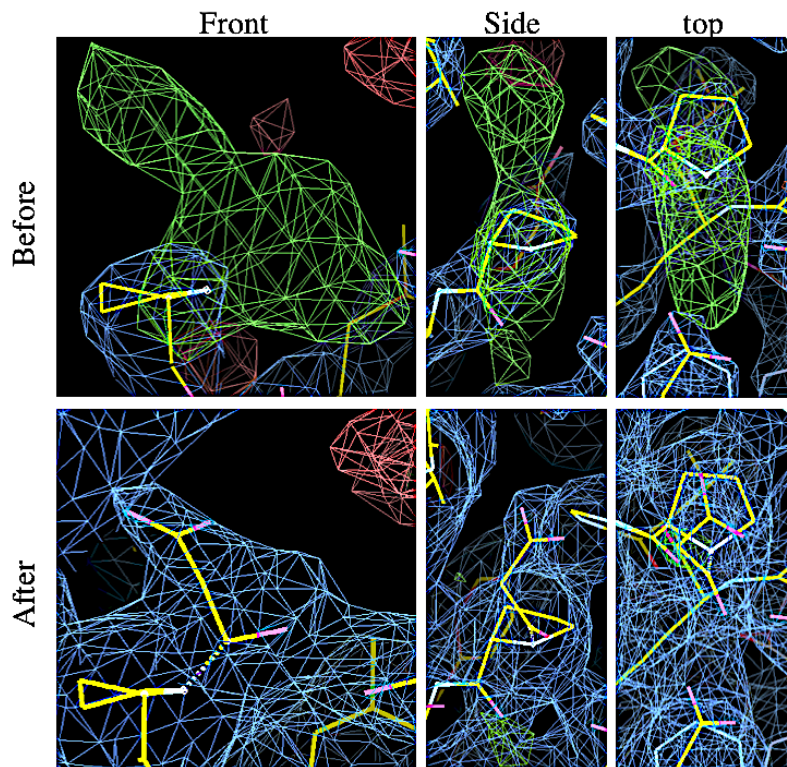


Figure 1. 36 Building 3-hydroxy-3-prolylpropionate in the active site of R73A mutants.
(contoured at rmsd 1.5 Å)

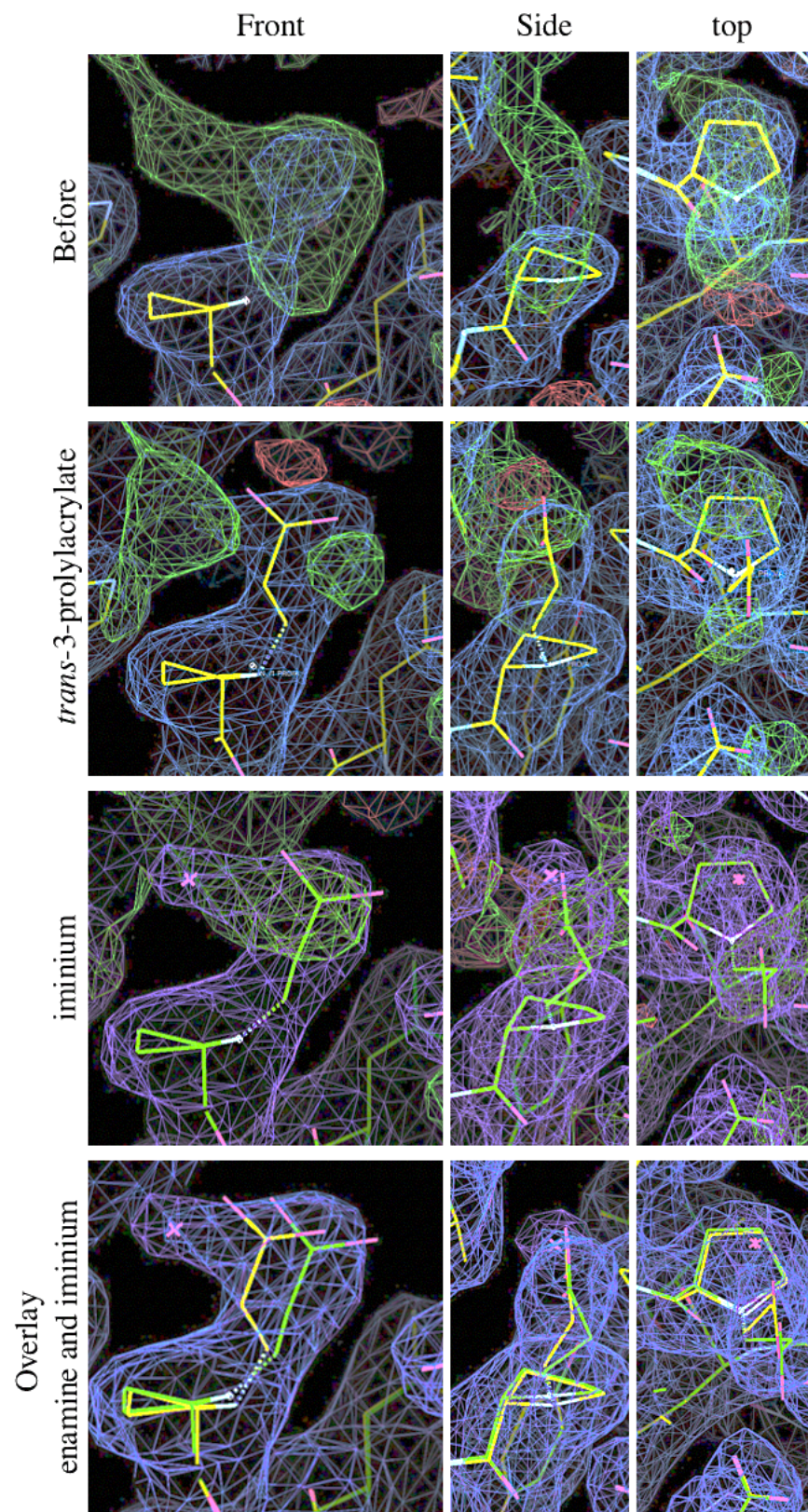


Figure 1. 37 *trans*-3-Prolylactylate in the active site of Y103F mutants in the form of enamine and iminium. (contoured at rmsd 1.5 Å), $2F_{\text{obs}}-F_{\text{calc}}$ (blue, pink, purple). $F_{\text{obs}}-F_{\text{calc}}$ (green)

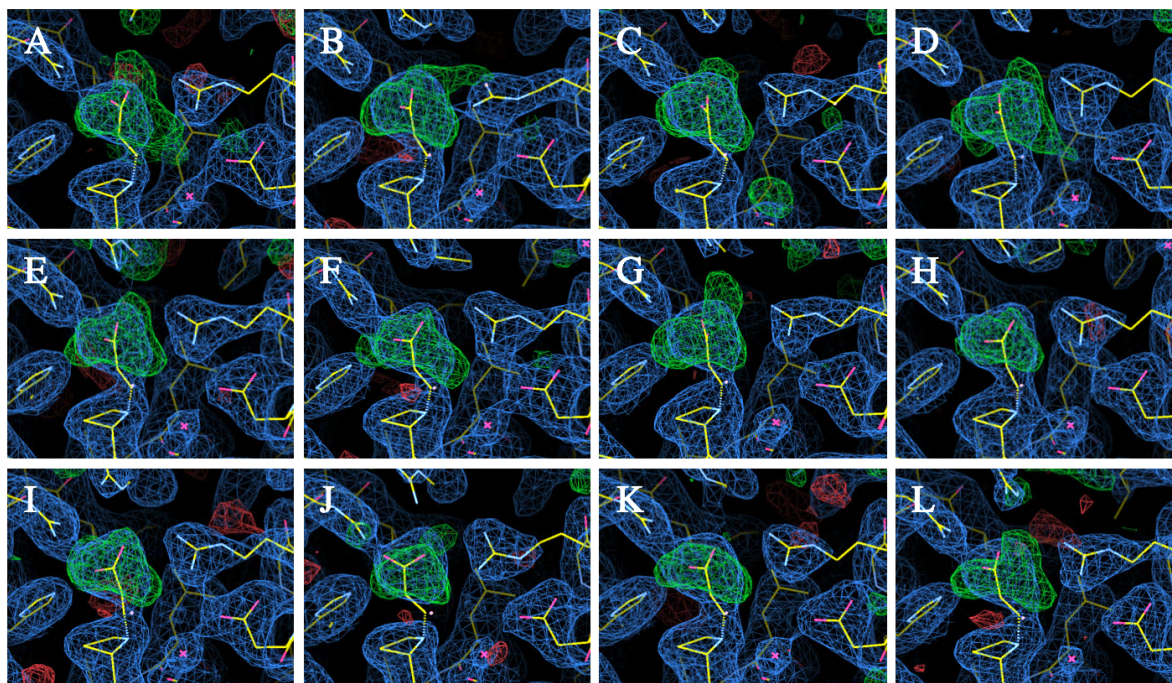


Figure 1. 38 trans-3-prolylacrylate in the active site of E114D mutants and the appearance of new positive electron density. (contoured at rmsd 1.5 Å), $2F_{\text{obs}} - F_{\text{calc}}$ (blue). $F_{\text{obs}} - F_{\text{calc}}$ (green)

The shape of the new electron density is ambiguous for individual chains, but when it is averaged over all chains, its shape suggested the presence of malonate semialdehyde. In order to confirm the presence of malonate semialdehyde in the active site, structures were refined with the two conformers of malonate semialdehyde. The positive electron density appeared after the refinement resembles the shape of the covalently bound 3-prolylacrylate intermediate (Figure 1.39 bottom row) (on average). All the chains in the ASU of the E114 crystal structure contain a different ratio of 3-prolylacrylate (enamine form) and malonate semialdehyde.

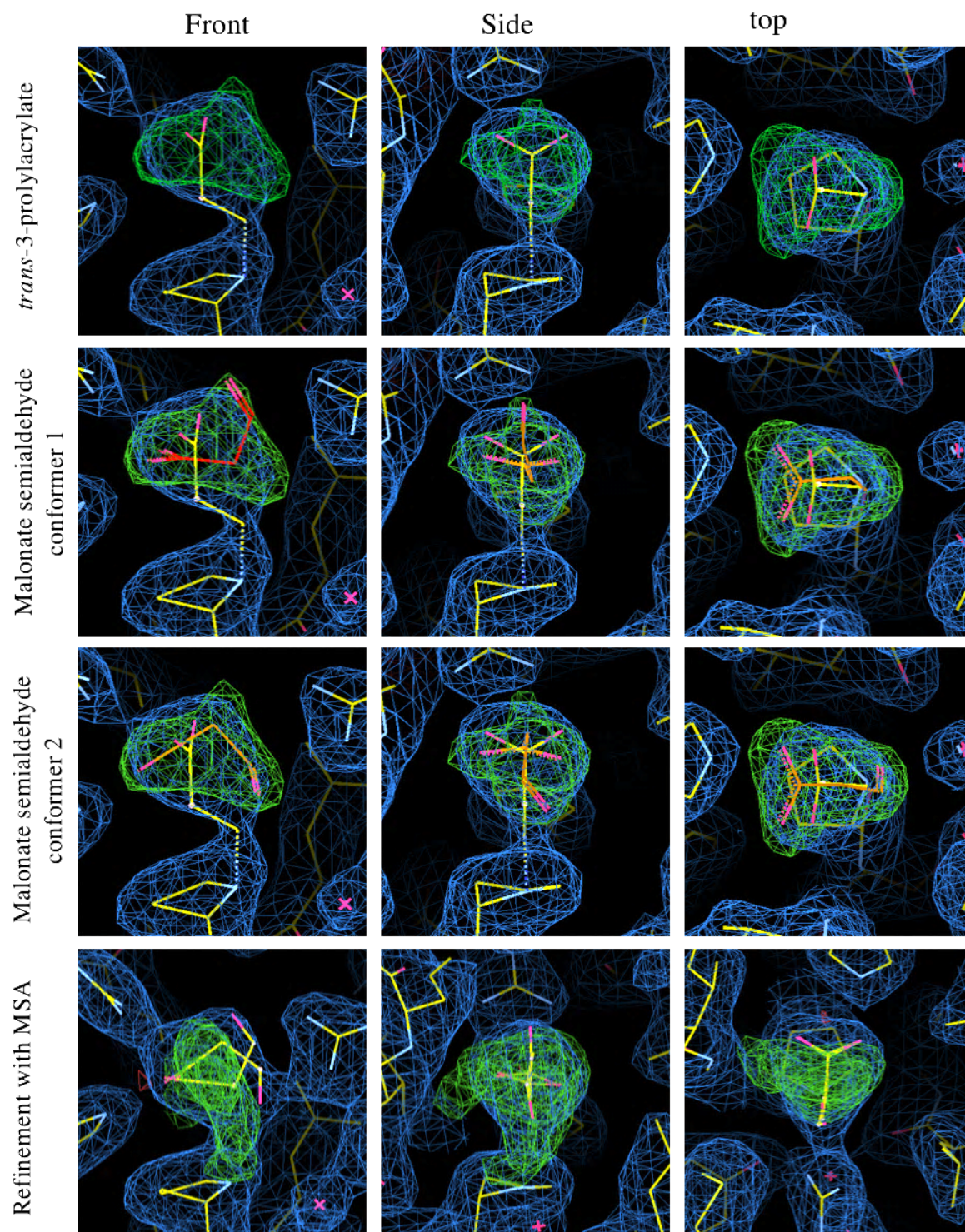


Figure 1. 39 3-prolylacrylate and malonate semialdehyde are both present in the active site of E114D mutants. (contoured at rmsd 1.5 Å), $2F_{\text{obs}}-F_{\text{calc}}$ (blue). $F_{\text{obs}}-F_{\text{calc}}$ (green)

1.2.8. Structures of native Cg10062 and its ligand-bound mutants

In summary, the structures of H28A and R73A mutants are more likely to contain the 3-hydroxy-3-prolylpropionate as an intermediate while the active site of Y103F and E114D mutants contain the 3-prolylacrylate, with the shape of the positive electron density in the Y103F most consistent with the iminium tautomer. Electron density for malonate semialdehyde, the final product of the reaction, is indicated in the active site of E114D.

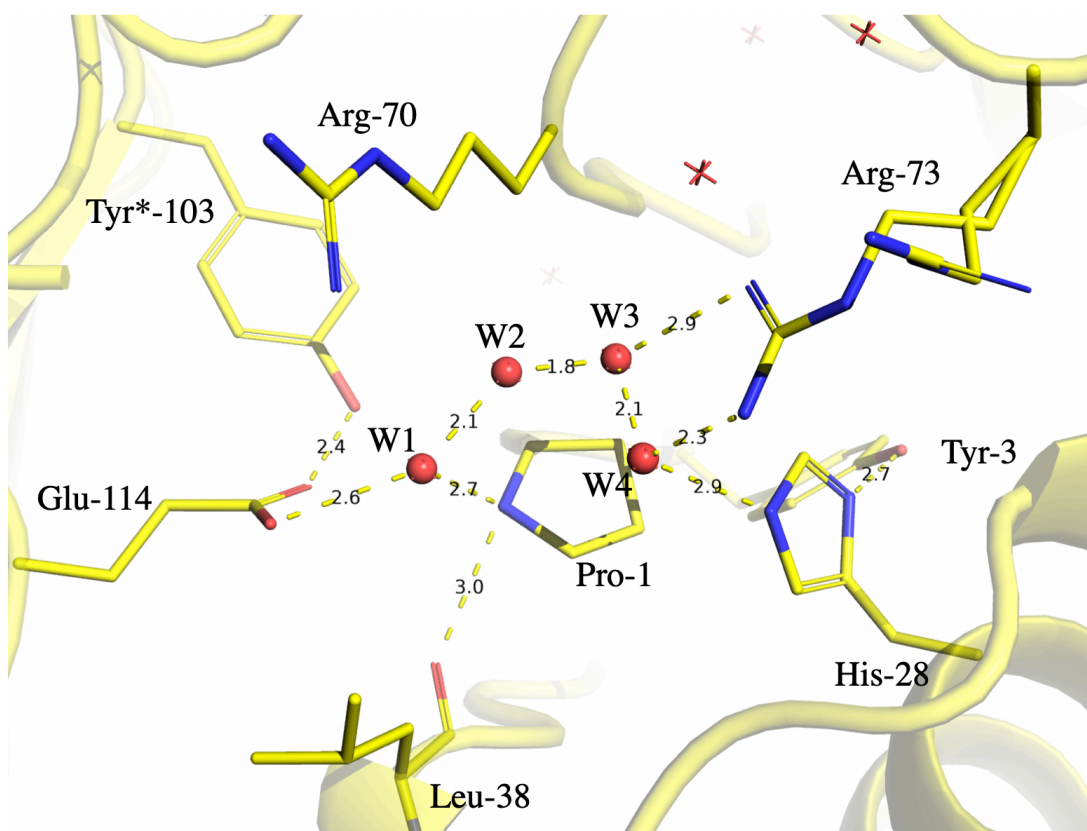


Figure 1. 40 Ordered water molecules in the active site of native Cg10062.

Investigating the active site of native Cg10062 shows four ordered water molecules (W1, W2, W3, and W4 in Figure 1.40). In cis-CaaD, Glu-114 and Tyr*-103 are identified as active site residues involved in water activation. The W1 water in the active site of native Cg10062 forms a hydrogen bond with Glu-114 and is also interacting with Tyr*-103 via Glu-114. It is possible that

W1 could be the water molecule that is activated for the hydration step. Arg-73 adopts two conformations wherein one form forms hydrogen bonds with W3 and W4 while in another form, it is facing away from the active site.

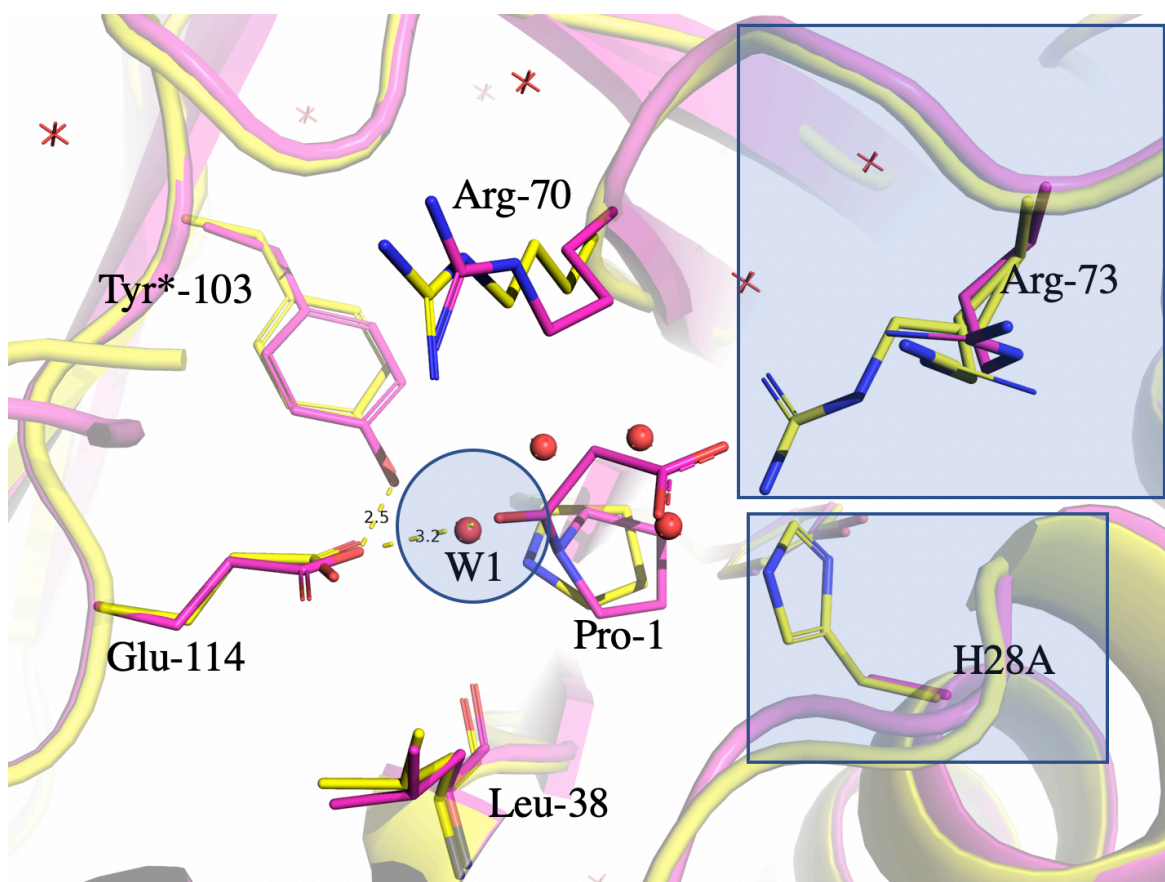


Figure 1. 41 Overlay the active site residues for H28A and native Cg10062. The hydroxyl group of 3-hydroxy-3-prolylpropionate in the structure of H28A and W-1 in native Cg10062 are nearly overlapping. Native Cg10062 (yellow), H28A mutant (pink).

Overlaying the active site residues of H28A (Figure 1.41) and R73A (Figure 1.42) to the native Cg10062 structure shows that the hydroxyl group from 3-hydroxy-3-prolylpropionate and wat-1 nearly overlap. Arg-73 in H28A is facing away from the active site and do not interact with the substrate. As mentioned before, His-28 and Arg-73 are proposed to be involved in substrate binding. Upon mutating His-28 to alanine, Arg-73 adopts a conformation where it is not interacting with the 3-hydroxy-3-prolylpropionate. Similarly, overlaying the active site residues from R73A

on the structure of the native Cg10062 shows similar interactions. The His-28 in the active site is more than 3.5 Å away from the carboxylic group of 3-hydroxy-3-prolylpropionate.

Arg-73 in the active site of H28A mutants and His-28 in the active site of R73A mutants do not interact with the carboxylic group of the 3-hydroxy-3-prolylpropionate. Therefore, either their substrate-binding role is critical before the hydration step, or it is a collaborative role. i.e., mutating one results in the loss of their interactions with the carboxylic group of the 3-hydroxy-3-prolylpropionate. Figure 1.43 shows the difference between the active site structures of H28A and R73A.

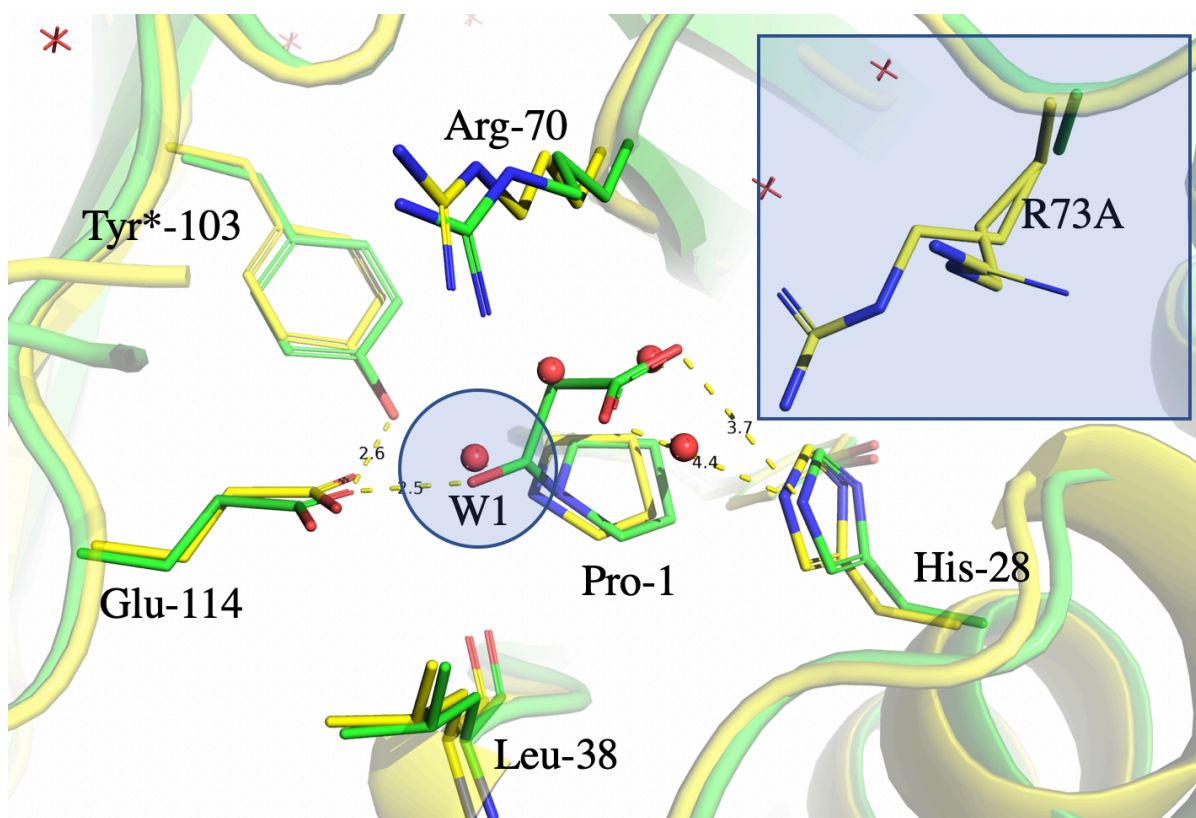


Figure 1. 42 Overlay of active site residues for R73A and native Cg10062. The hydroxyl group of 3-hydroxy-3-prolylpropionate in the structure of R73A and W1 water molecule in native Cg10062 structure nearly overlap. Native Cg10062 (yellow), R73A mutant (pink).

The shape of the electron density in the active site of Y103F is modeled with both enamine and iminium forms of 3-prolylacrylate. The Tyr*-103 is proposed to be involved in the water

activation for the hydration step. Therefore, in the absence of Tyr*-103, the hydration process would be slower, and the chance of observing 3-prolylacrylate in the active site will increase. The active site of the Y103F mutant is the only case where 3-prolylacrylate in the form of iminium could fit into the electron density, satisfying the planer geometry of the carbon-nitrogen double bond (Figure 1.44). Fitting an iminium form instead of the enamine form introduces an ordered water molecule in the active site, 2.2 Å away from the carboxylic group of 3-prolylacrylate. The role or the importance of this water molecule is unknown.

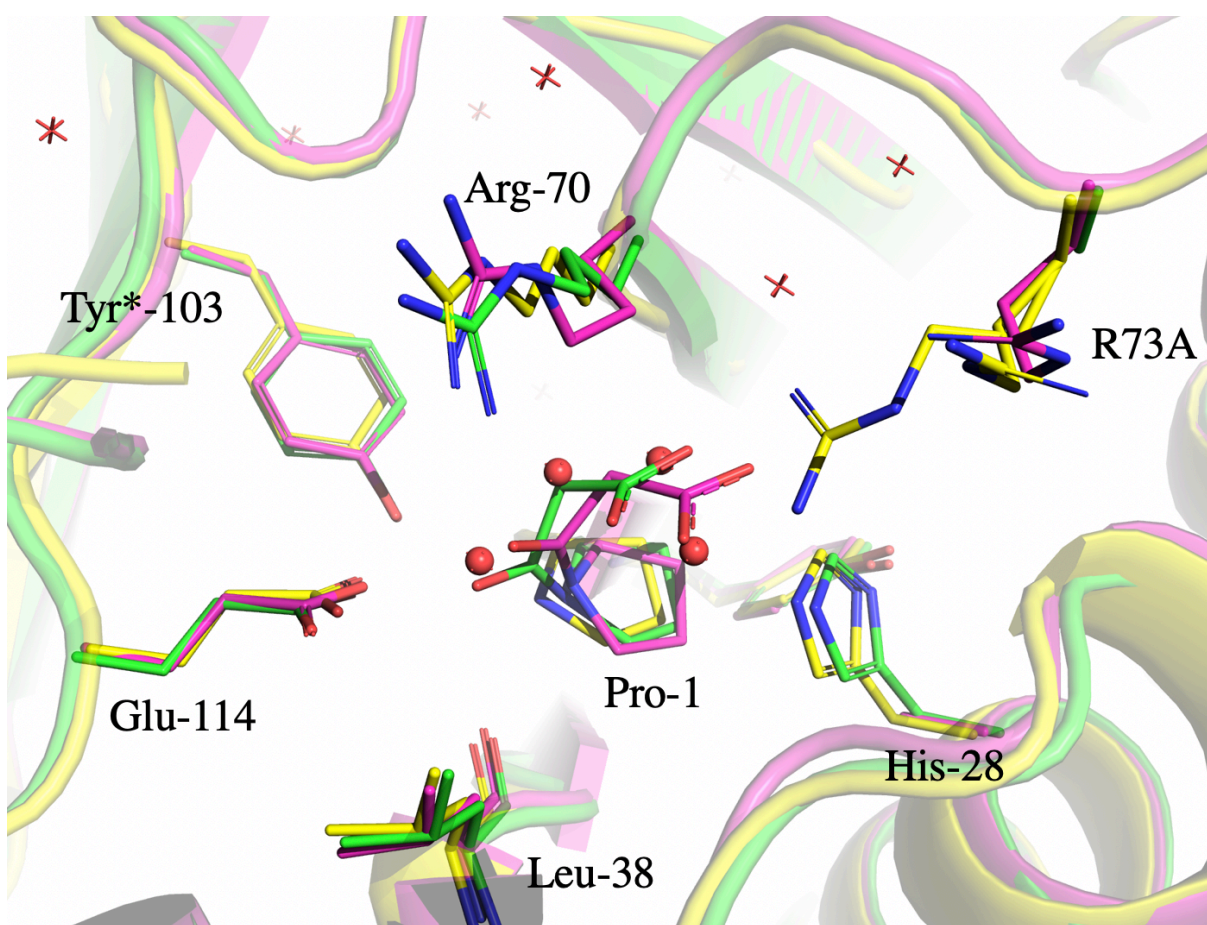


Figure 1. 43 The difference between the active sites of native Cg10062, H28A and R73A mutants.

In order to evaluate the likelihood of enamine vs. iminium forms, other active site residues were compared between Y103F and E114D. As discussed previously, in the crystal structure of E114D, the active site contains the enamine form and malonate semialdehyde at different ratios.

In E114D structures, Arg-73 is facing toward the active site. However, the Arg-73 trajectory in the Y103F active site is similar to H28A, facing away from the substrate. Because of the difference in trajectories of Arg-73 residues between Y103F and E114D, the iminium form is more likely to be the intermediate in the active site of the Y103F mutant.

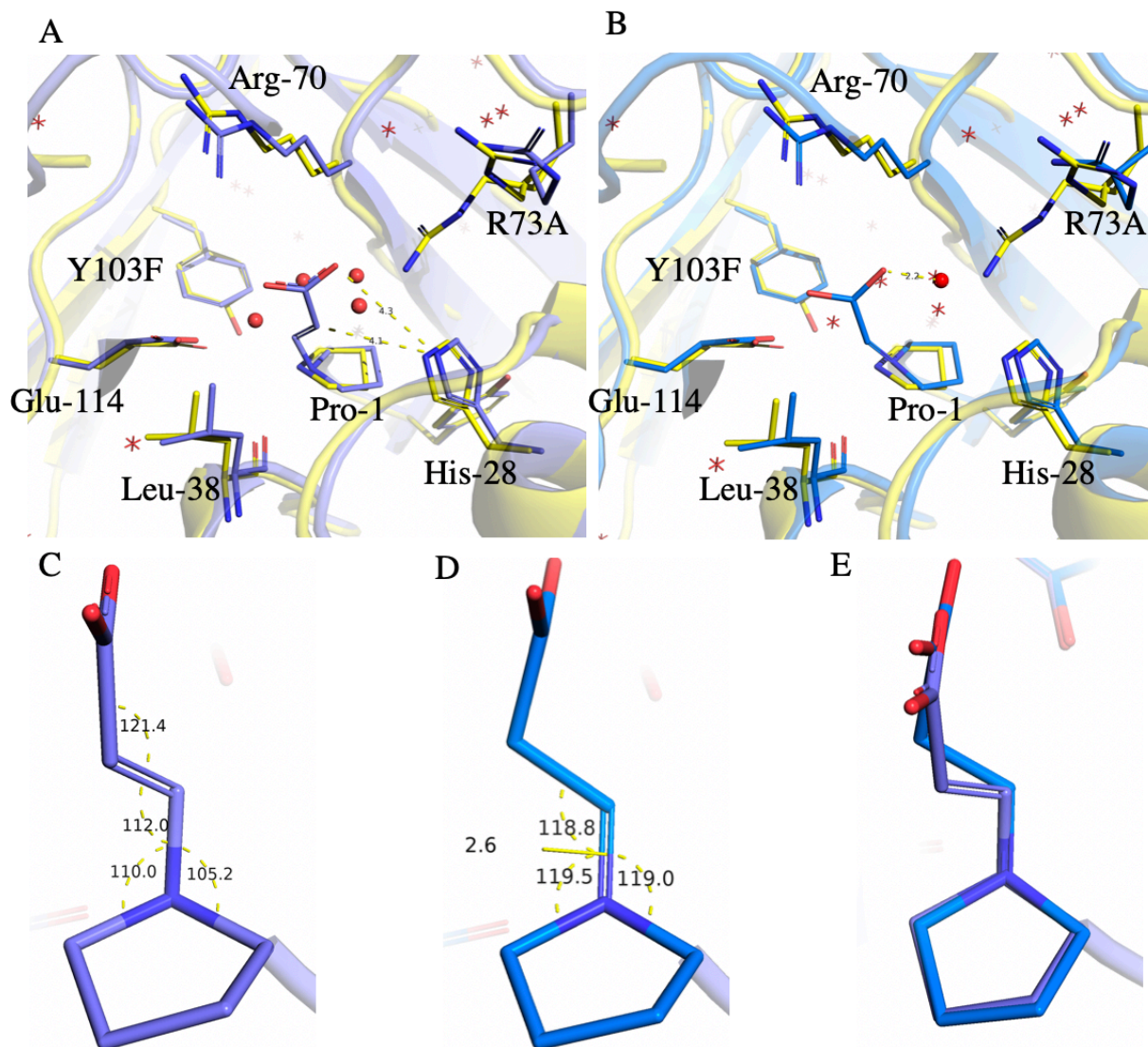


Figure 1. 44 3-Prolylacrylate in the active site of Y103F mutants. (A) enamine form overlaid on the native Cg10062, ordered water molecules in the native Cg10062 active site (red spheres), (B) iminium form overlaid on the native Cg10062, ordered water molecule in the Y103F active site (red sphere), ordered water molecules in the native Cg10062 active site (red crosses) (C) geometry of enamine form, (D) geometry of iminium form (dihedral angle is 2.6°), (E) enamine and iminium form of 3-prolylacrylate.

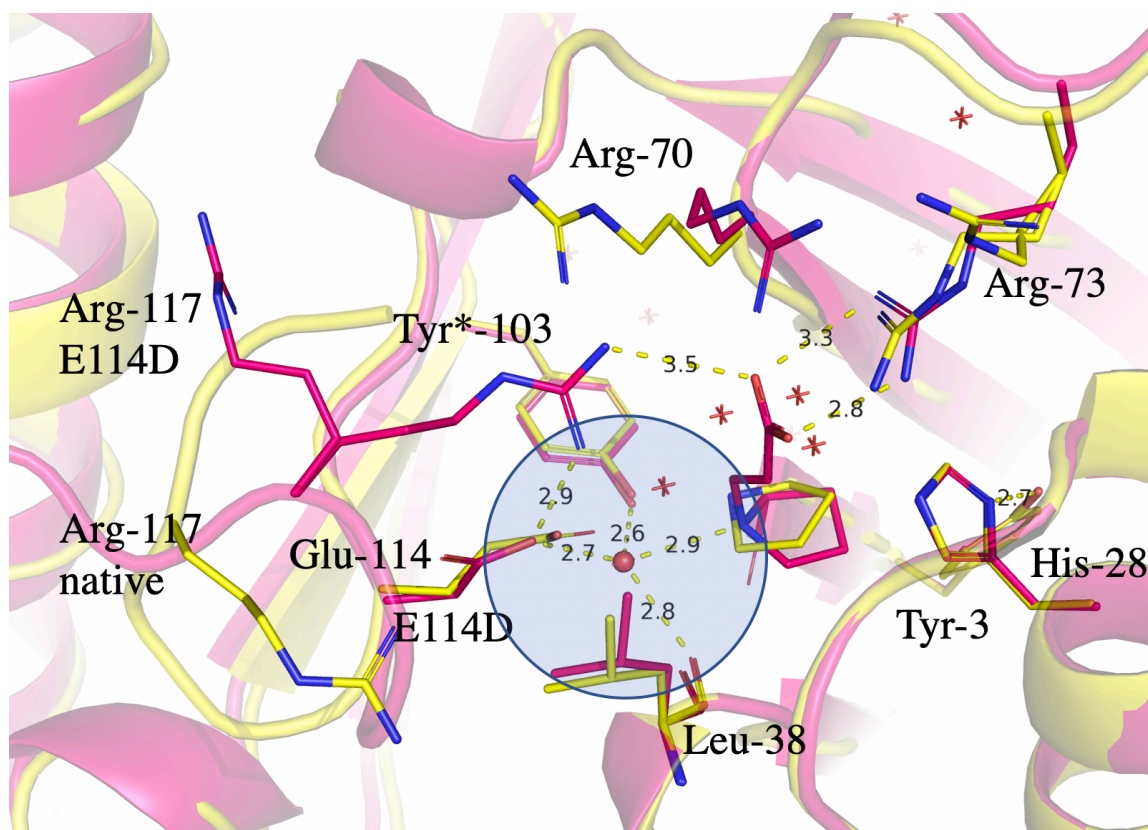


Figure 1. 45 The active site of E114D mutants overlaid on the structure of native Cg10062. The ordered water molecule in E114D active site (red sphere), ordered water molecules in native Cg10062 active site (red crosses)

Comparing the active site of the E114D mutants and native Cg10062 results in the following observations (Figure 1.45). (1) Arg-73 only has one trajectory; it is facing the substrate and forming hydrogen bonds (electrostatic interaction) with the carboxylate group of 3-prolylacrylate. (2) Arg-70 has a different trajectory when the native Cg10062 and other mutants are compared. (3) Arg-117, unlike any other mutants, has two conformations, it is facing toward the active site and forms a salt bridge with Asp-114 (E114D). (4) Aspartate is one carbon shorter than glutamate, and in the E114D active site, one ordered water molecule (Figure 1.46) is localized between the Pro-1 nitrogen and Asp-114. The biochemical characterization of E114D mutants does not show any decarboxylation product (see Table 1.2). The smaller size of aspartate vs. glutamate creates

space for the ordered water molecule. This water molecule acts as a hydrogen donor to form hydrogen bonds with aspartate and the backbone carbonyl group of Leu-38. However, it is a hydrogen bond acceptor with the Pro-1 nitrogen and Tyr*-103 hydroxyl group. The nitrogen of Pro-1, because of the presence of the ordered water molecule, stays protonated, and fails to form the iminium tautomer.

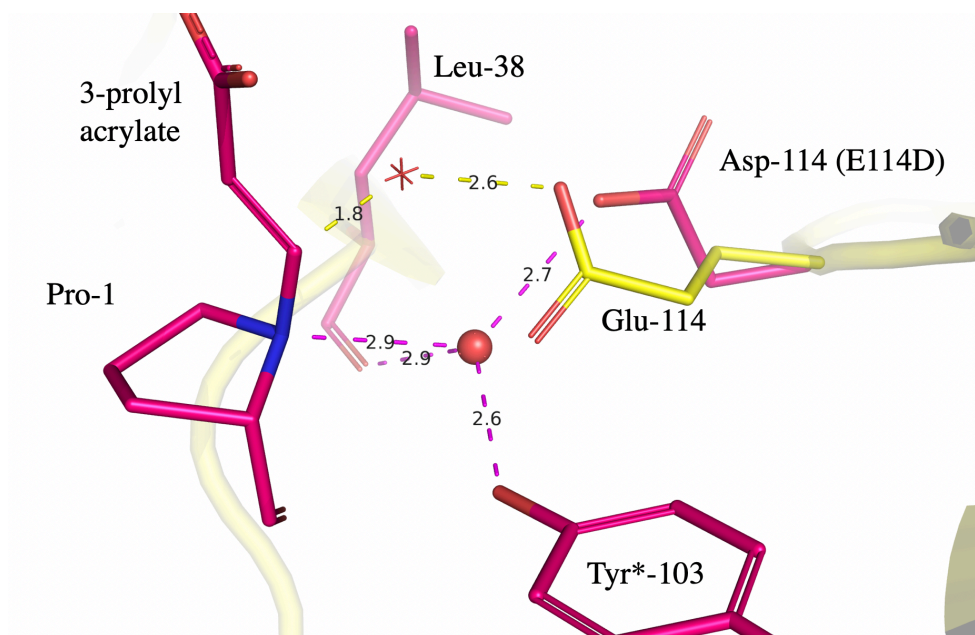


Figure 1. 46 The ordered water molecule in E114D active site. Ordered water molecule (W4) from native Cg10062 (red cross). E114D (magenta), Native Cg10062 (yellow).

As discussed earlier, decarboxylation is thought to be driven by the formation of the iminium species, recapitulating a mechanism common in other decarboxylases. In these enzymes, the cofactor pyridoxal phosphate (PLP) forms a Schiff base with an amino acid, creating the electron withdrawing group required to catalyze the decarboxylation.^{47,48} In E114D mutants where the Pro-1 nitrogen stays protonated and fails to form the iminium tautomer, decarboxylation is prevented. Although it is not clear whether this water molecule is the one to be activated for the hydration step or not, it is clear that this water molecule is preventing the formation of the iminium tautomer

and stopping the decarboxylation reaction. Notably, this water is by far the closest to the C-1 carbon of the intermediate in the structure, suggesting that it may be the water operant in the hydration reaction.

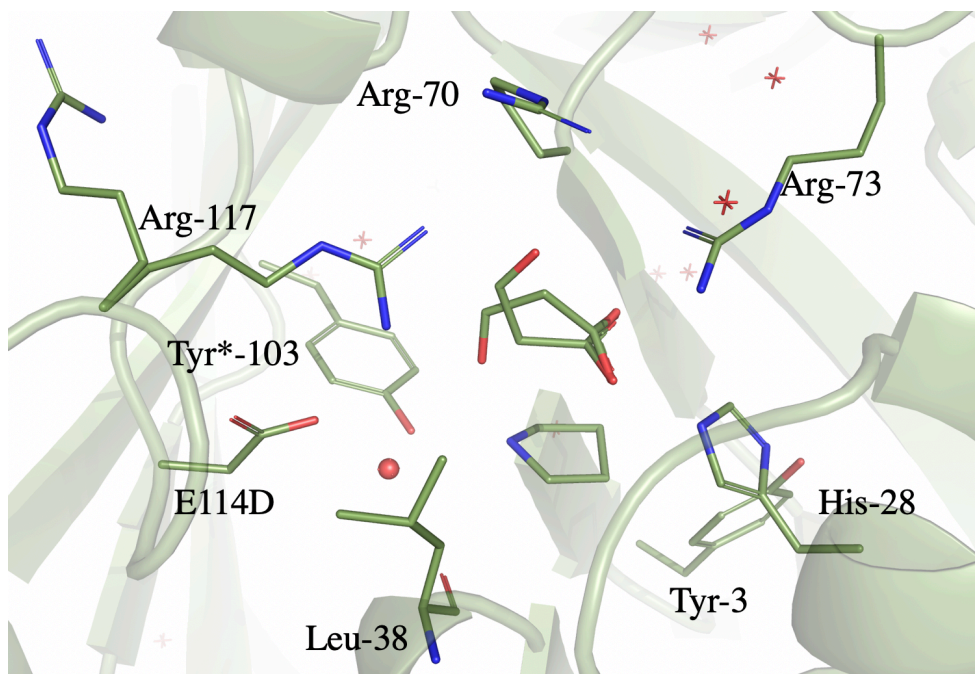


Figure 1. 47 Malonate semialdehyde in the active site of the E114D mutant.

Another molecule that is observed in the active site of the E114D mutant is the product malonate semialdehyde (Figure 1.47). This molecule is refined in two conformations. Malonate semialdehyde has a negative charge, and as shown previously, the presence of Arg-117 to the active site of E114D mutants might have an adverse effect on releasing the product and turnover of the enzyme. However, this effect has to be evaluated separately, for example, by biochemical characterization of the double mutant variants of E114D-R117A.

1.2.9. Catalytic cycle of Cg10062, hydration and decarboxylation

In summary, Cg10062 catalyzes both its hydration and decarboxylation reactions by forming covalent intermediates (Figure 1.48). Soaking crystals of various mutants of Cg10062 with the

substrate yielded different x-ray crystal structures where the active sites are occupied with an intermediate in its catalytic cycle.

Generally, the intermediates before the rate-limiting step are more likely to be observed in this technique. 3-Prolylacrylate (enamine) and malonate semialdehyde are the two dominant species observed in the active site of E114D mutants. This variant of Cg10062 stops the decarboxylation step by preventing the formation of iminium species (reactive toward hydration), and probably the native water activation pathway is also disturbed by the mutation, consistent with its decreased overall turnover. The E114D mutant generates a new pathway for hydration, without the formation of iminium. Although the E114D variants process only the hydration of propiolate, the rate of the reaction is more than 5-fold slower than the native enzyme. Generally, iminium species are a better substrate for the hydration, and in the absence of them, activated water has to go through a Michael addition of the α,β -unsaturated carboxylate intermediate. The Michael addition of activated water to α,β -unsaturated carboxylates could be the rate-limiting step for the modified catalytic cycle.

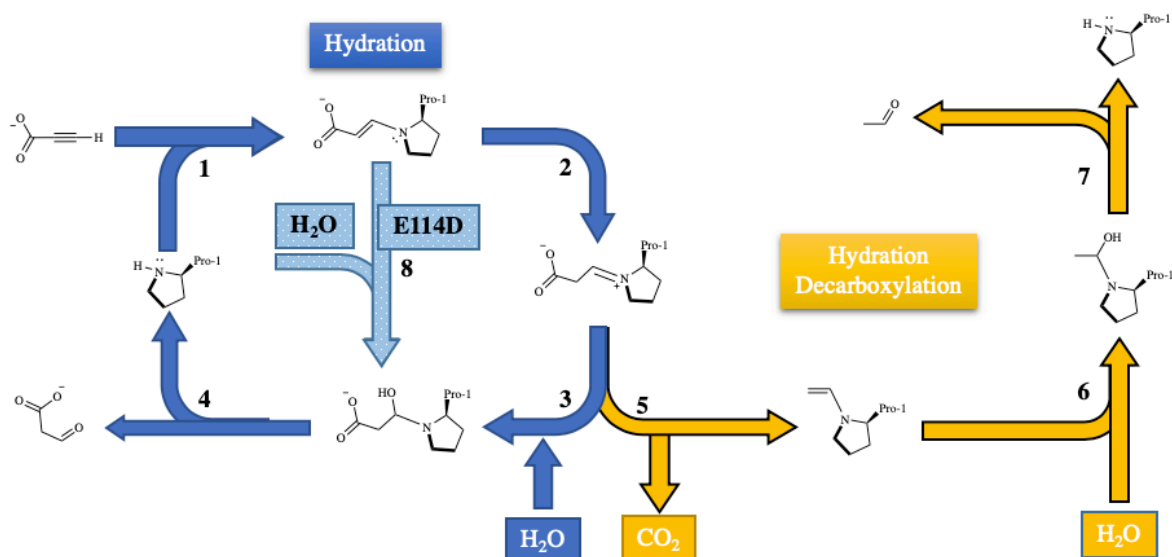


Figure 1. 48 The proposed catalytic cycle of Cg10062. (1) substrate binding, (2) tautomerization of enamine/imine, (3) hydration via activated water, (4) prolyl dissociation and product release (MSA), (5) decarboxylation of iminium intermediate, (6) hydration of decarboxylated intermediate, (7) prolyl dissociation and product release (acetaldehyde), (8) Variation of hydration in E114D mutants.

The Y103F variants have intermediate electron density similar to that of E114D, and based on the shape of the electron density, 3-prolylacrylate in enamine or iminium form is identified to be located in the active site. However, the arrangement of other active site residues, particularly Arg-73, suggests that the iminium form is more likely to be the dominant form in the active site of Y103F. Similarly, the addition of activated water could be considered to be the rate-limiting step for this variant since the enzyme is trapped in a pre-hydration state. On the other hand, in the active site of H28A and R73A mutants, all the residues responsible for water activation are intact, and the shape of the electron density matches the hydration product, 3-hydroxy-3-prolylpropionate. In both of these mutants, the addition of the activated water is no longer the rate-limiting step. However, the dissociation of the hydration product and release of malonate semialdehyde is more likely to be the rate-limiting step. Therefore, His-28 and Arg-73, other than playing a role in the binding of the substrate, might be involved in the dissociation of the final product and release of malonate semialdehyde. Among all these mutations, only E114D has evidence for the presence of malonate semialdehyde in the active site. Mutating Glu-114 to aspartate introduces another positively charged residue (Arg-117) into the active site, which increases the total positive charge of the active site cavity. The presence of malonate semialdehyde in the active site of E114D could be the result of the lower rate of product release due to the new trajectory of Arg-117.

1.3. Materials and methods

DNA cloning, expression, purification, and concentration of the native and mutant Cg10062 were performed by Ms. Amaya Sirinimal in Dr. Draths laboratory.

1.3.1 Protein Crystallization

All protein samples used in the crystallization process were concentrated to 18 mg/mL. Previously published crystallization conditions¹⁸ were attempted, and only flat multi-lattice crystals were grown (Figure 1.49) .

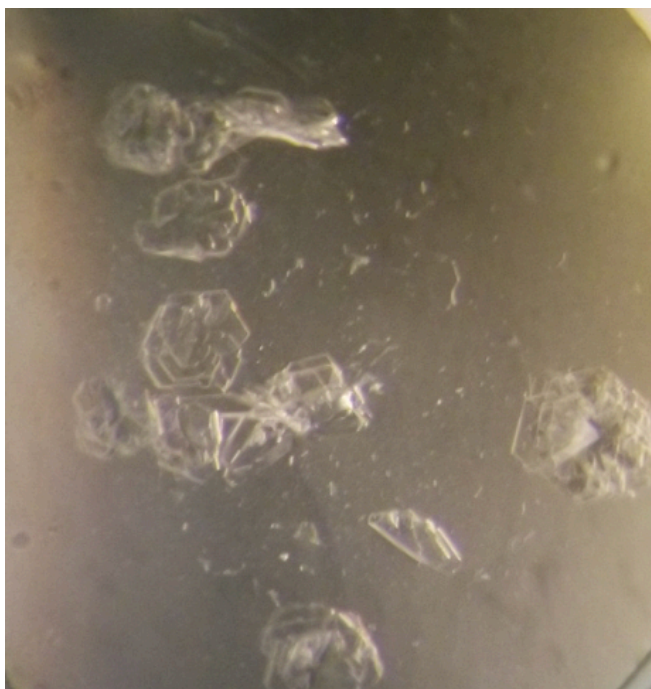


Figure 1. 49 Flat and multi-lattice crystals grown in 3% PEG 6000, 10mM Tris-SO₄, pH 4.0.

Native Cg10062 samples were screened using the hanging drop vapor diffusion method. The reservoir contained 50 μ L of commercially available crystallization screens (PegIon and Crystal Screen from Hampton Research). These screens are used after diluting with Millipore water. (i.e. PEG25 is a 25% (v/v) commercially available PegIon screen). The hanging drop consisted of 1 μ L

protein (18 mg/mL) and 1 μ L of reservoir solution. After 3 days (up to a week), several conditions yielded protein crystals bigger than 50 μ m (Table 1.10 lists the crystallization conditions for native Cg10062, and its mutations used to solve the x-ray crystal structures).

Table 1. 10 Crystallization condition for native Cg10062 and mutants with successful data collection.

Protein	Reservoir (well)
Native Cg10062	50 mM Magnesium chloride hexahydrate, 25 mM HEPES sodium (pH = 7.5), 7.5% (v/v) Polyethylene glycol 400
Cg10062-E114D	100 mM Potassium sulfate, 10% (w/v) Polyethylene glycol 3350
Cg10062-Y103F	100 mM Magnesium acetate tetrahydrate, 10% (w/v) Polyethylene glycol 3350
Cg10062-H28A	50 mM Ammonium acetate, 25 mM Sodium citrate tribasic dihydrate, (pH = 5.6), 7.5% (w/v), Polyethylene glycol 4000

All the conditions that yielded flat or three-dimensional crystals for various mutants of Cg10062 are reported in Table 1.11.

Table 1. 11 Screening crystallization conditions for Cg10062 mutants.

Protein	Reservoir (well)	Flat/3D
E114D	200 mM Ammonium sulfate	Flat
E114D	1 mM Cobalt (II) chloride hexahydrate, 10 mM MES monohydrate pH 6.5, 180 mM Ammonium sulfate	Flat
E114D	10 mM Sodium chloride, 10 mM HEPES pH 7.5, 160 mM Ammonium sulfate	Flat
E114D	150 mM Ammonium sulfate, 10 mM Tris pH 8.5, 1.2% v/v Glycerol	Flat
E114D	25 mM Sodium citrate tribasic dihydrate pH 5.6, 250 mM Ammonium phosphate monobasic	Flat
E114D	25 mM Sodium acetate trihydrate pH 4.6, 50 mM Sodium formate	Flat
E114D	25 mM Sodium acetate trihydrate pH 4.6, 50 mM Sodium chloride	Flat
E114D	50 mM Sodium chloride, 25 mM Sodium acetate trihydrate pH 4.6, 7.5% v/v (+/-)-2-Methyl-2,4-pentanedio	Flat
E114D	2.5 mM Cobalt (II) chloride hexahydrate, 25 mM Sodium acetate trihydrate pH 4.6, 250 mM 1,6-Hexanedio	Flat
E114D	100 mM Potassium sulfate pH 6.8, 10% w/v Polyethylene glycol 3,350	3D
E114D	100 mM Ammonium sulfate pH 6.0, 10% w/v Polyethylene glycol 3,350	Flat
E114D	50 mM Sodium malonate pH 4.0, 6% w/v Polyethylene glycol 3,350	Flat
E114D	2% v/v TacsimateTM pH 4.0, 6% w/v Polyethylene glycol 3,350	Flat
E114D	2% v/v TacsimateTM pH 5.0, 6% w/v Polyethylene glycol 3,350	Flat
E114D	4% v/v TacsimateTM pH 5.0, 10% w/v Polyethylene glycol 3,350	Flat

Table 1.11 (cont'd)

E114D	0.5% w/v Tryptone, 0,5 mM Sodium azide, 25 mM HEPES sodium pH 7.0, 10% w/v Polyethylene glycol 3,350	3D
E114D	0.2 M Magnesium chloride hexahydrate, 0.1 M HEPES sodium pH 7.5, 30% v/v Polyethylene glycol 400	3D
R73K	100 mM Ammonium chloride, 10% w/v Polyethylene glycol 3,350	3D
R73K	25 mM Citric acid, 25 mM BIS-TRIS propane, pH 5.0, 8% w/v Polyethylene glycol 3,350	3D
R73K	20 mM Citric acid, 30 mM BIS-TRIS propane, pH 6.4, 10% w/v Polyethylene glycol 3,350	3D
R73K	100 mM Ammonium sulfate, 10% w/v Polyethylene glycol 3,350	Flat
R73K	25 mM Sodium citrate tribasic dihydrate pH 5.6, 250 mM Ammonium phosphate monobasic	Flat
R73K	25 mM HEPES sodium pH 7.5, 200 mM Sodium phosphate monobasic monohydrate, 200 mM Potassium phosphate monobasic	Flat
R73K	50 mM Sodium chloride, 2.5% w/v Polyethylene glycol 6,000	Flat
Y103F	100 mM Magnesium chloride hexahydrate, 10% w/v Polyethylene glycol 3,350	Flat
Y103F	100 mM Calcium chloride dihydrate, 10% w/v Polyethylene glycol 3,350	Flat
Y103F	100 mM Ammonium chloride, 10% w/v Polyethylene glycol 3,350	Flat
Y103F	100 mM Magnesium sulfate heptahydrate, 10% w/v Polyethylene glycol 3,350	Flat
Y103F	100 mM Lithium acetate dihydrate, 10% w/v Polyethylene glycol 3,350	3D

Table 1.11 (cont'd)

Y103F	100 mM Magnesium acetate tetrahydrate, 10% w/v Polyethylene glycol 3,350	3D
Y103F	100 mM Magnesium acetate tetrahydrate, 10% w/v Polyethylene glycol 3,350	3D
Y103F	50 mM Sodium chloride, 2.5% w/v Polyethylene glycol 6,000	Flat
R70K	100 mM Potassium iodide, 10% w/v Polyethylene glycol 3,350	Flat
Y103A	100 mM Potassium citrate tribasic monohydrate, 10% w/v Polyethylene glycol 3,350	Flat
Y103A	100 mM Ammonium citrate dibasic, 10% w/v Polyethylene glycol 3,350	Flat
Y103A	25 mM HEPES sodium pH 7.5, 200 mM Sodium phosphate monobasic monohydrate, 200 mM Potassium phosphate monobasic	Flat
Y103A	50 mM Potassium sodium tartrate tetrahydrate, 25 mM Sodium citrate tribasic dihydrate pH 5.6, 50 mM Ammonium sulfate	3D
Y103A	8.2 % v/v 1,4-Dioxane	3D
Y103A	25 mM Sodium chloride, 25 mM HEPES pH 7.5, 400 M Ammonium sulfate	3D
H28A	25 mM Sodium citrate tribasic dihydrate pH 5.6, 250 mM Ammonium phosphate monobasic	Flat
H28A	50 mM Magnesium acetate tetrahydrate, 25 mM Sodium cacodylate trihydrate pH 6.5, 5% w/v Polyethylene glycol 8,000	3D
H28A	50 mM Sodium acetate trihydrate, 25 mM Sodium cacodylate trihydrate pH 6.5, 7.5% w/v Polyethylene glycol 8,000	3D

Table 1.11 (cont'd)

H28A	25 mM HEPES pH 7.5, 2.5% w/v Polyethylene glycol 8,000, 2% v/v Ethylene glycol	3D
H28A	4% v/v TacsimateTM pH 6.0, 10% w/v Polyethylene glycol 3,350	3D
H28A	100 mM Magnesium nitrate hexahydrate, 10% w/v Polyethylene glycol 3,350	3D
H28A	100 mM Ammonium acetate, 10% w/v Polyethylene glycol 3,350	3D
H28A	100 mM Ammonium nitrate, 10% w/v Polyethylene glycol 3,350	3D
H28A	100 mM Calcium chloride dihydrate, 10% w/v Polyethylene glycol 3,350	3D
H28A	100 mM Ammonium iodide, 10% w/v Polyethylene glycol 3,350	3D
H28A	4% v/v TacsimateTM pH 7.0, 10% w/v Polyethylene glycol 3,350	3D
H28A	1% v/v TacsimateTM pH 5.0, 50 mM Sodium citrate tribasic dihydrate pH 5.6, 8% w/v Polyethylene glycol 3,350	3D
H28A	50 mM Ammonium sulfate, 25 mM MES monohydrate pH 6.5, 7.5% w/v Polyethylene glycol monomethyl ether 5,000	3D
H28A	50 mM Ammonium sulfate, 25 mM Sodium cacodylate trihydrate pH 6.5, 7.5% w/v Polyethylene glycol 8,000	3D
H28A	100 mM Ammonium citrate tribasic pH 7.0, 10% w/v Polyethylene glycol 3,350	3D
E114A	25 mM Sodium citrate tribasic dihydrate pH 5.6, 250 mM Ammonium phosphate monobasic	Flat
E114A	2.5 mM Iron (III) chloride hexahydrate, 25 mM Sodium citrate tribasic dihydrate pH 5.6, 2.5% v/v Jeffamine ® M-600 ®	Flat

Table 1.11 (cont'd)

E114A	50 mM Magnesium chloride hexahydrate, 25 mM TRIS hydrochloride pH 8.5, 7.5% w/v Polyethylene glycol 4,000	3D
E114A	50 mM Magnesium acetate tetrahydrate, 25 mM Sodium cacodylate trihydrate pH 6.5, 5% w/v Polyethylene glycol 8,000	3D
E114A	100 mM Ammonium iodide, 10% w/v Polyethylene glycol 3,350	3D
E114A	100 mM Ammonium chloride, 10% w/v Polyethylene glycol 3,350	3D
E114A	100 mM Ammonium nitrate, 10% w/v Polyethylene glycol 3,350	3D
E114A	100 mM Sodium acetate trihydrate pH 7.0, 10% w/v Polyethylene glycol 3,350	3D
E114A	50 mM Ammonium sulfate, 25 mM MES monohydrate pH 6.5, 7.5% w/v Polyethylene glycol monomethyl ether 5,000	3D
R73A	100 mM Ammonium fluoride, 10% w/v Polyethylene glycol 3,350	3D
R73A	100 mM Ammonium chloride, 10% w/v Polyethylene glycol 3,350	3D
R73A	100 mM Ammonium nitrate, 10% w/v Polyethylene glycol 3,350	3D
R73A	100 mM Sodium malonate pH 7.0, 10% w/v Polyethylene glycol 3,350	3D
R73A	4% v/v TacsimateTM pH 6.0, 10% w/v Polyethylene glycol 3,350	3D

Majority of the 3D crystals reported in Table 1.11 were used in the soaking experiments.

1.3.2 Ligand soaking experiments and solving the structures

After obtaining crystals of Cg10062 and its mutations, several crystals were transferred to a cryoprotectant solution containing 20% to 40% glycerol in the reservoir solution in which the crystals were grown. These samples were used to determine an apo structure of the enzyme.

To obtain ligand bound structures of the Cg10062, apo crystals of the Cg10062 mutants were soaked in different ligand solutions. Stock solutions of the ligand molecules were prepared at 10 mM concentration in Millipore water. Later, 20% dilution of the stock solutions were made in the corresponding reservoir solution for every crystal. The final concentration of the ligand molecules in the soaking solution is 2 mM. Other than propionate, acetylenedicarboxylic acid is also used in the soaking experiments. However, in this study we only focus on the result of soaking experiments with propionate.

The crystals were mounted in nylon cryo-loops and flash frozen by immersion in liquid nitrogen. The diffraction data sets were collected at the Advanced Photon Source (APS) at the Argonne National Laboratories (Argonne, IL). The diffraction datasets with resolutions higher than 3 Å were indexed, refined and scaled using HKL2000. The resulting scaled maps then are used for molecular replacement using the PHENIX suite. Initially, the available structure of Cg10062 (PDBID:3n4g) is used to perform the molecular replacement for the native Cg10062. The output of molecular replacement is refined using the PHENIX suite and model manipulation and solvent addition is done using COOT. Structure of native Cg10062 is then used for the molecular replacement of remaining datasets. Using eLBOW (from PHENIX suite), the .cif files for all the ligands are generated. In the case of ligand-bound structures, after the convergence of refinements for the naked protein, first the ligand was refined in the active site and then solvent molecules were added to complete the structure.

The chemical bond parameters of prolyl species were estimated from the crystal structure of proline-derived enamines⁴⁹. The restraints files for each ligand were generated to define the bond length, bond angles, and dihedral torsion angle. In the final round of the refinement, the occupancies were refined for all the atoms.

REFERENCES

REFERENCES

- (1) Johnson, L. N.; Petsko, G. A. David Phillips and the Origin of Structural Enzymology. *Trends Biochem. Sci.* **1999**, 24 (7), 287–289. [https://doi.org/10.1016/S0968-0004\(99\)01423-1](https://doi.org/10.1016/S0968-0004(99)01423-1).
- (2) Go, N.; Noguti, T.; Nishikawa, T. Dynamics of a Small Globular Protein in Terms of Low-Frequency Vibrational Modes. *Proc. Natl. Acad. Sci.* **1983**, 80 (12), 3696–3700. <https://doi.org/10.1073/pnas.80.12.3696>.
- (3) Tsai, S.-C. S.; Ames, B. D. Structural Enzymology of Polyketide Synthases. *Methods Enzymol.* **2009**, 459, 17–47. [https://doi.org/10.1016/S0076-6879\(09\)04602-3](https://doi.org/10.1016/S0076-6879(09)04602-3).
- (4) Agarwal, P. K. Enzymes: An Integrated View of Structure, Dynamics and Function. *Microb. Cell Factories* **2006**, 5, 2. <https://doi.org/10.1186/1475-2859-5-2>.
- (5) Merlino, A.; Mazzarella, L.; Carannante, A.; Fiore, A. D.; Donato, A. D.; Notomista, E.; Sica, F. The Importance of Dynamic Effects on the Enzyme Activity X-RAY STRUCTURE AND MOLECULAR DYNAMICS OF ONCONASE MUTANTS. *J. Biol. Chem.* **2005**, 280 (18), 17953–17960. <https://doi.org/10.1074/jbc.M501339200>.
- (6) Gora, A.; Brezovsky, J.; Damborsky, J. Gates of Enzymes. *Chem. Rev.* **2013**, 113 (8), 5871–5923. <https://doi.org/10.1021/cr300384w>.
- (7) Poelarends, G. J.; Serrano, H.; Person, M. D.; Johnson, W. H.; Whitman, C. P. Characterization of Cg10062 from *Corynebacterium Glutamicum*: Implications for the Evolution of Cis-3-Chloroacrylic Acid Dehalogenase Activity in the Tautomerase Superfamily. *Biochemistry* **2008**, 47 (31), 8139–8147. <https://doi.org/10.1021/bi8007388>.
- (8) Davidson, R.; Baas, B.-J.; Akiva, E.; Holliday, G. L.; Polacco, B. J.; LeVieux, J. A.; Pullara, C. R.; Zhang, Y. J.; Whitman, C. P.; Babbitt, P. C. A Global View of Structure–Function Relationships in the Tautomerase Superfamily. *J. Biol. Chem.* **2018**, 293 (7), 2342–2357. <https://doi.org/10.1074/jbc.M117.815340>.
- (9) Poelarends, G. J.; Veetil, V. P.; Whitman, C. P. The Chemical Versatility of the β – α – β Fold: Catalytic Promiscuity and Divergent Evolution in the Tautomerase Superfamily. *Cell. Mol. Life Sci.* **2008**, 65 (22), 3606–3618. <https://doi.org/10.1007/s00018-008-8285-x>.
- (10) Poelarends, G. J.; Whitman, C. P. Evolution of Enzymatic Activity in the Tautomerase Superfamily: Mechanistic and Structural Studies of the 1,3-Dichloropropene Catabolic Enzymes. *Bioorganic Chem.* **2004**, 32 (5), 376–392. <https://doi.org/10.1016/j.bioorg.2004.05.006>.
- (11) Murzin, A. G. Structural Classification of Proteins: New Superfamilies. *Curr. Opin. Struct. Biol.* **1996**, 6 (3), 386–394. [https://doi.org/10.1016/S0959-440X\(96\)80059-5](https://doi.org/10.1016/S0959-440X(96)80059-5).

- (12) JENKINSt, J. R.; Cooper, R. A. Molecular Cloning, Expression, and Analysis of the Genes of the Homoprotocatechuate Catabolic Pathway of Escherichia Coli Ct. *J BACTERIOL* **1988**, *170*, 8.
- (13) Suzuki, M.; Sugimoto, H.; Nakagawa, A.; Tanaka, I.; Nishihira, J.; Sakai, M. Crystal Structure of the Macrophage Migration Inhibitory Factor from Rat Liver. *Nat. Struct. Biol.* **1996**, *3* (3), 259–266. <https://doi.org/10.1038/nsb0396-259>.
- (14) Lubetsky, J. B.; Swope, M.; Dealwis, C.; Blake, P.; Lolis, E. Pro-1 of Macrophage Migration Inhibitory Factor Functions as a Catalytic Base in the Phenylpyruvate Tautomerase Activity,. *Biochemistry* **1999**, *38* (22), 7346–7354. <https://doi.org/10.1021/bi990306m>.
- (15) Poelarends, G. J.; Serrano, H.; Person, M. D.; Johnson, William H.; Murzin, A. G.; Whitman, C. P. 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* **2004**, *43* (3), 759–772. <https://doi.org/10.1021/bi0355948>.
- (16) Poelarends, G. J.; Johnson, W. H.; Murzin, A. G.; Whitman, C. P. Mechanistic Characterization of a Bacterial Malonate Semialdehyde Decarboxylase: IDENTIFICATION OF A NEW ACTIVITY IN THE TAUTOMERASE SUPERFAMILY. *J. Biol. Chem.* **2003**, *278* (49), 48674–48683. <https://doi.org/10.1074/jbc.M306706200>.
- (17) Poelarends, G. J.; Almrud, J. J.; Serrano, H.; Darty, J. E.; Johnson, W. H.; Hackert, M. L.; Whitman, C. P. Evolution of Enzymatic Activity in the Tautomerase Superfamily. *Biochemistry* **2006**, *45* (25), 7700–7708. <https://doi.org/10.1021/bi0600603>.
- (18) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- (19) Subramanya, H. S.; Roper, D. I.; Dauter, Z.; Dodson, E. J.; Davies, G. J.; Wilson, K. S.; Wigley, D. B. Enzymatic Ketonization of 2-Hydroxymuconate: Specificity and Mechanism Investigated by the Crystal Structures of Two Isomerases. *Biochemistry* **1996**, *35* (3), 792–802. <https://doi.org/10.1021/bi951732k>.
- (20) Sun, H. W.; Bernhagen, J.; Bucala, R.; Lolis, E. Crystal Structure at 2.6-Å Resolution of Human Macrophage Migration Inhibitory Factor. *Proc. Natl. Acad. Sci.* **1996**, *93* (11), 5191–5196. <https://doi.org/10.1073/pnas.93.11.5191>.
- (21) Guo, Y.; Serrano, H.; Poelarends, G. J.; Johnson, W. H.; Hackert, M. L.; Whitman, C. P. Kinetic, Mutational, and Structural Analysis of Malonate Semialdehyde Decarboxylase from Corynebacterium Strain FG41: Mechanistic Implications for the Decarboxylase and Hydratase Activities. *Biochemistry* **2013**, *52* (28), 4830–4841. <https://doi.org/10.1021/bi400567a>.

- (22) Harayama, S.; Rekik, M.; Ngai, K. L.; Ornston, L. N. Physically Associated Enzymes Produce and Metabolize 2-Hydroxy-2,4-Dienoate, a Chemically Unstable Intermediate Formed in Catechol Metabolism via Meta Cleavage in *Pseudomonas Putida*. *J. Bacteriol.* **1989**, *171* (11), 6251–6258. <https://doi.org/10.1128/jb.171.11.6251-6258.1989>.
- (23) Harayama, S.; Lehrbach, P. R.; Timmis, K. N. Transposon Mutagenesis Analysis of Meta-Cleavage Pathway Operon Genes of the TOL Plasmid of *Pseudomonas Putida* Mt-2. *J BACTERIOL* **1984**, *160*, 5.
- (24) Rosengren, E.; Åman, P.; Thelin, S.; Hansson, C.; Ahlfors, S.; Björk, P.; Jacobsson, L.; Rorsman, H. The Macrophage Migration Inhibitory Factor MIF Is a Phenylpyruvate Tautomerase. *FEBS Lett.* **1997**, *417* (1), 85–88. [https://doi.org/10.1016/S0014-5793\(97\)01261-1](https://doi.org/10.1016/S0014-5793(97)01261-1).
- (25) Darty, J. E. Random and Rational Evolution of Tautomerase Superfamily Members : Analysis and Implications. Thesis, 2008.
- (26) Whitman, C. P. The 4-Oxalocrotonate Tautomerase Family of Enzymes: How Nature Makes New Enzymes Using a Beta-Alpha-Beta Structural Motif. *Arch. Biochem. Biophys.* **2002**, *402* (1), 1–13. [https://doi.org/10.1016/S0003-9861\(02\)00052-8](https://doi.org/10.1016/S0003-9861(02)00052-8).
- (27) Baas, B.-J.; Medellin, B. P.; LeVieux, J. A.; de Ruijter, M.; Zhang, Y. J.; Brown, S. D.; Akiva, E.; Babbitt, P. C.; Whitman, C. P. Structural, Kinetic, and Mechanistic Analysis of an Asymmetric 4-Oxalocrotonate Tautomerase Trimer. *Biochemistry* **2019**, *58* (22), 2617–2627. <https://doi.org/10.1021/acs.biochem.9b00303>.
- (28) van Hylckama Vlieg, J. E. T.; Janssen, D. B. Bacterial Degradation of 3-Chloroacrylic Acid and the Characterization of Cis- and Trans-Specific Dehalogenases. *Biodegradation* **1991**, *2* (3), 139–150. <https://doi.org/10.1007/BF00124488>.
- (29) Baas, B.-J.; Zandvoort, E.; Wasiel, A. A.; Quax, W. J.; Poelarends, G. J. 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* **2011**, *50* (14), 2889–2899. <https://doi.org/10.1021/bi200071k>.
- (30) Huddleston, J. P.; Wang, S. C.; Johnson, K. A.; Whitman, C. P. Resolution of the Uncertainty in the Kinetic Mechanism for the Trans-3-Chloroacrylic Acid Dehalogenase-Catalyzed Reaction. *Arch. Biochem. Biophys.* **2017**, *623–624*, 9–19. <https://doi.org/10.1016/j.abb.2017.05.004>.
- (31) Ang, T.-F.; Maiangwa, J.; Salleh, A. B.; Normi, Y. M.; Leow, T. C. Dehalogenases: From Improved Performance to Potential Microbial Dehalogenation Applications. *Molecules* **2018**, *23* (5), 1100. <https://doi.org/10.3390/molecules23051100>.

- (32) Guo, Y.; Serrano, H.; Johnson, W. H.; Ernst, S.; Hackert, M. L.; Whitman, C. P. Crystal Structures of Native and Inactivated Cis-3-Chloroacrylic Acid Dehalogenase: Implications for the Catalytic and Inactivation Mechanisms. *Bioorganic Chem.* **2011**, *39* (1), 1–9. <https://doi.org/10.1016/j.bioorg.2010.10.001>.
- (33) Almrud, J. J.; Poelarends, G. J.; Johnson, William H.; Serrano, H.; Hackert, M. L.; Whitman, C. P. Crystal Structures of the Wild-Type, P1A Mutant, and Inactivated Malonate Semialdehyde Decarboxylase: A Structural Basis for the Decarboxylase and Hydratase Activities. *Biochemistry* **2005**, *44* (45), 14818–14827. <https://doi.org/10.1021/bi051383m>.
- (34) Poelarends, G. J.; Serrano, H.; Johnson, W. H.; Hoffman, D. W.; Whitman, C. P. The Hydratase Activity of Malonate Semialdehyde Decarboxylase: Mechanistic and Evolutionary Implications. *J. Am. Chem. Soc.* **2004**, *126* (48), 15658–15659. <https://doi.org/10.1021/ja044304n>.
- (35) Poelarends, G. J.; Serrano, H.; Johnson, W. H.; Whitman, C. P. Inactivation of Malonate Semialdehyde Decarboxylase by 3-Halopropiolates: Evidence for Hydratase Activity. *Biochemistry* **2005**, *44* (26), 9375–9381. <https://doi.org/10.1021/bi050296r>.
- (36) Meer, J.-Y. van der; Poddar, H.; Baas, B.-J.; Miao, Y.; Rahimi, M.; Kunzendorf, A.; Merkerk, R. van; Tepper, P. G.; Geertsema, E. M.; Thunnissen, A.-M. W. H.; et al. Using Mutability Landscapes of a Promiscuous Tautomerase to Guide the Engineering of Enantioselective Michaelases. *Nat. Commun.* **2016**, *7* (1), 1–16. <https://doi.org/10.1038/ncomms10911>.
- (37) Poddar, H.; Rahimi, M.; Geertsema, E. M.; Thunnissen, A.-M. W. H.; Poelarends, G. J. Evidence for the Formation of an Enamine Species during Aldol and Michael-Type Addition Reactions Promiscuously Catalyzed by 4-Oxalocrotonate Tautomerase. *ChemBioChem* **2015**, *16* (5), 738–741. <https://doi.org/10.1002/cbic.201402687>.
- (38) van der Meer, J.-Y.; Biewenga, L.; Poelarends, G. J. The Generation and Exploitation of Protein Mutability Landscapes for Enzyme Engineering. *Chembiochem Eur. J. Chem. Biol.* **2016**, *17* (19), 1792–1799. <https://doi.org/10.1002/cbic.201600382>.
- (39) Baas, B.-J.; Zandvoort, E.; Wasiel, A. A.; Poelarends, G. J. Demethionylation of Pro-1 Variants of 4-Oxalocrotonate Tautomerase in Escherichia Coli by Co-Expression with an Engineered Methionine Aminopeptidase. *FEBS Open Bio* **2014**, *4*, 651–658. <https://doi.org/10.1016/j.fob.2014.07.003>.
- (40) Lukesch, M. S.; Pavkov-Keller, T.; Gruber, K.; Zangger, K.; Wiltschi, B. Substituting the Catalytic Proline of 4-Oxalocrotonate Tautomerase with Non-Canonical Analogues Reveals a Finely Tuned Catalytic System. *Sci. Rep.* **2019**, *9* (1), 1–9. <https://doi.org/10.1038/s41598-019-39484-9>.
- (41) Rahimi, M.; van der Meer, J.-Y.; Geertsema, E. M.; Poelarends, G. J. Engineering a Promiscuous Tautomerase into a More Efficient Aldolase for Self-Condensations of Linear

- Aliphatic Aldehydes. *ChemBioChem* **2017**, *18* (14), 1435–1441. <https://doi.org/10.1002/cbic.201700121>.
- (42) Rahimi, M.; van der Meer, J.-Y.; Geertsema, E. M.; Poddar, H.; Baas, B.-J.; Poelarends, G. J. Mutations Closer to the Active Site Improve the Promiscuous Aldolase Activity of 4-Oxalocrotonate Tautomerase More Effectively than Distant Mutations. *ChemBioChem* **2016**, *17* (13), 1225–1228. <https://doi.org/10.1002/cbic.201600149>.
- (43) Huddleston, J. P.; Johnson, W. H.; Schroeder, G. K.; Whitman, C. P. Reactions of Cg10062, a Cis-3-Chloroacrylic Acid Dehalogenase Homologue, with Acetylene and Allene Substrates: Evidence for a Hydration-Dependent Decarboxylation. *Biochemistry* **2015**, *54* (19), 3009–3023. <https://doi.org/10.1021/acs.biochem.5b00240>.
- (44) Robertson, B. A.; Schroeder, G. K.; Jin, Z.; Johnson, K. A.; Whitman, C. P. Pre-Steady-State Kinetic Analysis of Cis-3-Chloroacrylic Acid Dehalogenase: Analysis and Implications. *Biochemistry* **2009**, *48* (49), 11737–11744. <https://doi.org/10.1021/bi901349z>.
- (45) Schroeder, G. K.; Huddleston, J. P.; Johnson, W. H.; Whitman, C. P. A Mutational Analysis of the Active Site Loop Residues in Cis-3-Chloroacrylic Acid Dehalogenase. *Biochemistry* **2013**, *52* (24), 4204–4216. <https://doi.org/10.1021/bi4004414>.
- (46) Wang, S. C.; Person, M. D.; Johnson, William H.; Whitman, C. P. Reactions of Trans-3-Chloroacrylic Acid Dehalogenase with Acetylene Substrates: Consequences of and Evidence for a Hydration Reaction. *Biochemistry* **2003**, *42* (29), 8762–8773. <https://doi.org/10.1021/bi034598+>.
- (47) Momany, C.; Ernst, S.; Ghosh, R.; Chang, N.-L.; Hackert, M. L. Crystallographic Structure of a PLP-Dependent Ornithine Decarboxylase From *Lactobacillus* 30a to 3.0 Å Resolution. *J. Mol. Biol.* **1995**, *252* (5), 643–655. <https://doi.org/10.1006/jmbi.1995.0526>.
- (48) Rocha, J. F.; Pina, A. F.; Sousa, S. F.; Cerqueira, N. M. F. S. A. PLP-Dependent Enzymes as Important Biocatalysts for the Pharmaceutical, Chemical and Food Industries: A Structural and Mechanistic Perspective. *Catal. Sci. Technol.* **2019**, *9* (18), 4864–4876. <https://doi.org/10.1039/C9CY01210A>.
- (49) Bock, D. A.; Lehmann, C. W.; List, B. Crystal Structures of Proline-Derived Enamines. *Proc. Natl. Acad. Sci.* **2010**, *107* (48), 20636–20641. <https://doi.org/10.1073/pnas.1006509107>.

CHAPTER TWO

Investigating the Mechanism of Rice Branching Enzyme I and Structures of Rice Granule-Bound Starch Synthase I

This chapter aims to investigate the function of the rice branching enzyme I, using protein x-ray crystallographic methods and kinetic studies. New structures of granule-bound starch synthase I are presented as well. Rice branching enzyme is one of the enzymes involved in the biosynthesis of starch. The first section provides an introduction to the biosynthesis of starch/glycogen, and the role branching enzymes play in it. The second section provides new structures of granule-bound starch synthase I. The third section presents the results obtained for the investigation of rice branching enzyme and provides a discussion regarding the function of the rice branching enzyme. Finally, in the last section, experimental techniques and methods employed in this research are provided.

2.1. Introduction

Starch is a glucose polymer, and plants produce it as a primary energy storage molecule. Starch consists of glucose units connected via 1,4- and 1,6-glycosidic bonds.¹ Glucose molecules at high concentrations create osmotic pressure, which can harm the plant cells. However, plants store their glucose in the form of starch to prevent the damage.² Primarily, chloroplasts produce a specific type of starch called, transitory starch whose levels vary. The chloroplasts in leaves are responsible for the production of transitory starch. The other type of starch, called storage starch, mainly accumulates in roots, seeds, tubers, and fruits. Storage starch is essential for the survival of the plants during extreme conditions.^{3,4} The fine structure of starch varies by tissue and environmental conditions to accommodate the needs of the plant. Therefore, plants grown in a different environment produce unique starch molecules.⁵

On the other hand, glycogen is a carbohydrate storage molecule in bacteria and animals that also consists of alpha-1,4-linked glucose units with alpha-1,6 branch points. (Cyanobacteria

produce a different type of carbohydrate storage molecule which is more similar to starch).⁶ Bacteria synthesize their glycogen during the growth and stationary phases and use them when facing extreme conditions.⁷ In humans and animals, the liver is the primary organ for the production and storage of glycogen, but muscle cells also store glycogen for immediate energy demands.⁸ Based on the biochemical classification, glycogen and starch are carbon and energy storage molecules since they fulfill the following criteria. First, cells accumulate these compounds during times of energy excess; Second, cells use them when exogenous energy sources are not capable of providing maintenance energy levels for critical cell activities such as growth, division, and viability. Third, these compounds, upon decomposition, produce energy forms employable by the cell.⁹

In the past, since access to carbohydrate sources were limited, those species that adapted to store carbohydrates quickly had higher chances of survival. However, the abundance of carbohydrates in today's diet, has led to an obesity crisis.¹⁰ Studies^{11–13} show that abundantly stored glycogen in the liver decreases fat metabolism and, hence increasing stored fat. New dietary regimes like the keto diet focus on lowering the level of stored glycogen in the liver to enhance fat burning.

2.1.1. The structure of starch

Starch is one of the biggest biomolecules that exist, ranging from 0.1 to 50 μm in diameter.⁵ The particular organization of glucose molecules in starch allows it to form large molecules. Initially, glucose units are linked together by 1,4-glycosidic bonds. The unbranched, long stretching chains of glucose are sections of a starch granule that is called amylose. In starch, glucose units linked in α -anomeric form, unlike cellulose, another polymer of glucose with β -1,4-

glycosidic bonds.¹⁴ These glucose chains, in cellulose, are linear while they have helical conformation in starch. At specific intervals, these long glucose chains contain branches connected via a 1,6-glycosidic bond. The branches form left-handed double helices and are arranged in orderly patterns to create amylopectin regions. Alternating amylose (amorphous) and amylopectin (crystalline) regions are what make a starch granule semi-crystalline (Figure 2.1). Generally, starch granules consist of 18-28% amylose and 72-82% amylopectin.³ The difference in the fine structure of starch is because of different amylose/amylopectin composition, chain size in the branched region, the distance between the branches, arrangement of double helices formed by neighboring branches.¹⁵

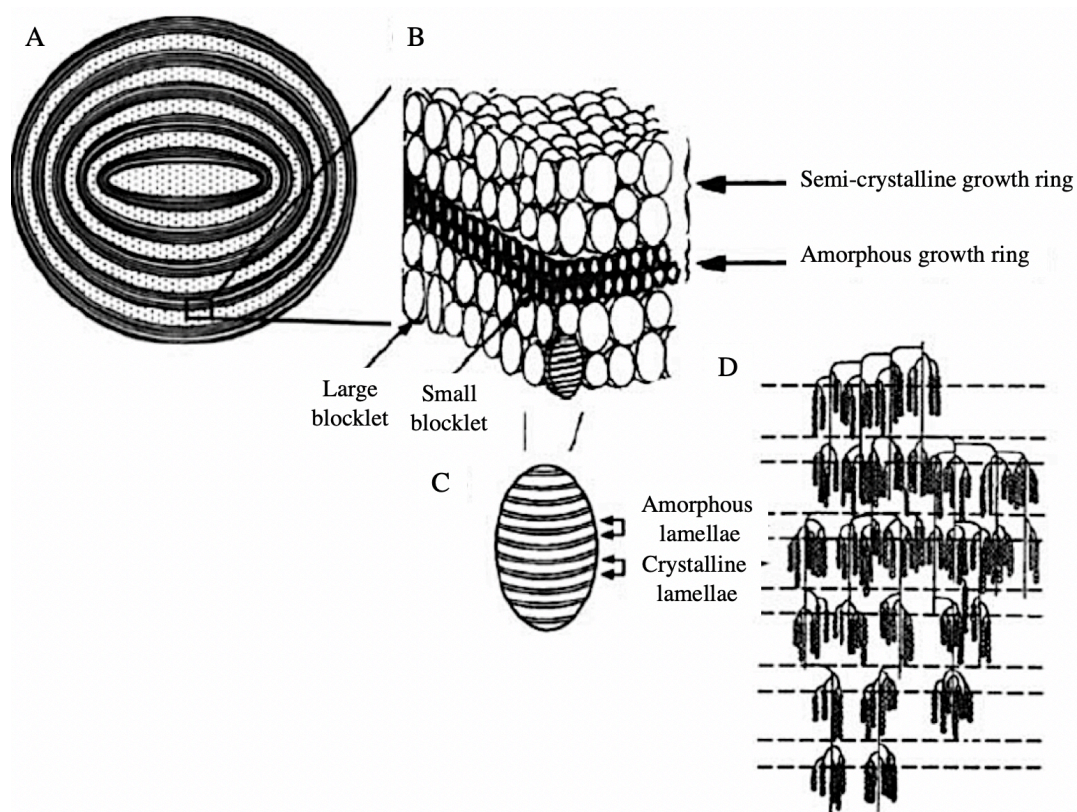


Figure 2. 1 Schematic diagram of starch granule structure.^{16,17} (A) Whole granule of starch, consisting of alternating semi-crystalline and amorphous growth rings, (B) a stack of large and small blocklets, (C) crystalline and amorphous lamellae in a blocklet, (D) ordered double helices within crystalline lamellae and amylopectin branch points within an amorphous lamella. (modified, see references)

Helical conformation (Figure 2.1, C) of a single glucose chain in amylose has a different geometry compared to the double helix formed by neighboring branches in amylopectin.¹⁸ The difference in the structure is one of the factors to determine which enzyme is interacting with amylose or amylopectin regions. A single-stranded glucose helix (Figure 2.2 C) has seven glucose per turn with a 0.805 nm pitch. On the other hand, double-helical structures (Figure 2.2 D) formed in amylopectin have six glucose per turn with a 2.13 nm pitch.³ Also, the cavity formed inside the helical structure has a bigger space in single-stranded helix compared to the double-stranded helix. The topological difference between single- and double-stranded helices of glucose chains have analytical importance besides their biological significance.

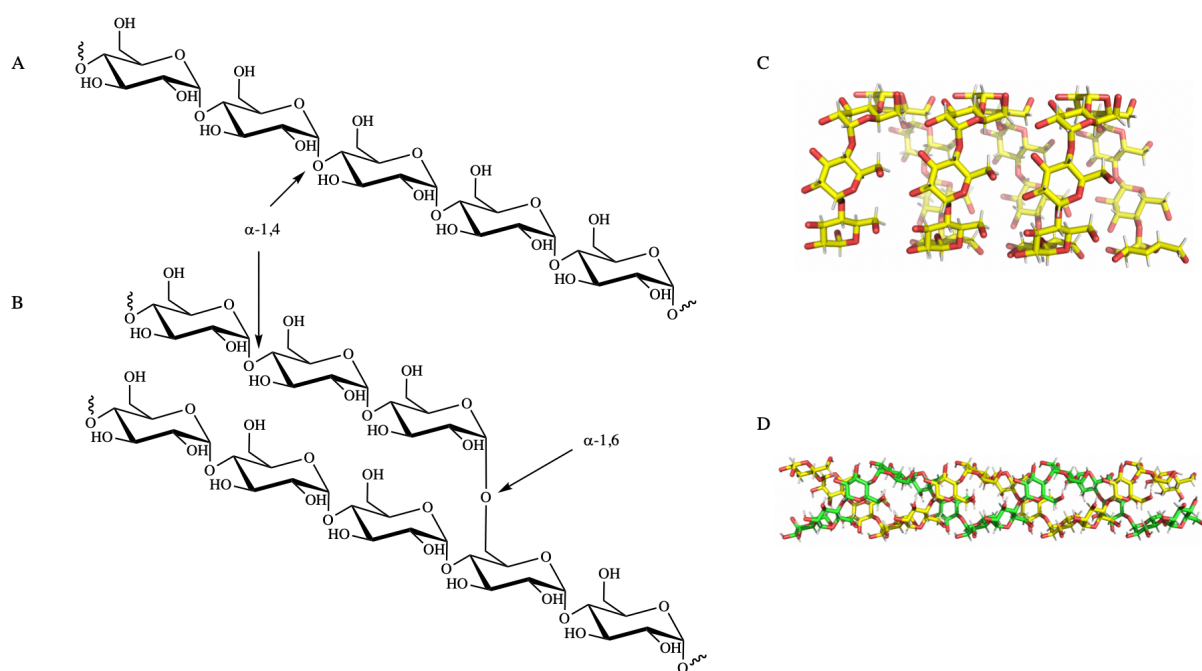


Figure 2. 2 Chemical structures of linear and branched glucose chains. (A) Chain of glucoses connected via α -1,4-glycosidic bond, (B) Branched glucose chain via α -1,6-glycosidic bond, (C) Helical conformation of single glucose chain, (D) Double helix formed by two adjacent glucose chains.

One of the well-known methods to study starch (amylose and amylopectin) is the iodine assay. In the iodine assay, triiodide ions form by the reaction of potassium iodide and iodine. Triiodide

ion absorbs at 287 and 350 nm, but in the presence of starch, it enters into the cavity of a single-stranded glucan helix. Triiodide-amylose complex has a strong absorption near 600nm.¹⁹⁻²¹

Glycogen, on the other hand, is formed by branching and extending the branches to form a natural dendrimer and does not have an ordered structure. Therefore, glycogen has a size limit, which is defined by the steric hindrance and accessibility of glucose chains for further branching.²²

2.1.2. Biosynthesis of starch

Biosynthesis of starch (or glycogen) consists of three steps.³ In the first step (Figure 2.3), ADP-glucose pyrophosphorylase (ADPGPPase) forms ADP-glucose from glucose-1-phosphate. ADP-glucose, the monomer for glucose polymerization, is the starting material for the second enzyme.⁹ The formation of ADP-glucose is the rate-limiting step for the biosynthesis of starch or glycogen, where the catalytic activity of the ADPGPPase is under allosteric regulation. For plants and bacteria, intermediates of the major carbon assimilatory pathway in the respective organism are the allosteric ligands for ADPGPPase.²³⁻²⁵ These intermediates can act as an activator or an inhibitor. The structure of ADPGPPase for different plants and bacteria is available, and its allosteric regulation has been studied in detail.²⁶⁻³¹ Besides, the crystal structure of ADPGPPase from potato was obtained in our lab.³²

In the second step (Figure 2.3), starch or glycogen synthase (SS or GS) extends the oligosaccharide chains. A Starch or glycogen synthase transfers a glucose unit from an ADP-glucose to the non-reducing end of an acceptor glucan chain and releases ADP.³³ Starch synthase belongs to the glycosyltransferase 5 (GT5) family and transfers the glycosyl units by retention of stereochemistry. i.e., ADP- α -1-glucose is used to form an α -1,4-glycosidic bond.^{34,35} Bacterial and archaeal synthase are also members of the GT5 family and share high sequence similarity and

structural resemblance to starch synthase. The common characteristic of GS and SS is that they all have a twin-Rossmann fold where the active site is located between the two Rossmann folds.^{34,35}

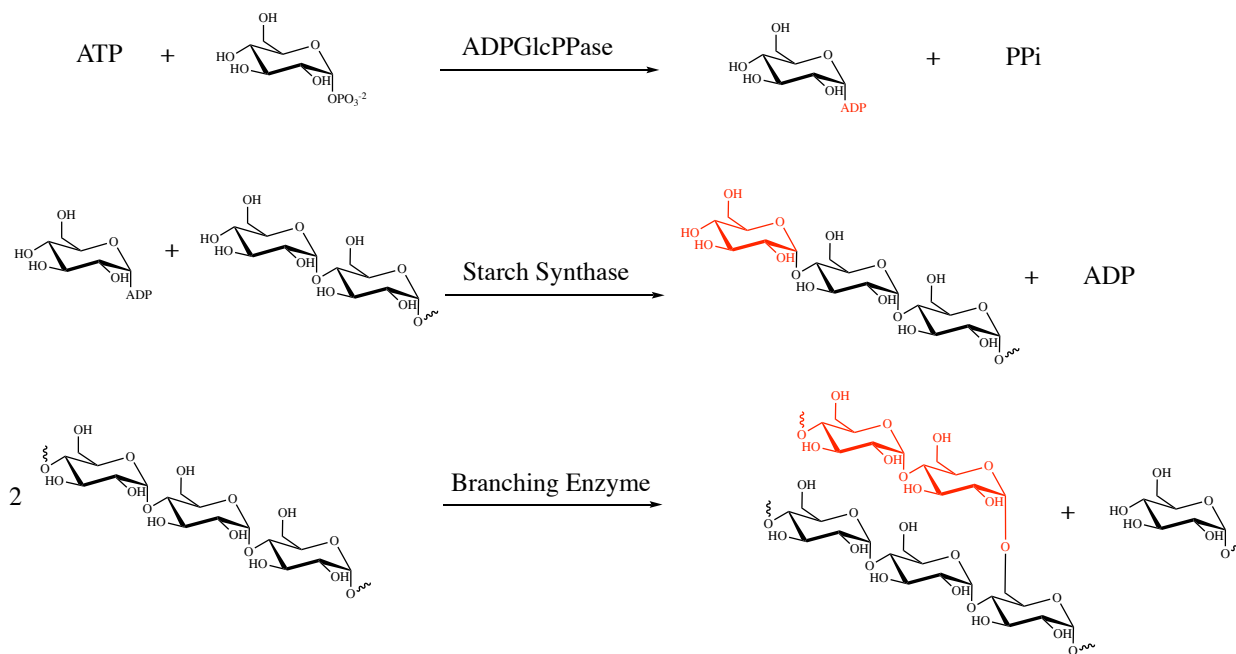


Figure 2. 3 Biosynthesis of starch (or glycogen).

The starch synthases in plant species, unlike glycogen synthases, have several isoforms. Soluble Starch Synthases (SSS), purified from the soluble portion of cell extracts, is divided into many subfamilies. Commonly studied SSS isoforms are annotated as SSI, SSIIa, SSIIb, SSIIc, SSIIIa, SSIIIb, SSIVa, and SSIVb.³⁶ These isoforms are classified based on the sequence similarity and their functions. Although all SSS catalyze the same enzymatic reaction, they have different preferences for the size of the substrate (chain length), and they extend their substrate up to a specific size. For example, SSI prefers shorter oligosaccharides and can only extend them to DP (degree of polymerization) of 8 to 12. SSII enzymes are known to extend their substrate up to DP of 13-25. SSIII and SSIV are involved in the initiation of starch biosynthesis and control the number of starch granules.³³

There exists another isoform of starch synthase, purified from the insoluble portion of the cell extract. These isoforms are granule-bound starch synthases, GBSSI, and GBSSII. The fundamental difference between SSS and GBSS is that SSS extends shorter oligosaccharides up to a limited size and is responsible for the formation of amylopectin.³⁶ However, GBSS stays bound to the granule of starch and extends oligosaccharides much longer to form amylose. Studies show that the elongation of oligosaccharides happens concurrently in the case of GBSS.³⁷ While GBSS's act processively, sequentially adding glucose units to a single growing chain, SSS's appear not to be processive, dissociating from the glucan chain upon each addition of a glucose unit.

The third step (Figure 2.3) in starch or glycogen biosynthesis is the branching of glucan chains. Branching enzymes (BE) are responsible for the formation of branches which forms the amylopectin region of the starch.⁹ First, the enzyme hydrolyzes the 1,4-glycosidic bond of the substrate (donor chain) and forms a covalent bond with the hydrolyzed oligosaccharide. Then the branching enzyme transfers the donor chain to a second oligosaccharide chain (acceptor) by forming a 1,6-glycosidic bond¹. CAZY (Carbohydrate Active enZymes) database classifies the branching enzyme as a Glycoside Hydrolases (GH), similar to amylases.^{34,35} The starch branching enzymes also have isoforms with different preferences toward the size of the transferred chain (donor chain) and the substrate for branching (amylose or amylopectin).¹

Besides ADPGPPase, SS, and SBE, plants have debranching enzymes that are involved in the biosynthesis of starch by removing incorrectly placed branches in order to maintain the structure of the starch.³⁸ Plants optimize the structure of the starch they produce by fine-tuning the expression level of all the enzymes involved in the biosynthesis of starch (ADPGPPase, isoforms of SSS, isoforms of GBSS, isoforms of BE, and DBE).³⁶ Figure 2.4 shows the variety of these enzymes from different species. There are also the starch kinases and phosphorylases which are

involved in the biosynthesis of starch.⁴⁰ Plants utilize the vast diversity of these enzymes to produce starch molecules that are structurally optimized for the tissue and environmental condition. Understanding the function of all these enzymes individually and when working together will enhance our understanding of starch biosynthesis.

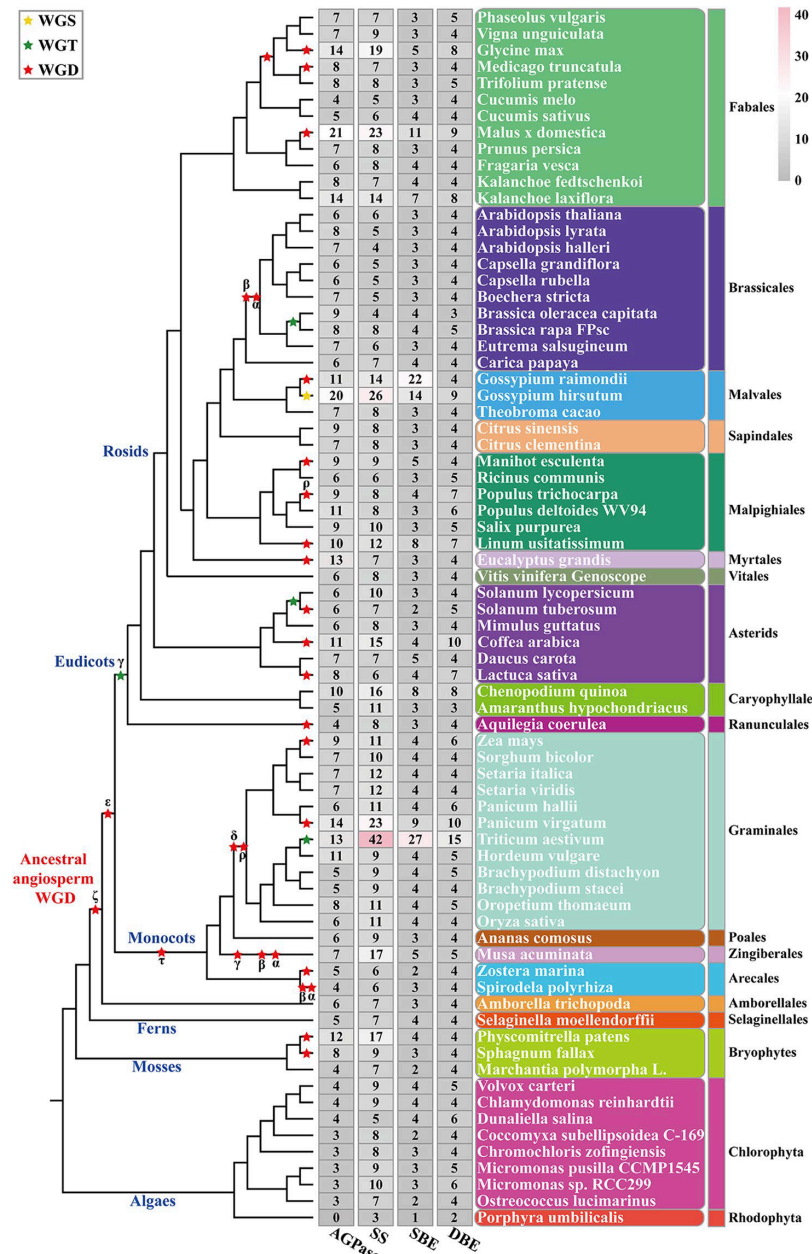


Figure 2. 4 Organization of ADPGPPase, SS, SBE, and DBE enzymes in plants. Stars represent all polyploidy events. Whole genome duplication, WGD (red), Whole genome triplication, WGT (green), Whole genome sextuplication, WGS (yellow). Total number of isoforms of the four-core enzyme families³⁹.

2.1.3. Application and relevance

Understanding the biosynthesis of starch has many benefits, from industry to health. For decades, starch has been used in food, textile, and paper industries. In food industries, starch is used as a thickener and a binder.^{41,42} In the cooking process, amylopectin helices melt, and by absorbing water, starch granules swell. Then, complete disintegration leads to dissolving in water. Later as it cools down, linear chains start aggregating and forming a gel.^{43,44} This process makes starch tolerant of the freeze-thawing processes. Controlling the gelatinization of starch is one of the essential goals in food industries. Various forms of modified starch, waxy, and resistant starch have related properties favored by the food industries. Waxy starch is amylose-deficient and synthesized by removing GBSS genes, which are necessary for the production of amylose. This type of starch gelatinizes readily.^{45,46} On the other hand, resistant starch is 50 - 90% amylose (compared to native starch with 20 - 30% amylose), which makes it useful in the production of sweets. Besides, resistant starch has applications in adhesives, paper, and textile industries. Resistant starch, unlike the native form of starch, is digested in the small intestine and transformed into short-chain fatty acids that are useful for colon health.⁴⁷⁻⁴⁹ Researchers try to manipulate the enzymes that are involved in the production of essential precursors for starch biosynthesis, such as adenylate kinase.⁵⁰ Also, there have been studies on post-harvest chemical modification of starch.⁵¹⁻⁵³ Recently, modified starch molecules are being studied for developing drug delivery systems,⁵⁴⁻⁵⁷ biodegradable materials,^{54,58-61} and the production of biofuels.⁶²⁻⁶⁵

2.1.4. Starch Synthases and GBSS

GBSS belongs to the GT5 (glycosyltransferase 5) family and, similar to starch synthases, processes the elongation of oligosaccharide chains.³⁵ GBSS is responsible for the biosynthesis of

long glucose chains of the amylose, whereas other starch synthases only extend short chains in the amylopectin region. This enzyme catalyzes the formation of alpha-1,4-glycosidic bonds continuously.⁹

There are 17 structures of starch/glycogen synthases available in the Protein Data Bank⁶⁶ (PDB) from eight species (*Oryza sativa*, *Hordeum vulgare*, *Arabidopsis thaliana*, *Pyrococcus abyssi*, *Rhizobium radiobacter*, *Escherichia coli*, *Cyanobacterium* sp., *C. Cyanophora paradoxa* and yeast). All these enzymes share a similar structure, a twin Rossmann-fold, where the active site is located between two Rossmann folds. The C-terminal Rossmann fold is quite similar among all these enzymes as demonstrated in figure 2.5. However, the N-terminal Rossmann fold varies significantly by the presence of different loops. An amino acid sequence (KXGGL), located in the active site, is conserved among these enzymes.⁶⁷ (Figure 2.5 C)

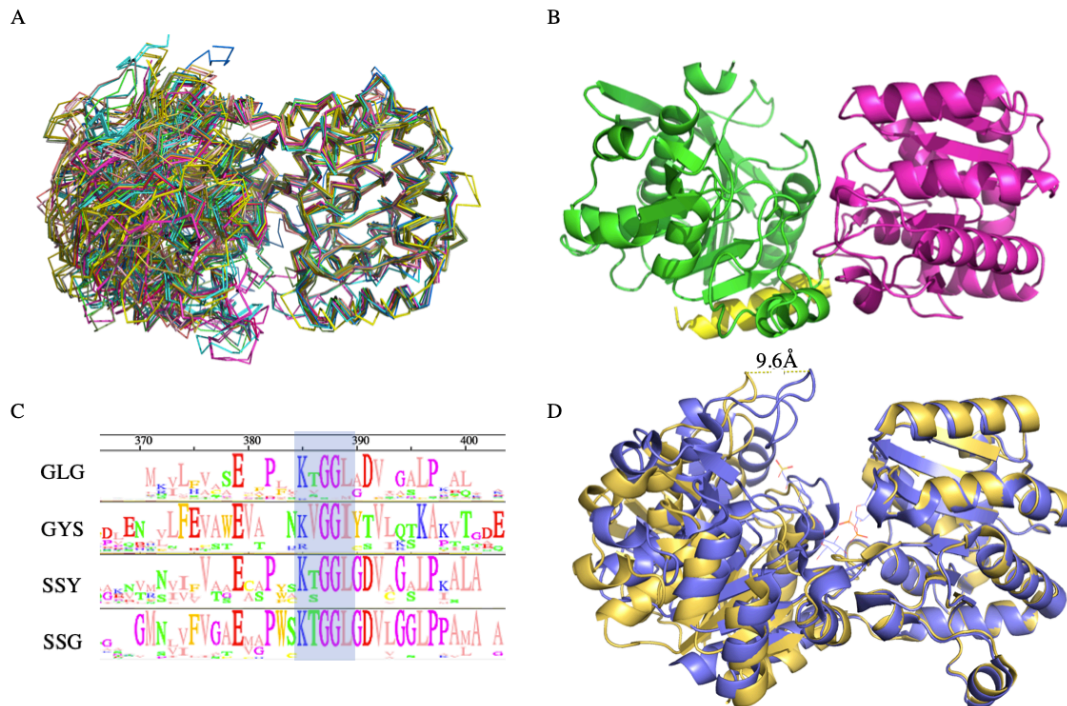


Figure 2. 5 Structural features in SS, GS, and GBSS. (A) Overlay of OsGBSSI, EcGS, RrGS, PaGS, HvSSI, CyGBSS, CpGBSSI, and AtSSIV. (B) rice GBSSI, two Rossmann folds (C-terminus in pink and N-terminus in green), C-terminus α -helix. (C) Conserved KXGGL active site sequence, UniProt classification: GLG and GYS for glycogen synthases, SSY for soluble starch synthases, and SSG for granule bound starch synthases.

A Study⁶⁸ from our research group showed that the two domains form open and closed conformations. *E. coli* glycogen synthase crystallizes in the open conformation in the absence of its substrates. However, in the presence of its substrates or substrate mimics it crystallizes in the closed form. Two Rossmann folds in *E. coli* glycogen synthase (N-terminus and C-terminus) are rotated 15.2° (Figure 2.5 D) in the closed form relative to its open conformer.⁶⁸

Considering the crystal structures of *Escherichia coli* glycogen synthase, several essential amino acids are identified. In *Escherichia coli* glycogen synthase, Lys-305, Arg-300, and the backbone amide from Gly-18 and Thr-382 form hydrogen bonds with the phosphate group of ADP. Tyr-355 forms a π - π stacking interaction with the purine ring. All these interactions, including the hydrogen bonds between the ribose sugar and Lys-15 and Asp-21, hold the ADP portion of ADP-1-glucose in the active site. As a result, the glucose unit is positioned at the bottom of the active site. The glucose unit forms hydrogen bonds with Glu-377, Asn-162, and His-161.⁶⁸

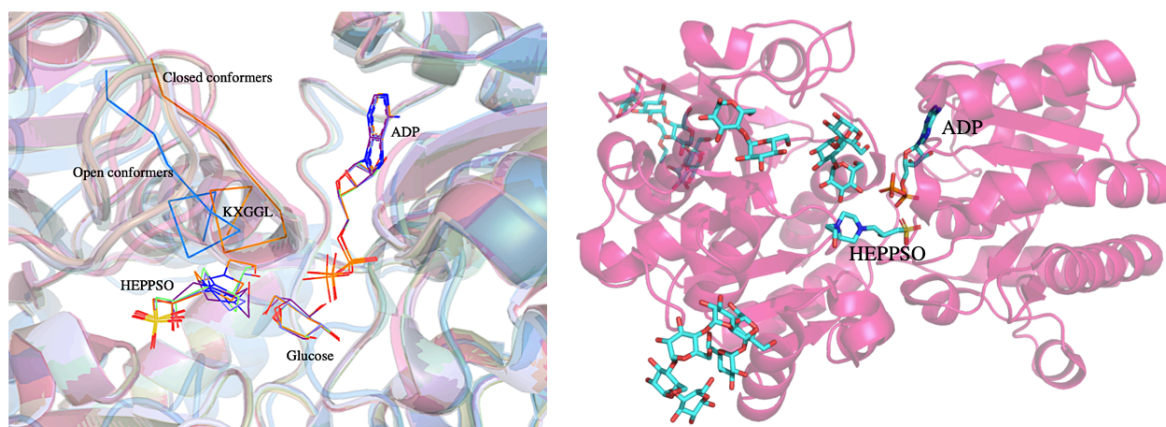


Figure 2. 6 Relative location of KXGGL motif in the active site of EcGS open and closed conformers (left) and surface oligosaccharide binding sites (right)

As mentioned earlier, GT5³⁵ enzymes transfer the glucose unit by retention of chirality. The location of glucose at the bottom of the active site leaves only one face of the C1 carbon available for a nucleophilic attack. A proposed mechanism includes a double-displacement S_N² reaction

where a nucleophile provided by the enzyme replaces the ADP group. Then, the hydroxyl group from the non-reducing end of an oligosaccharide replaces the enzyme to complete the reaction. A second mechanism postulates the formation of a carbocation by the dissociation of the phosphate group of the ADP unit.⁶⁸ Besides the active site, the *E. coli* glycogen synthase structure shows several surface binding sites for oligosaccharides (Figure 2.6). These sites are generally in the N-terminal half, whereas ADP-1-glucose interacts with the C-terminal half.

The GBSS enzyme is less studied than glycogen synthases and two structures of rice GBSSI were available in the form of apo and also complexed with ADP.⁶⁹ In recent years, two more GBSS structure were added to protein data bank from *Cyanobacterium sp.* and *Cyanophora paradoxa*.⁶⁷

Rice GBSSI contains a disulfide bond which connects the two halves of the enzyme. Fujimoto and colleagues investigated the importance of the disulfide bond in rice GBSSI. C337V mutant of rice GBSSI do not have the disulfide bond and the resulting protein is insoluble. Generally, plants that belong to the grass family, such as rice, wheat, rye, barley, sorghum, and corn, have the disulfide bond.⁶⁹ Starch synthases and glycogen synthases have also been studied biochemically. Since the byproduct of the reaction is ADP, it can be coupled to pyruvate kinase and lactate dehydrogenase to oxidize NADH.⁷⁰ (Figure 2.7)

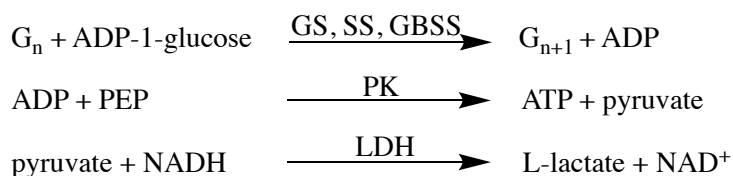


Figure 2. 7 An enzyme coupled assay for biochemical study.

There are many studies using the enzyme coupled assay for biochemical characterization of SS and GS.^{70–74} Recombinant proteins of GS and SS can be used for these assays. However, in all cases where GBSS is biochemically studied, GBSS was harvested from the plant bound to the starch granule.^{75–77} Zeeman and colleagues identified a small protein from *Arabidopsis thaliana*

which is involved in starch biosynthesis, particularly the production of amylose. PTST (Protein Targeting to STarch) is conserved in plant species. Knockout mutations of PTST in *Arabidopsis thaliana* produce of an amylose-free starch. Besides, these mutants had less starch bound GBSS protein. Initial sequence analysis of this specific protein showed that it contains a CBM48 (carbohydrate binding module 48) domain and a coiled-coil domain. Further mutational studies identified an α -helix on the GBSS from *Arabidopsis thaliana* that is likely to interact with the coiled-coil domain of PTST. They proposed that PTST is an anchoring protein which upon interacting with GBSS provides a CBM domain for the enzyme and the resulting complex binds to starch. However, PTST dissociates after GBSS is bound to starch (amylose).⁷⁸ CBM binding domains are common among enzymes interacting with carbohydrates. All starch and glycogen branching enzymes and debranching enzyme are equipped with a CBM domain.

2.1.5. The Branching Enzymes

Most branching enzymes belong to the GH-13 (glycosylhydrolase-13) family, similar to α -amylase enzymes,¹ though a few branching enzymes are GH-57 enzymes. These enzymes process the oligosaccharides, synthesized by the starch synthase isoforms, to form branched oligosaccharides. They form a glycoprotein intermediate by hydrolysis of an α -1,4-glycosidic bond. The reducing end of the glycan (donor chain) is covalently attached to an aspartate residue in the active site. Branching enzymes transfer the donor chain to an acceptor glycan by forming an α -1,6-glycosidic bond. The outcome of this process is a branched glycan, which is the basic unit of the amylopectin region.⁷⁹

Branching enzymes consist of a central $(\beta/\alpha)_8$ -TIM barrel catalytic domain (Figure 2.8) annotated as an amylase domain. Generally, branching enzymes have a CBM48 (carbohydrate-

binding module 48) in the N-terminus and a β -sandwich fold (amylose C-terminal domain) in the C-terminus. Bacterial glycogen branching enzyme might have several CBM or β -sandwich domains proceeding with the CBM48 domain.⁸⁰ The general CBM48-amylose catalytic domain-amylose C-terminus architecture is common among the GH13 family, where rice branching enzymes belong. A completely different fold also exists for branching enzymes that belong to the GH57 family.⁸¹

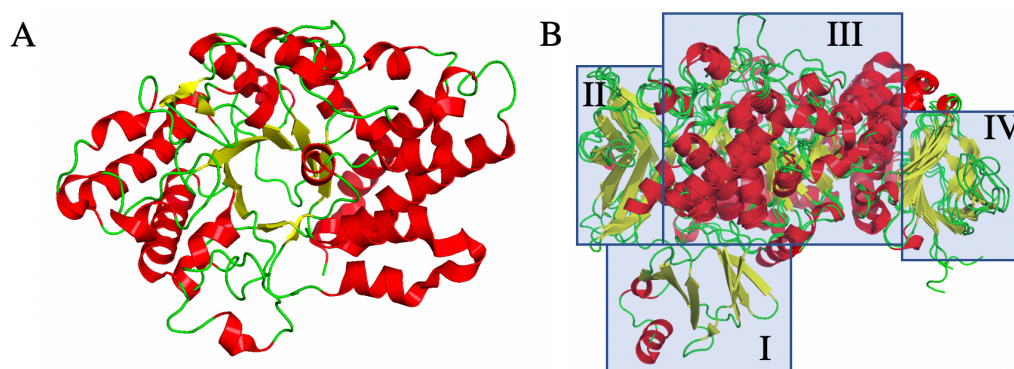


Figure 2. 8 Architecture of branching enzymes in GH13 family. (A) central $(\beta/\alpha)_8$ -barrel catalytic domain, (B) four structural motifs of branching enzymes, N-terminus β -sandwich domain common to bacterial glycogen branching enzymes I, CBM48 II, amylase catalytic domain III, C-terminus β -sandwich domain IV. (PDBID: 5gqy, 4lpc, 5clw, 3vu2 for cyanobacteria, E. coli, human, and rice respectively)

All available branching enzyme structures show a diverse range of surface binding sites. Since the substrate of branching enzymes is long oligosaccharides (amylose and amylopectin), many surface binding sites are available. Figure 2.9 shows the distribution of surface binding sites in branching enzymes.^{82–85} The characteristic of these surface bound oligosaccharides is not entirely understood, and it is unknown whether they are part of an acceptor chain or the donor chain. The active site residue for the branching enzymes is an aspartate located in one of the β -strands of the TIM barrel fold. Many of the surface-bound oligosaccharides are more than 10 Å away from the active site. There are even oligosaccharides on the opposite side of the enzymes. Another feature

on the surface of the branching enzymes is the diversity of the loop structures that commonly interact with the surface-bound oligosaccharides. All branching enzymes have four distinct conserved regions⁸⁶ (Figure 2.12) I, II, III, IV. Region II contains the active site aspartate residue.

When a branching enzyme binds to the first oligosaccharide (donor chain) and hydrolyzes it to form the covalent-bound intermediate, the exact size of the hydrolysis product varies. The biochemical characterization of the branched products shows a different distribution (chain length distribution or CLD). Nevertheless, there are several limits (Figure 2.10) to the size of the donor chain in branching enzymes. Generally, branching enzymes do not transfer chains smaller than maltohexaose. Besides, if the donor chain has been already branched, the second hydrolysis does not happen less than six glucose units away from the branching point. Finally, if the acceptor chain has already been branched, the second branching does not happen less than six glucose units away from the previous branch point.⁸⁷

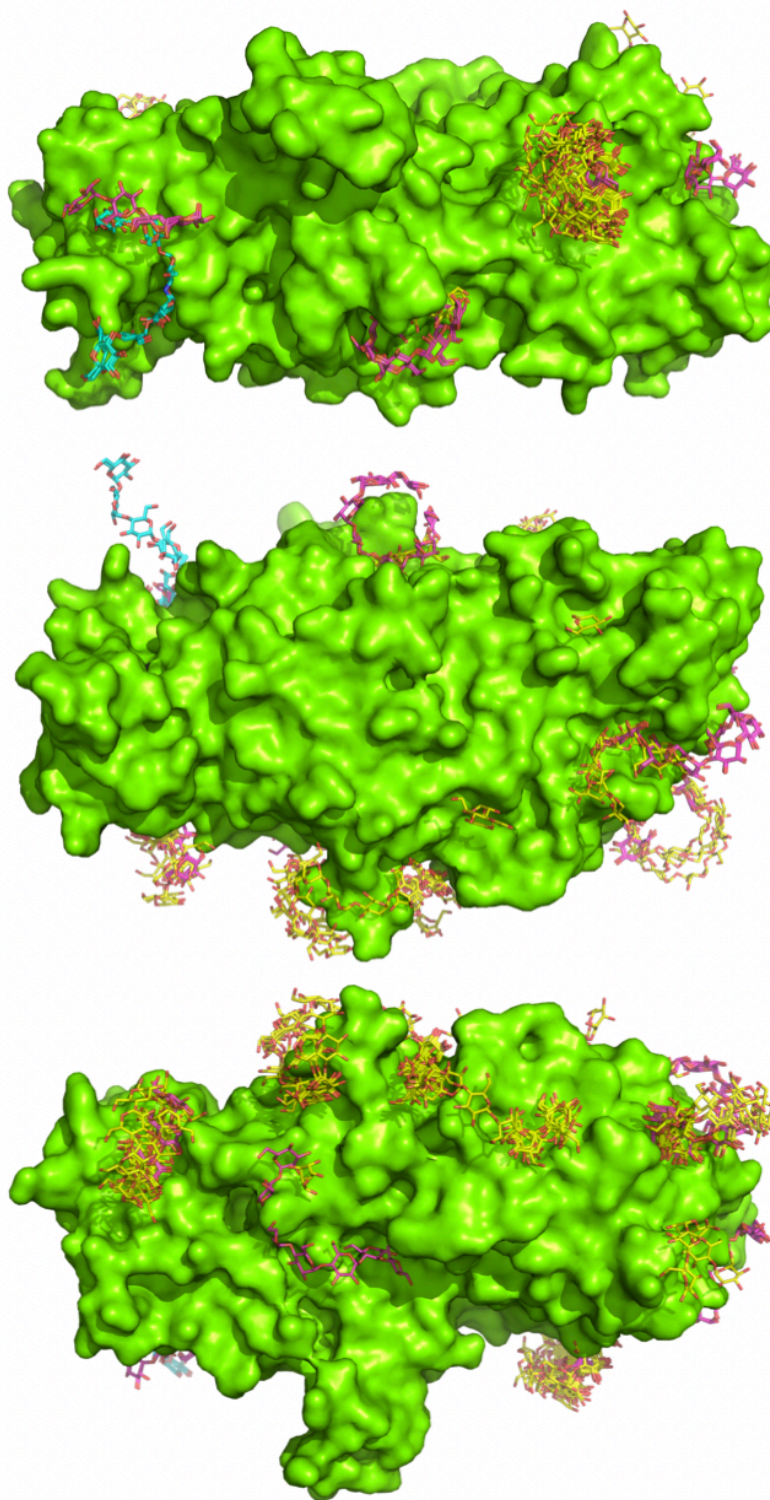


Figure 2. 9 Surface binding sites in branching enzymes. Three different views of a branching enzyme and distribution of surface binding sites. EcBE (green surface), oligosaccharides (EcBE: yellow, CyBE: magenta, HsBE: blue)

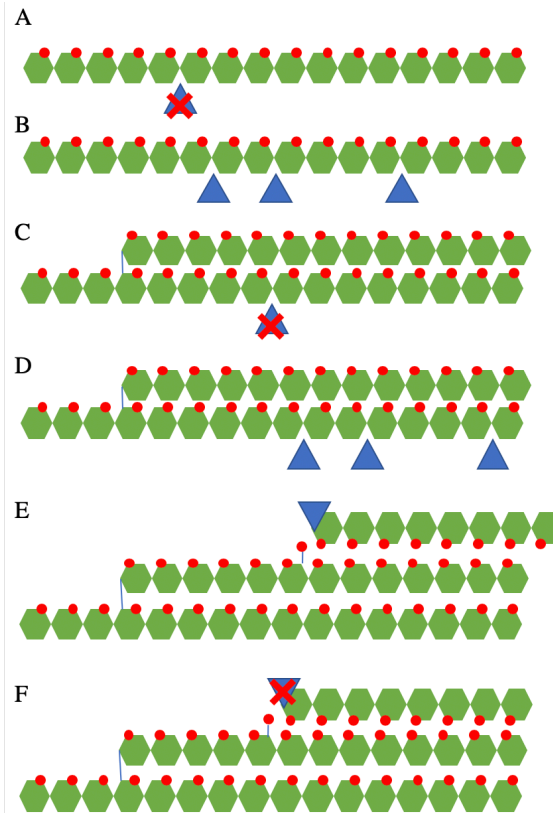


Figure 2.10 Limitations of branching enzymes. (A) less than six glucose units to non-reducing end of donor chain, no hydrolysis, (B) at least six glucose units to non-reducing end of donor chain, successful hydrolysis, (C) less than six glucose units to the branch point on a branched donor chain. No hydrolysis. (D) at least six glucose units to the branch point on a branched donor chain, successful hydrolysis. (E) at least six glucose units to the branching point of an acceptor chain, successful transfer, (F) less than six glucose units to the branch point of an acceptor chain.

Similar to starch synthases, branching enzyme has many isoforms in plant species where they differ only in the chain length distribution. Only 22 branching enzymes from different species and isoforms have been biochemically studied, and their CLD profiles are known. Figure 2.11 shows a classification of them based on the similarities of their CLD.

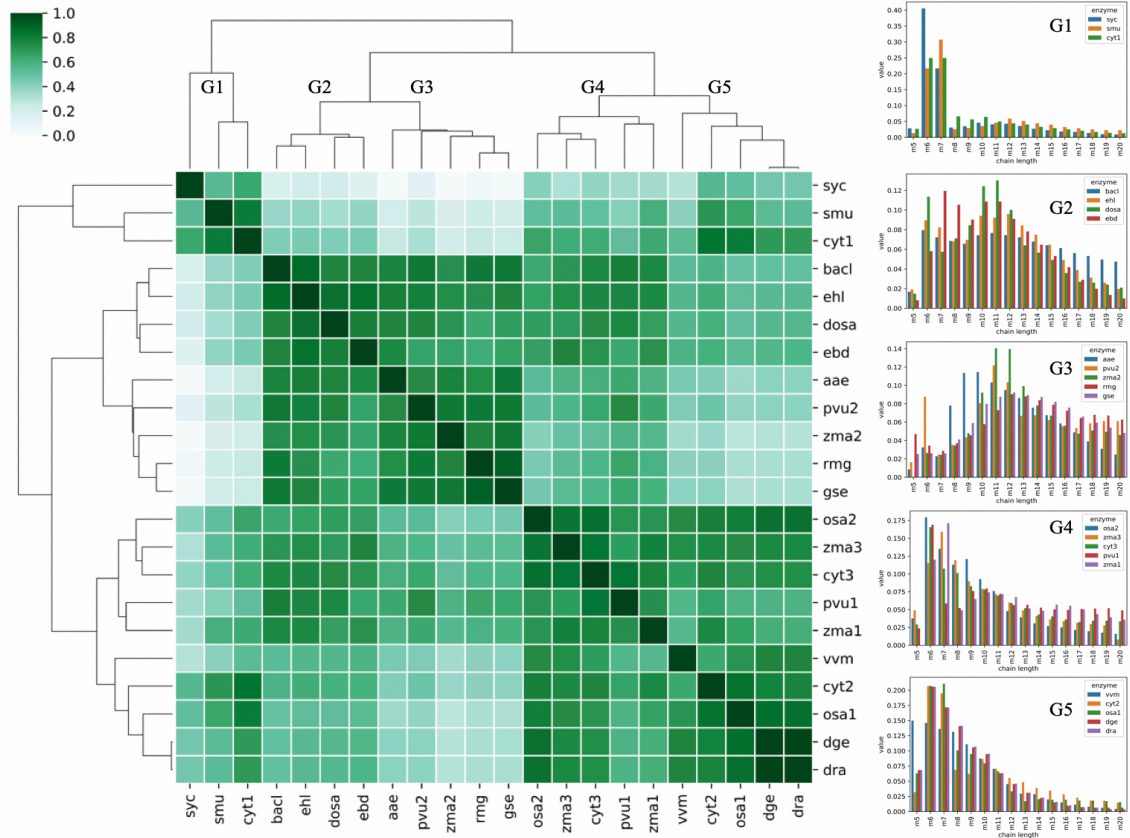


Figure 2. 11 A hierarchical classification of chain length distribution for 22 branching enzymes. Bacterial (aae⁸⁸, bac1⁸⁹, dge⁹⁰, dra⁹⁰, rmg⁹¹, smu⁹², syc⁹³, vvm⁹⁴, ehl⁹⁵, gse⁹⁶, cyt1⁹³, cyt2⁹³, cyt3⁹³, ebd⁹⁷), Plants (osa1⁸⁶, pvu1⁹⁸, pvu2⁹⁸, zma1⁹⁹, zma2⁹⁹, zma3⁹⁹, osa2⁸⁶, dosa⁸⁶). Left: Similarity of the chain length distribution for 22 branching enzyme CLD profiles (0, least similar. 1, identical). Right: five distinct clusters of CLD profiles obtained by hierarchical clustering method.

Some enzymes transfer chains of six or seven glucose units, whereas other branching enzymes transfer a wider range of chain lengths. Plants optimize the structure of starch they make by fine-tuning the expression of different branching enzyme isoforms.¹⁰⁰

Rice has four branching enzyme isoforms, RBE1, RBEIIa, RBEIIb and RBEIII. The RBE1 chain length distribution shows that it transfers more hexamers and then chains of size eleven, while RBEII and RBEIII transfer maltohexaose and maltoheptaose units dominantly.⁸⁶

Sequence analysis of these isoforms reveals two insertions at the 147 and 541 regions (Figure 2.12, light blue). The first insertion (541 region) is eleven amino acids long and is present on

RBEIIa and RBEIIb. The second difference is in the loop starting at residue number 147. This loop is two amino acids longer for RBEI but the whole loop is 10 amino acid long.⁸⁶ There is a big difference between RBEIIa and IIb versus RBEI. RBEI transfers longer chains, IIb is almost exclusive for M6, while IIa is kind of in the middle.



Figure 2. 12 Sequence analysis of four rice branching enzyme isoforms. Conserved regions of I, II, III, IV common to all branching enzymes. CBM48 domain (blue),

2.2. Structure of rice GBSSI enzyme

Several new structures of rice GBSSI have been obtained. ADP-1-glucose-bound GBSSI in the closed conformation, GBSSI (reduced form) in its open conformation, open conformer of rice GBSSI and ADP complex, open conformer of rice GBSSI and UDP complex. The soaking of open conformer with ADP-1-glucose resulted in the GBSSI-ADP complex, where ADP is found in two different conformations in the interdomain space.

2.2.1. Structure of ADP-glucose bound Rice GBSSI (closed conformer)

The x-ray crystal structure of native rice GBSSI with ADP-glucose (Figure 2.13) was solved to 2.49Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.21 and 0.27, respectively. The asymmetric unit contains one protein chain.

The tertiary structure is similar to previously published structures. Both N-terminal and C-terminal domains have a Rossmann fold with six parallel β -strands forming a continuous β -sheet. Six α -helices surround the β -sheets. After obtaining the apo crystals, soaking them with ADP-1-glucose resulted in a ligand-bound structure.

Figure 2.13 shows the hydrogen bonds interacting with ADP-1-glucose. In the structure of the closed conformer, there is a disordered loop that includes residue 91 to 110. The equivalent of this loop from *e. coli* glycogen synthase is interacting with an oligosaccharide, and it is involved in guiding the non-reducing end of the substrate into the active site.

As mentioned earlier, starch synthases have open and closed conformers where the two Rossmann folds move toward and away from each other. Table x shows the interdomain gap for all the starch/glycogen synthase structures that are available.

Table 2.1 shows the interdomain gap for all the available starch/glycogen synthase structures (the rice GBSSI ADP-1-glucose complex (closed conformer) is used as reference)

Table 2. 1 Summary of available starch/glycogen synthase structures.

Enzyme	Species	Active site ligands	Interdomain Gap ^a	PDBID
GBSSI	<i>Oryza sativa</i>	-	1.8	3vue
GS	<i>Rhizobium radiobacter</i>	-	6.3, 6.5	1rzv
GS - C7S:C408S	<i>Escherichia coli</i>	-	3.9	3dlj
GS	<i>Pyrococcus abyssi</i>	-	4.2, 4.9, 6.0, 4.7, 5.0, 5.9	2bis, 3l01, 3fro
SSI	<i>Hordeum vulgare</i>	-	5.1	4hln
GS	<i>Rhizobium radiobacter</i>	ADP ^b	7.1, 7.3	1rzu
GBSSI	<i>Oryza sativa</i>	ADP	1.5	3vuf
GBSS	<i>Cyanobacterium sp.</i>	Acarbose ^c , ADP	1.7, 2.2, 2.6	6gnf
GBSSI	<i>Cyanophora paradoxa</i>	acarbose, ADP	3.2, 3.0	6gng
GS	<i>Escherichia coli</i>	ADP, Glucose	1.0, 1.1, 1.0	2qzs, 2r4t, 2r4u
GS - E377A	<i>Escherichia coli</i>	ADP, HEPPSO	0.8	3cop
GS - E377A	<i>Escherichia coli</i>	ADP, oligosaccharides	1.2	3cx4
GS	<i>Escherichia coli</i>	ADP and DGM ^d	0.8	3guh
GS	<i>Pyrococcus abyssi</i>	2-dehydroxyGlucose	1.9	3fro
SSIV	<i>Arabidopsis thaliana</i>	acarbose, ADP	0.7, 0.8	6gne

^a the distance between the alpha carbons of X residue in the conserved sequence of KXGGL. Rice GBSSI-ADP-glucose complex structure is used as reference.

^b Adenosine diphosphate

^c Anti-diabetic drug

^d Carbocation species of glucose after the dissociation of ADP

The glycogen synthase from *E. coli* was crystallized in the closed form only in the presence of both of the substrates. The disulfide bond in rice GBSSI locks the enzyme in the closed form, and as a result, crystallization of the apo protein in the closed form is possible. Crystallizing the apo form is essential for a soaking experiment. The ADP-1-glucose bound GBSSI is obtained by soaking the apo crystals in the substrate solution.

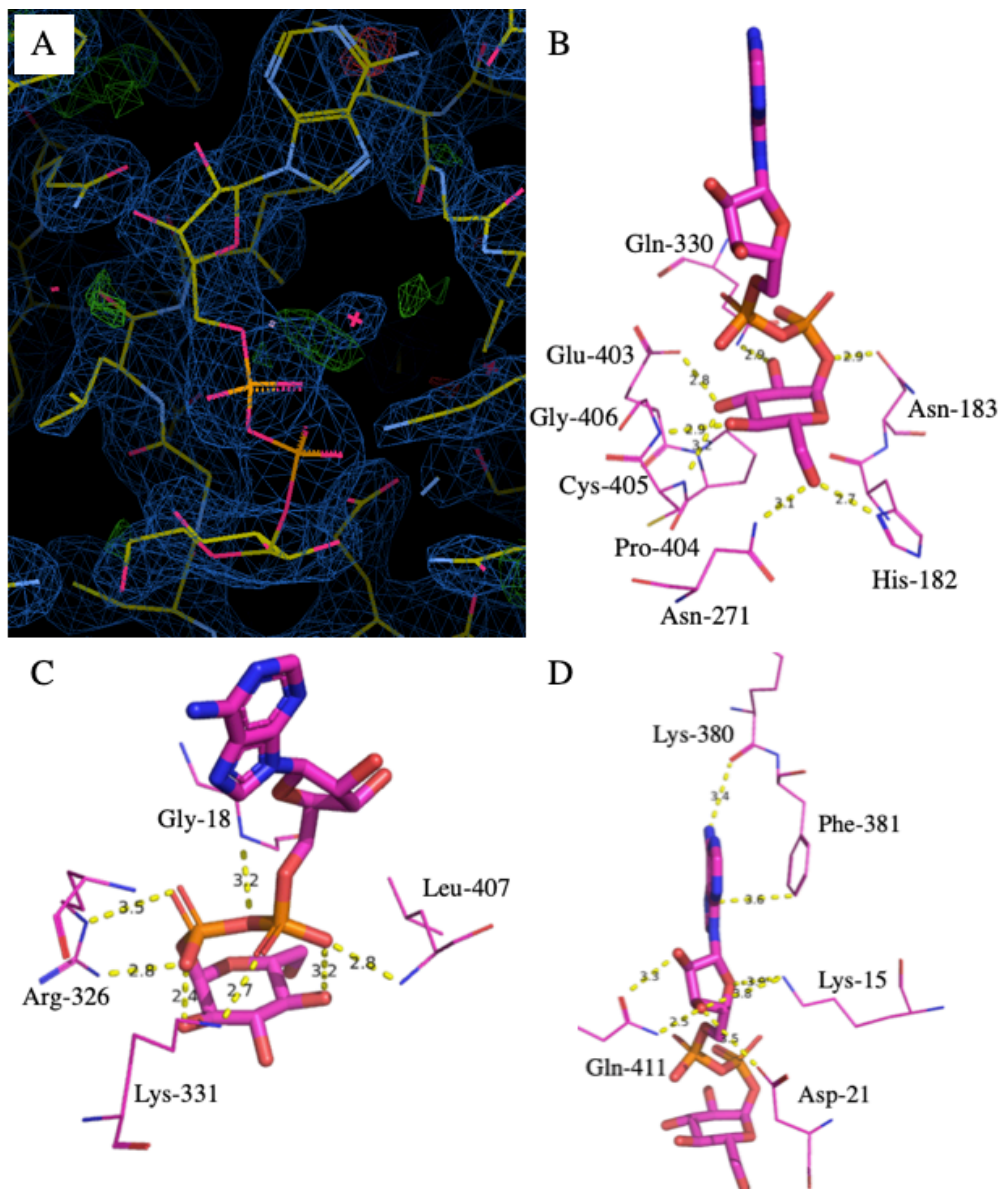


Figure 2. 13 ADP-glucose bound Rice GBSSI, (A) ADP-glucose electron density. (B) hydrogen bond interactions between the enzyme and glucose unit. (C) hydrogen bond interactions between the enzyme and diphosphate. (D) hydrogen bond interactions between the enzyme and adenine unit.

Table 2. 2 Data collection and refinement statistics for GBSSI-ADP-1-glucose complex.

Data statistics	
Space group	P 4 3 2
No. of chains/ASU	1
Unit cell dimensions (Å, °)	a = 153.771, b = 153.771, c = 153.771 $\alpha = 90, \beta = 90, \gamma = 90$
Resolution range (Å) (outer shell)	48.63 - 2.494 (2.583 - 2.494)
Unique reflections (outer shell)	22292 (2172)
Completeness (%) overall (outer shell)	99.87 (100.00)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	22274 (2172)
Reflections used for R-free	2000 (195)
R-work	0.2126 (0.2686)
R-free	0.2719 (0.3261)
Number of non-hydrogen atoms/ solvent	3882/49
Protein residues	486
RMS (bonds) (Å)	0.005
RMS (angles) (Å)	1.15
Ramachandran plot (%) favored/allowed/outliers	94.61/4.98/0.41
Rotamer outliers (%)	0.00
Average B-factor/macromolecules/solvent	59.09/59.16/55.94

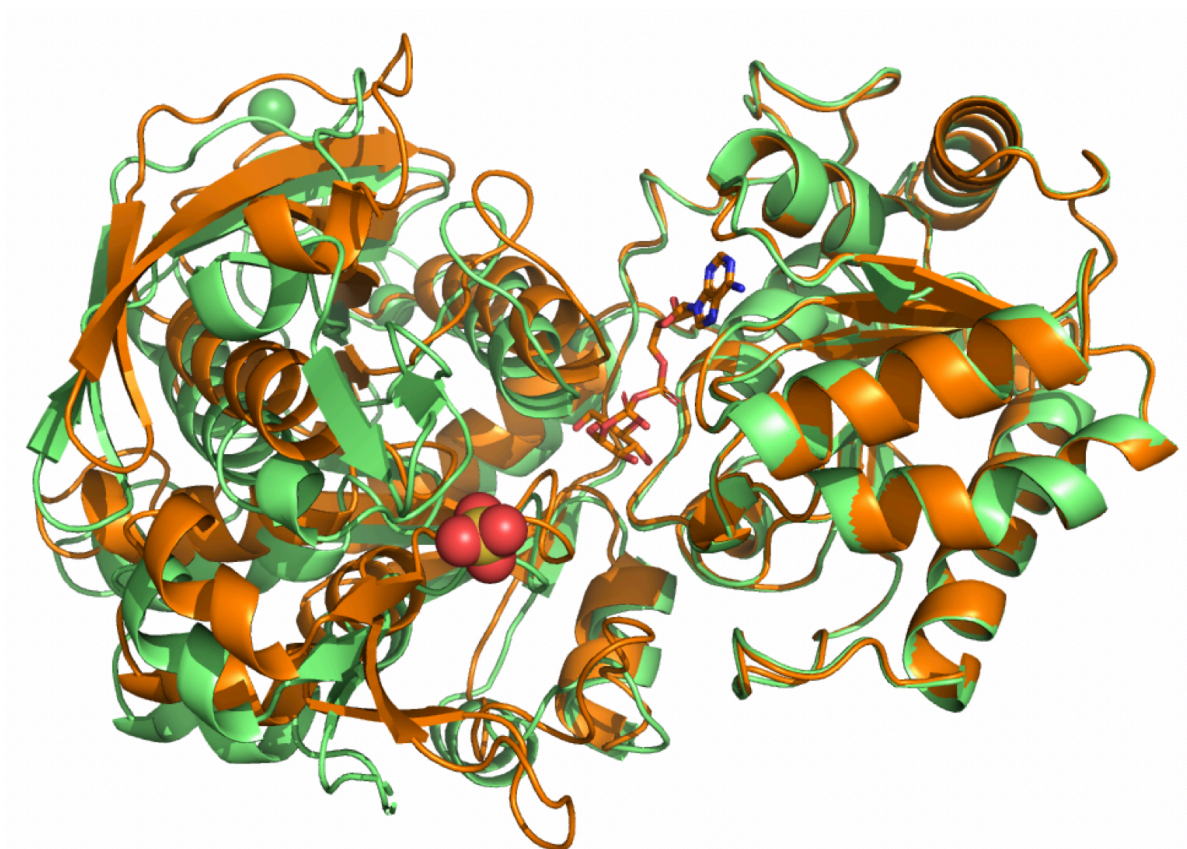


Figure 2. 14 Overlay of open and closed conformers of rice GBSSI.

2.2.2. Structure of Rice GBSSI (open conformer)

The x-ray crystal structure of open conformer of rice GBSSI, the reduced form (Figure 2.14) was solved to 1.59 Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.18 and 0.20, respectively. The asymmetric unit contains one protein chain. Tertiary structure is similar to previously published structures except for the interdomain gap. The N-terminus has moved away from the C-terminus by 3.5 Å. The disordered loop from the closed conformer is ordered and forms a β -hairpin fold.

Table 2. 3 Data collection and refinement statistics for GBSSI-ADP-1-glucose complex.

Data statistics	
Space group	P 31 2 1
No. of chains/ASU	1
Unit cell dimensions (Å, °)	a = 85.598, b = 85.598, c = 140.585 $\alpha = 90, \beta = 90, \gamma = 120$
Resolution range (Å) (outer shell)	31.76 - 1.594 (1.651 - 1.594)
Unique reflections (outer shell)	79761 (7610)
Completeness (%) overall (outer shell)	99.52 (95.64)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	79698 (7552)
Reflections used for R-free	2002 (193)
R-work	0.1805 (0.3303)
R-free	0.2089 (0.3332)
Number of non-hydrogen atoms/ solvent	4520/ 468
Protein residues	513
RMS (bonds) (Å)	0.010
RMS (angles) (Å)	1.32
Ramachandran plot (%) favored/allowed/outliers	97.05/ 2.75 / 0.20
Rotamer outliers (%)	0.23
Average B-factor/macromolecules/solvent	28.68/ 26.74/ 45.41

2.2.3. Structure of Rice GBSSI (open conformer) and ADP complex 1

The x-ray crystal structure of the open conformer of rice GBSSI with ADP (Figure 2.15) was solved to 1.6 Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.18 and 0.22, respectively. The asymmetric unit contains one protein chain. Tertiary structure is similar to the apo open conformer. The soaking experiment with ADP-1-glucose resulted in an ADP bound structure. Either the glucose unit of the ADP-1-glucose is disordered, or it is hydrolyzed. The binding site of ADP is similar to the binding site of ADP-1-glucose in the closed conformer.

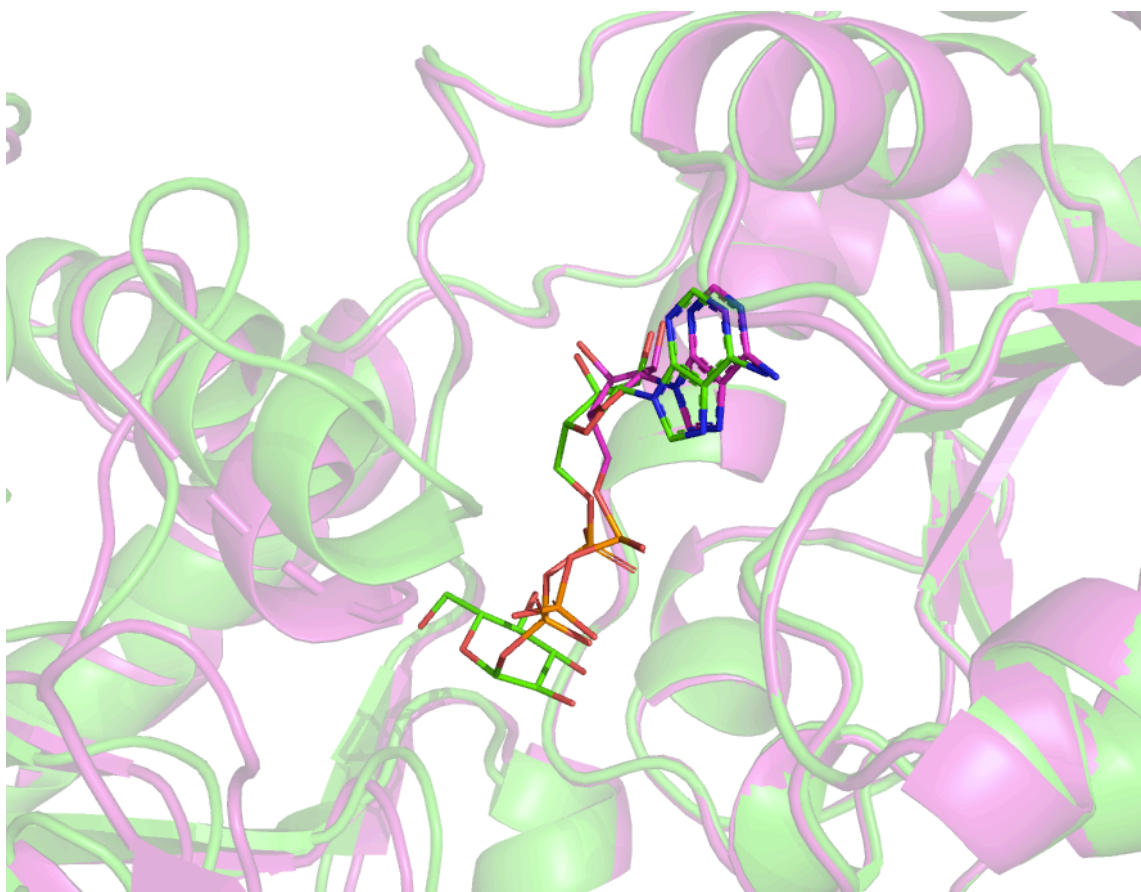


Figure 2. 15 ADP and GBSSI (open conformer) complex overlaid on closed form GBSSI and ADP-1-glucose complex. Closed form (green), open form (magenta)

Table 2. 4 Data collection and refinement statistics for GBSSI-ADP-1-glucose complex.

Data statistics	
Space group	P 31 2 1
No. of chains/ASU	1
Unit cell dimensions (Å, °)	a = 86.511, b = 86.511, c = 140.918 $\alpha = 90, \beta = 90, \gamma = 120$
Resolution range (Å) (outer shell)	29.29 - 1.601 (1.659 - 1.601)
Unique reflections (outer shell)	80556 (7921)
Completeness (%) overall (outer shell)	99.64 (99.36)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	80542 (7919)
Reflections used for R-free	2009 (195)
R-work	0.1852 (0.3003)
R-free	0.2207 (0.3096)
Number of non-hydrogen atoms/ solvent	4595/ 539
Protein residues	507
RMS (bonds) (Å)	0.007
RMS (angles) (Å)	1.28
Ramachandran plot (%) favored/allowed/outliers	97.60/ 1.80/ 0.60
Rotamer outliers (%)	0.00
Average B-factor/macromolecules/solvent	28.94/ 26.64/ 44.95

2.2.4. Structure of Rice GBSSI (open conformer) and ADP complex 2

The x-ray crystal structure of the open conformer of rice GBSSI with ADP (Figure 2.16) was solved to 1.7 Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.18 and 0.21, respectively. The asymmetric unit contains one protein chain. The ADP molecule has different conformation in the binding site. Similarly, the glucose unit of ADP-1-glucose is not observed in this structure. The absence of the glucose unit is either because of disordered glucose or the hydrolysis of the ADP – glucose bond.

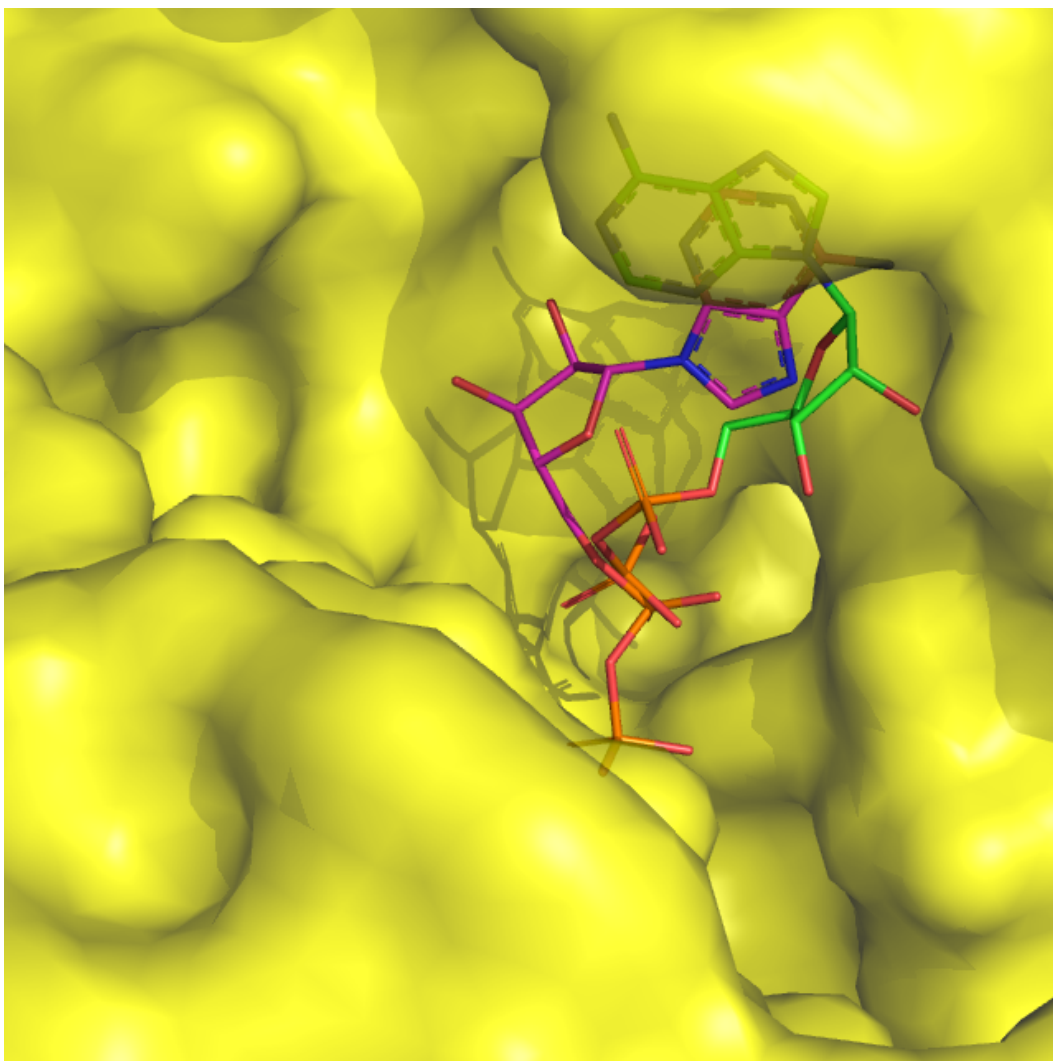


Figure 2. 16 Different binding mode for the ADP in the active site of GBSSI (open conformer). New binding mode (green)

Table 2. 5 Data collection and refinement statistics for GBSSI-ADP-1-glucose complex.

Data statistics	
Space group	P 31 2 1
No. of chains/ASU	1
Unit cell dimensions (Å, °)	a = 86.508, b = 86.508, c = 141.498 $\alpha = 90, \beta = 90, \gamma = 120$
Resolution range (Å) (outer shell)	31.99 - 1.7 (1.761 - 1.7)
Unique reflections (outer shell)	67521 (6623)
Completeness (%) overall (outer shell)	99.24 (98.25)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	67498 (6613)
Reflections used for R-free	2006 (201)
R-work	0.1816 (0.2812)
R-free	0.2121 (0.2979)
Number of non-hydrogen atoms/ solvent	4545/ 552
Protein residues	504
RMS (bonds) (Å)	0.007
RMS (angles) (Å)	1.24
Ramachandran plot (%) favored/allowed/outliers	97.79/2.01/ 0.20
Rotamer outliers (%)	0.48
Average B-factor/macromolecules/solvent	31.92/ 29.44/ 47.41

2.2.5. Structure of Rice GBSSI (open conformer) and UDP complex

The x-ray crystal structure of open conformer of rice GBSSI with UDP (Figure 2.17) was solved to 1.7 Å resolution by molecular replacement methods and refined to R and R_{free} values of 0.18 and 0.21, respectively. The asymmetric unit contains one protein chain. The binding mode for UDP is similar to the new ADP binding mode.

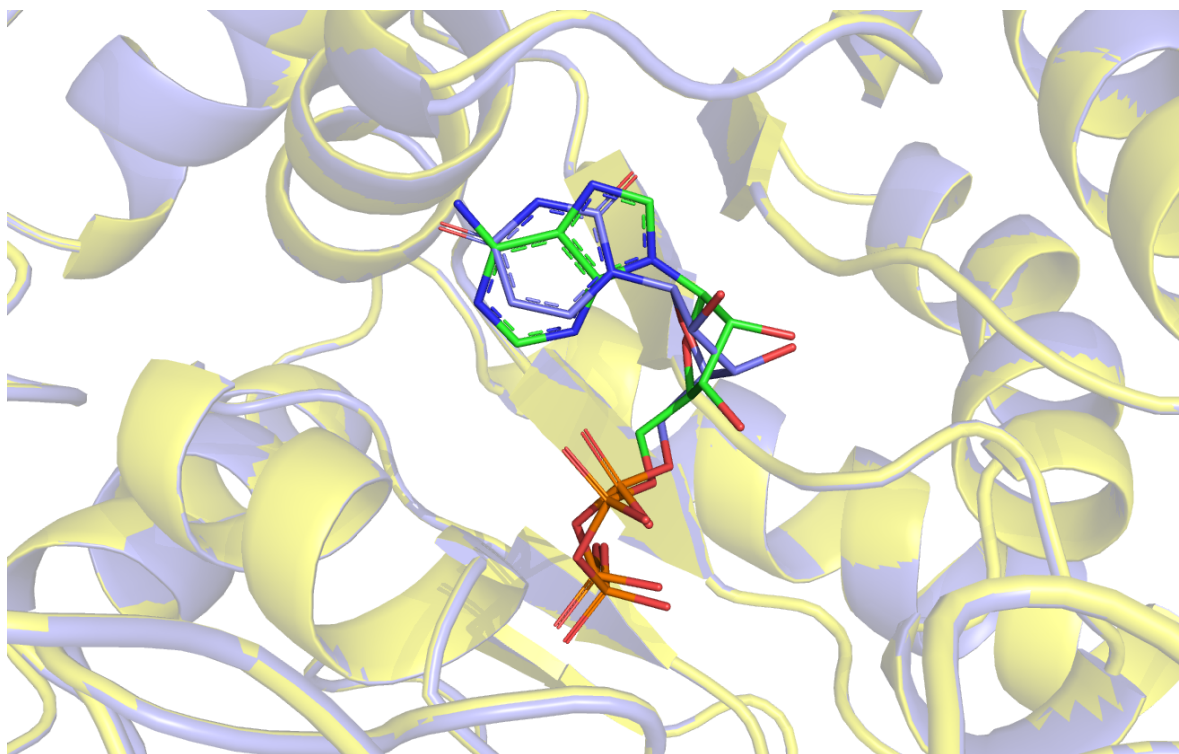


Figure 2. 17 Binding mode for ADP and UDP are similar. ADP (green), UDP (purple).

Table 2. 6 Data collection and refinement statistics for GBSSI-UDP complex.

Data statistics	
Space group	P 31 2 1
No. of chains/ASU	1
Unit cell dimensions (Å, °)	a = 86.508, b = 86.508, c = 141.498 $\alpha = 90, \beta = 90, \gamma = 120$
Resolution range (Å) (outer shell)	31.99 - 1.7 (1.761 - 1.7)
Unique reflections (outer shell)	67521 (6623)
Completeness (%) overall (outer shell)	99.24 (98.25)
Mean I/sigma(I)	
R-merge/R-meas/ R-pim	
Refinement	
Reflections used in refinement	67498 (6613)
Reflections used for R-free	2006 (201)
R-work	0.1810 (0.2827)
R-free	0.2129 (0.2853)
Number of non-hydrogen atoms/ solvent	4543/ 542
Protein residues	504
RMS (bonds) (Å)	0.007
RMS (angles) (Å)	1.23
Ramachandran plot (%) favored/allowed/outliers	97.78/1.61/ 0.60
Rotamer outliers (%)	0.72
Average B-factor/macromolecules/solvent	31.70/ 29.27/ 48.36

2.3. Rice Branching Enzyme

2.3.1. Structure of rice branching enzyme and maltododecaose complex

The crystal structure of the rice branching enzyme maltododecaose complex was obtained by a former student (Remie Fawez). Of the three M12 molecules bound to rBE1 (Figure 2.18), two occupy sites identified in the M5-bound rBE1 structure previously described⁸² (labeled sites 1 and 3)

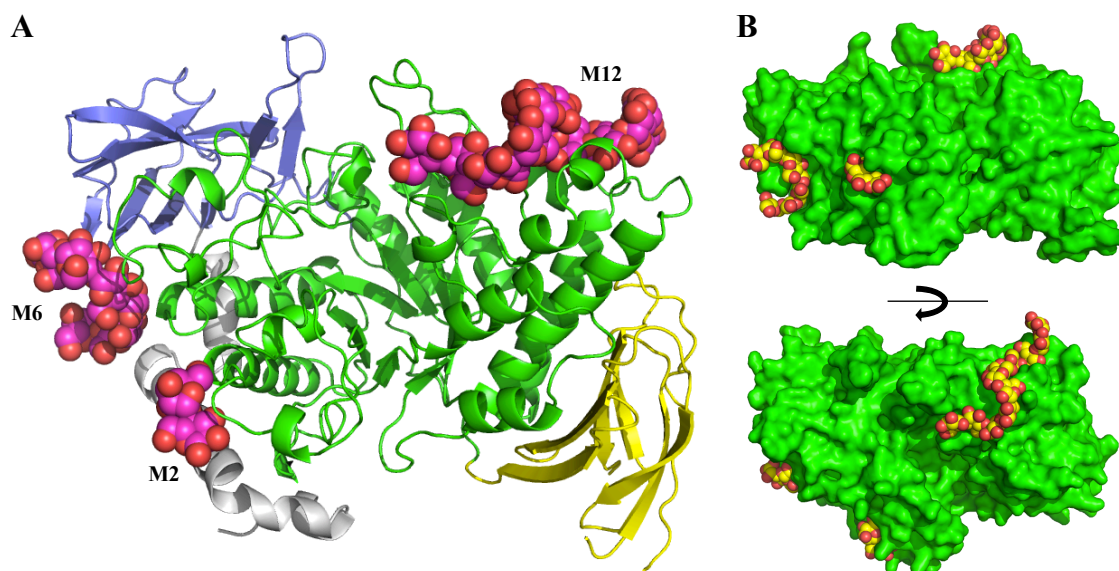


Figure 2. 18 Surface binding sites in rice branching enzyme. (A) Cartoon representation of rBEI and three bound glucans. N-terminal domain (Gray), CBM48 domain (Slate Blue), Catalytic Core domain (Green), C-terminal domain (Yellow) and Glucans (Red and Magenta) depicted as space filling models. (B) Surface representation of rBEI (green) with bound glucans depicted as space filling models (C atoms, yellow, O atoms, red). Top, oriented as in A, bottom, rotated approximately 90° along the horizontal axis.

Six glucose units are ordered in site 1, and only two glucose units are ordered in site 3, but in both cases, the interactions between glucan and protein are similar to those seen in the previous structures (Figure 2.19). The Third molecule occupies a binding site heretofore not identified in

Branching Enzyme (site 4) and has all 12 glucose units ordered. It begins at the edge of the catalytic domain, advances toward the active site traversing the width of the catalytic domain. The glucan adopts a helical conformation (Figure 2.20), with six glucose units per turn. The majority of this glucan has a helical conformation similar to one chain of a glycogen double helix, but it deviates from this conformation as it approaches the active site.¹⁰¹

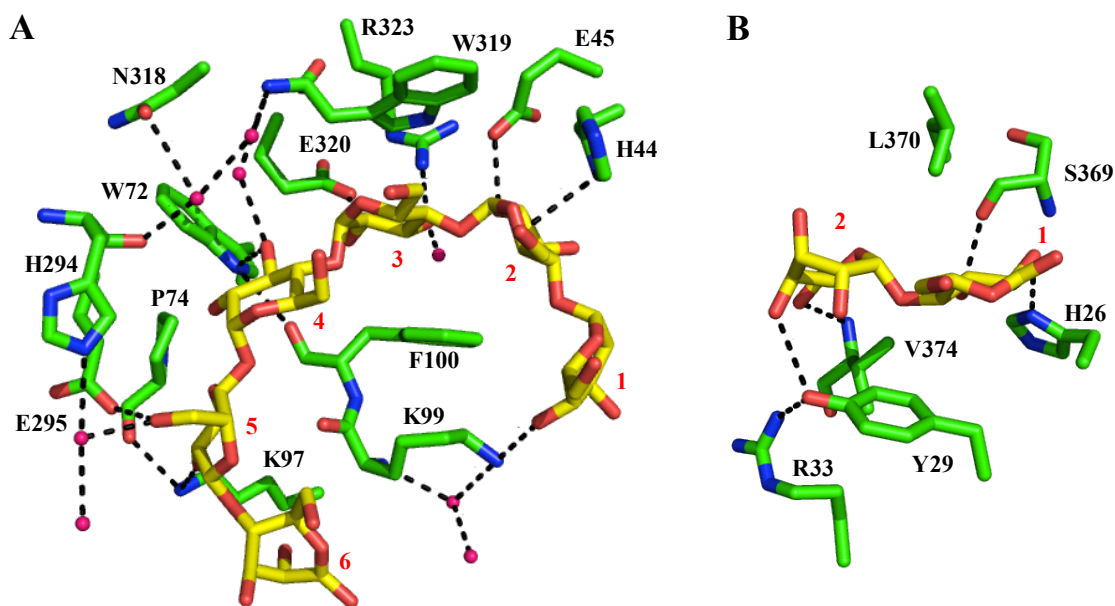


Figure 2. 19 Residues of rBEI interacting in site 1. (A) and site 3 (B). Glucans (Yellow), Residues (Green)

The surface of rBEI is predisposed toward binding the helical conformation, with aromatic stacking interactions (Tyr487 and Trp535) that serve to project the glucan away from the surface, and hydrogen-bonding interactions with the glucans that directly contact the surface (Figure 2.21).

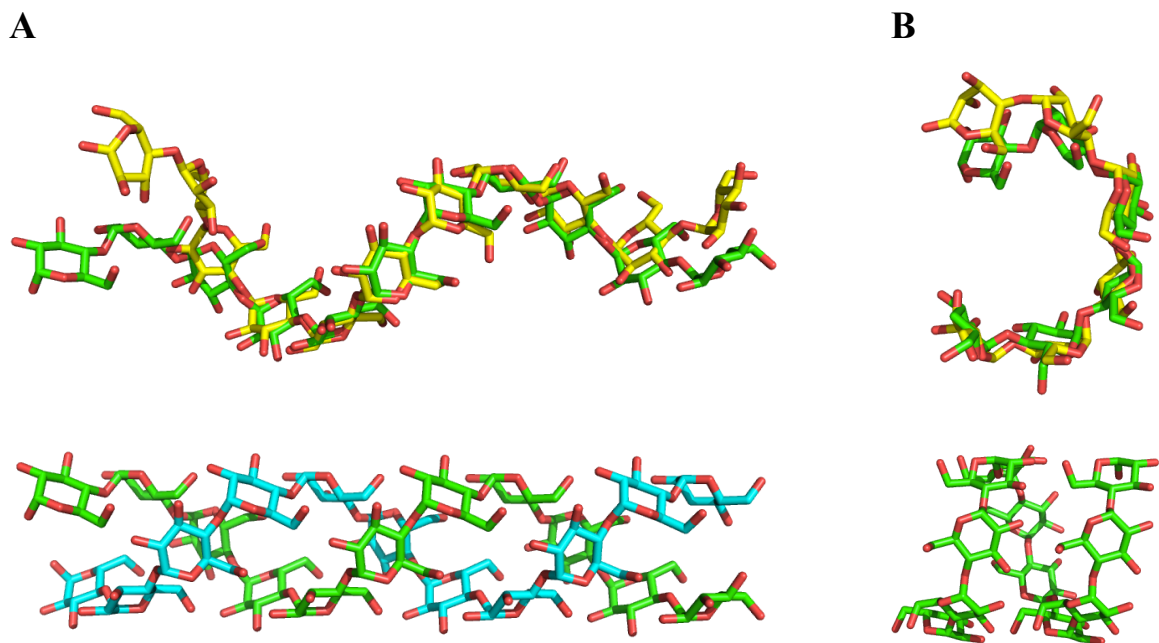


Figure 2. 20 Helical features of bound glucans to rBEI., rBEI bound glucans (Yellow), Reference glucans (Green) (A) Top, M12 from site 4 overlaid on one strand of a glycogen or amylopectin-like double helix. Bottom, the original double helix. (B) Top, M6 from site 1 overlaid on a model of an amylose single helix. Bottom, the amylose single helix model

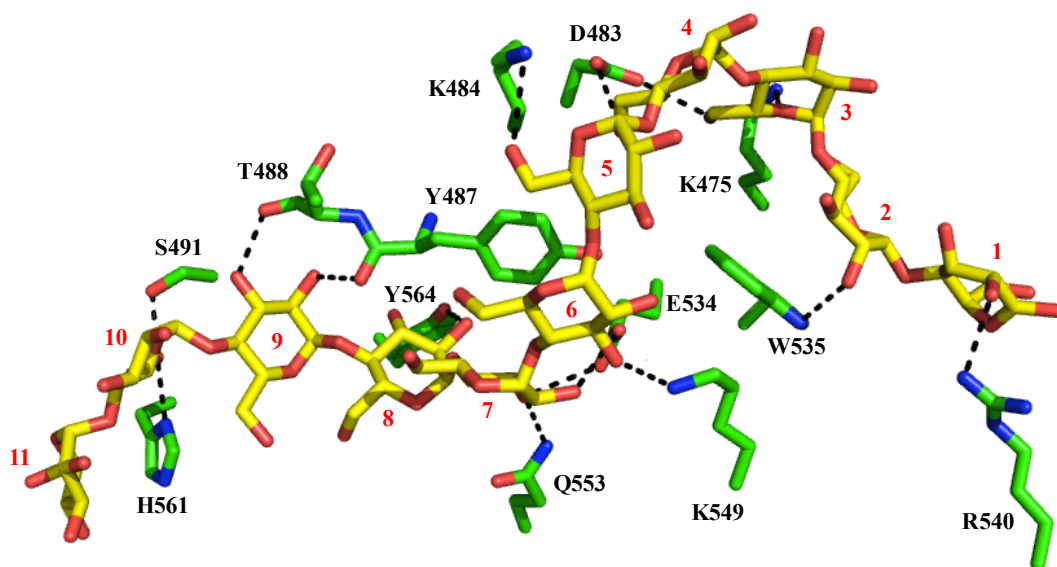


Figure 2. 21 Detailed interactions between rBEI and M12 occupying site 4.

The residues that directly interact with M12 in this binding site are highly conserved in plant and animal BE's, with 11 of the 19 residues that directly interact with M12 identical in virtually all plant and animal BE's (Figure 2.22, sequence alignment showing, plant BE1s, BEII's, human, drosophila, yeast, cyanobacteria).

		Residues 496 - 501	Residues 528 - 566
Plant BEI	Oryza ativa BEI	QSI V GDKTIAFLLMDK EM YTGMSDLQPASPTIN	NEFGHP EW IDFPRE-----GNNWSYDKCRRQ W SLVDT DH L-RYKY
	Zea mays BEI	QSI V GDKTIAFLLMDK EM YTGMSDLQPASPTID	NEFGHP EW IDFPRE-----GNNWSYDKCRRQ W SLVDT DH L-RYKY
	Solanum tuberosum BEI	QSI V GDKTIAFLLMDK EM YSGMSCLTDA5PVVD	NEFGHP EW IDFPRE-----GNNWSYDKCRRQ W NLADSE HL -RYKF
	Aegilops tauschii BEI	QSI V GDKTMAFLLMDK EM YTGMSDLQPASPTID	NEFGHP EW IDFPRE-----GNNWSYDKCRRQ W SLADID HL -RYKY
	Oryza ativa BEIIb	QAL V GDKTIAFWLMDK DM YDFMALDRPSTPTID	NEFGHP EW IDFPRA PQV -LPNGKFIPGNNNSYDKCRRRFDLGDAD YL -RYRG
Plant BEII	Zea mays BEIIb	QAL V GDKTIAFWLMDK DM YDFMALDRPSTPTID	NEFGHP EW IDFPRA PQV -LPNGKFIPGNNNSYDKCRRRFDLGDAD YL -RYRG
	Solanum tuberosum BEIIb	QAL V GDKTIAFWLMDK DM YDFMALDRPSTPTID	NEFGHP EW IDFPRA EQH -LSDG5VIPGNNSYDKCRRRFDLGDAD YL -RYRG
	Aegilops tauschii BEIIb	QAL V GDKTIAFWLMDK DM YDFMALDRPSTPTID	NEFGHP EW IDFPRA PQT -LPTGKVLPGNNNSYDKCRRRFDLGDAD FL -RYRG
	Arabidopsis thaliana BE2.2	QAL V GDKTIAFWLMDK DM YDFMAVDRA TPR VD	NEFGHP EW IDFPRA TDQH -LPDGRVIAGNNNSYDKCRRRFDLGDAD YL -RYRG
Animal and Yeast	Human BE	QAL V GKSLAFWLMD AE MYTN MSV LTPFTPVID	NEFGHP EW LD FP RK-----GNNESYHYARR QF HLTDD LL -RYKF
	Drosophila BE	QAL V GDKTIAFWLMD EM YTH MS TLSDSSV I ID	NEFGHP EW LD FP RV-----GNNDSYHYARR QW NLVD DD LL-KYKY
	Yeast BE	QAL V GDKTLAFWLMD AA MYTDM TV LKEPS I VID	NEFGHP EW LD FP NV-----NNGDSYKYARR QF NLAD DP LL-RYQN
		*****-*****-*****	*****-*****-*****
Prokaryotes	cyanothecae BE1	EIV H GKSNM L GKM-----PG-DEW	MEFGQWSEW-----NVW-GDLE W HL L LQ-YEP H Q
	cyanothecae BE2	EVV H LKGHLVN K M-----PG-DEW	MEFGQTSEW-----QEF-LDID W Y L LQ-YEP H KG
	Escherichia coli	EVV H GKKSILDRM-----PG-DAW	NEFAQGREW-----NHD-ASLD W HL L E-GGD N WH H G

Figure 2. 22 Sequence alignment for some plant BE1s, BEII's, human, drosophila, yeast, cyanobacteria and Escherichia coli in site 4.

However, little conservation is seen in bacterial BE's for this surface, indicating this binding site to be common to the eukaryotic BE's and distinct from the bacterial enzymes, including the starch-making cyanobacterial enzymes with bacteria-like BE's such as Cyanothecae (GH13_9).¹⁰²

Besides, when the structures of M12-bound BE1 are compared to the apo and M5-bound BE1 structures, little structural change is observed, with the exception of the flexible loop between residues 468 and 474 (This loop is either disordered or found in two distinct conformations in the other BEI structures). M12 binding causes this loop to adopt a conformation not seen previously.

Numerous residues in the loop make interactions with M12, necessitating the loop to adopt the orientation seen in the M12-bound structure. This loop (Figure 2.23) interacts with the glucose moiety found closest to the active site and may act as a "door" into the active site. No other large conformational changes are seen in the M12 binding site when all three rBEI structures are compared.

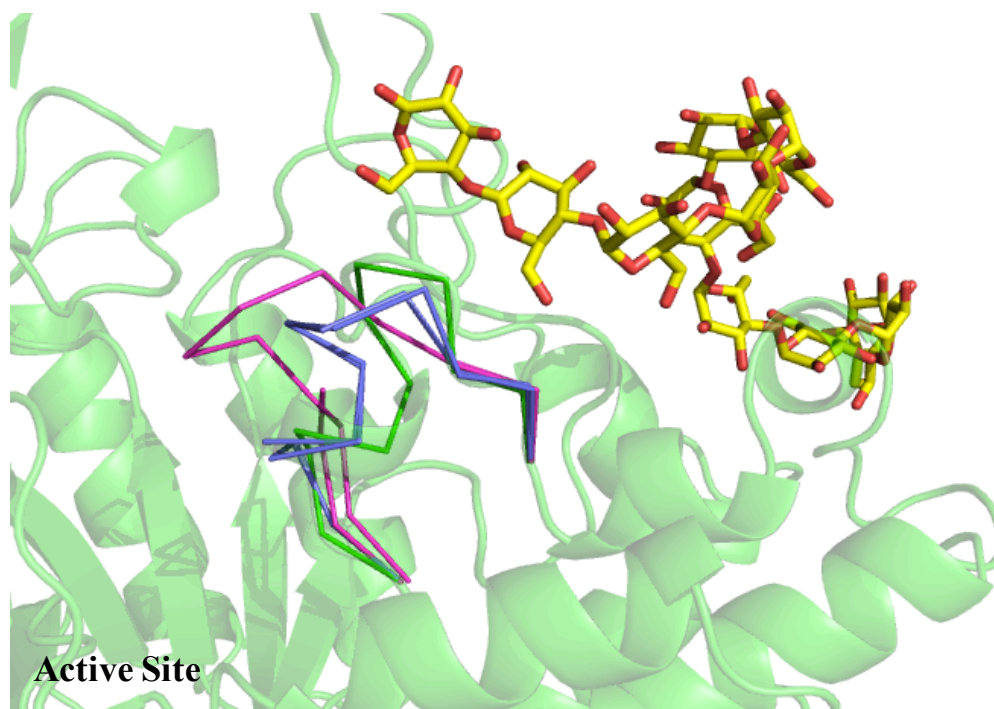


Figure 2. 23 Disordered loop (residues 468 and 474) adopts a new conformation upon site 4 M12 binding. Apo rBEI-chain A and B-3AMK (Magenta), M5 bound rBEI-chain A and B-3VU2 (Slate Blue) and M12 bound rBEI (Green)

2.3.2. Mutagenesis of rice branching enzyme

Iodine-assay is used for kinetic studies and evaluating the activities. Wildtype rice branching enzyme I branches more than 80% of the linear amylose substrate in less than 10 min. Mutating the active site aspartate, D344A and D344M, leads to an inactive enzyme. Also, residues in the CBM domain were mutated. Residues from region 1 and region 2 (highlighted in Figure 2.24) are mutated as well. Besides, different regions of the enzyme were mutated to find mutations that change the CLD profile. (Figure 2.25, 2.26, 2.27, 2.28) Region 1 mutants have lower activity than the wildtype, and they follow an opening toward the CBM domain (figure 2.24). For a glucan to enter the active site while it is bound to the CBM domain, it has to go through the opening between two loops (blue color in Figure 2.29). Interestingly, one of these loops is the same insertion that was mentioned before (the difference between the isoforms of rice branching enzyme).

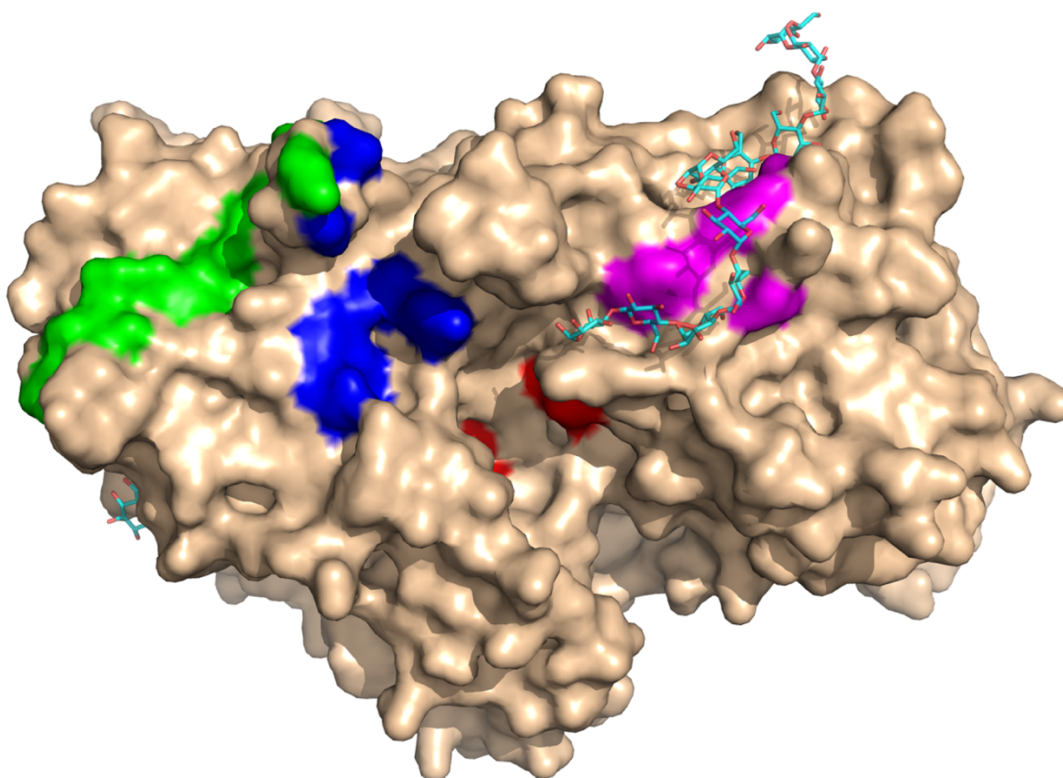


Figure 2. 24 Mutation map. Active site (red), region 1 (blue), CBM domain (green), region 2 (magenta)

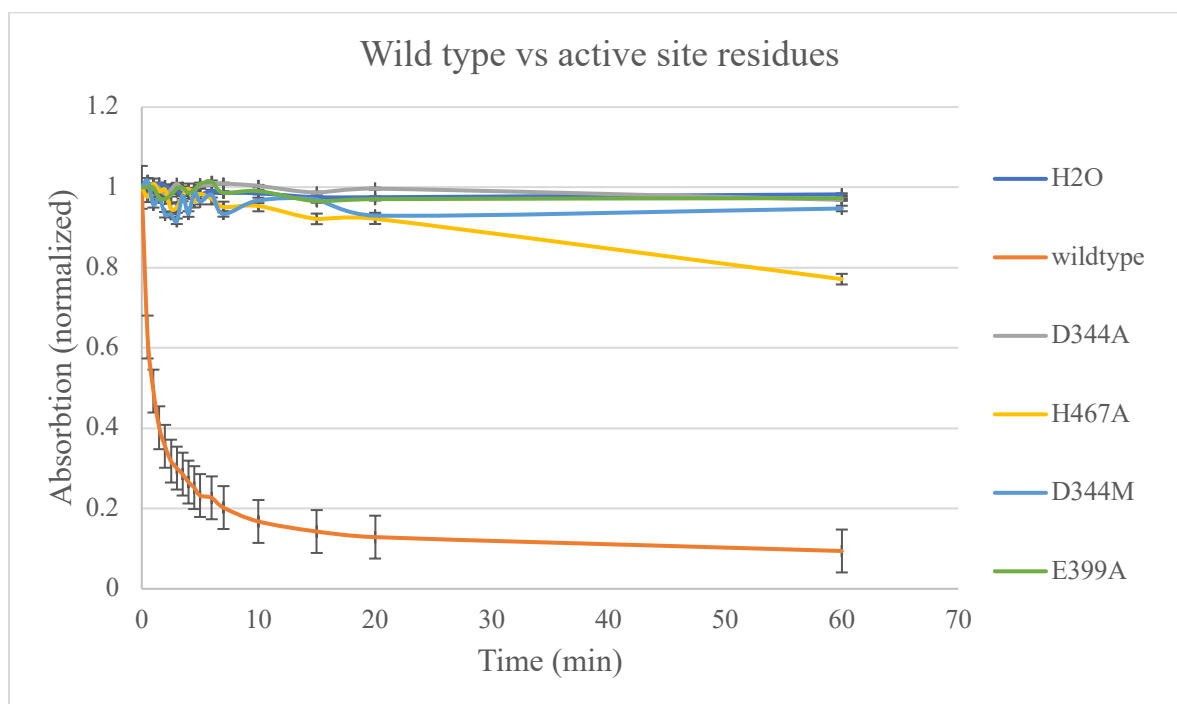


Figure 2. 25 Relative activity of the wildtype enzyme and active site mutants.

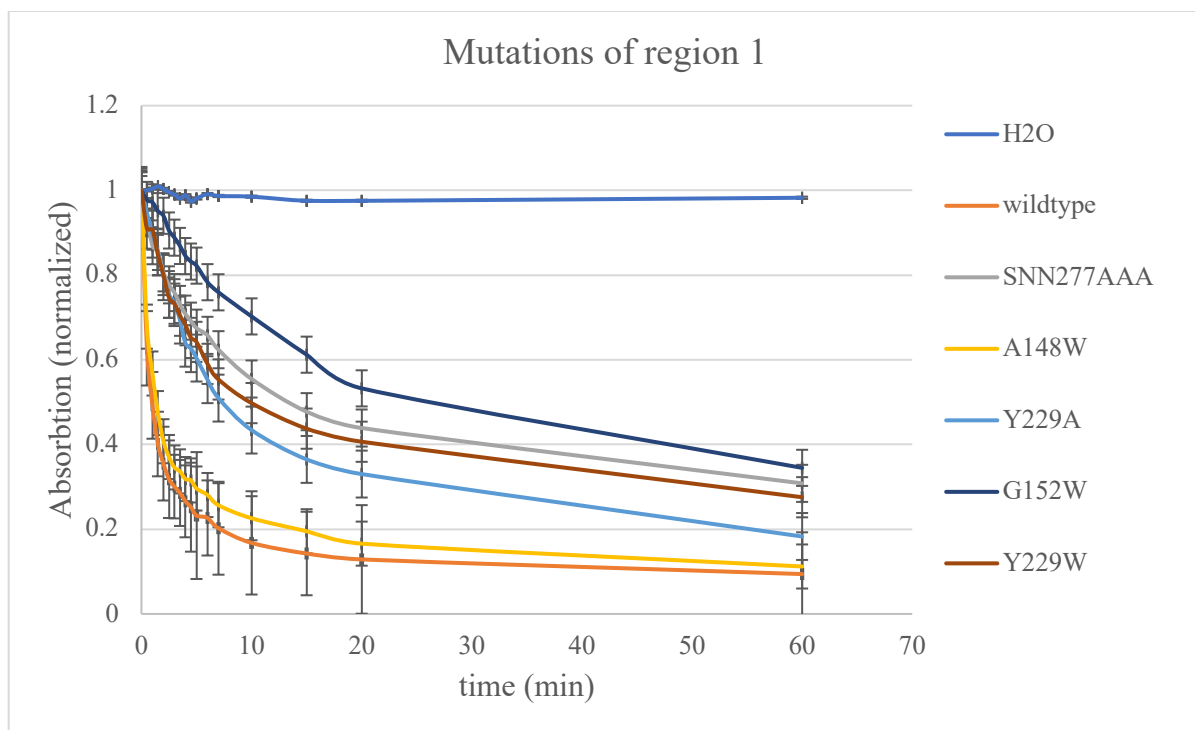


Figure 2. 26 Relative activity of the wildtype enzyme and region 1 mutants.

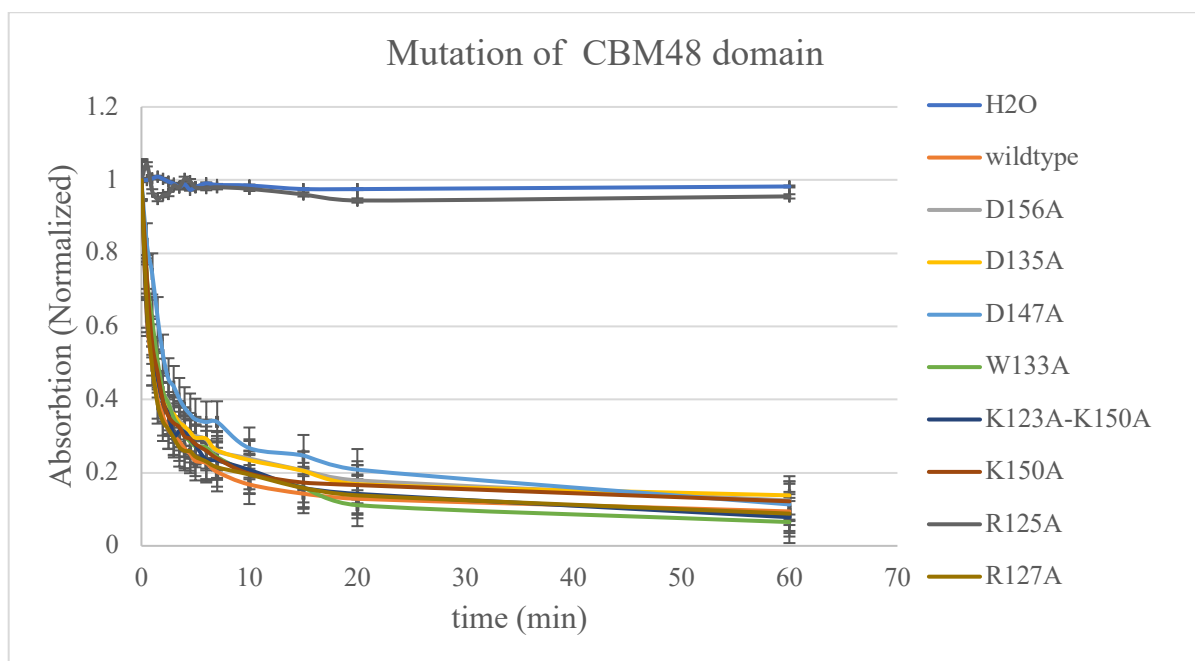


Figure 2. 27 Relative activity of the wildtype enzyme and CBM domain mutants.

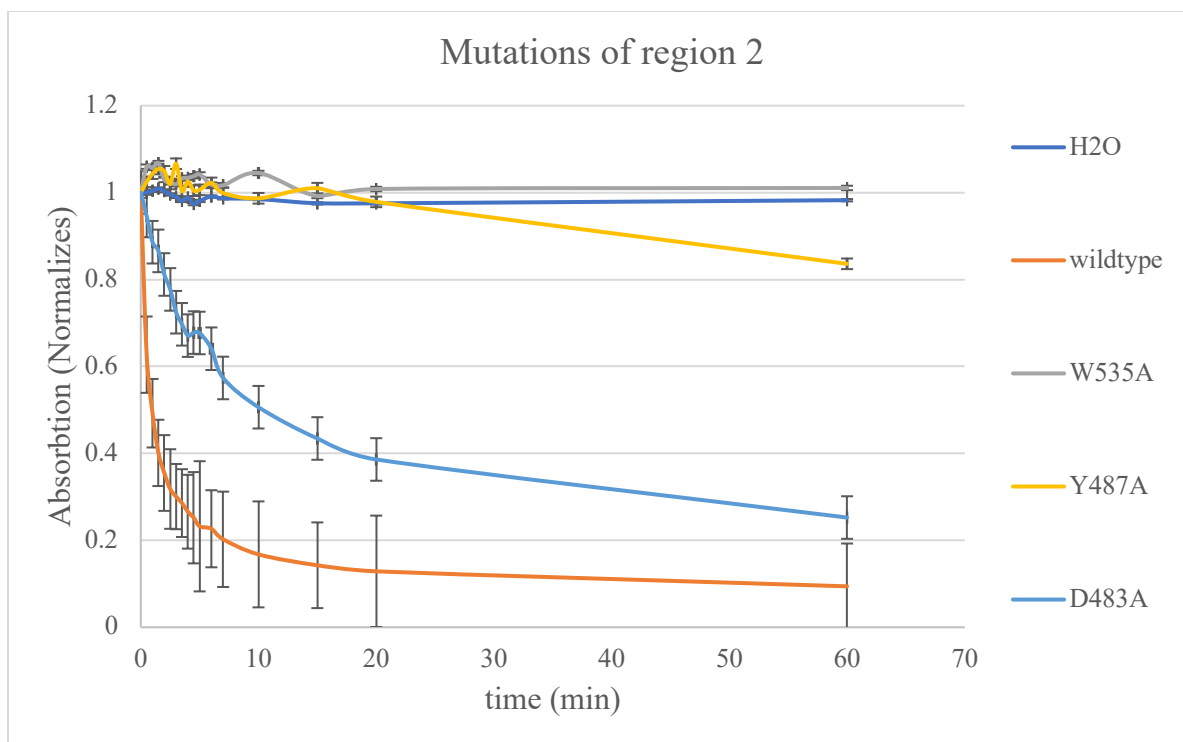


Figure 2. 28 Relative activity of the wildtype enzyme and region 2 mutants.

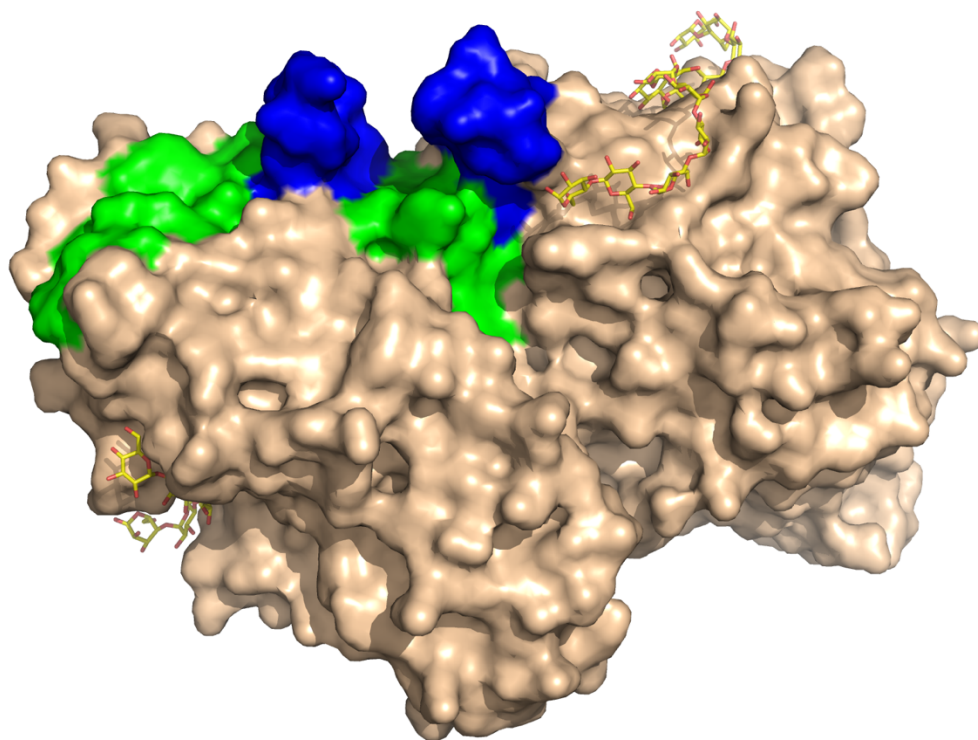


Figure 2. 29 Connecting the active site to the CBM48 domain.

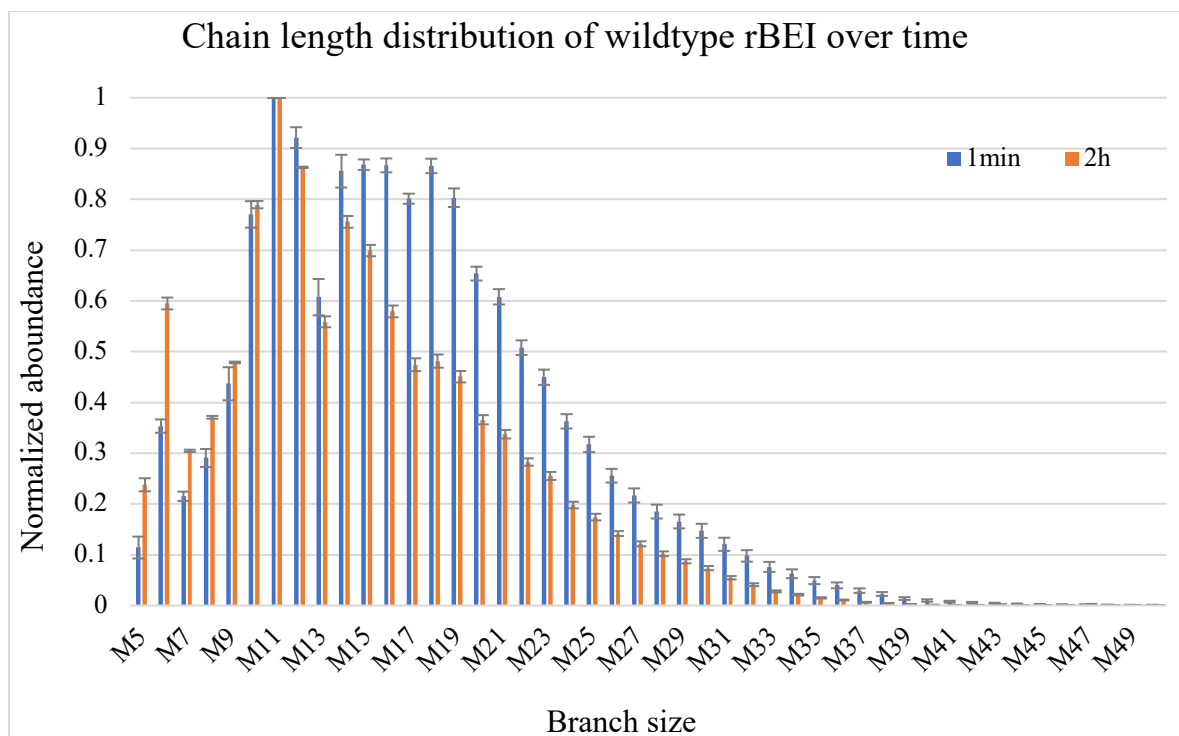


Figure 2. 30 Chain length distribution changes over time.

Mutations of region 2 generated inactive enzymes. These mutations are at least 10Å away from the active site. The strong conservation of the interacting residues and the relative proximity to the active site suggest that the M12 binding site may play a key role in the activity of the enzyme. This possibility was evaluated by site-directed mutagenesis, activity, and transfer chain specificity assays. Unlike the activity assays, transfer chain specificity for this enzyme is rather a complicated experiment, and it needs to be carefully designed. The duration of the transfer chain specificity assay is a critical parameter since the observed specificity is affected by the time that the enzyme is branching a substrate. As shown in Figure 2.30, longer reaction times invariably produce more short branched chains and fewer longer chains.

The reason for this observation is that the long chains produced by rice BEI are also substrates for secondary transfers. A Secondary transfer is the process in which an already branched substrate is hydrolyzed again. For example, if the enzyme transfers a chain of size 18 and later the enzyme

hydrolyzes this branched product to fragments of 12 and 6 glucose units, the initial product does not contribute in the chain length distribution profile. The initial chain of size 18 will be recognized as 6 or 12. The secondary transfers are limited to the branches that are still long enough to be substrates of the enzyme (size 12, considering the limitations of branching enzymes)

The side effect of these secondary transfers is a statistical equilibrium. If the reaction continues up to saturation, the majority of the long branches will be used again as donor chains and become shorter. However, since there is a size limit for the transfer (six glucose units), chains smaller than 12 glucose will not be subjected to the secondary transfers. Therefore, upon saturation, longer chains ratio will drop, and the ratio of shorter chains (shorter than M12) will increase where M11 is the pivot point. Single point mutations might affect the CLD profile, but under this statistical equilibrium, the effect of the mutations will be covered.

Mutated enzymes with lower activity must, therefore, be calibrated with enzymes that have higher activity by varying the reaction time in the assay. This calibration was accomplished by terminating the transfer chain specificity assay for each mutant only when the iodine assay absorption reached 50% of the initial absorption at 660 nm.

Table 2.7 summarizes the activities of rBEI mutations. Though several point mutants (W535A, Y487A, and D483A) showed significant loss of activity, none of the mutants significantly impacted the branch chain specificity (Figure 2.32, middle panel). We also identified a large insertion of eleven residues in the 525-553 loop found in all BEII enzymes.⁸⁶

This loop is relatively close to the M12 binding site, and five conserved residues in this loop make interactions with M12, suggesting that it plays an important role in M12 binding. In an effort to study differences between BEI and BEII isoforms, an eleven-residue insertion found in rBEIIb was introduced to the 525-553 loop in BEI (Figure 2.31). Though a significant loss of activity was

observed, no change in branch chain specificity was identified, though one of the most significant differences in BEI versus BEII activity is the preference of BEII enzymes for the transfer of shorter (6-7 units) chains relative to BEI enzymes.⁸⁶

Table 2. 7 Activities of rice branching enzyme and its mutants.

Mutation	Relative Activity
Wild type	100% $\pm$ 2.59
Control (No Enzyme)	0.07% $\pm$ 0.00
D344A – Active Site	0.11% $\pm$ 0.01
H467A – Active Site	0.68% $\pm$ 0.03
W535A – M11Binding Site	0.11% $\pm$ 0.01
Y487A – M11Binding Site	0.61% $\pm$ 0.05
D483A – M11Binding Site	21.97% $\pm$ 1.47
G152W – Donor Chain Site	5.81% $\pm$ 1.01
Y229W – Donor Chain Site	16.61% $\pm$ 5.76
Y229A – Donor Chain Site	20.81% $\pm$ 2.28
SNN277AAA – Donor Chain Site	27.51% $\pm$ 1.01
A148W – Donor Chain Site	83.06% $\pm$ 3.22
Loop 143	31.26% $\pm$ 1.43
Loop 541	25.10% $\pm$ 1.61
Loop 143 and Loop 541	2.13% $\pm$ 0.26
D156A – CBM Domain	83.06% $\pm$ 3.53
D135A – CBM Domain	81.72% $\pm$ 3.40
D147A – CBM Domain	51.80% $\pm$ 6.32

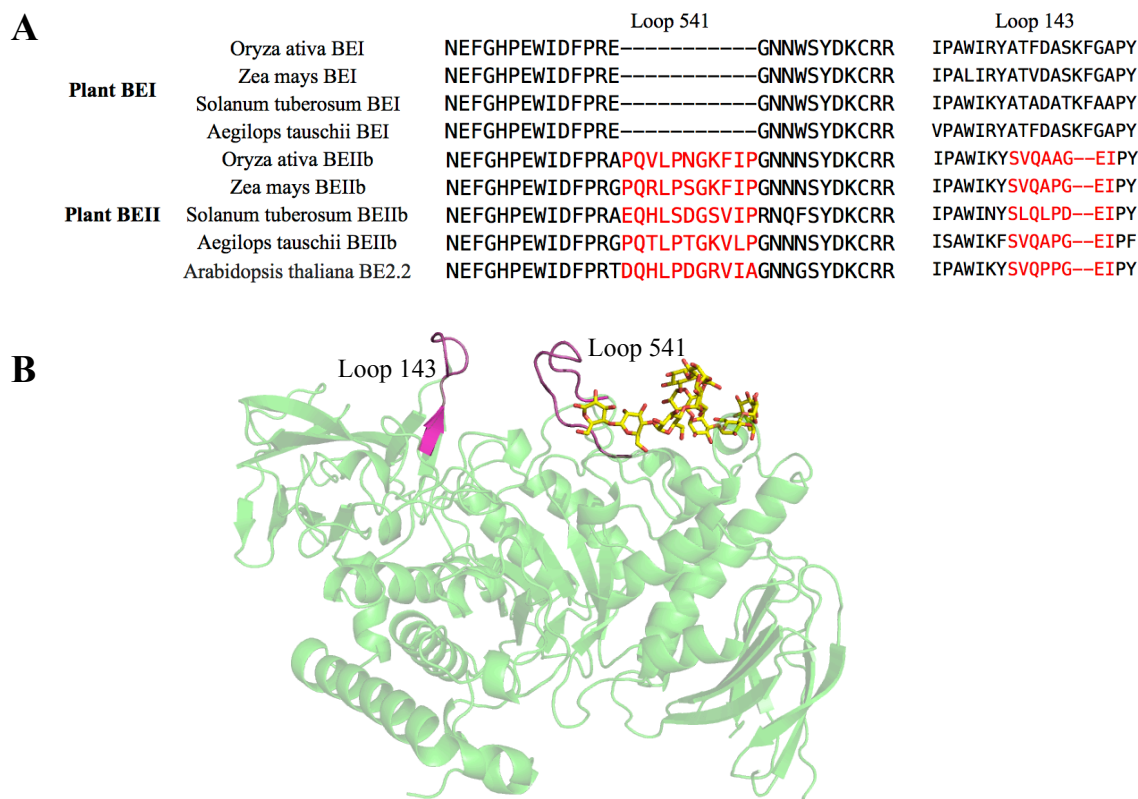


Figure 2. 31 Isoform defining loops of BEIs and BEIIs. (A) Sequence Alignment of the two loops that distinguish BEIs and BEIIs. (B) Location of the loops on M12-bound rBEI. rBEI (Green), Loop 143 and 541 (Magenta), M12 (stick model, C, yellow, all other atoms colored as above).

A second loop (encompassing residues 146-156), found proximal to the 525-553 loop, also had significant sequence deviation between BEI and BEII enzymes (Figure 2.12). Dramatic differences were seen in the activity when both of the loops in BEI are replaced with the loops found in rBEIIb. First, the overall activity was significantly decreased when amylose was used as a substrate (Table 2.7). Second, the branch chain specificity of BEI was converted from a preference for longer (11-12 glucose units) chains to an almost exclusive preference for chains of six-seven glucose units (Figure 2.32, bottom panel), similar to that seen for rBEIIb. We, therefore, conclude that these two loops work together to control the branch chain specificity in rice, and likely other plant, branching enzymes.

The unique function and specificity of branching enzymes; their role in synthesizing and modifying growing polymeric substrates and their relatively imprecise, though widely divergent transfer chain specificities, depending on species or even isoform, make them relatively unusual enzymes, given the relatively high specificity for substrate and product of most enzymes. Though several branching enzyme structures are known,^{82–85,99,103–105} many of which are bound to malto-oligosaccharides, the structural details that give rise to the unique characteristics of BE's remained mysterious.

M12 is the largest oligosaccharide to be observed at atomic resolution bound to a BE and elaborates a binding surface stretching from the far edge of the enzyme and almost entering the active site. Numerous mutations confirm the importance of the binding surface for the catalytic activity of the enzyme, and the residues interacting with M12 are highly conserved in eukaryotic (GH13 sub-family 8) BE's. Together this leads to the hypothesis that the M12 binding surface defines a part of either the donor or acceptor chain binding site. The Y487A mutation is quite far from the active site and displays very low activity ($0.61\% \pm 0.05$ of wild type rBEI).

If we suppose that M12 represents the donor chain binding site, the tyrosine can be involved in interactions between the donor chain and the enzyme only if rBEI is transferring chains longer than 10 glucose units. The fact that the fraction of transferred chains smaller than 11 glucose still accounts for almost 15% of all transferred chains is inconsistent with only $0.61\% \pm 0.05$ activity. In addition, the fact that none of the mutations to this binding surface resulted in any noticeable change in transfer chain length also argues against its involvement in donor chain binding^{85,104,106}).

Further, a recent crystal structure of Maltoheptaose (M7)-bound *Cyanothece* BE (sp. ATCC 51142)⁸⁵ shows for the first time, a donor chain bound in the active site of a BE. Many of the residues that define the donor chain binding in the M7-bound *Cyanothece* BE structure are

conserved in bacterial and eukaryotic BE's, including rBEI, suggesting that all BE's use a similar donor chain binding surface. The M7 binding surface follows a trajectory distinct from that of the rBEI M12 surface (Figure 2.33).

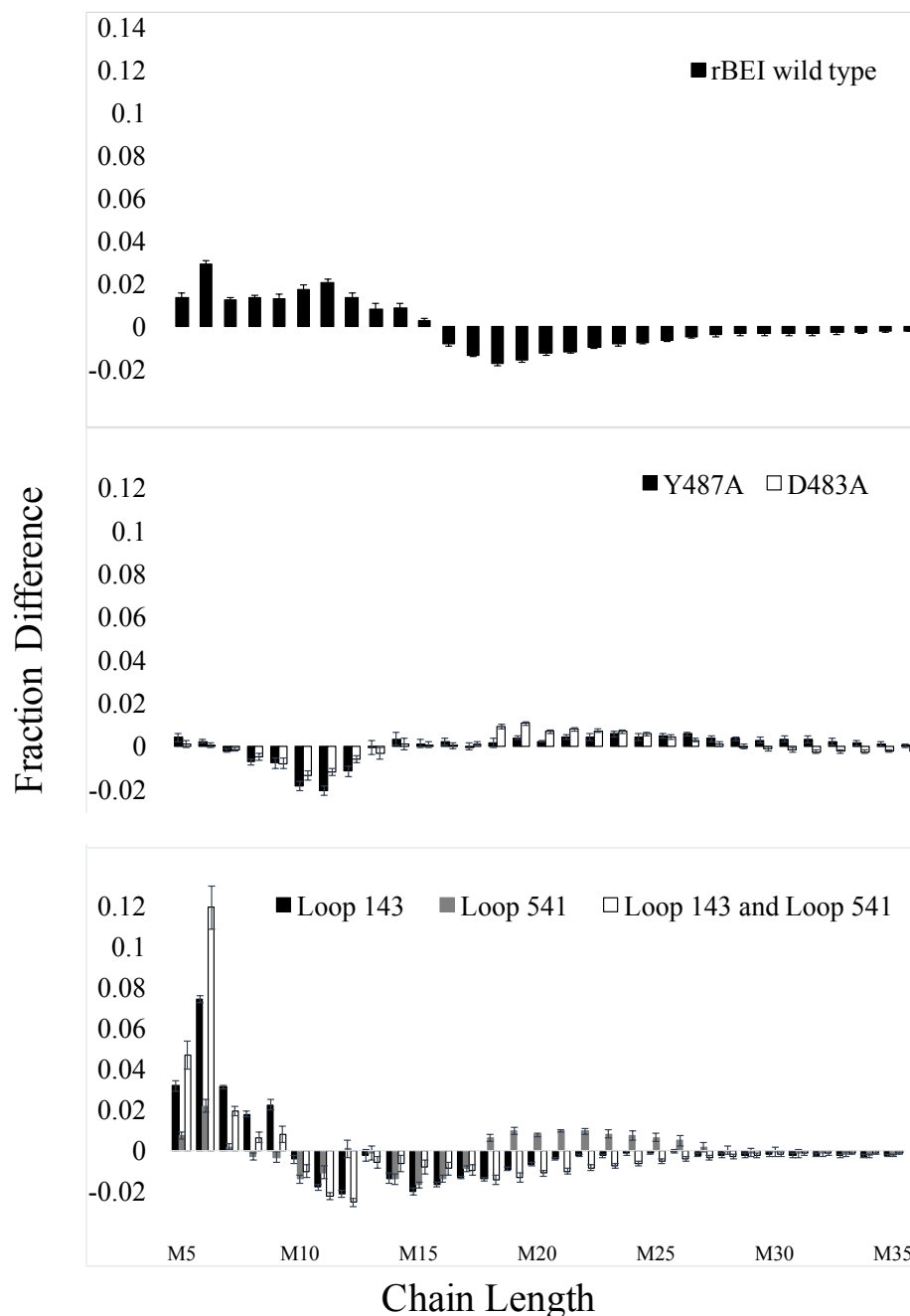


Figure 2. 32 Transfer Chain Specificity Assay. Fraction differences of transferred chains by wild type rBEI in 2 hrs. vs 1 min. (top panel), Fraction differences of transferred chains by wild type rBEI vs Y487A and D483A (middle panel), Fraction differences of transferred chains by wild type rBEI vs rBEII loop replacements (bottom panel).

This leads naturally to the conclusion that the M12 binding surface represents part of the acceptor chain binding site. Using the M7-bound *Cyanothece* BE structure as a guide, an M7 was modeled into the putative donor chain binding surface of M12-bound rBEI (Figure 2.33). As shown, though there is no overlap between the putative donor and M12 binding sites, they are proximal. Interestingly, the flexible 525-553 loop is located between the putative donor and acceptor chain binding sites. This is the same loop that is substantially extended in BEIIb isoforms. As previously discussed, replacement of the rBEI loop sequence with that of rBEIIb reduced the activity of the enzyme, but this change alone had no perceptible effect on branch chain specificity.

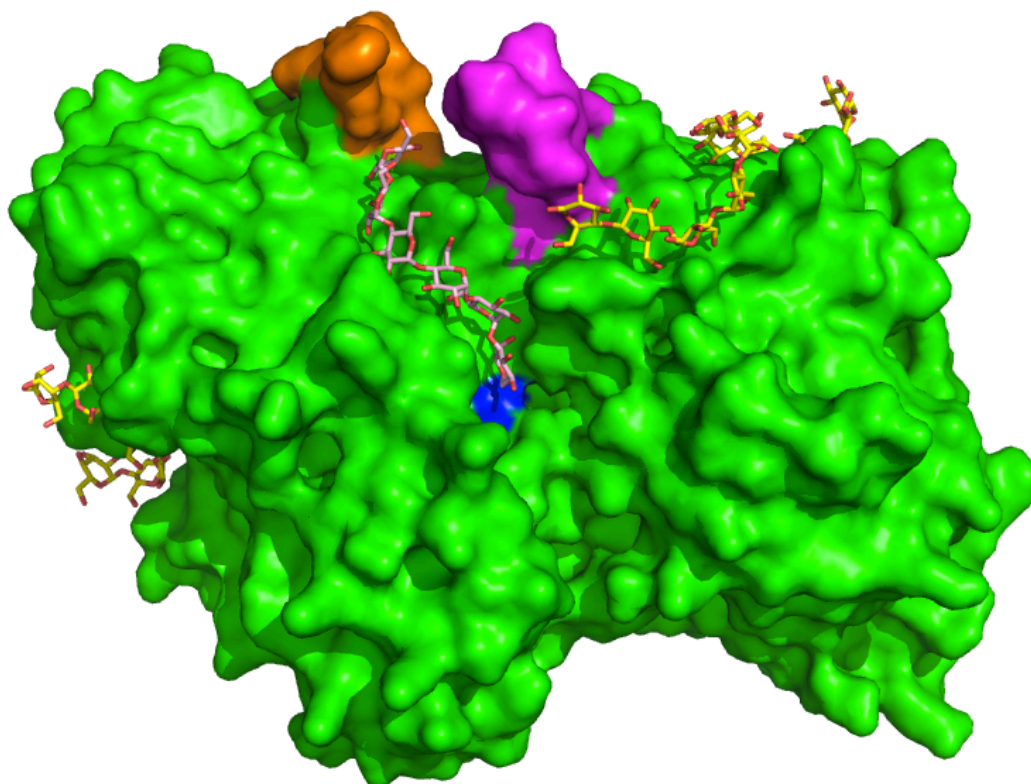


Figure 2. 33 Overlay of M7 bound in *Cyanothece* BE and M12-bound rBEI, rBEI (Green), glucans bound to rBEI (C, yellow, all other atoms as above), M7 bound to *Cyanothece* BE (C, Pink), active site (Blue), Loop 143 (Orange), Loop 541 (Magenta)

However, simultaneously exchanging both this loop and the 146-156 loop substantially altered branch chain specificity, essentially converting rBEI into an rBEIIb in its product specificity. This

second loop lies on the opposite side of the donor chain binding site (Figure 2.33) such that the two loops surround the non-reducing end of the donor chain, exactly where they would be expected to be to play a role in branch chain specificity. This serves to confirm that the donor chain trajectory is very similar in rBE's to that of *Cyanothece* BE, making it likely that all eukaryotic BE's share a common donor chain binding surface. Further, it seems that loops on both sides of the donor chain are required for controlling donor chain length. We hypothesize that the longer 525 loop found in BEII enzymes reaches over the donor chain binding surface, interacts with the 146-156 loop and the end of an M6 or M7 donor chain to select for shorter donor chains. It is interesting to note that a different loop in *Cyanothece* occupies the space of the 525 loop, interacts with the non-reducing end glucose unit, and likely provides some of the specificity for shorter glucan chains seen in the *Cyanothece* enzyme. This loop is not conserved in other bacterial enzymes, many of which have branch chain specificities distinct from that of *Cyanothece*. The proximity of the 525 loop to both donor and acceptor chains, and the fact that residues in this loop make direct interactions with the putative M12 acceptor chain in the M12-bound rBEI structure suggest the possibility that there is allosteric communication between donor and acceptor binding sites such that binding of one does not inhibit the binding of the other in the active site. Their proximity also suggests that there may be interaction between donor and acceptor chains when both are bound, as suggested for pullulanases.

In conclusion, with the insights gained from the recent donor-chain bound *Cyanothece* structure, and the acceptor-chain bound rBEI structure described here, combined with the mutagenesis results that define the structures responsible for controlling donor chain specificity in plant isoforms, an atomic resolution picture of this critical component in the dynamic biosynthesis of the starch granule finally begins to emerge.

2.4. Materials and methods

DNA of rice branching enzyme was purchased from Riken institute (accession number AK65121). Q5 mutagenesis (New England Biolab) kit was used in all mutagenesis. Primers for all the mutagenesis were purchased from IDT (integrated DNA technologies) and the sequence of the primers are listed in table 2.9. All PCR cycles follow the following setup.

Table 2. 8 General PCR cycle for mutagenesis

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
25 Cycles	98°C	10 seconds
	50–72°C*	10–30 seconds
	72°C	20–30 seconds/kb
Final Extension	72°C	2 minutes
Hold	4–10°C	

Table 2. 9 PCR primers for mutagenesis (Q5 style)

Mutation	Forward primer	Reverse primer	Annealing temperature °C
G152W	CTCTAAATTTTGGGCTCCATATGATG	GCATCAAAAGTTGCATAACGAATCC	55
S228R	CATGGAACATCGCTACTATGCTTC	ATTGCCATTAAC TGAAC TGAAC TG	54
Y229W	GGAACATTCTGCTGCTTCTTTTGG	ATGATTGCCATTAAC TGAAC TGTGTTG	57
W535A	CCATCCAGAAGCGATTGACTTTCCAAGAG	CCAAACTCATTGCCCATAAAATTTAAGTAGCC	59
D147A	GCAACTTTTGCTGCCTCTAAATTTG	ATAACGAATCCATGCGGGAATACG	56
D156A	GCTCCATATGCTGGTGTACAC	TCCAAATTTAGAGGCATCAAAAGTTGC	57
M40L	CCAAAAATGCCTGATTGAAAAACATG	TCGAGGTATCTTTTTATCC	52
M280V	GAGTAATAATGTGACCGATGGTC	GCATGGCTATGGACAACA	53
W133A	GGGTGGAGCAGCGGTTGATCGTATTC	CCATGCCTAAAGCGAAATTTAAC	57
A148W	AACTTTTGATTGGTCTAAATTTGGAGCTCCATATG	GCATAACGAATCCATGCG	55
D344A	TTCCGATTTGCGGGGGTTACATC	GCCATCAAACATGAATTC	54
D344M	CTTCCGATTTATGGGGGTACATCAATGC	CCATCAAACATGAATTCG	54
P443S	CCGCAAATGGTCTATGAGTGAAATAG	TCCTCTTTGTTCTTCAGG	54
A669T	AGTACCAGAAACCAATTTCAACAACCGCC	CCTGGCATTCCTCGGGA	60

Table 2.9 (cont'd)

W88A	GTTCAATAACGCGAATGGTGCAAAACATAAGATG	TCACCAATGAGCTGTGCT	57
D135A	AGCATGGGTTGCTCGTATTCCCG	CCACCCCATGCCTAAAG	58
Y229A	GGAACATTCCGCCTATGCTTCTTTTG	ATGATTGCCATTAACCTGAAC	54
SNN277AAA	TGCTGTGACCGATGGTCTAAATG	GCAGCCGCATGGCTATGGACAAC	60
LOOP143	GCGGGCGAAATTCCATATGATGGTGTACACTG	CGCCTGCACAGAATAACGAATCCATGCGGG	64
LOOP541-1	CGGCAGCGTGCTGCCGGGCAACAACTGGAGCTATG	TTCGGCAGAGACTGCGGTTCTCTTGAAAAGT CAATCCATTC	69
LOOP541-2	AACGGCAAATTTATCCCGGGC	CGGCAGCACCTGCGGTTC	60
D344N	CTTCCGATTTAACGGGGTTACATC	CCATCAAACATGAATTCG	51
H467A	TGCCGAGAGCGCGGATCAGTCCATTG	TAGGCAATGCATTTTCTG	55
W88I	GTTCAATAACATCAATGGTGCAAAACATAAGATG	TCACCAATGAGCTGTGCT	56
K127A	TTCCAAGGTTGCGTTTCGCTTTAGGCATGGGG	TTGTGAGGGATGGCAGGC	63
K150A	TGATGCCTCTGCCTTTGGAGCTCC	AAAGTTGCATAACGAATCC	57
W133A	GGGTGGAGCAGCGGTTGATCGTATTCC	CCATGCCTAAAGCGAAATTTAACC	60
E399A	TATTGTTGCTGCAGATGTTTCGGG	GTTGCTTCCGGCAAGAGT	57

2.4.1. Protein expression and purification

All proteins were expressed using BL21 bacterial strains. First, BL21 cells made competent using CaCl_2 solutions and later transformed by heat shock process. Manufacturer protocols were employed for these procedures. Transformed cells were grown to early log phase before IPTG induction. Temperature, time and optical density were optimized for each protein. The composition of LB (Luria broth) is NaCl (5 g), tryptone (10 g), and yeast extract (5 g).

GBSSI protein was expressed at 20°C from optical density of 0.5-0.7. Expression was continued overnight, and the following day cells were collected by centrifuge (5000 rpm). Cells were resuspended in PBS (25 mL for 1L LB) and sonicated for 6 min (15 sec on, 45 sec off) at 40% power. Lysed cells were centrifuged for 30 min at 10000 rpm to separate the supernatant from insoluble fraction. Supernatant were diluted with equal volume of buffer (500 mM NaCl, 40mM disodium phosphate pH 8.0, and 40 mM imidazole) and then it was loaded to a nickel column for affinity purification via His-tag. Figure x shows the SDS-page gel image for the wildtype GBSSI.

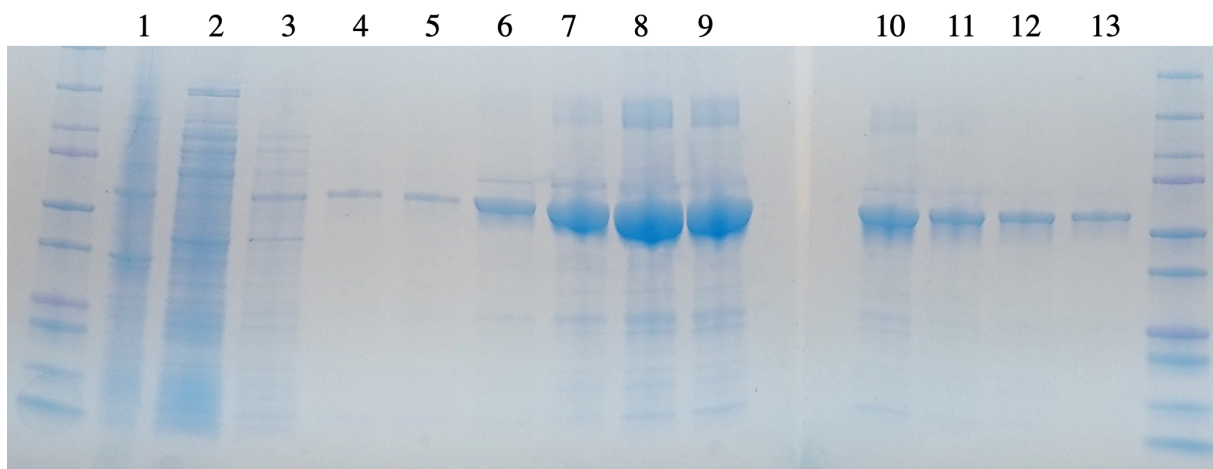


Figure 2. 34 SDS-page gel image for nickel column His-tag affinity purification of GBSSI (59 kDa). Lanes 1 through 13 are: insoluble fraction, run through, wash 1-3 (all washes are 20 mM imidazole in PBS), elution 1-8 (200 mM imidazole in PBS)

GBSSI was further purified using size-exclusion chromatography prior to crystallization. The buffer used to elute run the size exclusion chromatography has 200 mM NaCl, 20 mM HEPPSO pH 8.0. Purified protein is then concentrated for crystallization.

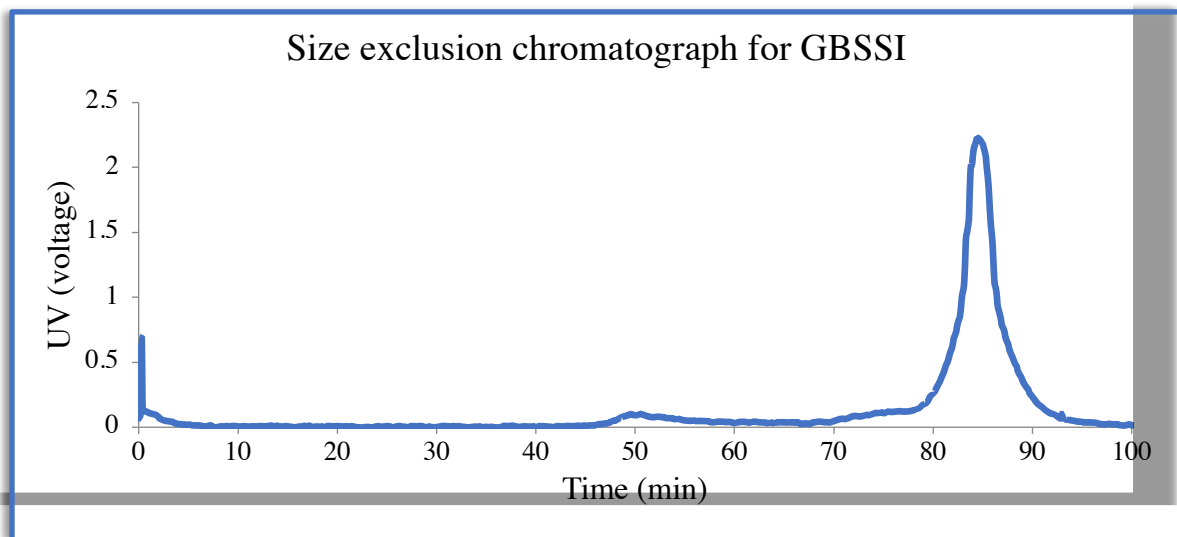


Figure 2. 35 Size exclusion chromatograph for GBSSI

2.4.2. Crystallization of rice GBSSI

Six commercially available screening conditions (Crystal Screen, PEGION, Index, Salt Rx from Hampton research and Wizard I, II, III, IV) were used to find the crystallization condition for rice GBSSI.

Rice GBSSI (6.5 mg/mL) was crystallized by vapor diffusion hanging drop technique. Crystallization drop consists of 1 μ L protein solution and 1 μ L reservoir solution. The open conformer was crystallized without the purification by size exclusion chromatography. Samples eluted from Ni-column affinity purification were directly screened for crystallization. (0.01 M $MgCl_2$, 0.1 M HEPES sodium pH 7.0, 15% polyethylene glycol 3350, and 5 mM $NiCl_2$)

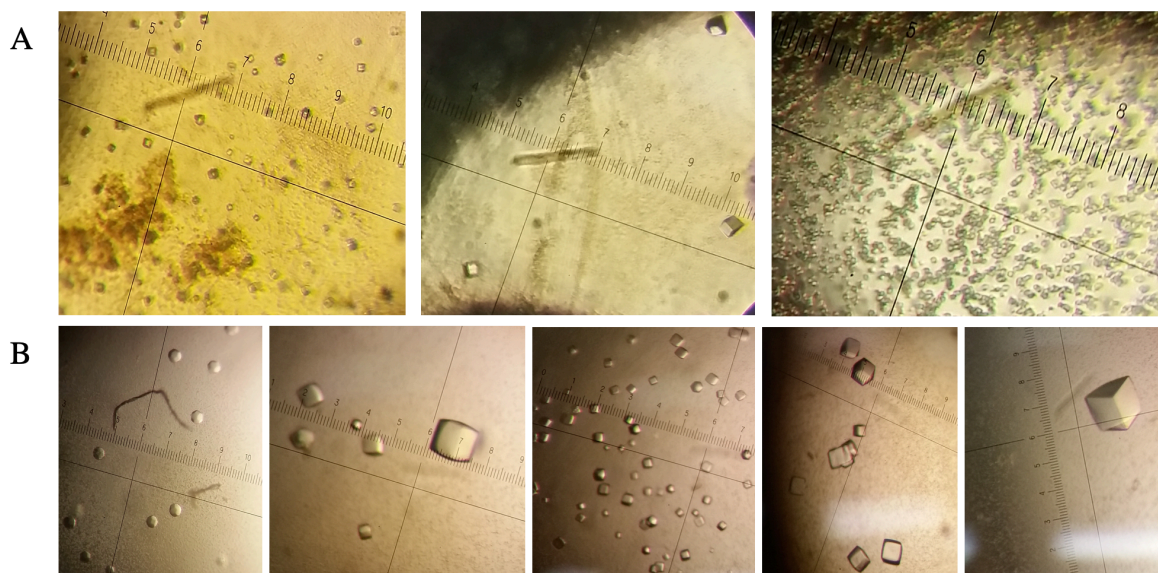


Figure 2. 36 Crystals of GBSSI. (A) before optimization, (B) after optimization

Table 2. 10 Crystallization conditions for GBSSI

	Buffer (0.1M)	Precipitant	Additive
1	BisTrisPropane, pH=6.0-7.5	Ammonium Sulfate 1.2-1.7M	-
2	Tris, pH=7.5-9.0	Ammonium Sulfate 1.2-1.7M	-
3	HEPES Sodium pH=6.5-8.0	Ammonium Sulfate 1.4-2.4M	2% PEG400
4	Tris HCl pH=7.5-9.0	10%-35% w/v PEG 4000	Lithium Sulfate 0.2M
5	MES pH=6.0-7.5	Ammonium Sulfate 1.3-1.8M	10% v/v Dioxane
6	MES pH=6.0-7.5	18%-33% PEGME 5000	Ammonium Sulfate 0.2M

2.4.3. Protein expression and purification of rice branching enzyme

Expression and purification of rice branching enzyme: rice branching enzyme was expressed from BL21 bacterial strain. Expression temperature was optimized at 16°C. Similar to GBSSI, rice branching enzyme was purified using his-tag affinity column. No further purification was needed for the activity and chain length distribution assay.

2.4.4. Chain length distribution assay for rice branching enzyme

Activity and CLD profile: Amylose (substrate) stock solution was made by dissolving 50 mg amylose in 2 mL water and 0.5 mL 10% NaOH. The working solution of amylose was made by mixing 1-part amylose stock solution, 1-part 10x ammonium citrate buffer pH 8.0 (20 mM), and 8-part Millipore water. pH was adjusted to 8.0 by adding an appropriate amount of concentrated HCl. Samples were centrifuged for 5 min at 10000 rpm and aliquoted in 900 μ L portions. 50 μ g of the enzyme was added to the mixture. Every 30 sec, 15 μ L of the reaction was added to 985 μ L iodine solution (2.6 g KI, 0.26 g I₂, in 10 mL water as a stock solution. The working solution is 25-fold diluted). The UV absorbance for triiodide-oligosaccharide complex was measured using UV-spectrometer at a wavelength of 660 nm.

For the branching (CLD profile) assay, the branching reactions were stopped at 50% of initial absorption by boiling for 2 minutes and changing pH to 4. Then IsoAmylase (debranching enzyme) was added to cleave α -1,6-glycosidic bonds. Debranched products were desalted and they were analyzed by liquid chromatography (BH-Amide as a stationary phase and ammonium acetate/acetonitrile as mobile phase) and Quadrupole Time-of-Flight Mass spectroscopy (QToF-MS) to identify the abundance of various branches based on their chain length.

REFERENCES

REFERENCES

- (1) Tetlow, I. J.; Emes, M. J. A Review of Starch-Branching Enzymes and Their Role in Amylopectin Biosynthesis. *IUBMB Life* **2014**, *66* (8), 546–558. <https://doi.org/10.1002/iub.1297>.
- (2) Joanna, K.; Michał, P.; Anna, P. Osmotic Properties of Polysaccharides Solutions. *Solubility Polysacch.* **2017**. <https://doi.org/10.5772/intechopen.69864>.
- (3) Buléon, A.; Colonna, P.; Planchot, V.; Ball, S. Starch Granules: Structure and Biosynthesis. *Int. J. Biol. Macromol.* **1998**, *23* (2), 85–112. [https://doi.org/10.1016/S0141-8130\(98\)00040-3](https://doi.org/10.1016/S0141-8130(98)00040-3).
- (4) Lloyd, J. R.; Kossmann, J. Transitory and Storage Starch Metabolism: Two Sides of the Same Coin? *Curr. Opin. Biotechnol.* **2015**, *32*, 143–148. <https://doi.org/10.1016/j.copbio.2014.11.026>.
- (5) French, D. Fine Structure of Starch and Its Relationship to the Organization of Starch Granules. *澱粉科学* **1972**, *19* (1), 8–25. <https://doi.org/10.5458/jag1972.19.8>.
- (6) Suzuki, E.; Onoda, M.; Colleoni, C.; Ball, S.; Fujita, N.; Nakamura, Y. Physicochemical Variation of Cyanobacterial Starch, the Insoluble α -Glucans in Cyanobacteria. *Plant Cell Physiol.* **2013**, *54* (4), 465–473. <https://doi.org/10.1093/pcp/pcs190>.
- (7) Strange, R. E. Bacterial “Glycogen” and Survival. *Nature* **1968**, *220* (5167), 606–607. <https://doi.org/10.1038/220606a0>.
- (8) Baldwin, K. M.; Fitts, R. H.; Booth, F. W.; Winder, W. W.; Holloszy, J. O. Depletion of Muscle and Liver Glycogen during Exercise. Protective Effect of Training. *Pflugers Arch.* **1975**, *354* (3), 203–212. <https://doi.org/10.1007/bf00584644>.
- (9) Iglesias, A. A.; Preiss, J. Bacterial Glycogen and Plant Starch Biosynthesis. *Biochem. Educ.* **1992**, *20* (4), 196–203. [https://doi.org/10.1016/0307-4412\(92\)90191-N](https://doi.org/10.1016/0307-4412(92)90191-N).
- (10) Malomo, K.; Ntlholang, O. The Evolution of Obesity: From Evolutionary Advantage to a Disease. *Biomed. Res. Clin. Pract.* **2018**, *3* (2). <https://doi.org/10.15761/BRCP.1000163>.
- (11) Owen, O. E. Ketone Bodies as a Fuel for the Brain during Starvation. *Biochem. Mol. Biol. Educ.* **2005**, *33* (4), 246–251. <https://doi.org/10.1002/bmb.2005.49403304246>.
- (12) Manninen, A. H. Metabolic Effects of the Very-Low-Carbohydrate Diets: Misunderstood “Villains” of Human Metabolism. *J. Int. Soc. Sports Nutr.* **2004**, *1* (2), 7–11. <https://doi.org/10.1186/1550-2783-1-2-7>.

- (13) Paoli, A.; Canato, M.; Toniolo, L.; Bargossi, A. M.; Neri, M.; Mediati, M.; Alesso, D.; Sanna, G.; Grimaldi, K. A.; Fazzari, A. L.; et al. The ketogenic diet: an underappreciated therapeutic option? *Clin. Ter.* **2011**, *162* (5), e145-53.
- (14) Verbančič, J.; Lunn, J. E.; Stitt, M.; Persson, S. Carbon Supply and the Regulation of Cell Wall Synthesis. *Mol. Plant* **2018**, *11* (1), 75–94. <https://doi.org/10.1016/j.molp.2017.10.004>.
- (15) Patindol, J.; Siebenmorgen, T.; Wang, Y.-J. Impact of Environmental Factors on Rice Starch Structure: A Review. *Starch - Starke* **2014**, 67. <https://doi.org/10.1002/star.201400174>.
- (16) Figure 3 Schematic representation of a starch granule https://www.researchgate.net/figure/Schematic-representation-of-a-starch-granule-A-semi-crystalline-and-amorphous-growth_fig1_51780443 (accessed Nov 12, 2019).
- (17) Kim, H.-S.; Kim, B.-Y.; Baik, M.-Y. Application of Ultra High Pressure (UHP) in Starch Chemistry. *Crit. Rev. Food Sci. Nutr.* **2012**, *52*, 123–141. <https://doi.org/10.1080/10408398.2010.498065>.
- (18) Psilodimitrakopoulos, S.; Amat-Roldan, I.; Loza-Alvarez, P.; Artigas, D. Estimating the Helical Pitch Angle of Amylopectin in Starch Using Polarization Second Harmonic Generation Microscopy. *J. Opt.* **2010**, *12* (8), 084007. <https://doi.org/10.1088/2040-8978/12/8/084007>.
- (19) Awtrey, A. D.; Connick, R. E. The Absorption Spectra of I₂, I₃⁻, I⁻, IO₃⁻, S₄O₆⁼ and S₂O₃⁼. Heat of the Reaction I₃⁻ = I₂ + I⁻. *J. Am. Chem. Soc.* **1951**, *73* (4), 1842–1843. <https://doi.org/10.1021/ja01148a504>.
- (20) Robin, M. B. Optical Spectra of Benzamide—Triiodide Ion Complexes : A Model of the Starch—Iodine Complex. *J. Chem. Phys.* **1964**, *40* (11), 3369–3377. <https://doi.org/10.1063/1.1725009>.
- (21) Teitelbaum, R. C.; Ruby, S. L.; Marks, T. J. On the Structure of Starch-Iodine. *J. Am. Chem. Soc.* **1978**, *100* (10), 3215–3217. <https://doi.org/10.1021/ja00478a045>.
- (22) Calder, P. C. Glycogen Structure and Biogenesis. *Int. J. Biochem.* **1991**, *23* (12), 1335–1352. [https://doi.org/10.1016/0020-711X\(91\)90274-Q](https://doi.org/10.1016/0020-711X(91)90274-Q).
- (23) Ballicora, M. A.; Iglesias, A. A.; Preiss, J. ADP-Glucose Pyrophosphorylase: A Regulatory Enzyme for Plant Starch Synthesis. *Photosynth. Res.* **2004**, *79* (1), 1–24. <https://doi.org/10.1023/B:PRES.0000011916.67519.58>.
- (24) Comino, N.; Cifuentes, J. O.; Marina, A.; Orrantia, A.; Eguskiza, A.; Guerin, M. E. Mechanistic Insights into the Allosteric Regulation of Bacterial ADP-Glucose

- Pyrophosphorylases. *J. Biol. Chem.* **2017**, 292 (15), 6255–6268. <https://doi.org/10.1074/jbc.M116.773408>.
- (25) Hwang, S.-K.; Salamone, P. R.; Okita, T. W. Allosteric Regulation of the Higher Plant ADP-Glucose Pyrophosphorylase Is a Product of Synergy between the Two Subunits. *FEBS Lett.* **2005**, 579 (5), 983–990. <https://doi.org/10.1016/j.febslet.2004.12.067>.
- (26) Fu, Y.; Ballicora, M. A.; Leykam, J. F.; Preiss, J. Mechanism of Reductive Activation of Potato Tuber ADP-Glucose Pyrophosphorylase. *J. Biol. Chem.* **1998**, 273 (39), 25045–25052. <https://doi.org/10.1074/jbc.273.39.25045>.
- (27) Ball, K.; Preiss, J. Allosteric Sites of the Large Subunit of the Spinach Leaf ADPglucose Pyrophosphorylase. *J. Biol. Chem.* **1994**, 269 (40), 24706–24711.
- (28) Ballicora, M. A.; Frueauf, J. B.; Fu, Y.; Schürmann, P.; Preiss, J. Activation of the Potato Tuber ADP-Glucose Pyrophosphorylase by Thioredoxin. *J. Biol. Chem.* **2000**, 275 (2), 1315–1320. <https://doi.org/10.1074/jbc.275.2.1315>.
- (29) Bhayani, J. A.; Hill, B. L.; Sharma, A.; Iglesias, A. A.; Olsen, K. W.; Ballicora, M. A. Mapping of a Regulatory Site of the Escherichia Coli ADP-Glucose Pyrophosphorylase. *Front. Mol. Biosci.* **2019**, 6. <https://doi.org/10.3389/fmolb.2019.00089>.
- (30) Diez, M. D. A.; Aleanzi, M. C.; Iglesias, A. A.; Ballicora, M. A. A Novel Dual Allosteric Activation Mechanism of Escherichia Coli ADP-Glucose Pyrophosphorylase: The Role of Pyruvate. *PLOS ONE* **2014**, 9 (8), e103888. <https://doi.org/10.1371/journal.pone.0103888>.
- (31) Ebrecht, A. C.; Solamen, L.; Hill, B. L.; Iglesias, A. A.; Olsen, K. W.; Ballicora, M. A. Allosteric Control of Substrate Specificity of the Escherichia Coli ADP-Glucose Pyrophosphorylase. *Front. Chem.* **2017**, 5. <https://doi.org/10.3389/fchem.2017.00041>.
- (32) Jin, X.; Ballicora, M. A.; Preiss, J.; Geiger, J. H. Crystal Structure of Potato Tuber ADP-Glucose Pyrophosphorylase. *EMBO J.* **2005**, 24 (4), 694–704. <https://doi.org/10.1038/sj.emboj.7600551>.
- (33) Yang, Z.; Wang, Y.; Xu, S.; Xu, C.; Yan, C.-J. Molecular Evolution and Functional Divergence of Soluble Starch Synthase Genes in Cassava (*Manihot Esculenta* Crantz). *Evol. Bioinforma. Online* **2013**, 9, 239–249. <https://doi.org/10.4137/EBO.S11991>.
- (34) Coutinho, P. M.; Deleury, E.; Davies, G. J.; Henrissat, B. An Evolving Hierarchical Family Classification for Glycosyltransferases. *J. Mol. Biol.* **2003**, 328 (2), 307–317. [https://doi.org/10.1016/s0022-2836\(03\)00307-3](https://doi.org/10.1016/s0022-2836(03)00307-3).
- (35) Campbell, J. A.; Davies, G. J.; Bulone, V.; Henrissat, B. A Classification of Nucleotide-Diphospho-Sugar Glycosyltransferases Based on Amino Acid Sequence Similarities. *Biochem. J.* **1997**, 326 (Pt 3), 929–939. <https://doi.org/10.1042/bj3260929u>.

- (36) Ohdan, T.; Francisco, P. B.; Sawada, T.; Hirose, T.; Terao, T.; Satoh, H.; Nakamura, Y. Expression Profiling of Genes Involved in Starch Synthesis in Sink and Source Organs of Rice. *J. Exp. Bot.* **2005**, *56* (422), 3229–3244. <https://doi.org/10.1093/jxb/eri292>.
- (37) Denyer, K.; Clarke, B.; Hylton, C.; Tatge, H.; Smith, A. M. The Elongation of Amylose and Amylopectin Chains in Isolated Starch Granules. *Plant J.* **1996**, *10* (6), 1135–1143. <https://doi.org/10.1046/j.1365-313X.1996.10061135.x>.
- (38) Møller, M. S.; Henriksen, A.; Svensson, B. Structure and Function of α -Glucan Debranching Enzymes. *Cell. Mol. Life Sci. CMLS* **2016**, *73* (14), 2619–2641. <https://doi.org/10.1007/s00018-016-2241-y>.
- (39) Qu, J.; Shutu, X.; Zhang, Z.; Chen, G.; Zhong, Y.; Liu, L.; Zhang, R.; Xue, J.; Guo, D. Evolutionary, Structural and Expression Analysis of Core Genes Involved in Starch Synthesis. *Sci. Rep.* **2018**, *8*. <https://doi.org/10.1038/s41598-018-30411-y>.
- (40) Claassens, A. P. Investigation of Starch Metabolism Genes and Their Interactions. **2013**.
- (41) SCHOCH, T. J.; ELDER, A. L. Starches in the Food Industry. In *USE OF SUGARS AND OTHER CARBOHYDRATES IN THE FOOD INDUSTRY*; Advances in Chemistry; AMERICAN CHEMICAL SOCIETY, 1955; Vol. 12, pp 21–34. <https://doi.org/10.1021/ba-1955-0012.ch002>.
- (42) Egharevba, H. O. Chemical Properties of Starch and Its Application in the Food Industry. *Chem. Prop. Starch* **2019**. <https://doi.org/10.5772/intechopen.87777>.
- (43) Park, S. H.; Na, Y.; Kim, J.; Kang, S. D.; Park, K.-H. Properties and Applications of Starch Modifying Enzymes for Use in the Baking Industry. *Food Sci. Biotechnol.* **2017**, *27* (2), 299–312. <https://doi.org/10.1007/s10068-017-0261-5>.
- (44) Martens, B. M. J.; Gerrits, W. J. J.; Bruininx, E. M. A. M.; Schols, H. A. Amylopectin Structure and Crystallinity Explains Variation in Digestion Kinetics of Starches across Botanic Sources in an in Vitro Pig Model. *J. Anim. Sci. Biotechnol.* **2018**, *9* (1), 91. <https://doi.org/10.1186/s40104-018-0303-8>.
- (45) Bettge, A. D.; Giroux, M. J.; Morris, C. F. Susceptibility of Waxy Starch Granules to Mechanical Damage. *Cereal Chem.* **2000**, *77* (6), 750–753. <https://doi.org/10.1094/CCHEM.2000.77.6.750>.
- (46) Yasui, T.; Matsuki, J.; Sasaki, T.; Yamamori, M. Amylose and Lipid Contents, Amylopectin Structure, and Gelatinisation Properties of Waxy Wheat (*Triticum Aestivum*) Starch. *J. Cereal Sci.* **1996**, *24* (2), 131–137. <https://doi.org/10.1006/jcrs.1996.0046>.

- (47) Zhao, X.; Andersson, M.; Andersson, R. Resistant Starch and Other Dietary Fiber Components in Tubers from a High-Amylose Potato. *Food Chem.* **2018**, *251*, 58–63. <https://doi.org/10.1016/j.foodchem.2018.01.028>.
- (48) Nugent, A. P. Health Properties of Resistant Starch. *Nutr. Bull.* **2005**, *30* (1), 27–54. <https://doi.org/10.1111/j.1467-3010.2005.00481.x>.
- (49) Sajilata, M. G.; Singhal, R. S.; Kulkarni, P. R. Resistant Starch—A Review. *Compr. Rev. Food Sci. Food Saf.* **2006**, *5* (1), 1–17. <https://doi.org/10.1111/j.1541-4337.2006.tb00076.x>.
- (50) Regierer, B.; Fernie, A. R.; Springer, F.; Perez-Melis, A.; Leisse, A.; Koehl, K.; Willmitzer, L.; Geigenberger, P.; Kossmann, J. Starch Content and Yield Increase as a Result of Altering Adenylate Pools in Transgenic Plants. *Nat. Biotechnol.* **2002**, *20* (12), 1256–1260. <https://doi.org/10.1038/nbt760>.
- (51) Aggarwal, P.; Dollimore, D. The Effect of Chemical Modification on Starch Studied Using Thermal Analysis. *Thermochim. Acta* **1998**, *324* (1), 1–8. [https://doi.org/10.1016/S0040-6031\(98\)00517-6](https://doi.org/10.1016/S0040-6031(98)00517-6).
- (52) Chen, Q.; Yu, H.; Wang, L.; Abidin, Z. ul; Chen, Y.; Wang, J.; Zhou, W.; Yang, X.; Khan, R. U.; Zhang, H.; et al. Recent Progress in Chemical Modification of Starch and Its Applications. *RSC Adv.* **2015**, *5* (83), 67459–67474. <https://doi.org/10.1039/C5RA10849G>.
- (53) Moad, G. Chemical Modification of Starch by Reactive Extrusion. *Prog. Polym. Sci.* **2011**, *36* (2), 218–237. <https://doi.org/10.1016/j.progpolymsci.2010.11.002>.
- (54) Elvira, C.; Mano, J. F.; San Román, J.; Reis, R. L. Starch-Based Biodegradable Hydrogels with Potential Biomedical Applications as Drug Delivery Systems. *Biomaterials* **2002**, *23* (9), 1955–1966. [https://doi.org/10.1016/S0142-9612\(01\)00322-2](https://doi.org/10.1016/S0142-9612(01)00322-2).
- (55) Malafaya, P. B.; Elvira, C.; Gallardo, A.; Román, J. S.; Reis, R. L. Porous Starch-Based Drug Delivery Systems Processed by a Microwave Route. *J. Biomater. Sci. Polym. Ed.* **2001**, *12* (11), 1227–1241. <https://doi.org/10.1163/156856201753395761>.
- (56) Santander-Ortega, M. J.; Stauner, T.; Loretz, B.; Ortega-Vinuesa, J. L.; Bastos-González, D.; Wenz, G.; Schaefer, U. F.; Lehr, C. M. Nanoparticles Made from Novel Starch Derivatives for Transdermal Drug Delivery. *J. Controlled Release* **2010**, *141* (1), 85–92. <https://doi.org/10.1016/j.jconrel.2009.08.012>.
- (57) Simi, C. K.; Emilia Abraham, T. Hydrophobic Grafted and Cross-Linked Starch Nanoparticles for Drug Delivery. *Bioprocess Biosyst. Eng.* **2007**, *30* (3), 173–180. <https://doi.org/10.1007/s00449-007-0112-5>.

- (58) Godbole, S.; Gote, S.; Latkar, M.; Chakrabarti, T. Preparation and Characterization of Biodegradable Poly-3-Hydroxybutyrate–Starch Blend Films. *Bioresour. Technol.* **2003**, 86 (1), 33–37. [https://doi.org/10.1016/S0960-8524\(02\)00110-4](https://doi.org/10.1016/S0960-8524(02)00110-4).
- (59) Gontard, N.; Guilbert, S. Bio-Packaging: Technology and Properties of Edible and/or Biodegradable Material of Agricultural Origin. In *Food Packaging and Preservation*; Mathlouthi, M., Ed.; Springer US: Boston, MA, 1994; pp 159–181. https://doi.org/10.1007/978-1-4615-2173-0_9.
- (60) Lu, D. R.; Xiao, C. M.; Xu, S. J. Starch-Based Completely Biodegradable Polymer Materials. *Express Polym. Lett.* **2009**, 3 (6), 366–375. <https://doi.org/10.3144/expresspolymlett.2009.46>.
- (61) Röper, H.; Koch, H. The Role of Starch in Biodegradable Thermoplastic Materials. *Starch - Stärke* **1990**, 42 (4), 123–130. <https://doi.org/10.1002/star.19900420402>.
- (62) Lang, Q.; Yin, L.; Shi, J.; Li, L.; Xia, L.; Liu, A. Co-Immobilization of Glucoamylase and Glucose Oxidase for Electrochemical Sequential Enzyme Electrode for Starch Biosensor and Biofuel Cell. *Biosens. Bioelectron.* **2014**, 51, 158–163. <https://doi.org/10.1016/j.bios.2013.07.021>.
- (63) Schmidt, L. D.; Dauenhauer, P. J. Hybrid Routes to Biofuels. *Nature* **2007**, 447 (7147), 914–915. <https://doi.org/10.1038/447914a>.
- (64) Subramanian, S.; Barry, A. N.; Pieris, S.; Sayre, R. T. Comparative Energetics and Kinetics of Autotrophic Lipid and Starch Metabolism in Chlorophytic Microalgae: Implications for Biomass and Biofuel Production. *Biotechnol. Biofuels* **2013**, 6 (1), 150. <https://doi.org/10.1186/1754-6834-6-150>.
- (65) Tanadul, O.; VanderGheynst, J. S.; Beckles, D. M.; Powell, A. L. T.; Labavitch, J. M. The Impact of Elevated CO₂ Concentration on the Quality of Algal Starch as a Potential Biofuel Feedstock. *Biotechnol. Bioeng.* **2014**, 111 (7), 1323–1331. <https://doi.org/10.1002/bit.25203>.
- (66) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, 28 (1), 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- (67) Nielsen, M. M.; Ruzanski, C.; Krucewicz, K.; Striebeck, A.; Cenci, U.; Ball, S. G.; Palcic, M. M.; Cuesta-Seijo, J. A. Crystal Structures of the Catalytic Domain of Arabidopsis Thaliana Starch Synthase IV, of Granule Bound Starch Synthase From CLg1 and of Granule Bound Starch Synthase I of Cyanophora Paradoxa Illustrate Substrate Recognition in Starch Syntheses. *Front. Plant Sci.* **2018**, 9. <https://doi.org/10.3389/fpls.2018.01138>.

- (68) Sheng, F.; Jia, X.; Yep, A.; Preiss, J.; Geiger, J. The Crystal Structures of the Open and Catalytically Competent Closed Conformation of Escherichia Coli Glycogen Synthase. *J. Biol. Chem.* **2009**, *284*, 17796–17807. <https://doi.org/10.1074/jbc.M809804200>.
- (69) Momma, M.; Fujimoto, Z. Interdomain Disulfide Bridge in the Rice Granule Bound Starch Synthase I Catalytic Domain as Elucidated by X-Ray Structure Analysis. *Biosci. Biotechnol. Biochem.* **2012**, *76* (8), 1591–1595. <https://doi.org/10.1271/bbb.120305>.
- (70) Wayllace, N. Z.; Valdez, H. A.; Merás, A.; Ugalde, R. A.; Busi, M. V.; Gomez-Casati, D. F. An Enzyme-Coupled Continuous Spectrophotometric Assay for Glycogen Synthases. *Mol. Biol. Rep.* **2012**, *39* (1), 585–591. <https://doi.org/10.1007/s11033-011-0774-6>.
- (71) Skryhan, K.; Cuesta-Seijo, J. A.; Nielsen, M. M.; Marri, L.; Mellor, S. B.; Glaring, M. A.; Jensen, P. E.; Palcic, M. M.; Blennow, A. The Role of Cysteine Residues in Redox Regulation and Protein Stability of Arabidopsis Thaliana Starch Synthase 1. *PLOS ONE* **2015**, *10* (9), e0136997. <https://doi.org/10.1371/journal.pone.0136997>.
- (72) Keeling, P. L.; Bacon, P. J.; Holt, D. C. Elevated Temperature Reduces Starch Deposition in Wheat Endosperm by Reducing the Activity of Soluble Starch Synthase. *Planta* **1993**, *191* (3), 342–348. <https://doi.org/10.1007/BF00195691>.
- (73) Wilson, W. A.; Pradhan, P.; Madhan, N.; Gist, G. C.; Brittingham, A. Glycogen Synthase from the Parabasalian Parasite Trichomonas Vaginalis: An Unusual Member of the Starch/Glycogen Synthase Family. *Biochimie* **2017**, *138*, 90–101. <https://doi.org/10.1016/j.biochi.2017.04.016>.
- (74) Martín, M.; Wayllace, N. Z.; Valdez, H. A.; Gomez-Casati, D. F.; Busi, M. V. Improving the Glycosyltransferase Activity of Agrobacterium Tumefaciens Glycogen Synthase by Fusion of N-Terminal Starch Binding Domains (SBDs). *Biochimie* **2013**, *95* (10), 1865–1870. <https://doi.org/10.1016/j.biochi.2013.06.009>.
- (75) Hanashiro, I.; Itoh, K.; Kuratomi, Y.; Yamazaki, M.; Igarashi, T.; Matsugasako, J.; Takeda, Y. Granule-Bound Starch Synthase I Is Responsible for Biosynthesis of Extra-Long Unit Chains of Amylopectin in Rice. *Plant Cell Physiol.* **2008**, *49* (6), 925–933. <https://doi.org/10.1093/pcp/pcn066>.
- (76) Raemakers, K.; Schreuder, M.; Suurs, L.; Furrer-Verhorst, H.; Vincken, J.-P.; de Vetten, N.; Jacobsen, E.; Visser, R. G. F. Improved Cassava Starch by Antisense Inhibition of Granule-Bound Starch Synthase I. *Mol. Breed.* **2005**, *16* (2), 163–172. <https://doi.org/10.1007/s11032-005-7874-8>.
- (77) Liu, D.-R.; Huang, W.-X.; Cai, X.-L. Oligomerization of Rice Granule-Bound Starch Synthase 1 Modulates Its Activity Regulation. *Plant Sci.* **2013**, *210*, 141–150. <https://doi.org/10.1016/j.plantsci.2013.05.019>.

- (78) Seung, D.; Soyk, S.; Coiro, M.; Maier, B. A.; Eicke, S.; Zeeman, S. C. PROTEIN TARGETING TO STARCH Is Required for Localising GRANULE-BOUND STARCH SYNTHASE to Starch Granules and for Normal Amylose Synthesis in Arabidopsis. *PLOS Biol.* **2015**, *13* (2), e1002080. <https://doi.org/10.1371/journal.pbio.1002080>.
- (79) Hernández, J. M.; Gaborieau, M.; Castignolles, P.; Gidley, M. J.; Myers, A. M.; Gilbert, R. G. Mechanistic Investigation of a Starch-Branching Enzyme Using Hydrodynamic Volume SEC Analysis. *Biomacromolecules* **2008**, *9* (3), 954–965. <https://doi.org/10.1021/bm701213p>.
- (80) Zmasek, C. M.; Godzik, A. Phylogenomic Analysis of Glycogen Branching and Debranching Enzymatic Duo. *BMC Evol. Biol.* **2014**, *14*, 183. <https://doi.org/10.1186/s12862-014-0183-2>.
- (81) Murakami, T.; Kanai, T.; Takata, H.; Kuriki, T.; Imanaka, T. A Novel Branching Enzyme of the GH-57 Family in the Hyperthermophilic Archaeon *Thermococcus Kodakaraensis* KOD1. *J. Bacteriol.* **2006**, *188* (16), 5915–5924. <https://doi.org/10.1128/JB.00390-06>.
- (82) Chaen, K.; Noguchi, J.; Omori, T.; Kakuta, Y.; Kimura, M. Crystal Structure of the Rice Branching Enzyme I (BEI) in Complex with Maltopentaose. *Biochem. Biophys. Res. Commun.* **2012**, *424* (3), 508–511. <https://doi.org/10.1016/j.bbrc.2012.06.145>.
- (83) Feng, L.; Fawaz, R.; Hovde, S.; Sheng, F.; Nosrati, M.; Geiger, J. Crystal Structures of Escherichia Coli Branching Enzyme in Complex with Cyclodextrins. *Acta Crystallogr. Sect. Struct. Biol.* **2016**, *72*. <https://doi.org/10.1107/S2059798316003272>.
- (84) Feng, L.; Fawaz, R.; Hovde, S.; Gilbert, L.; Choi, J.; Geiger, J. Crystal Structures of E. Coli Branching Enzyme Bound to Linear Oligosaccharides. *Biochemistry* **2015**, *54*. <https://doi.org/10.1021/acs.biochem.5b00228>.
- (85) Hayashi, M.; Suzuki, R.; Colleoni, C.; Ball, S. G.; Fujita, N.; Suzuki, E. Bound Substrate in the Structure of Cyanobacterial Branching Enzyme Supports a New Mechanistic Model. *J. Biol. Chem.* **2017**, *292* (13), 5465–5475. <https://doi.org/10.1074/jbc.M116.755629>.
- (86) Mizuno, K.; Kobayashi, E.; Tachibana, M.; Kawasaki, T.; Fujimura, T.; Funane, K.; Kobayashi, M.; Baba, T. Characterization of an Isoform of Rice Starch Branching Enzyme, RBE4, in Developing Seeds. *Plant Cell Physiol.* **2001**, *42* (4), 349–357. <https://doi.org/10.1093/pcp/pce042>.
- (87) Pfister, B.; Lu, K.-J.; Eicke, S.; Feil, R.; Lunn, J. E.; Streb, S.; Zeeman, S. C. Genetic Evidence That Chain Length and Branch Point Distributions Are Linked Determinants of Starch Granule Formation in Arabidopsis. *Plant Physiol.* **2014**, *165* (4), 1457–1474. <https://doi.org/10.1104/pp.114.241455>.

- (88) Takata, H.; Ohdan, K.; Takaha, T.; Kuriki, T.; Okada, S. Properties of Branching Enzyme from Hyperthermophilic Bacterium, *Aquifex Aeolicus*, and Its Potential for Production of Highly-Branched Cyclic Dextrin. *J. Appl. Glycosci.* **2003**, *50* (1), 15–20. <https://doi.org/10.5458/jag.50.15>.
- (89) Lee, C.-K.; Le, Q.-T.; Kim, Y.-H.; Shim, J.-H.; Lee, S.-J.; Park, J.-H.; Lee, K.-P.; Song, S.-H.; Auh, J. H.; Lee, S.-J.; et al. Enzymatic Synthesis and Properties of Highly Branched Rice Starch Amylose and Amylopectin Cluster. *J. Agric. Food Chem.* **2008**, *56* (1), 126–131. <https://doi.org/10.1021/jf072508s>.
- (90) Palomo, M.; Kralj, S.; van der Maarel, M. J. E. C.; Dijkhuizen, L. The Unique Branching Patterns of *Deinococcus* Glycogen Branching Enzymes Are Determined by Their N-Terminal Domains. *Appl. Environ. Microbiol.* **2009**, *75* (5), 1355–1362. <https://doi.org/10.1128/AEM.02141-08>.
- (91) Yoon, S.-A.; Ryu, S.-I.; Lee, S.-B.; Moon, T.-W. Purification and Characterization of Branching Specificity of a Novel Extracellular Amylolytic Enzyme from Marine Hyperthermophilic *Rhodothermus Marinus*. *J. Microbiol. Biotechnol.* **2008**, *18* (3), 457–464.
- (92) Kim, E.-J.; Ryu, S.-I.; Bae, H.-A.; Huong, N. T.; Lee, S.-B. Biochemical Characterisation of a Glycogen Branching Enzyme from *Streptococcus Mutans*: Enzymatic Modification of Starch. *Food Chem.* **2008**, *110* (4), 979–984. <https://doi.org/10.1016/j.foodchem.2008.03.025>.
- (93) Suzuki, R.; Koide, K.; Hayashi, M.; Suzuki, T.; Sawada, T.; Ohdan, T.; Takahashi, H.; Nakamura, Y.; Fujita, N.; Suzuki, E. Functional Characterization of Three (GH13) Branching Enzymes Involved in Cyanobacterial Starch Biosynthesis from *Cyanobacterium* Sp. NBRC 102756. *Biochim. Biophys. Acta* **2015**, *1854* (5), 476–484. <https://doi.org/10.1016/j.bbapap.2015.02.012>.
- (94) Jo, H.-J.; Park, S.; Jeong, H.-G.; Kim, J.-W.; Park, J.-T. *Vibrio Vulnificus* Glycogen Branching Enzyme Preferentially Transfers Very Short Chains: N1 Domain Determines the Chain Length Transferred. *FEBS Lett.* **2015**, *589* (10), 1089–1094. <https://doi.org/10.1016/j.febslet.2015.03.011>.
- (95) Thiemann, V.; Saake, B.; Vollstedt, A.; Schäfer, T.; Puls, J.; Bertoldo, C.; Freudl, R.; Antranikian, G. Heterologous Expression and Characterization of a Novel Branching Enzyme from the Thermoalkaliphilic Anaerobic Bacterium *Anaerobranca Gottschalkii*. *Appl. Microbiol. Biotechnol.* **2006**, *72* (1), 60–71. <https://doi.org/10.1007/s00253-005-0248-7>.
- (96) Takata, H.; Takaha, T.; Okada, S.; Takagi, M.; Imanaka, T. Cyclization Reaction Catalyzed by Branching Enzyme. *J. Bacteriol.* **1996**, *178* (6), 1600–1606.

- (97) Binderup, K.; Preiss, J. Glutamate-459 Is Important for Escherichia Coli Branching Enzyme Activity. *Biochemistry* **1998**, *37* (25), 9033–9037. <https://doi.org/10.1021/bi980199g>.
- (98) Hamada, S.; Nozaki, K.; Ito, H.; Yoshimoto, Y.; Yoshida, H.; Hiraga, S.; Onodera, S.; Honma, M.; Takeda, Y.; Matsui, H. Two Starch-Branching-Enzyme Isoforms Occur in Different Fractions of Developing Seeds of Kidney Bean. *Biochem. J.* **2001**, *359* (Pt 1), 23–34.
- (99) Kuriki, T.; Stewart, D. C.; Preiss, J. Construction of Chimeric Enzymes out of Maize Endosperm Branching Enzymes I and II: Activity and Properties. *J. Biol. Chem.* **1997**, *272* (46), 28999–29004. <https://doi.org/10.1074/jbc.272.46.28999>.
- (100) Cornejo-Ramírez, Y. I.; Martínez-Cruz, O.; Del Toro-Sánchez, C. L.; Wong-Corral, F. J.; Borboa-Flores, J.; Cinco-Moroyoqui, F. J. The Structural Characteristics of Starches and Their Functional Properties. *CyTA - J. Food* **2018**, *16* (1), 1003–1017. <https://doi.org/10.1080/19476337.2018.1518343>.
- (101) Buleon, A.; Colonna, P.; Planchot, V.; Ball, S. Starch Granules: Structure and Biosynthesis. *Int. J. Biol. Macromol.* **1998**, *23* (2), 85–112. [https://doi.org/10.1016/S0141-8130\(98\)00040-3](https://doi.org/10.1016/S0141-8130(98)00040-3).
- (102) Suzuki, E.; Suzuki, R. Distribution of Glucan-Branching Enzymes among Prokaryotes. *Cell. Mol. Life Sci.* **2016**, *73* (14), 2643–2660. <https://doi.org/10.1007/s00018-016-2243-9>.
- (103) Abad, M.; Binderup, K.; Rios-Steiner, J.; Arni, R.; Preiss, J.; Geiger, J. The X-Ray Crystallographic Structure Of Escherichia Coli Branching Enzyme. *J. Biol. Chem.* **2002**, *277*, 42164–42170. <https://doi.org/10.1074/jbc.M205746200>.
- (104) Pal, K.; Kumar, S.; Sharma, S.; Garg, S. K.; Alam, M. S.; Xu, H. E.; Agrawal, P.; Swaminathan, K. Crystal Structure of Full-Length Mycobacterium Tuberculosis H37Rv Glycogen Branching Enzyme: Insights of N-Terminal β -Sandwich in Substrate Specificity and Enzymatic Activity. *J. Biol. Chem.* **2010**, *285* (27), 20897–20903. <https://doi.org/10.1074/jbc.M110.121707>.
- (105) Noguchi, J.; Chaen, K.; Vu, N. T.; Akasaka, T.; Shimada, H.; Nakashima, T.; Nishi, A.; Satoh, H.; Omori, T.; Kakuta, Y.; et al. Crystal Structure of the Branching Enzyme I (BEI) from Oryza Sativa L with Implications for Catalysis and Substrate Binding. *Glycobiology* **2011**, *21* (8), 1108–1116. <https://doi.org/10.1093/glycob/cwr049>.
- (106) Feng, L.; Fawaz, R.; Hovde, S.; Gilbert, L.; Chiou, J.; Geiger, J. H. Crystal Structures of Escherichia Coli Branching Enzyme in Complex with Linear Oligosaccharides. *Biochemistry* **2015**, *54* (40), 6207–6218. <https://doi.org/10.1021/acs.biochem.5b0>