USING GENETIC CODE EXPANSION AND RATIONAL DISULFIDE BOND DESIGN TO ENGINEER IMPROVED ACTIVITY AND (THERMO)STABILITY OF *RHODOCOCCUS OPACUS* CATECHOL 1,2-DIOXYGENASE

By

Joshua Lister

A thesis submitted to the Department of Chemistry and Biomolecular Sciences

In conformity with the requirements for

the degree of Master of Science

University of Ottawa

Ottawa, Ontario, Canada

(January 15, 2024)

Abstract

Catechol 1,2-Dioxygenase from *Rhodococcus opacus* is a type of intradiol dioxygenase enzyme that catalyzes the conversion of catechol to cis, cis muconic acid. This enzymatic conversion has the potential to be useful in a number of different applications such as treating wastewater contaminated with aromatic compounds to creating a greener method to produce cis, cis muconic acid which can be used to make a number of industrially important base chemicals. However, for enzymes to be used in industrial conditions, they must be highly stable. The experimental chapters in this thesis explore whether this enzyme can be stabilized to meet industrial requirements while minimizing any loss in catalytic activity. Through the studies described in Chapter 2, a mutant enzyme was generated through disulfide bond engineering with significantly improved thermostability. However overall catalytic activity was reduced. Toward addressing this loss of catalytic activity, in Chapter 3, attempts were made to implement state-of-the-art genetic code expansion strategies to increase catalytic activity of the enzymes. However, these attempts were unsuccessful. Finally, Chapter 4 describes how future stability engineering could be optimized using design pipelines similar to the one developed in this study. Additionally, it describes possible additional optimizations toward making the application of these enzymes cost effective in the near future.

Acknowledgements

"If I have seen further, it is by standing on the shoulders of giants." -Isaac Newton

The undertaking of a Master's degree is no small feat; for two years you pour your passion, drive, skills, and energy into this commitment. It is a difficult journey, there have been times when I felt on top of the world and times when I questioned if this was even the right path for me. However, these past two years have been some of the most rewarding of my life and I'm truly grateful to have undertaken this journey. While I am incredibly proud and humbled to have made it to this point, no one succeeds alone. I could not have reached this point without the support of countless people. While there are too many people to list here, there are a few I would like to specifically mention.

Firstly, I would like to mention my incredible supervisor, mentor, and friend; Michele Loewen. When I first started in your lab, I was an undergrad with limited experience and knowledge of how to conduct research. I truly believe that the time I've spent in your lab has not only made me a better scientist, but also a better person. Your constant encouragement helped inspire me to push forward and do my very best, and your feedback and wisdom have helped me to hone my skills and develop both personally and professionally. You've been an incredible role model over the years, and you've taught me how to work hard while still getting the most out of life.

Next, I'd like to thank Antony St-Jacques. Your tough love forced me to learn how to learn and work independently. You were always available for help and guidance.

I'd also like to give a specific thank you to Kelly Robinson. I remember the first day I walked into the lab, you were the first person I met. I was honestly not very good in the lab, or at working independently but you helped teach me different techniques, how to take good notes, how to organize my stuff, and helped me hone so many skills to develop into the scientist I am today.

To the rest of the team Fang, Chengsong, Trent, Emma, Ricardo, Ben, Joanna, Yutong, Djay, Tanya, and Luana. You've not only been work colleagues, but you've also been like a second family to me.

To my family and friends, thank you so much for all your love and support. You kept me sane when I was feeling overwhelmed and helped support me during the ups and downs of my degree.

To my grandmother Wauneta, though your dementia caused you to keep forgetting what I did. Every time you asked me what I was doing in school you got so excited and proud of me and it always inspired me to feel proud of myself, even when I wasn't. Though your memory was failing you, you maintained your spunky personality and always were able to give good life advice and make everyone laugh. I will forever treasure the time we had together.

Thank you to everyone I've crossed paths with on this journey, I'll treasure these memories forever!

Table of Contents

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

List of Tables

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

List of Abbreviations

3,4-PCD	protocatechuate 3,4-dioxygenase
1,2-CTD	catechol 1,2-dioxygenase
aaRS	aminoacyl tRNA synthetase
*Aci*1,2-CTD	catechol 1,2-dioxygenase from *Acinetobacter* sp. ADP1
APX2	engineered ascorbate peroxidase
BH32	computationally designed enzyme for the Morita-Baylis-Hillman reaction
ccMA	cis, cis muconic acid
CTD	catechol dioxygenase
CW	Combined Washes
DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)
DTT	dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FT	Flowthrough
GCE	Genetic Code Expansion
HQ12O	hydroxyquinol 1,2-dioxygenase
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-d-thiogalactopyranoside
ncAA	non-canonical amino acid
NmH	Nδ-methylhistidine
*Pseu*1,2-CTD	catechol 1,2-dioxygenase from *Pseudomonas arvilla*
RF1	Release Factor 1
*Rho*3-CCD	3-chlorocatechol 1,2-dioxygenases from *R. opacus*
*Rho*1,2-CTD	catechol 1,2-dioxygenase from *R opacus*
SEC	Size Exclusion Chromatography
Supe	Supernatant
T_{50}	Temperature of half-inactivation
$t_{1/2}$	Half-Life
T_m	Melting Temperature
TTN	Total Turnover Number
WT	wild type

Chapter 1

Literature Review

1.1 Aromatic compounds and their catabolism by microorganisms

1.1.1 Aromatic compounds

Some microorganisms possess the exceptional ability to utilize aromatic compounds as their sole energy and carbon source. This capability is crucial to maintaining the global carbon cycle[1]. Aromatic compounds are planar, fully conjugated, ring-shaped carbon-based molecules possessing $(4n+2)\pi$ electrons, where n is a non-negative integer. These compounds are exceptionally stable due to the delocalization of their π orbitals.

Aromatic compounds are formed by a variety of biogeochemical processes. They are found as lignin components, aromatic amino acids and xenobiotic compounds[2]. As such these compounds are widely distributed in nature and range in size from low-molecular-weight compounds, such as phenols, to biopolymers such as lignin (the second most abundant polymer in nature)[3].

In addition, the stability of aromatic compounds has contributed to the widespread production and usage of naturally occurring and non-natural aromatic compounds in a variety of industrial applications. A report from Shell Chemical lists numerous uses such as the production of clothing, packaging, paints, adhesives, unbreakable windows, plywood, computer casings, compact discs, dyes, agrochemicals, pharmaceuticals and many more. Industrial usage leads to the distribution of stable, non-natural aromatic compounds in the environment. Catechol is a benzenediol comprising of a benzene core carrying two hydroxy substituents ortho to each other[4]. Catecholic compounds are present in the wastewaters from many industrial processes, such as the production of oil and oil products, furnace coke, steel, explosives, paint, cork, and

fiberglass, the production and recycling of rubber goods, the textile industry, and some branches of the food and beverage industry[5–7].

1.1.2 Catabolism of aromatic compounds in microorganisms

There are diverse catabolic pathways employed by microorganisms to degrade aromatic compounds including both aerobic and anaerobic strategies. In principle, microorganisms utilize a distinct catabolic pathway to degrade each type of aromatic compound. However, the catabolic pathways of these compounds all involve 2 key steps and usually proceed via one of four intermediates.

In the first key step, the thermodynamically stable benzene ring is activated. In the case of aerobic degradation, this occurs through the action of an oxygenase via the incorporation of oxygen-containing substituents[8]. This first key step results in a hydroxylated ring intermediate. The second key step is the subsequent cleavage of this hydroxylated ring intermediate by a dioxygenase to give a cleaved ring intermediate[8]. The ring cleavage product is then degraded to give smaller units that will enter the tri-carboxylic acid cycle (Figure 1.1). Compounds containing more than one aromatic ring are degraded via interactions of the strategies used to degrade monocyclic compounds.

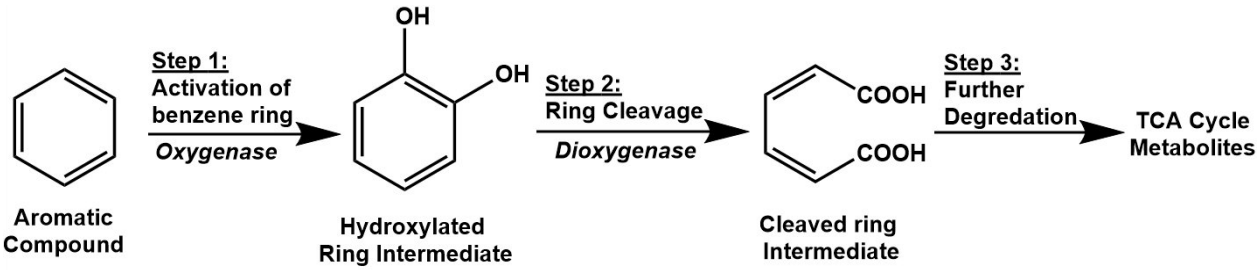

Figure 1.1: General pathway of aerobic aromatic compound catabolism in microorganisms. The catabolism of benzene, a common aromatic, is shown as an example to showcase the key steps and intermediates in this process. Note: in the ring cleaving step, there are two possible straight chain intermediates that can be formed, however, only one of these is shown for improved clarity.

The catabolic pathways of aromatic compounds usually proceed via one of four intermediates catechol, protocatechuate, gentisate, or hydroquinone[1] (Figure 1.2A). The first key intermediate, catechol, occurs in the catabolism of benzene[9], benzoate[10], phenol[11], and derivatives thereof. Catecholic intermediates that arise in the degradation of polycyclic aromatics such as naphthalene[12] and biphenyl[13] may be considered as substituted catechols. Interestingly, catechols also occur in the degradation of diterpenoids and steroids in what may be termed "cryptic aromatic catabolism"[1]. Catechol is subjected to both intradiol and extradiol modes of cleavage (Figure 1.2B) . The second key intermediate, Protocatechuate, occurs in the catabolism of hydroxybenzoates[14,15], phthalates[16], and vanillyl alcohols[17,18]. Protocatechuate is subject to three different modes of cleavage: intradiol cleavage, 2,3-extradiol cleavage, and 4,5-extradiol cleavage (Figure 1.2C). The third key intermediate, gentisate, occurs in the catabolism of salicylate[19], 3-hydroxybenzoic acid[20], and anthranilate[21]. Homogentisate, an important derivative of gentisate is involved in the catabolism of phenylalanine, tyrosine, and tryptophan[1]. Gentisate is only subject to extradiol cleavage (Figure 1.2D). The fourth key intermediate, hydroxyquinol, occurs in the catabolism of *p*-nitrophenol[22], resorcinol[23], 4-chlorophenol[24], 4-hydroxysalicylic acid[25], 2-chlorophenol[26] and 2,4-dichlorophenol[26]. The related 6-chlorohydroxyquinol occurs in the catabolism of 2,6-dichlorophenol and 2,4,6-trichlorophenol[26]. Hydroxyquinol is further degraded by intradiol cleavage (Figure 1.2E). Several other pathways involving derivates of these intermediates or involving different intermediates have been reported but are not shown here.

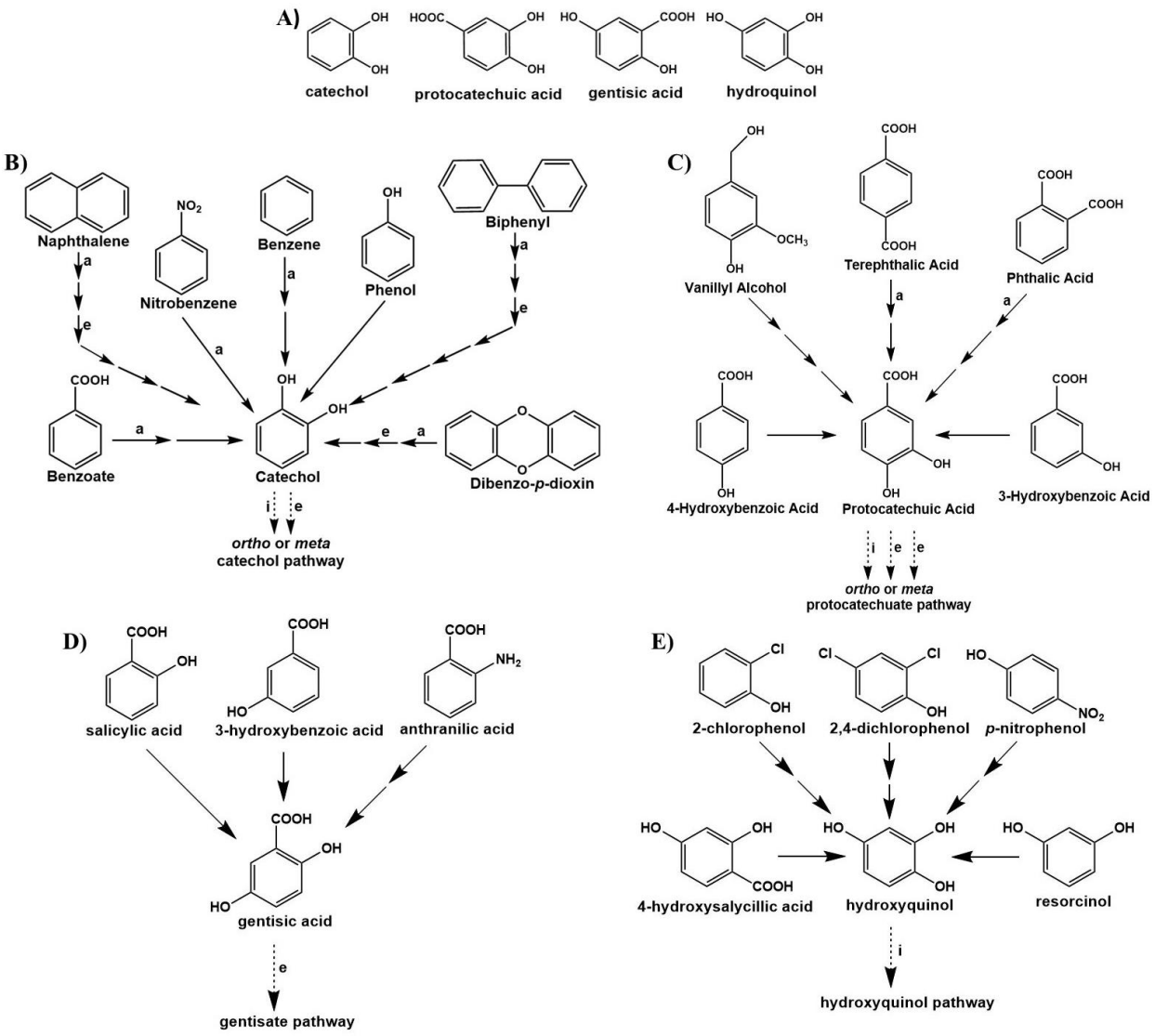

Figure 1.2: Pathways for the catabolism of aromatic compounds leading to one of the four ring-cleavage intermediates (adapted from[1]). A) structures of 4 key intermediates in microbial aromatic compound catabolism. B)-E) catabolic pathways leading to intermediates: B) catechol; C) protocatechuic acid; D) gentisic acid; E) hydroxyquinol. Each solid arrow indicates a single enzyme-catalyzed reaction. Dotted arrows indicate multiple enzyme-catalyzed reactions. Reactions labelled with an "a" are catalyzed by a Rieske non-heme iron oxygenase, reactions labelled with an "i" are catalyzed by an intradiol dioxygenase, reactions labelled with an "e" are catalyzed by an extradiol or extradiol-type dioxygenase.

1.2 Problems with, and potential uses of catechol

1.2.1 Catechol and other phenolic compounds are common pollutants

It is well established that phenolic compounds such as catechol are found in the environment both as natural and artificial monoaromatic products. They are among the most common air and water pollutants[27].

These compounds are present in the wastewaters of many industrial processes such as the production of oil and oil products, furnace coke, steel, explosives, paint, cork, and fiberglass, the production and recycling of rubber goods, the textile industry, and some branches of the food and beverage industry[5–7]. For example, the concentration of catechol in wastewater varies from a few to 2000 mg/L in wastewater from coal carbonization and gasification; at low-temperature wastewater from coal carbonization, its concentration may be as high as 5300 mg/L[28–30]. Phenol concentrations in some wastewaters can reach 10 g/L[31].

The negative impact of these compounds on the environment and human health has been exhaustively reported and discussed in the scientific literature[32]. Phenolic compounds such as catechol have been listed as priority pollutants by the US EPA[29,33]. Catechol is a toxic and persistent water pollutant in the environment[34], it is fatally toxic to fish at concentrations of 5-25 mg/L, inhibits biological growth in microorganisms, and has been classified as a carcinogenic risk to humans by the International Agency for Research on Cancer[34–36]. In order to protect against hazards posed by catechol to public health and the environment, aqueous waste contaminated with catechol must be treated with an efficient, cost-effective and environmentally benign technique before being discharged[33,37].

1.2.2 Catechol can be converted to a high-value-added product

Catechol can be readily converted to cis, cis-muconic acid (ccMA), a high-value-added bio-product. Because of its stereochemical configuration along with the reactive dicarboxylic groups and the conjugated double bonds, ccMA can undergo an large array of reactions (see Figure 1.3). Recently, this molecule has garnered increasing interest owing to its potential applications in the manufacture of new functional resins, food additives, agrochemicals, and pharmaceuticals[38]. Among its other uses, muconic acid can be easily converted into commercially important bulk chemicals such as adipic acid, terephthalic acid, and trimellitic acid[38].

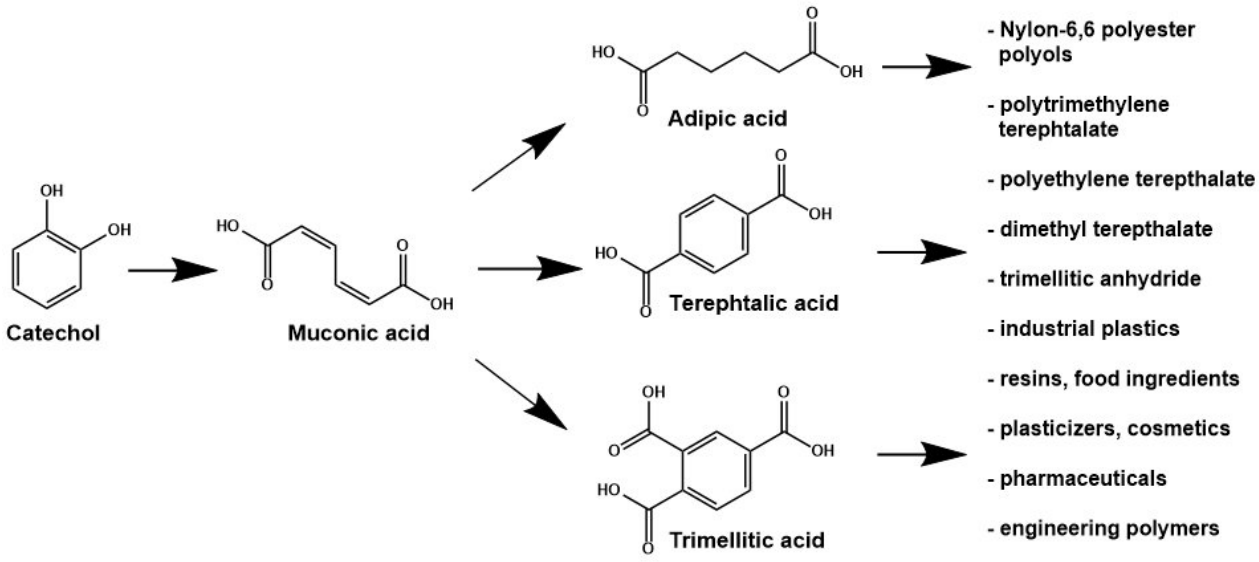

Figure 1.3: Summary of commercial products that can be produced from ccMA

1.3 Ring-Cleaving Dioxygenases

1.3.1 Ring-cleaving dioxygenase enzymes in aromatic compound catabolism

The four key intermediates involved in aromatic compound catabolism intermediates contain at least two hydroxyl substituents which may be next to (as in the case of catechol, protocatechuate and hydroquinone) or opposite to each other (as in the case of gentisate) around the ring. Where the hydroxyl groups are next to each other, the ring will either be cleaved between them (ortho cleavage) or to one side of them (meta cleavage). Where the hydroxyl groups are opposite to each other, the ring is cleaved immediately to one side of one of the hydroxyl groups[1].

These different modes of cleavage are performed by dioxygenase enzymes from one of two distinct classes: intradiol and extradiol dioxygenases[39] (see Figure 1.4). Intradiol dioxygenases employ non-heme Fe(III) to cleave the aromatic ring ortho to (between) the hydroxyl substituents via the incorporation of two atoms of molecular oxygen into the substrate[40–42]. Extradiol dioxygenases utilize non-heme Fe(II) to catalyze the ring fission meta (adjacent) to the hydroxyl groups[43,44].

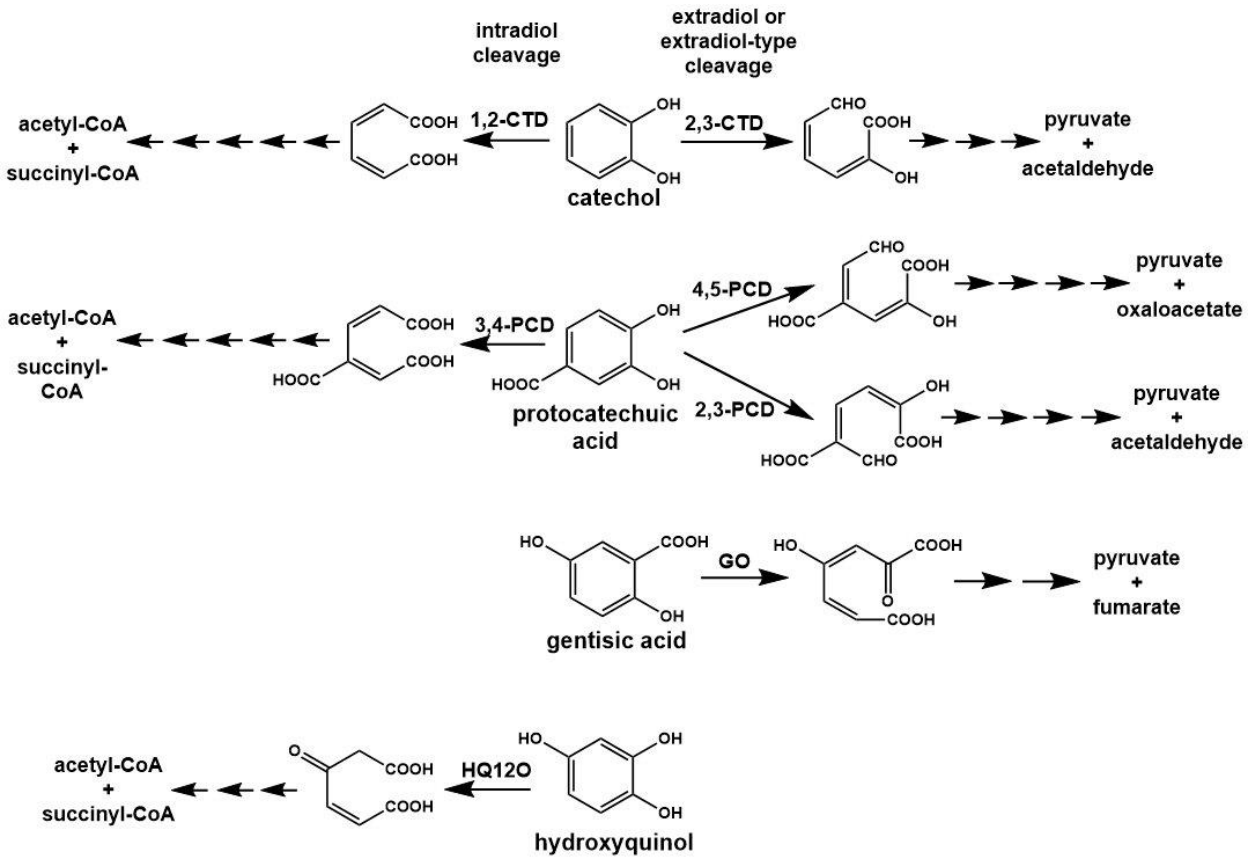

Figure 1.4: Downstream products arising from intradiol and extradiol cleavage of the 4 key ring cleavage intermediates: catechol, protocatechuic acid, gentisic acid, and hydroquinol (adapted from[1]). Each arrow indicates a single enzyme-catalyzed reaction. Depending on the positioning of the hydroxyl groups in the intermediate, they can be cleaved by either an intradiol dioxygenase, an extradiol (or extradiol-type) dioxygenase or some intermediates can be cleaved by both. Following this cleavage, the straight chain intermediate will be further catabolized into a form that can enter the TCA cycle. [1]

Acronyms:

1,2-CTD= catechol 1,2-dioxygenase

2,3-CTD= catechol 2,3-dioxygenase

3,4-PCD= protocatechuate 3,4-dioxygenase

4,5-PCD= protocatechuate 4,5-dioxygenase

2,3-PCD= protocatechuate 2,3-dioxygenase

GO= gentisate 1,2-dioxygenase

HQ12O= 2-hydroxyquinol 1,2-dioxygenase

1.3.2 Classification of Ring-Cleaving Dioxygenases

Intradiol and extradiol enzymes do not share significant sequence or structural similarities, indicating they belong to evolutionarily distinct classes of proteins (Table 1.1). Sequence and structural analyses further indicate that all intradiol dioxygenases characterized to date share a common evolutionary lineage. Thus, despite their different subunit compositions, the catalytic domains of 3,4-PCD (protocatechuate 3,4-dioxygenase) and 1,2-CTD (catechol 1,2-dioxygenase) share a common structural fold[45]. Moreover, these enzymes share key conserved residues with the HQ12O (hydroxyquinol 1,2-dioxygenase), including the four endogenous iron coordinating ligands[46].

In contrast, extradiol and extradiol-type dioxygenases belong to at least three evolutionarily independent families[47,48]. Type I extradiol dioxygenases belong to the vicinal oxygen chelate superfamily[49,50]. Type I extradiol dioxygenases include both one- and two-domain enzymes. Type II extradiol dioxygenases belong to an unknown superfamily and include both one- and two-domain enzymes. Type III extradiol dioxygenases belong to the cupin superfamily[48].

Table 1.1: Families of ring-cleavage enzymes based on structural folds

Type	Superfamily	Prototypic members	Subunit
Intradiol		*Pseu*3,4-PCD	$(\alpha\ \beta)_{12}$
		*Rho*1,2-CTD	α_2
		*Rho*3-CCD	α_2
		*Aci*1,2-CTD	α_2
		*Noc*HQ12O	α_2
Extradiol I	Vicinal oxygen chelate	*Bur*DHBD	α_8
		*Pseu*2,3-CTD	α_4
		Bul2,3-HPCD	α_4
		*Rho*DHBD-III	α_2
Extradiol II	Unknown	*Esc*2,3-DHPPD	α_4
		*Pseu*GLD	α_3
		*Sph*4,5-PCD	$\alpha_2\beta_2$
		*Com*4,5-PCD	$\alpha_4\beta_4$
		*Pseu*1,6-APD	$\alpha_2\beta_2$
Extradiol III	Cupin	*Pseu*GO	α_4
		*Noc*HND	α_6

*Pseu*3,4-PCD protocatechuate 3,4-dioxygenase from *P. putida* B-10[51]
*Rho*1,2-CTD catechol 1,2-dioxygenase from *R. opacus*[52]
*Rho*3-CCD 3-chlorocatechol 1,2-dioxygenases from *R. opacus*[53]
*Aci*1,2-CTD catechol 1,2-dioxygenase from *Acinetobacter* sp. ADP1[45]
*Noc*HQ12O hydroxyquinol 1,2-dioxygenase from Nocardioides simplex 3E[54]
*Bur*DHBD 2,3-dihydroxybiphenyl 1,2-dioxygenase from *Burkholderia* sp. LB400[55]
*Pseu*2,3-CTD catechol 2,3-dioxygenase from *P. putida* mt-2[56]
Bul2,3-HPCD homoprotocatechuate 2,3-dioxygenase from *B. fuscum*[57]
*Rho*DHBD-III 2,3-dihydroxybiphenyl 1,2-dioxygenase III from *R. globerulus* P6[58]
*Esc*2,3-DHPPD 2,3-dihydroxyphenylpropionate 1,2-dioxygenase from *E. coli*[59]
*Pseu*GTD gallate dioxygenase from *P. putida* KT2440[60]
*Sph*4,5-PCD protocatechuate 4,5-dioxygenase from *Sphingomonas paucimobilis* SYK-6[61]
*Com*4,5-PCD protocatechuate 4,5-dioxygenase from *Comomonas testosteroni* T-2[62]
*Pseu*1,6-APD 2-aminophenol 1,6-dioxygenase from *P. pseudoalcaligenes* JS45[63]
*Pseu*GO gentisate dioxygenase from *P. testosteroni*[64]
*Noc*HND 1-hydroxy-2-naphthoate dioxygenase from *Nocardioides* sp. KP7[65]

1.4 Catechol 1,2-Dioxygenase from *Rhodococcus opacus*

The specific ring-cleaving dioxygenase used in this research is catechol 1,2-dioxygenase from the gram-positive bacterium *Rhodococcus opacus (Rho1,2-CTD). Rhodococcus opacus is* capable of catabolizing a broad range of feedstocks, and recent technological advances have further demonstrated its potential for lignin and industrial waste valorization[52]. Previous studies have shown that *R. opacus* has some existing capacity to utilize lignin and industrial wastes for the production of lipids and lipid-based products, however, the efficiency of lignin bioconversion needs to be further enhanced[66]. Thus, *Rho*1,2-CTD provided a strong starting point for this project.

1.4.1 Catalytic mechanism of catechol 1,2-dioxygenase

*Rho*1,2-CTD is a type of intradiol dioxygenase that uses catechol as its primary substrate. The catalytic mechanism of *Rho*1,2-CTD has been proposed as follows; intradiol cleavage commences via a substrate activation mechanism, in which the catechol substrate is activated by iron(III) to give an iron(II) semiquinone. The semiquinone reacts directly with dioxygen to give a hydroperoxide intermediate, which then undergoes Criegee rearrangement via acyl migration to muconic anhydride (cis, cis-muconic acid)[67]. An alternative mechanism has also been proposed wherein the electron-deficient acyl group migrates via a benzene-oxide-oxepin interconversion[68] (Figure 1.5).

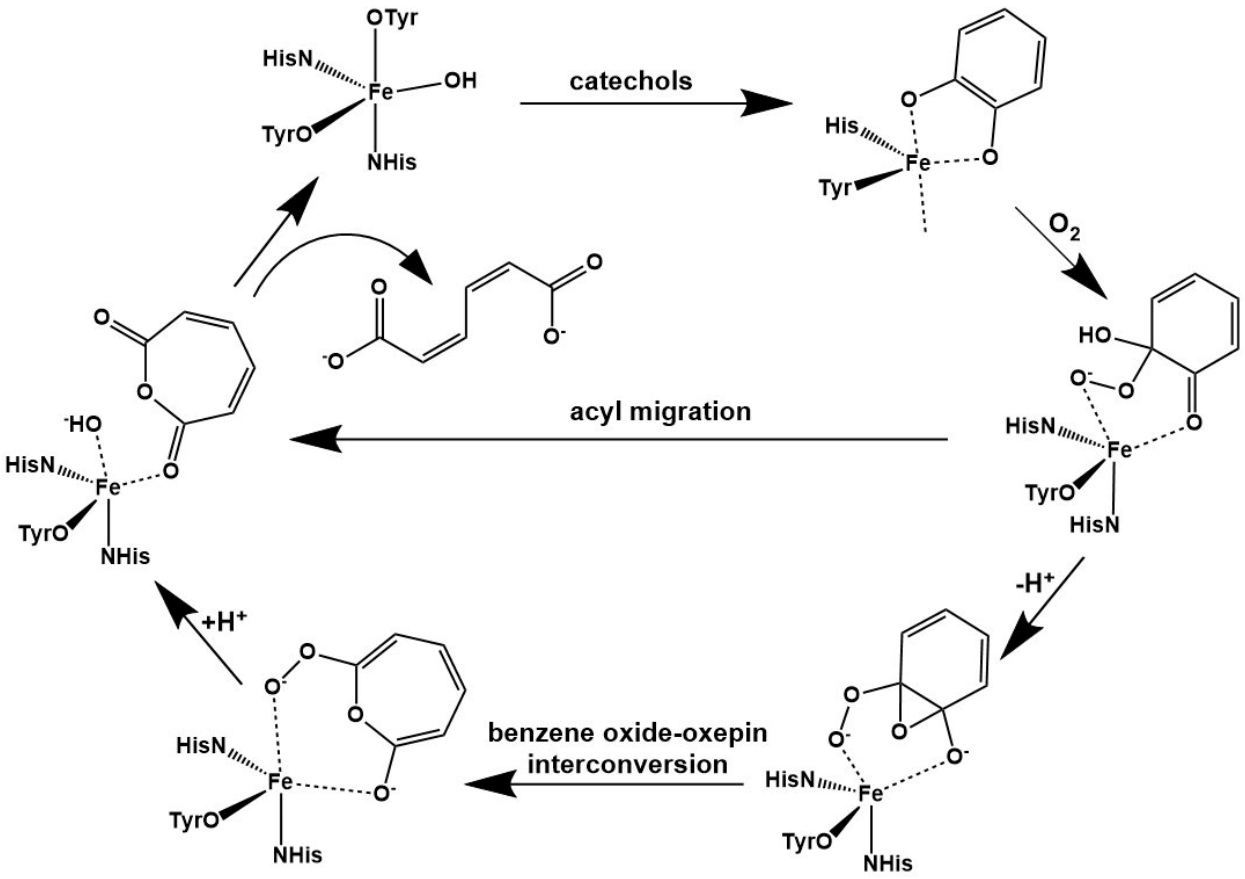

Figure 1.5: Scheme of catechol 1,2-dioxygenase catalytic mechanism displaying the 2 proposed mechanisms for the degradation of catechol (adapted from[52]).

1.4.2 Structural information and comparison of catechol 1,2-dioxygenase to other intradiol dioxygenases

Structural comparison of known crystal structures shows that intradiol dioxygenases present two variations of a common underlying architecture[1]. These include: 1) a dimer comprising two subunits of related structure. 2) Protomers being a single subunit.

*Rho*1,2-CTD has the latter structure (Figure 1.6). Enzymes sharing this architecture are homodimeric. The subunits of these proteins are approximately 300 residues and are folded into two domains: a catalytic domain that replicates the basic core structure shared by all intradiol dioxygenases, and an N-terminal domain that mediates dimerization[1]. The N-terminal domain consists of roughly 100 residues that fold into five helices[1]. The catalytic domain consists of a

non-heme iron(III) cofactor ligated by four amino acid side chains: two histidine and two tyrosine residues (His220, His222, Tyr162, Tyr196 in the case of *Rho*1,2-CTD)[51]. A fifth ligand, a water molecule/hydroxide ion completes a trigonal bipyramidal structure[52]. An extended segment links the dimerization domain to the catalytic domain. The quaternary structure is a Fe(III)α_2 homodimer containing a novel helical zipper motif at the interface of the two subunits, and contains two molecules of bound phospholipid[52].

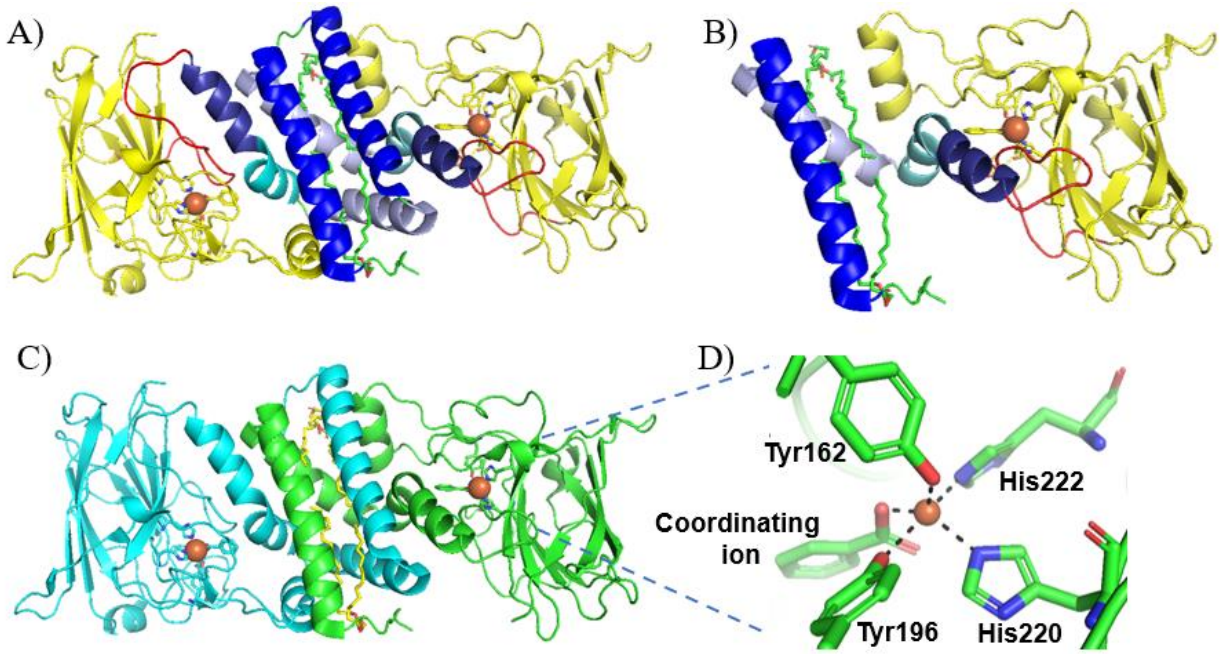

Figure 1.6: Structures of *Rho*1,2-CTD (PDB ID: 3HGI) coloured to emphasize structural features. A) and B) emphasize common structural features Rho1,2-CTD shares with other intradiol dioxygenases. A) Biological assembly, B) Single subunit view. The helixes of the dimerization domain are shown in different shades of blue. The catalytic domain is shown in yellow. The extended segment linking the dimerization domain to the catalytic domain is shown in red. Molecules of bound phospholipid are shown in green. Catalytic Fe (III) is shown in orange. C) and D) showcase homodimers and location of active site. C) Biological assembly. Subunit 1 is coloured in green, subunit 2 is coloured in blue. Molecules of bound phospholipid are coloured in yellow. Catalytic Fe (III) is shown in orange. D) Close up of catalytic site. A non-heme iron(III) cofactor is ligated by four amino acid side chains: two histidine's (His220 and His222) and two tyrosine residues (Tyr162 and Tyr196). A fifth ligand, a coordinating ion (water/hydroxide) completes a trigonal bipyramidal structure. Note in this image the water/hydroxide coordinating ion is replaced with a benzoate. Structures were visualized using PyMOL[69].

The fold of *Rho*1,2-CTD is similar to those of other 1,2-CTDs, with large structural homology (see Figure 1.7). However, there are some unique differences[52]. The structure of *Rho*1,2-CTD presents the highest homology with *Rho*3-CCD. The main structural differences are found in the dimerization domain, where *Rho*1,2-CTD is 28 residues longer than *Rho*3-CCD. The major differences with the Gram-negative *Aci*1,2-CTDs occur in both domains (see Figure 1.8). Compared to prototypical gram-negative 1,2-CTDs, in the dimerization domain of *Rho*1,2-CTD, helix 5 is shorter and strand 1 is missing. In the catalytic domain of *Rho*1,2-CTD, a new short helix is present between positions 124-128, helix 7 is missing, and the C-terminus is 27-30 residues shorter than other 1,2-CTDs from Gram-negative bacteria. Moreover, residues 23-32 in *Rho*1,2-CTD present a random coil region as opposed to an alpha-helix structure as in other 1,2-CTDs. Additionally, it extends in a different direction; the 27 N-terminal residues in *Aci*1,2-CTD and *Pseu*1,2-CTD[70] form the first helix (and connection to the second) that points in the direction of the other subunit, whereas in *Rho*1,2-CTD residues 23-32 are directed in the opposite direction, towards the catalytic domain.

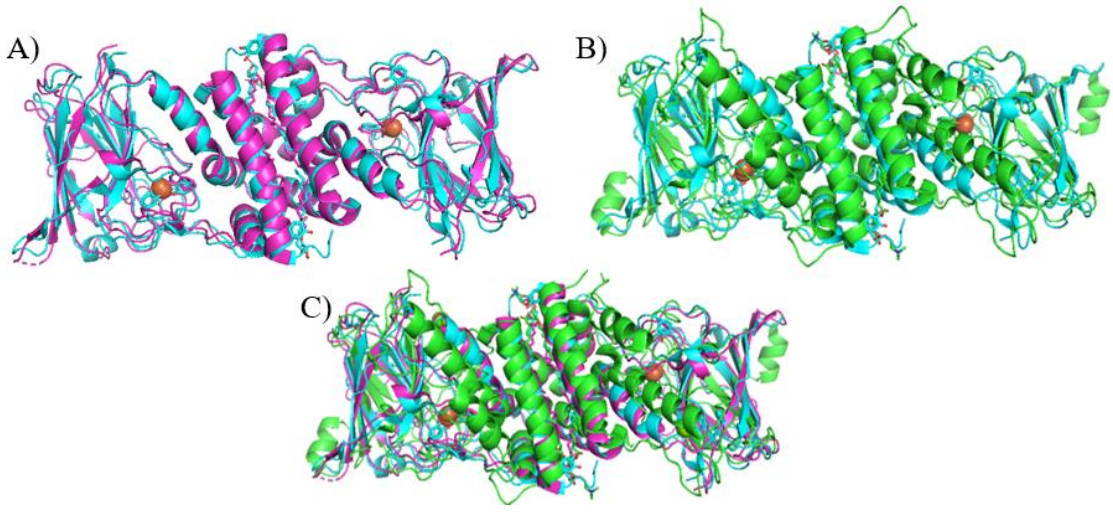

Figure 1.7: Structural alignment of *Rho*1,2-CTD (in cyan with *Aci*1,2-CTD (green) and *Rho*3-CCD (pink). A) structural alignment of *Rho*3-CCD to *Rho*1,2-CTD (backbone RMSD alignment 1.429 Å) B) Alignment of *Aci*1,2-CTD to *Rho*1,2-CTD (backbone RMSD alignment 3.845 Å) C) Alignment of *Rho*3-CCD and *Aci*1,2-CTD to *Rho*1,2-CTD. Structures were aligned and visualized with PyMOL[69].

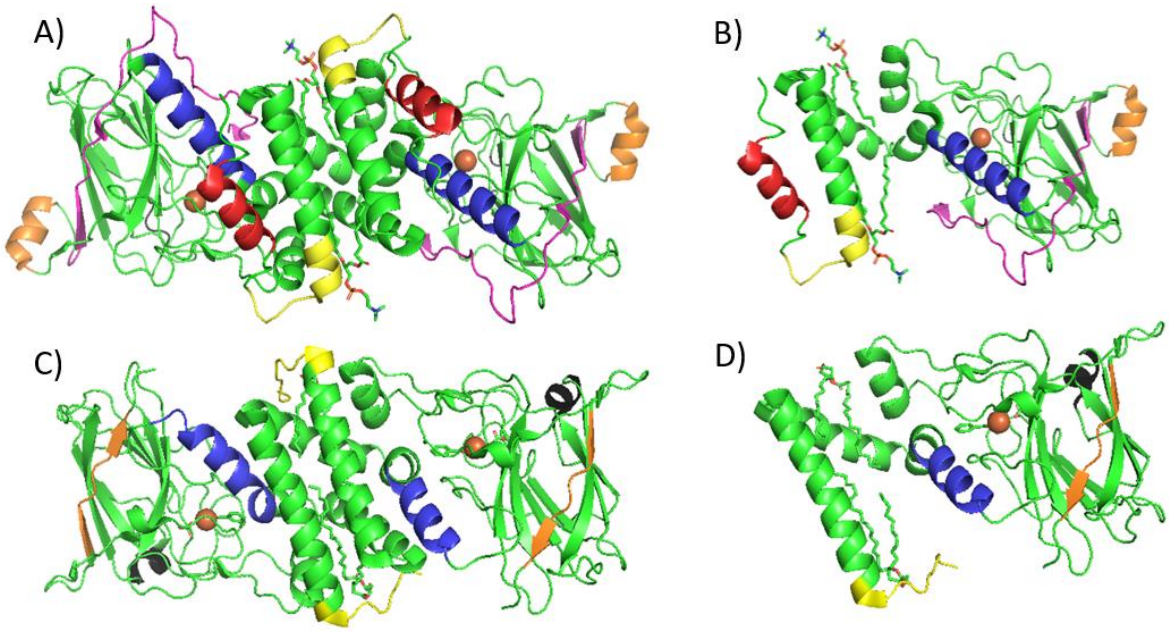

Figure 1.8: Cartoon structures highlighting main structural differences between *Rho*1,2-CTD and a prototypical gram-negative 1,2-CTD; *Aci*1,2-CTD. A) Biological assembly of *Aci*1,2-CTD B) Single subunit of *Aci*1,2-CTD C) Biological assembly of *Rho*1,2-CTD D) Single subunit of *Rho*1,2-CTD. Residues 8-19 (shown in red) form an alpha helix in *Aci*1,2-CTD, whereas they form a random coil region in *Rho*1,2-CTD (not pictured). Residues 23-32 (shown in yellow) points towards the other subunit in *Aci*1,2-CTD, whereas they point to the catalytic domain in *Rho*1,2-CTD. Helix 5 (shown in blue) is shorter in *Rho*1,2-CTD than in *Aci*1,2-CTD. Residues 124-128 (shown in black) forms a new short, small helix in *Rho*1,2-CTD. Helix 7 (shown in orange) is present in *Aci*1,2-CTD but not present in *Rho*1,2-CTD. The C-term of *Aci*1,2-CTD (shown in pink) is 30 residues longer than *Rho*1,2-CTD. Structures were visualized using PyMOL[69].

1.5 Potential applications of Rho1,2-CTD

1.5.1 Remediation of phenolic-contaminated wastewater

Most investigations reporting on phenolic compound removal from wastewater have focused on physical, chemical, and biological processes. Physical processes such as adsorption transfer of pollutants from a liquid to a solid phase requires further treatment of the by-products. Chemical processes such as photocatalysis, Fenton, photo-Fenton, and Ozonation processes are quite efficient, but come with a high price tag. Additionally, chemical processes are often incomplete and create toxic by-products[34,71–73].

In contrast, biological processes have multiple technical advantages such as greater cost-effectiveness, simplicity of implementation and reliability, and high capacity for degredation[72].

Several studies have explored the possibility of using *Rho*1,2-CTD in the process of remediating wastewater contaminated with phenol, benzoate, fluorocatechol, bromocatechol, chlorocatechol, methylcatechol, herbicides (diuron), polychlorinated biphenyls, and chloroethanes[74,75]. Studies show that *Rho*1,2-CTD in particular is highly active compared to other catechol 1,2-dioxygenases from other bacteria of the genus *Rhodococcus* and has activity against a large range of phenolic compounds[52,74]. However, while *Rho*1,2-CTD shows strong promise in this application, it has been reported that the enzyme loses 55 and 90 % activity upon heating at the optimal reaction temperature (40 °C) for 1 and 4 h respectively[74]. The pulp and paper industry, one of the biggest sources of these pollutants generates high-temperature processes and wastewaters, which are normally cooled down to about 40 °C[76]. Thus, to be used in the remediation of wastewater, *Rho*1,2-CTD must be more stable at this temperature.

1.5.2 Cis, cis-muconic acid production

There is an enormous and growing demand for muconic acid and the platform chemicals that can be produced from it. However, current production methods for these chemicals are unsustainable. Take for example adipic acid. Global production of this chemical may account for ~10 % of global nitrous oxide emissions, a potent greenhouse gas[77]. Meanwhile, current production methods for terephthalic acid result in 2.4-3.1kg CO_2-equiv per kg of purified terephthalic acid[78,79], which with a global production of 81.56 million tonnes in 2021[80] is calculated to release 195-253 gigatons of CO_2, equivalent to about 35 % of Canada's overall emissions in 2020[81].

Bearing these statistics in mind, the biological production of muconic acid seems a promising and eco-friendly approach. While muconic acid is not endogenously produced from carbohydrates by any known organism, muconic acid can be found during the catabolism and detoxification of aromatic compounds such as *Rhodococcus*[82]. To date, numerous attempts have been made to employ microorganisms to produce ccMA through the conversion of various biomass feedstock[83–85]. Attempts have also been made to develop microbial cell factories through genetic modifications, metabolic engineering by constructing artificial biosynthetic pathways, and process optimizations of metabolic pathways[86–88]. However, while these attempts have seen some promising results, such bioconversions require multiple enzyme-catalyzed reactions, have inadequate recovery yields and their purification fold appeared to be inadequate due to other metabolic intermediates[89].

To circumvent the current issues associated with attempts at ccMA biosynthesis using living organisms, an alternative bio-based synthesis of ccMA through catechol by single enzymatic conversion could be a superior strategy. Catechol is one of the most common intermediates in aromatic compounds metabolism. Meanwhile, 1,2-CTD from *R. opacus* is highly active and readily converts catechol to ccMA, making it a promising candidate for the proposed biosynthesis of ccMA. This new approach has the potential to eliminate many of the problems faced in previous attempts to produce ccMA via biotechnological approaches. This proposed bioconversion requires only one enzyme-catalyzed reaction and there aren't any metabolic intermediates that would interfere with purification. However, while this proposed biosynthesis can circumvent many issues, it still runs into one notable issue; *Rho*1,2-CTD is insufficiently stable for use in industrial applications.

1.6 Enzyme engineering

1.6.1 History and current state of enzyme engineering

Enzymes (also know as biocatalysts) have greatly contributed to a wide range of industrial applications, such as production, processing and improving the quality attributes of animal feed, beverages, detergent, food, pharmaceutical and textile products[90–94]. However, natural enzymes have limitations such as lower catalytic efficiency and insufficient stability in industrial conditions, as well as issues with selectivity or specificity. In order to address these issues and meet the increasing demand for biocatalysts at the industrial level, enzyme engineering is now routinely practiced in order to enhance enzyme activity and to meet the requirements for complex industrial applications[93,95].

The primary method applied to achieve enzyme engineering is the modification of amino acid sequences by gene alteration. This can be done through modifying existing genes or creating novel ones. The major strategies used to this end are: rational design, directed evolution, semi-rational design, and *de novo* design. These techniques can be employed by themselves or in combination to significantly enhance enzyme properties[96]. The focus of this thesis is on employing engineering strategies that expand on the concept of rational design.

Rational design was the earliest developed approach to protein engineering starting in 1978 and has remained useful over the subsequent decades[97,98]. This strategy is largely completed *in silico*, thus it requires information about the target protein such as amino acid sequence, three-dimensional (3D) structure, and catalytic mechanisms. This strategy can be summarized as follows: 1) algorithms are used that integrate structural, thermodynamic and functional data to predict parameters for enzyme stabilization and to localize potential locations for stabilization; 2) variants undergo site-directed mutagenesis targeted at potential locations for

stabilization; 3) variants are screened to determine which mutants have the target property. (Figure 1.9)

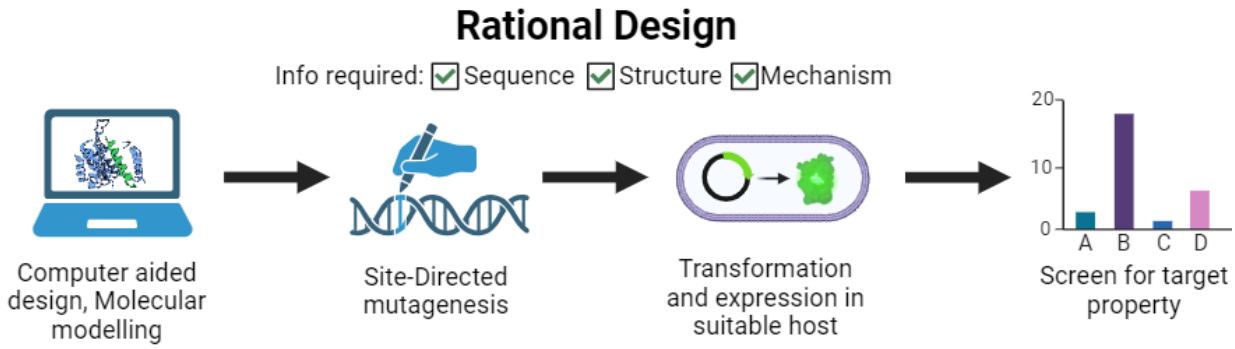

Figure 1.9: Infographic summarizing the rational protein design process

1.6.2 Rational disulfide bond engineering

In proteins, disulfide bonds are formed by the oxidation of two cysteines[104,105]. The rational design of disulfide bonds is a strategy that can be used to enhance thermostability. Previous studies have shown that disulfide bonds linking nonadjacent cysteines play crucial roles in the thermostabilization, recognition, or activation of proteins[106,107]. Predominantly, disulfide bonds maintain the conformations of globular structures, thereby enhancing their stability[108,109]. Due to the improvement in stability conferred by disulfide bonds, many studies have sought to utilize engineered disulfide bonds to increase the stability of proteins. Indeed, the engineering of disulfide bonds into proteins has been successfully applied to improve the stability of various enzymes such as alkaline α-amylase[110] and 1,4-β-endoglucanase[111].

However, not all engineered disulfides have provided an increase in stability; there are a number of reports of destabilizing disulfides[112–114]. As such, it is important to be vigilant when engineering disulfide bonds into protein structures. Along these lines, in recent years several software's such as Yosshi[115], DbD2[116], and SSBondPre[117] have been developed that can predict potential residue pairs for the construction of disulfide bonds. While useful, these computational

tools predict too many suitable locations to test experimentally. Additional criteria are needed to limit the number of locations[118].

1.6.3 Rational genetic code expansion

In the case of engineering improved stability in proteins, experience has shown that there is often an activity-stability trade off, wherein an increase in protein stability is accompanied by a concomitant decrease in protein activity[119–122]. Based on this, it is anticipated that efforts to increase the thermostability of *Rho*1,2-CTD, could also lead to a loss of activity at a given temperature. Genetic Code Expansion (GCE) has recently been shown to serve as a useful method for increasing the catalytic activity of enzymes[123,124]. This emerging technology offers the potential of unlocking new chemistries which have the potential to improve catalytic properties.

Genetic code expansion is a technique used to expand the repertoire of amino acids that can be incorporated into proteins during translation. The natural genetic code consists of 64 different codons, where these 64 codons allow the incorporation of only 20 canonical amino acids and 3 unique stop codons. Genetic code expansion enables the incorporation of non-canonical amino acids (ncAAs) by reprogramming rarely-used codons (usually the amber stop codon) to introduce the non-canonical amino acids. By expanding on the repertoire of amino acids that can be incorporated into a protein, the use of GCE can expand upon the existing presentation of the chemistries currently provided by the 20 naturally occurring canonical amino acids. This allows for novel applications such as photocrosslinking, site specific fluorescent labelling, and optically-controlled protein activity switches[125].

To achieve incorporation of ncAA's into a protein structure, modified tRNAs, codons and tRNA synthetases are introduced into the cell on plasmids and the new ncAA is introduced in the media (see Figure 1.10 for details).

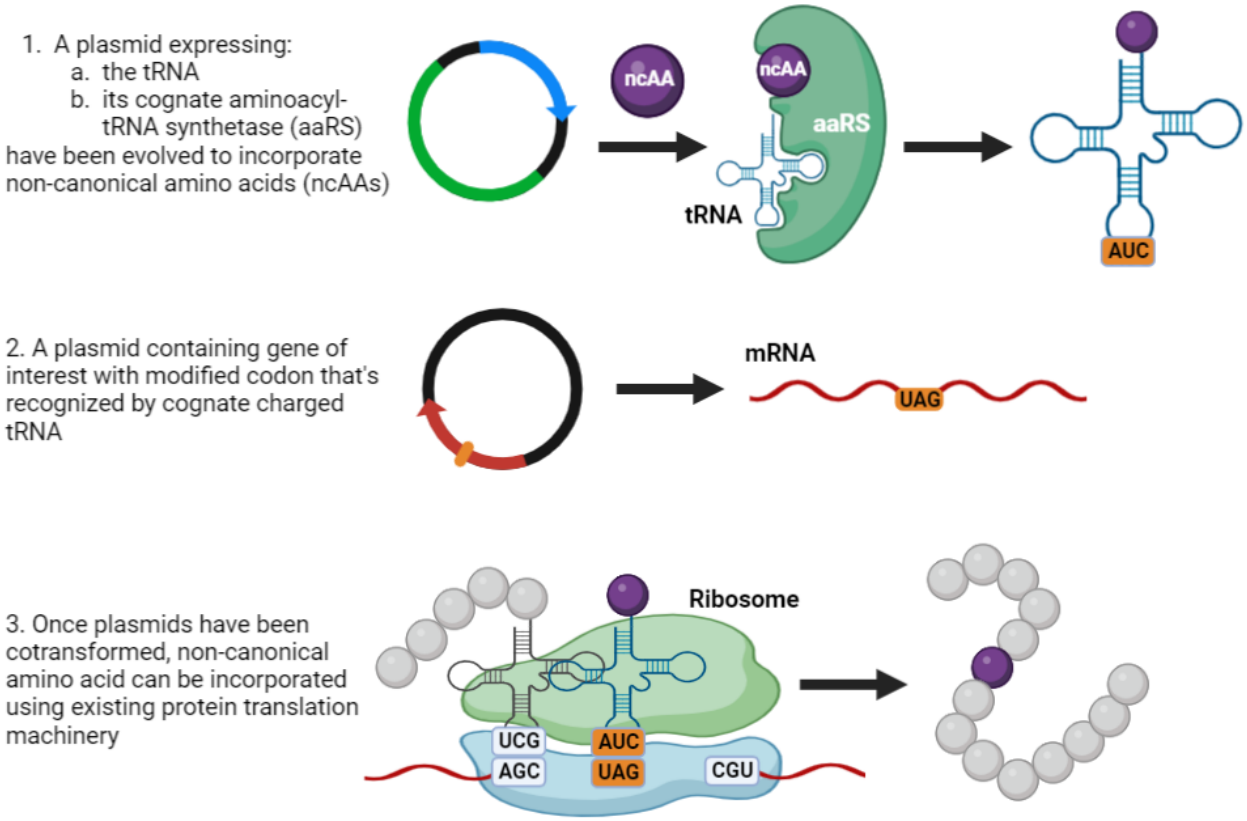

Figure 1.10: Overview of the methodology underlying genetic code expansion technologies. A cell is co-transformed with two plasmids. The first, a plasmid encoding a tRNA and its cognate aminoacyl-tRNA synthetase (aaRS) that have been evolved to incorporate non-canonical amino acids (ncAAs). The second, a plasmid containing the gene of interest with modified codon that's recognized by cognate charged tRNA. With these plasmids co-transformed, the protein containing the non-canonical amino acid can be produced using existing protein translation machinery.

In the context of *Rho*1,2-CTD, an ncAA with potential to improve its' catalytic activity is Nδ-methylhistidine (NmH). The NmH residue is a slight modification of the histidine structure where the Nδ of the histidine is bonded to a methyl group. This modification is proposed to do several things[126]: (1) it fixes the tautomeric form of the imidazole ring; (2) it ensures the neutral

charge state of the proximal ligand throughout the catalytic cycle by preventing the possible formation of anionic ligand via the deprotonation of the Nδ hydrogen; and (3) it prevents the buildup of imidazolate character through N-H bond polarization.

Previous studies that have substituted NmH into enzyme active sites have demonstrated marked improvements in catalytic activity when replacing canonical histidine residues with NmH[123,124]. In one report, researchers showed that the installation of NmH as the proximal heme ligand in myoglobin led to a 3.7-fold increase in catalytic efficiency [123]. In another report, researchers determined that the introduction of a non-canonical Me-His23 nucleophile substantially increased the rate of hydrolytic turnover in a computationally designed enzyme for the Morita-Baylis-Hillman reaction (BH32) [124].

The mechanisms underlying the benefits provided by substitution with NmH aren't fully understood at this time. Mutagenesis studies have shown that imparting imidazolate character onto the axial residue affects Fe-His bond strength, redox potential, and iron electronic properties. Regardless of the mechanism, the histidine's in *Rho*1,2-CTD play a similar role as in BH32 and engineered myoglobin as axial iron-coordinating residues. Based on this, it is hypothesized that substitution of one or both native histidine's in *Rho*1,2-CTD with NmH could improve its' catalytic properties.

Section 1.7 Objectives

The overarching goal of this project is to improve the thermostability of Rho1,2-CTD, while limiting the loss of catalytic activity that comes with stabilization. To improve the stability, disulfide bonds will be rationally designed into the structure of *Rho*1,2-CTD. To counteract any loss in activity that comes with an increase in stability, GCE will be used to improve the catalytic activity of the active site.

Hypothesis 1: The rational design of disulfide bonds into the structure of *Rho*1,2-CTD will improve its thermostability, while decreasing its activity.

Hypothesis 2: The substitution of native histidine's 220 and/or 222 with NmH will improve the catalytic properties, namely the catalytic efficiency (k_{cat}/K_m) of *Rho*1,2-CTD.

Chapter 2

Rational design of disulfide bonds to increase (thermo)stability of *Rho*1,2-CTD

2.1 Abstract

Catechol 1,2-Dioxygenase is a versatile enzyme with several potential applications. However, due to its low thermostability, its industrial potential isn't being met. In this study, the thermostability of a mesophilic Catechol 1,2-Dioxygenase from the species *Rhodococcus opacus* (*Rho*1,2-CTD) was enhanced via the rational design and engineering of disulfide bonds into the structure of the protein. Engineered designs were developed using computational software's to predict potential residue pairs for disulfide bond engineering, followed by application of a set of selection criteria designed to prioritize mutants with the highest likelihood of success. Several mutant enzymes that were designed demonstrated significantly improved protein thermostability, with a minimal loss in catalytic activity. The mutant K96C-D278C displayed a 4.7°C increase in T_{50}, 635% increase in half-life, 5.5°C increase in T_m, and over an 8-fold increase in total turnover number.

2.2 Introduction

Cis, cis-muconic acid (ccMA) is a high-value bio-product. Because of its stereochemical configuration, reactive dicarboxylic groups and the conjugated double bonds, ccMA can undergo an enormous array of reactions. ccMA can be readily converted into commercially important bulk chemicals such as adipic acid, terephthalic acid, and trimellitic acid[38]. In addition, ccMA has potential applications in the manufacture of emerging, performance-enhanced materials such as new functional resins, plastics, agrochemicals, and pharmaceuticals[38,127].

Traditionally, chemical processes for ccMA production rely on non-renewable petroleum-based feedstock and high concentrations of heavy metal catalysts[128,129]. These processes result in problems such as environmental pollution, petroleum depletion and/or high cost of required product separation processes. Strikingly, the current industrial process for producing adipic acid (which can be produced from muconic acid), requires high energy input and produces large amounts of nitrous oxide as a byproduct, accounting for between 5-8 % of global nitrous oxide (N_2O) emissions. N_2O is commonly thought to cause global warming, ozone depletion, acid rain, and smog[130–132].

In light of this, recently, the development of biotechnological processes for ccMA production has been pursued. Recently several attempts were made to develop microbial cell factories to generate ccMA from lignin biomass through genetic modifications, metabolic engineering via construction of artificial biosynthetic pathways, and process optimizations of metabolic pathways[38,87,88]. While these have resulted in promising increases in the yield of ccMA, nevertheless, such bioconversions require multiple enzyme-catalyzed reactions, the recovered yields were inadequate, and their purity appeared to be inadequate due to other metabolic intermediates[89]. Additionally, there was observed limited breakdown of lignin-derived aromatics during fermentation, likely due their levels being too low to upregulate pathways associated with their metabolism in the microbe, an effect that is further compounded by the presence of excess alternate carbon sources in their growth medium.

In order to avoid these limitations, the biosynthesis of ccMA through catechol by single enzymatic conversion using recombinant bacteria could be a superior strategy to achieve maximum production of ccMA. By employing this enzyme to synthesize ccMA from lignin-

derived aromatic compounds, the complications involved with multi-step microbial cell factories can be avoided.

Studies show that *Rho*1,2-CTD in particular is highly active compared to other catechol 1,2-dioxygenases from other bacteria of the genus Rhodococcus and has activity against a large range of phenolic compounds[52,74] (see Figure 2.1A). This enzyme belongs to a family of enzymes known as intradiol dioxygenases. This group uses a non-heme iron (III) to cleave the carbon-carbon bond between phenolic hydroxyl groups by inserting an oxygen molecule to yield muconic acid as the product[52] (see Figure 2.1B).

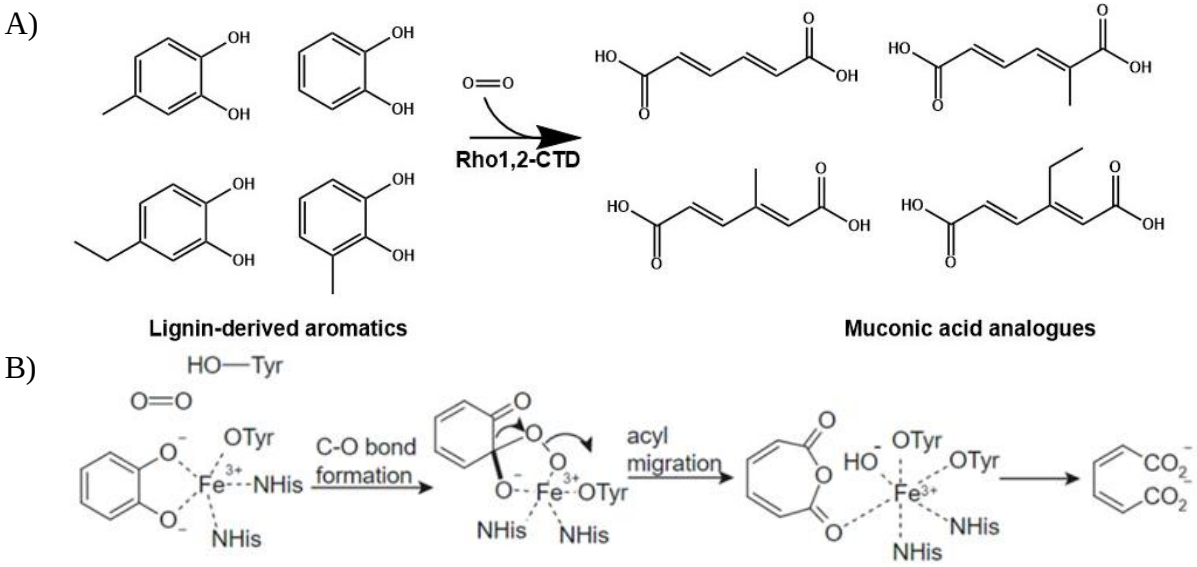

Figure 2.1: Activity of *Rho*1,2-CTD. A) *Rho*1,2-CTD has a broad spectrum of activity against lignin-derived aromatic compounds. B) *Rho*1,2-CTD Proposed mechanism against catechol.

The quaternary structure of Rho1,2-CTD is a Fe(III)α_2 homodimer. The structure contains a helical zipper motif at the interface of the two subunits, and contains two molecules of bound phospholipid[52]. Regarding the active site, the non-heme iron (III) is coordinated by two tyrosine's (Tyr 162 and Tyr196), two histidine's (His 220 and His 222), and a water or hydroxide ion[52] (see Figure 2.2).

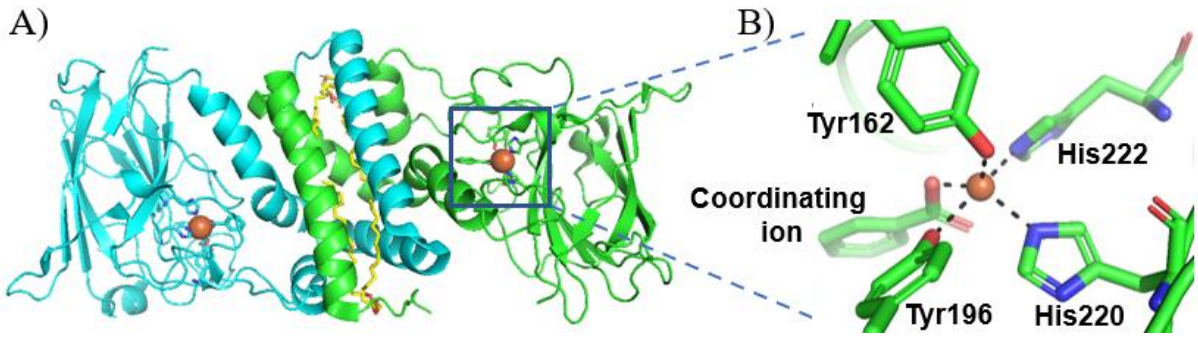

Figure 2.2: Cartoon representations of *Rho*1,2-CTD. A) Cartoon ribbon representation showing biological assembly. The two subunits and the phospholipid chains are coloured differently in order to showcase the quaternary structure of the protein. B) Cartoon representation of the iron coordination sphere in the enzyme active site. The iron coordinating residues are labeled and the intermolecular interactions in the coordination sphere are shown with dashed black lines.

While the application of this enzyme in a single-step conversion would solve many of the problems associated with fermentative production, it has been shown that that the enzyme loses 55 and 90 % activity upon heating at the optimal reaction temperature for 1 and 4h respectively[74]. Therefore, in order to make this process more viable, it is important to discover new stable catechol dioxygenases or improve the thermostability of existing catechol dioxygenases by employing state-of-the-art protein engineering strategies.

Engineering of disulfide bonds in proteins is a fairly common strategy in rational protein design which has been successfully applied to improve the stability of various enzymes such as alkaline α-amylase[110] and 1,4-β-endoglucanase[111]. In recent years, several prediction software packages for disulfide bond design have been released such as DisulfideByDesign2[116], Yosshi[117], and SSBondPre[117]. However, these software packages predict too many suitable locations to test experimentally. Additional criteria are needed to limit the number of locations[118].

This study improves upon the design workflow developed by a previous study[133], by implementing a set of selection criteria with which computationally predicted designs can be prioritized for improved stability, using *Rho*1,2-CTD as a prototype.

2.3 Results

2.3.1 Disulfide bond design strategy

In this study, a total of 56 potential residue pairs were predicted by at least one of three software's including DbD2[116], Yosshi[115,117], SSBondPre[117], based on geometric parameters, bioinformatic analysis, or machine learning respectively. To reduce the number of potential residue pairs to a more manageable number, a set of selection criteria hypothesized to select those with a higher likelihood of success were applied (Figure 2.3). Initially, 12 residue pairs were excluded because the distance between the two resides in the primary sequence was less than 10 amino acids, as this had the potential to produce structural conflicts in native secondary structure[134,135]. Upon further analysis, 14 residue pairs located within 5Å of the iron coordinating residues (His220, His222, Tyr196, Tyr162) in the catalytic centre were omitted to preserve the integrity of the active site. Following this, B-factors from crystal structure 3HHY[52] were analyzed showing that 15 of the proposed bonds were made up of residues that were among the most rigid 25% in the protein, which were then were eliminated. Finally, structural and sequence information was analyzed using DynaMut2, a web server that combines Normal Mode Analysis (NMA) and graph-based signatures to predict the effects of missense mutations on protein stability[136]. From this criterion, 6 residue pairs that were predicted to be destabilizing with a high level of confidence were rejected. This left a total of 9 potential residue pairs, 7 intrasubunit bonds and 2 intersubunit bonds (Figure 2.4). Intrasubunit disulfide bonds are denoted with a dash between the residue pair making up the bond while intersubunit bonds are denoted with a slash. In addition to the 9 single disulfide mutants, two double disulphide mutants (K96C-D278C + A35C/Q211C, and K96C-D278C + H85C-A252C) were created based on the outcomes from the single disulfide mutants.

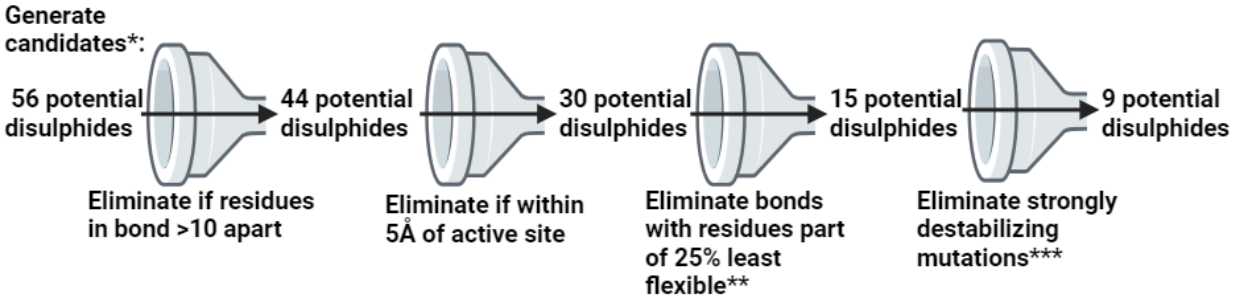

Figure 2.3: Diagram outlining design pipeline used for refining the list of potential residue pairs.

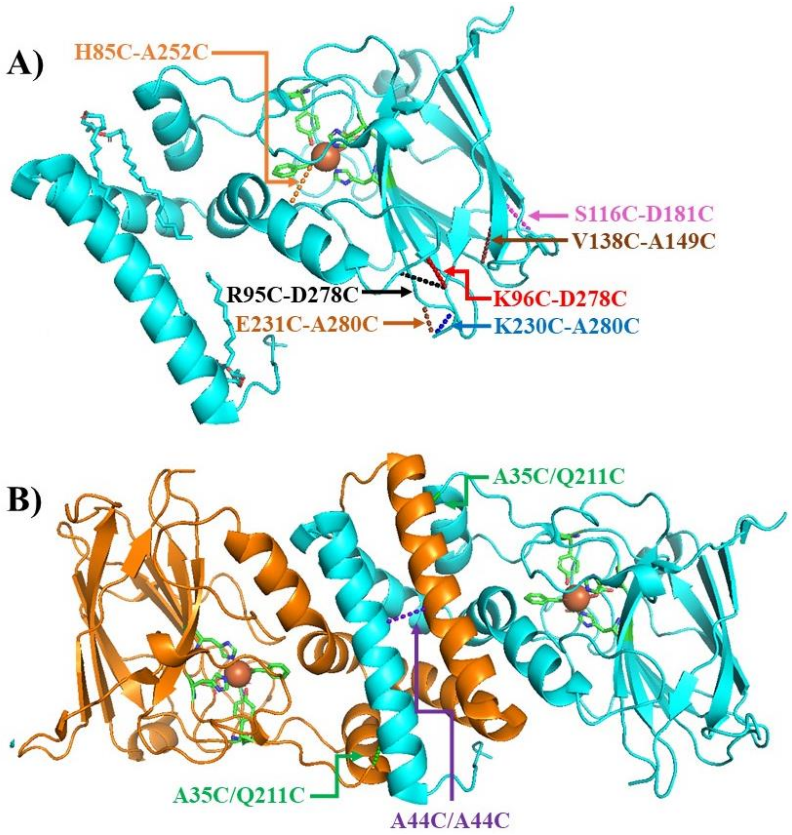

Figure 2.4: PyMOL images of *Rho*1,2-CTD with locations of disulfide bonds selected for experimental testing labelled. Structures are shown in cartoon representation. Subunit 1 is coloured in cyan, while subunit 2 is coloured in orange. Active site residues are coloured in green, and shown in stick representation. Iron (III) cofactors are shown as orange spheres. Locations of disulfides are shown as dashed lines connecting the Cα's of the residues making up the bond. A) Single subunit view with locations of intra-subunit bonds labelled. B) Biological assembly, with locations of inter-subunit bonds labelled. In addition to the single bond mutants, two double disulfide mutants were assessed (DB1: K96C-D278C + A35C/Q211C, DB2: K96C-D278C + H85C-A252C). Structures were visualized using PyMOL[69] and coordinates from PDB ID# 3HGI.

2.3.2 Protein expression and purification

The wild-type (WT) *Rho*1,2-CTD coding sequence was cloned into a pET-31b(+) vector and transformed into *E. coli* BL21DE3 cells. Following induction of protein expression, the accumulated His-tagged enzyme was purified using Ni-NTA immobilized metal affinity chromatography (IMAC). As can be seen in Figure 2.5 there are intense bands in the pellet and supernatant lanes around the expected molecular weight of the enzyme (33 kDa). Comparing the relative intensity of these 2 bands indicates a majority of enzyme was in the supernatant following cell lysis. The elution lanes show large, intense bands corresponding to the M.W. of *Rho*1,2-CTD. These lanes are clean, indicating that a majority of other proteins and cell debris were removed during purification, though the presence of some faint bands suggests the protein is not entirely pure following metal-affinity purification.

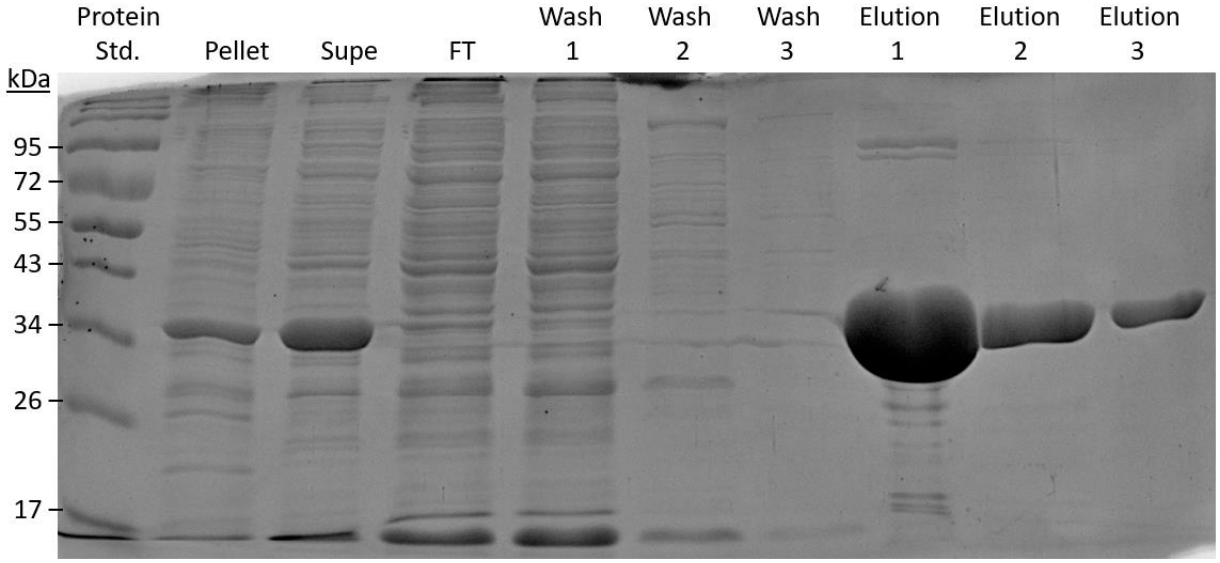

Figure 2.5: SDS-PAGE gel showing metal-affinity purification of wild-type *Rho*1,2-CTD. Gel was stained with Coomassie blue and imaged using a Bio-Rad ChemiDoc™ MP imager. The pellet (Pellet) and supernatant (Supe) were obtained after lysed cells were centrifuged. Flowthrough (FT) was obtained after running the supernatant through the nickel column. Washes were obtained by running wash buffer with an imidazole concentration of 20 mM through the column. Elution's were obtained by running elution buffer with an imidazole concentration of 250 mM through the column.

Following metal affinity purification, the enriched enzyme was further purified by size-exclusion chromatography (SEC) (Figure 2.6). The SEC elution profile shows a large, sharp peak in absorbance occurring between 67 and 73 minutes. The fairly steep and regular nature of the peak suggested that the purification was likely successful with a single homogeneous protein eluted in this time frame. Fractions collected during this time frame were pooled and evaluated by SDS-PAGE compared to the immobilized metal affinity column-enriched sample (Figure 2.6). A single band in each fraction was observed at around the expected molecular weight of 33 kDa, showing that no contaminants were detected by SDS-PAGE. The final yield was approximately 69.6 mg·L^{-1}. Most of the mutants with exception of V138C-A149C had similar yields (Table 2.1)

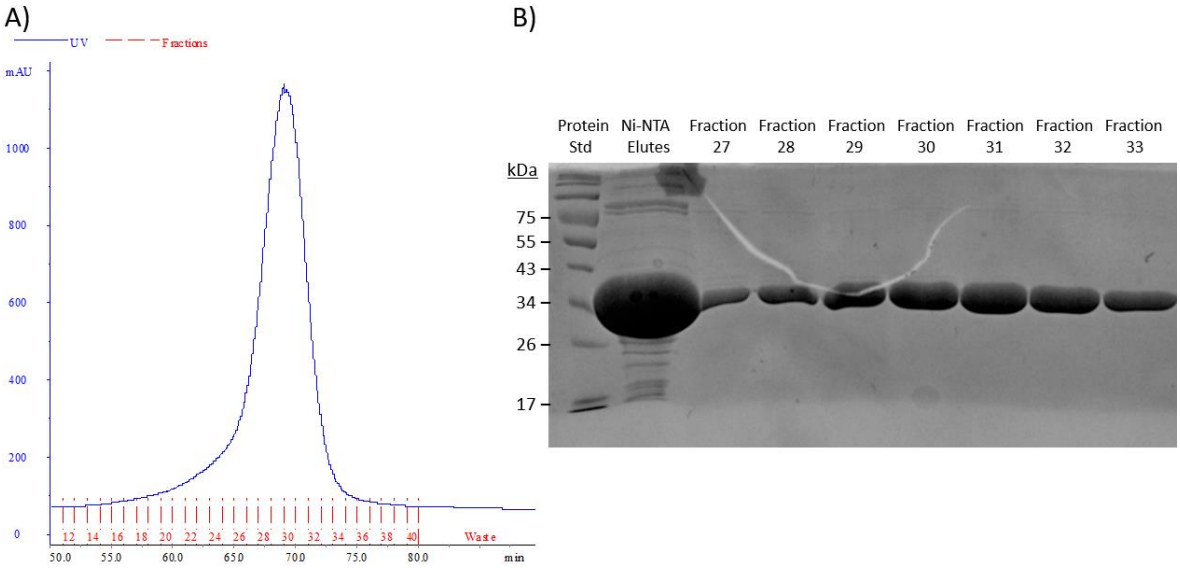

Figure 2.6: Size-exclusion purification of wild-type *Rho*1,2-CTD. Size exclusion was performed using a HiLoad 16/600 Superdex 200 size exclusion column on AKTA Pure 25L FPLC. A) Size exclusion purification curve. Absorbance data was collected at 280 nm. Red numbers indicate fraction numbers. B) SDS-PAGE gel of fractions 27-33. Gel was stained with Coomassie blue and imaged using a Bio-Rad ChemiDoc^{TM} MP imager. Ni-NTA Elutes refers to elutions 1-3 obtained from metal-affinity purification, which were combined and run on the gel as a reference.

Table 2.1: Yields of Rho1,2-CTD variants after purification

Mutant	Yield ($mg \cdot L^{-1}$)
WT	69.6
K96C-D278C	62.6
A35C/Q211C	60.7
A44C/A44C	58.3
S116C-D181C	96.3
H85C-A252C	59.3
E231C-A280C	62.8
K230C-A280C	88.3
V138C-A149C	18.1
R95C-D278C	76.9
DB 1	91.0
DB 2	33.6

2.3.3 Thermostability analysis of *Rho*1,2-CTD and the mutants

Protein stability comes in two flavours: i) thermodynamic stability, which is related to low-levels of unfolded and partially-folded states in equilibrium with the native, functional protein, and ii) kinetic stability, which is related to a high free-energy barrier "separating" the native state from the non-functional forms (unfolded states, irreversibly-denatured protein)[137] (see Figure 2.7)

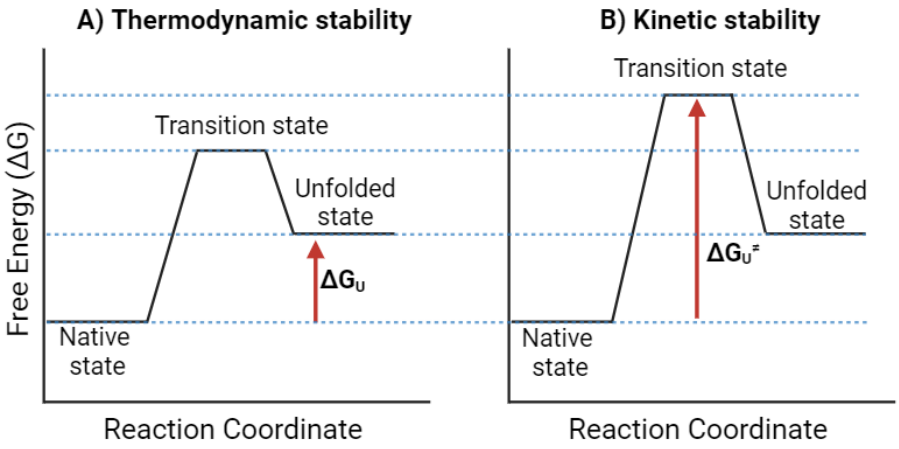

Figure 2.7: Infographic showcasing the two types of protein stability (adapted from[138]). There are two different types of protein stability: 1) Thermodynamic stability and 2) Kinetic stability. The thermodynamic stability of proteins is defined as the difference in free energy (ΔG_U) between the native and unfolded states of a protein. A protein with high thermodynamic stability will have an equilibrium that favours the native state. The kinetic stability of proteins refers to the size of the activation energy barrier ($\Delta G_U^{\ddagger}$) for unfolding. A protein with high kinetic stability will have a high activation energy barrier that confines them to their native state.

Measures of both [137]the kinetic and thermodynamic stability of wild-type *Rho*1,2-CTD and the mutants were determined and compared. Three different measures of stability were determined using different assays to get a robust overall picture of thermostability.

2.3.3.1 Kinetic Stability

To measure kinetic stability, two assays were performed, one to determine the temperature of half-inactivation (T_{50}) and the other to determine the half-life ($t_{1/2}$). T_{50} is used as to quantify the loss of catalytic activity as temperature increases. The T_{50} value represents the temperature at which the enzyme loses 50 % of its activity after 15 minutes of incubation compared to a control. As can be seen from Figure 2.8 and Table 2.2, several mutants show improved kinetic stability compared to the wild-type. Of particular note, the two best-performing single disulfide mutants K96C-D278C and A35C/Q211C had T_{50} values of 45.9 ± 0.6 °C and 44.7 ± 2.4 °C, which were 4.6 °C and 3.17 °C higher than that of wild-type respectively. Double disulfide mutants 1 and 2 demonstrated T_{50} values of 44.4 °C and 45.9 °C respectively, which is a 3.2 °C and a 4.7 °C increase. Overall, 5 mutants demonstrated increases of more than 1 °C, 3 demonstrated an increase or decrease of less than 1 °C, and 3 demonstrated decreases of more than 1 °C.

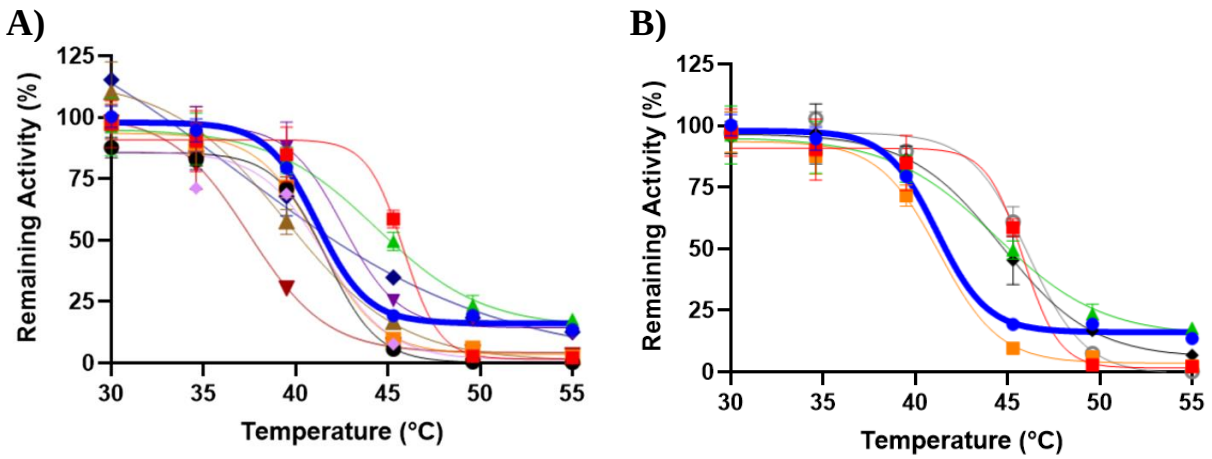

Figure 2.8: Thermal inactivation profiles of wild-type (WT) and mutant *Rho*1,2-CTDs showcasing results from T_{50} testing. Enzymes were incubated at various temperatures for 15mins, and their residual activity was tested for residual activity at 25°C in triplicate*. Activity at each temperature is reported as a percentage relative to activity 25°C. These values are plotted along a Boltzmann sigmoidal curve and the value of T_{50} is obtained by determination of the inflection point on the curve. Error bars show the standard error. A) Data for all single disulfide mutants, B) Data for double disulfide mutants and the single disulfide mutants they are comprised of.

*Note for A35C/Q211C the data presented is from duplicate experiments

● =WT ■ = K96C-D278C, ▲ = A35C/Q211C, ▼= A44C/A44C ◆ = S116C-D181C, ● = R95C-D278C, ■ = H85C-A252C, ▲ = E231C-A280C, ▼ = V138C-A149C, ◆ = K230C-A280C, ◆ = DB1, ○ = DB2

Table 2.2: Summary of T_{50} data for WT and mutant *Rho*1,2-CTDs.
Values represent the mean of three replicates with the standard error

Enzyme	T_{50} (°C)	Δ T_{50} (°C)
Wild Type	41.1 ± 0.4	0
DB2	45.9 ± 0.8	+ 4.8
K96C-D278C	45.9 ± 0.6	+ 4.8
A35C/Q211C	44.7 ± 2.4	+ 3.6
DB1	44.4 ± 0.7	+ 3.3
A44C/A44C	43 ± 1.4	+ 1.4
R95C-D278C	41.9 ± 0.4	+ 0.7
S116C-D181C	41.6 ± 0.1	+ 0.5
H85C-A252C	41.3 ± 0.5	- 0.1
E231C-A280C	39.4 ± 0.6	- 1.7
K230C-A280C*	38	- 2.7
V138C-A149C	37.5 ± 0.4	- 3.7

*Note: data for mutant K230C-A280C should be interpreted with caution as this mutant did not fit the Boltzmann sigmoidal model used to obtain the T_{50}.

To confirm these changes in kinetic stability, the $t_{1/2}$ values were determined. The half-life value represents the amount of time it takes for the enzyme to lose 50 % of its activity at a given temperature, in this case, 40 °C. As can be seen in Figure 2.9 and Table 2.3, the two best-performing single disulfide mutants were K96C-D278C, and H85C-A252C which demonstrated a 635 % and 264 % increase in half-life compared to the wild-type respectively. DB1 and DB2 demonstrated 501 %, and 322 % increases respectively. Overall, five mutants showed an improvement in $t_{1/2}$ compared to the wild-type enzyme, while 5 showed a decrease compared to the wild-type.

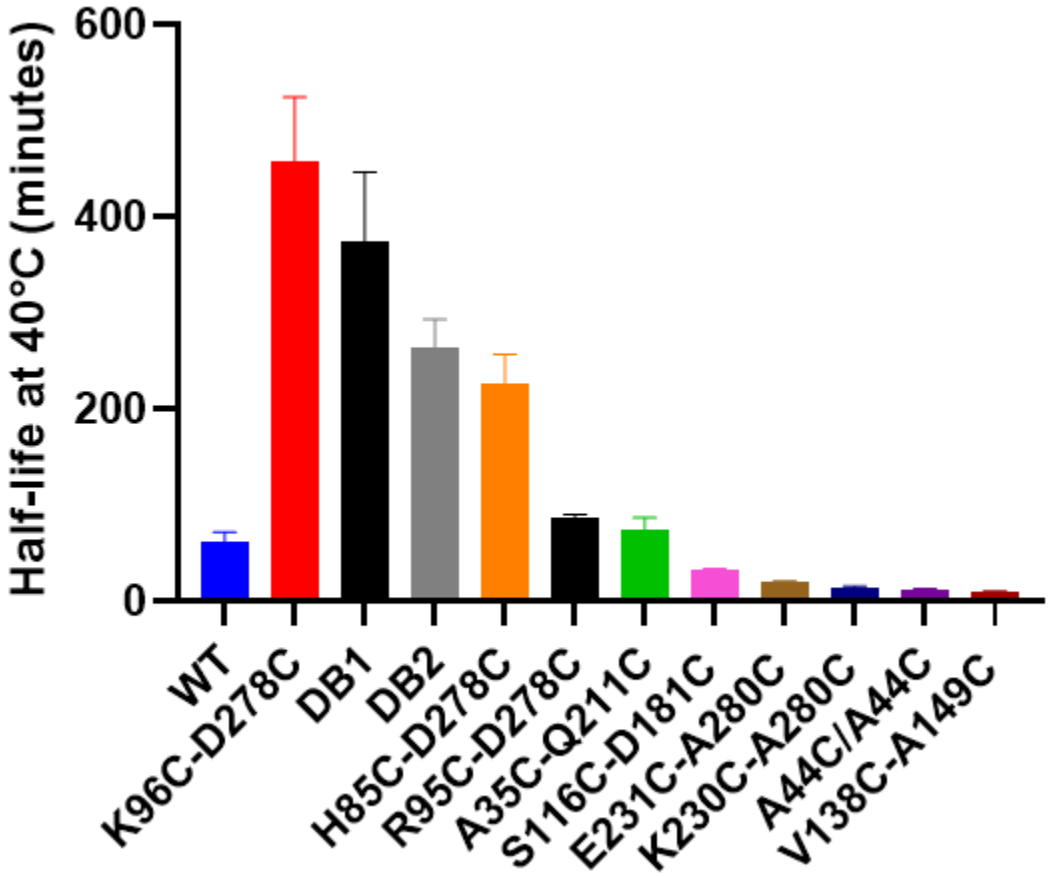

Figure 2.9: Half-life data of WT and variant *Rho*1,2-CTD enzymes. Enzymes were incubated at 40°C and their residual activities were assayed 40°C at different time intervals. The raw data were fitted to first-order plots and analyzed with the first order rate constants (k_d) determined by linear regression of ln(residual activity) versus incubation time. The half life ($t_{1/2}$) represents the time required to the residual activity to be reduced to half and was calculated using the equation $t_{1/2} = \ln 2/k_d$. Values reported represent the mean of three technical replicates with error bars representing standard error.

Table 2.3: Half-life values and deactivation constants for *Rho*1,2-CTD and mutants
Values represent the mean of three replicates, reported with the S.E.

Enzyme	$t_{1/2}$ (minutes)	k_d (min^{-1})
WT	62.3 ± 9.4	0.011 ± 0.002
K96C-D278C	458 ± 66	0.0015 ± 0.0002
DB 1	375 ± 72	0.0019 ± 0.0004
DB 2	263 ± 30	0.0026 ± 0.0003
H85C-D252C	227 ± 30	0.0030 ± 0.0004
R95C-D278C	87.0 ± 3.4	0.0080 ± 0.0003
A35C/Q211C	74 ± 12	0.0093 ± 0.002
S116C-D181C	31.5 ± 2.0	0.022 ± 0.001
E231C-A280C	19.6 ± 1.3	0.035 ± 0.002
K230C-A280C	14.4 ± 1.2	0.048 ± 0.004
A44C/A44C	12.0 ± 0.8	0.058 ± 0.004
V138C-A149C	10.0 ± 0.4	0.069 ± 0.003

k_d represents the first order rate constant for unfolding.

2.3.3.2 Thermodynamic Stability

To determine thermodynamic stability, the melting temperature (T_m) of the enzymes was measured. T_m is a measure of structural stability and the protein unfolding process. The T_m value represents the temperature at which 50 % of the protein is unfolded.

To determine the T_m for the Rho1,2-CTD variants a more recently developed technique called Differential Scanning Fluorimetry was used which is explained in Figure 2.10.

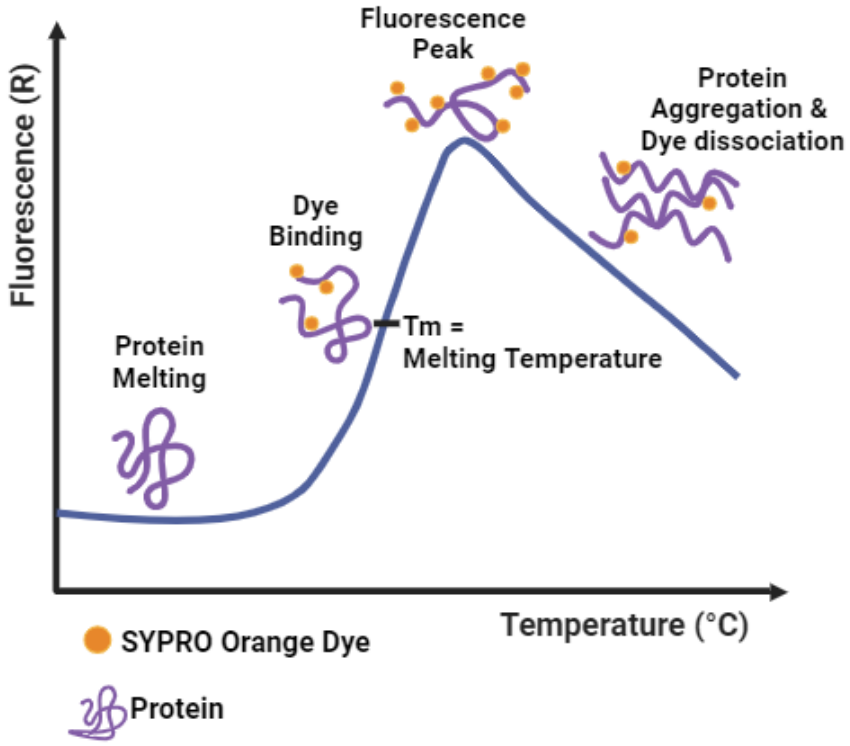

Figure 2.10: Diagram outlining how Differential Scanning Fluorimetry works. Differential Scanning Fluorimetry works by heating the protein of interest in a solution containing SYPRO Orange dye. SYPRO Orange has an affinity for hydrophobic surfaces and fluoresces upon binding. As the protein unfolds, more of the hydrophobic interior is exposed, causing more SYPRO Orange dye to bind and increasing the fluorescence signal. The increasing fluorescence signal is used to determine the progress of unfolding.

Figure 2.11 and Table 2.4 show that 7/9 single disulfide mutants and both double disulfide mutants display improved thermodynamic stability over the wild type. One of the single disulfide mutants, A35C/Q211C, and DB1 displayed increases of more than 10 °C. An additional 3 single disulfide mutants (H85C-A252C, K96C-D278C, E231C-A280C), and DB2 displayed increases of more than 5 °C. Three single disulfide mutants (A44C/A44C, R95C-D278C and V138C-A149C) displayed increases of between 1-3 °C. Two mutants (K230C-A280C, S116C-D181C) displayed decreases in T_m.

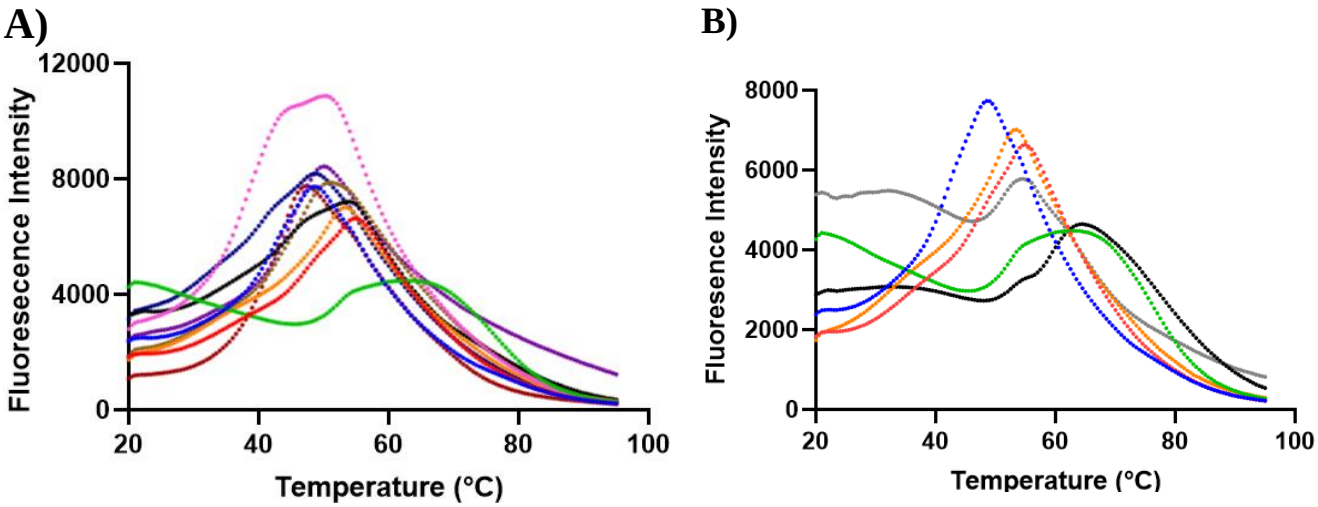

Figure 2.11: Thermal unfolding curves of WT and mutant *Rho*1,2-CTD mutants measured by Differential Scanning Fluorimetry. Enzymes were placed in a PCR plate with SYPRO Orange dye. Progression of unfolding is measured by the increase in fluorescence signal. Thermal shift assays were performed using a Bio-Rad iQ5 qPCR machine. Temperature was increased from 20 to 95 °C in increments of 0.5 °C with fluorescence signal measured at each increment. T_m represents the temperature where 50 % of the protein is denatured and was determined at the temperature where the global minimum for dT/dF of raw fluorescence is. Figure shows raw fluorescence data with plotted values representing the mean of three technical replicates. A) Data for all single disulfide mutants, B) Data for double disulfide mutants and the single disulfide mutants they are comprised of. ● =WT ● = K96C-D278C, ● = A35C/Q211C, ●= A44C/A44C ● = S116C-D181C, ● = R95C-D278C, ● = H85C-A252C, ● = E231C-A280C, ● = V138C-A149C, ● = K230C-A280C, ● = DB1, ● = DB2

Table 2.4: Melting temperatures of WT and mutant *Rho*1,2-CTDs
Values represent the mean of three replicates, reported with S.E. Melting temperature was determined using a Thermal Shift Assay[139]. The melting temperature was taken as the minimum of the first derivative of fluorescence. In the case where there were 2 distinct minima, two melting points are reported.

Sample	T_m 1 (°C)	T_m 2 (°C)
WTP	41.8 ± 0.2	46.2 ± 0.2
DB Mutant	52.7 ± 0.2	59.5
A35C/Q211C	52.3 ± 0.2	
DB #2	51.2 ± 0.2	
H85C-A252C	50.7 ± 0.2	
K96C-D278C	47.3 ± 0.2	52.67 ± 0.2
E231C-A280C	47.0	
A44C/A44C	44.5	
R95C-D278C	44.0	
V138C-A149C	43.7 ± 0.2	
K230C-A280C	40.5	46.5
S116C-D181C	38.0	

2.3.4 Kinetic Testing

To test the effect the disulfide bonds had on the enzyme activity, Michaelis-Menten analyses were performed on each of the enzyme variants. As expected, most of the mutants had decreased catalytic efficiency compared to wild-type *Rho*1,2-CTD. Of the mutants that performed best in terms of thermostability; K96C-D278C has a 28% decrease in catalytic efficiency, H85C-A252C had a 11% decrease. Surprisingly, A35C/Q211C had an 11% increase and decrease. Surprisingly, DB2 had a 40% increase in catalytic efficiency. Though, these values should be treated with caution as there is some overlap of the standard errors. (Figure 2.12 and Table 2.5).

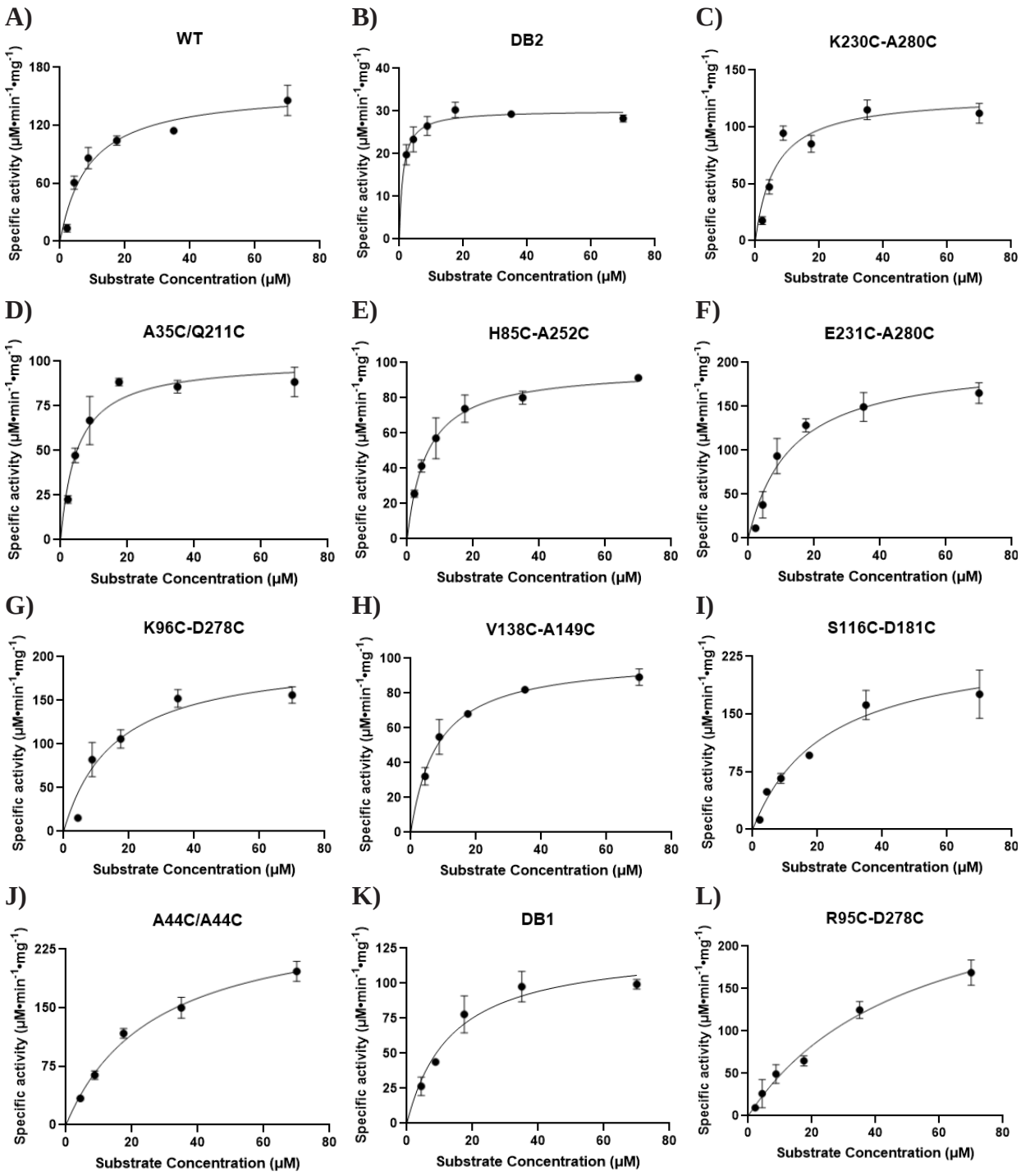

Figure 2.12: Michaelis-Menten kinetics plots of wild-type (WT) and variant *Rho*1,2-CTD enzymes. Activity of *Rho*1,2-CTD was measured spectrophotometrically to measure the initial rate of product formation at differing substrate concentrations. These data were fit to the Michaelis-Menten equation using GraphPad Prism (ver. 9.1.12) to extract the kinetic parameters. Specific activity was calculated by dividing the initial rate of product formation by the enzyme concentration. Data plots specific activity vs substrate concentration and represents the mean of three technical replicates with error bars showing the standard error. A-K) Plots of the different enzyme variants tested: A) WT, B) DB2 (K96C-D278C +H85C-A252C), C) K230C-A280C, D) A35C/Q211C, E) H85C-A252C, F) E231C-A280C, G) K96C-D278C, H) V138C-A149C, I) S116C-D181C, J) A44C/A44C, K) DB1 (K96C-D278C + A35C/Q211C), L) R95C-D278C.

Table 2.5: Summary of Michaelis-Menten kinetic data for WT and mutant *Rho*1,2-CTDs. The values for k_{cat} and K_m were obtained by fitting the data to the Michaelis-Menten equation using GraphPad Prism (version 9.1.12).

Mutant	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1}• μM^{-1})	R^2 value
Wild Type	9 ± 2	157 ± 14	18 ± 5	0.8102
DB #2	1.2 ± 0.5	30.2 ± 0.3	26 ± 12	0.5776
K230C-A280C	6 ± 2	128 ± 13	21 ± 7	0.8029
A35C/Q211C	5 ± 1	100 ± 5	21 ± 5	0.8129
H85C-A252C	6.0 ± 0.5	97 ± 4	16 ± 2	0.8567
E231C-A280C	12 ± 5	204 ± 19	16 ± 6	0.8584
K96C-D278C	16 ± 3	202 ± 5	13 ± 3	0.8124
V138C-A149C	8 ± 2	101 ± 1	12 ± 3	0.8647
S116C-D181C	23 ± 6	244 ± 43	10 ± 3	0.8974
A44C/A44C	27 ± 5	272 ± 25	10 ± 2	0.944
DB #1	14 ± 3	126 ± 7	9 ± 2	0.8207
R95C-D278C	49 ± 2	296 ± 20	6.0 ± 0.5	0.9197

2.3.5 Total Turnover Number

The continuous replacement of enzymes and other proteins appropriates up to half the maintenance energy budget in microorganisms and plants; thus, high enzyme replacement rates cut the productivity of biosystems[140]. Therefore, the expected product yield of a biocatalyst during its useful lifetime is an important consideration when designing biocatalytic processes[141]. One important indicator of lifetime biocatalyst productivity is the dimensionless total turnover number (TTN).

The TTN of an enzyme represents the number of molecules of substrate converted to product by one molecule of enzyme before it's permanently deactivated. It can also be thought of as the number of catalytic cycles the enzyme performs before it denatures.

In this study, an approximation was used to obtain the static TTN by dividing the value of k_{cat} by the enzyme deactivation constant k_d. Of the single disulfide mutants, 4 had improved TTNs

compared to wild type while 5 displayed decreases. The best performing mutant K96C-D278C demonstrated an over 800% increase in TTN. The first double disulfide mutant, DB1 displayed a lesser increase over wild-type than mutant K96C-D278C with a 381% increase. The TTN of DB2 was virtually unchanged from that of wild type (see Table 2.6).

Table 2.6: Summary of static TTN values for wild-type and mutant *Rho*1,2-CTDs.
Static TTN values were obtained by dividing the turnover number (k_{cat}) by the enzyme deactivation constant (k_d). The values of k_{cat} and k_d were obtained from Michaelis-Menten testing and half-life testing respectively.

Mutant	k_{cat} (min^{-1})	k_d (min^{-1})	TTN
Wild Type	157 ± 14	0.011 ± 0.002	14,164 ± 2491
K96C-D278C	202 ± 5	0.0015 ± 0.0002	133,664 ± 19,518
R95C-D278C	296 ± 20	0.0080 ± 0.0003	96,920 ± 11,905
DB #1	126 ± 5	0.0019 ± 0.0004	68,216 ± 13,548
H85C-A252C	97 ± 4	0.0030 ± 0.0004	36,766 ± 5773
S116C-D181C	244 ± 43	0.022 ± 0.001	26,106 ± 6014
A35C/Q211C	100 ± 5	0.0093 ± 0.002	12,597 ± 2527
DB #2	30.2 ± 0.3	0.0026 ± 0.0003	11,440 ± 1330
E231C-A280C	204 ± 19	0.035 ± 0.002	9726 ± 1315
A44C/A44C	272 ± 25	0.058 ± 0.004	5655 ± 703
K230C-A280C	128 ± 13	0.048 ±0.004	3621 ± 531
V138C-A149C	101 ± 1	0.069 ± 0.003	1747 ± 79

2.3.6 Verification of disulfide bond formation

In order to verify the proper formation of disulfide bonds within the protein, the DTNB method was used (5,5'-dithiobis-(2-nitrobenzoic acid). This method works by adding to DTNB to both a reduced and an untreated sample of protein. DTNB reacts with free sulfhydryl groups to yield a mixed disulfide and TNB (2-nitro-5-thiobenzoic acid). TNB is a coloured species and can be easily detected via absorbance reading at 412nm. Therefore, disulfide bond verification can be done be down by comparing the absorbance signals of the reduced and untreated samples of protein. Absorbance values were converted to free sulfhydryl concentration using a standard curve (Figure 2.13)

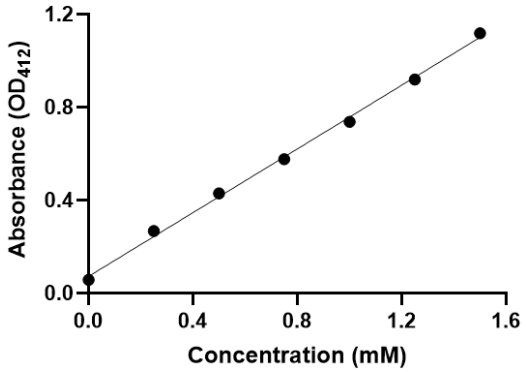

Figure 2.13: Standard curve of thiol concentration vs OD_{412} for the DTNB assay. The standard curve was generated by performing adding DTNB to a set of thiol standards and measuring the OD_{412}. A line of best fit was added using the simple linear regression function in GraphPad Prism (version 9.1.12) and gave the equation $y = 0.6856x + 0.007667$. Values represent the mean of three replicates. Error bars for 95% CI were too small to be visualized.

Two free sulfhydryl groups were detected in each disulfide mutant tested under reducing conditions. Four free sulfhydryl groups were detected in each double disulphide mutant under reducing conditions. For DB2, K96C-D278C and H85C-A252C, there was less than 10% of the signal under non-reducing conditions compared to reducing conditions, suggesting more than 90% of enzymes had properly formed disulfide bonds. For DB1 there was 11% of the signal under non-reducing conditions and for A35C/Q211C there was 19%, suggesting 89%, and 81% of enzymes had properly formed disulfide bonds respectively (see Table 2.7 and 2.8).

Table 2.7: Raw OD_{412} values for wild-type and mutant *Rho*1,2-CTDs in non-reduced and reduced conditions using the DTNB method.

Mutant	Reduced Samples OD_{412}			Non-Reduced Samples OD_{412}		
	1	2	3	1	2	3
DB2	0.453	0.446	0.454	0.086	0.083	0.086
A35C/Q211C	0.392	0.401	0.402	0.126	0.14	0.136
DB1	0.822	0.785	0.832	0.15	0.15	0.152
H85C-A252C	0.413	0.428	0.425	0.106	0.108	0.11
K96C-D278C	0.423	0.422	0.403	0.1	0.102	0.105

Table 2.8: Free sulfhydryl titration of wild-type and mutant *Rho*1,2-CTDs in non-reduced and reduced conditions using the DTNB method.
NR=Non-Reduced, R=Reduced

Enzyme	Reduction state (NR/R)	Expected Number of free CYS	Enzyme Concentration (mM)	Expected CYS concentration (mM)	Observed CYS concentration (mM)
K96C-D278C	NR	0	0.270	0	0.045 ± 0.002
K96C-D278C	R	2	0.270	0.539	0.50 ± 0.01
H85C-A252C	NR	0	0.247	0	0.054 ± 0.002
H85C-A252C	R	2	0.247	0.495	0.511 ± 0.007
A35C/Q211C	NR	0	0.270	0	0.092 ± 0.006
A35C/Q211C	R	2	0.270	0.541	0.476 ± 0.005
DB1	NR	0	0.366	0	0.116 ± 0.001
DB1	R	4	0.366	1.46	1.08 ± 0.02
DB2	NR	0	0.127	0	0.020 ± 0.001
DB2	R	4	0.127	0.507	0.553 ± 0.004

DB1 = K96C-D278C + A35C/Q211C
DB2 = K96C-D278C + H85C-A252C

2.4 Discussion

The introduction of disulfide bonds into protein structure is a common strategy used to improve the thermostability of enzymes[142–144]. However, if not done carefully, poorly-designed disulfide bonds can result in reduced catalytic activity, inferior thermostability and misfolding[112–114]. This study demonstrated the design of a pipeline to streamline this process which produces mutants with dramatically improved thermostability.

2.4.1 Disulfide bond design strategy

In this study, a balance had to be struck between trying to maximize success while keeping the scope of work required to a feasible level. To this end, a design strategy was developed to leverage newly available computational prediction tools.

The first step involved identifying potential residue pairs. Toward this, rather than visual identification of potential locations for disulfide bonds, 3 different computer programs were used

to predict potential locations: DbD2[116], Yosshi[115], and SSBondPre[117]. Each of these programs utilized a different method to identify disulfide bonds; DBD2 used geometric constraints, Yosshi used bioinformatic analysis, and SSBondPre used a machine learning model. Since these 3 programs each use a different algorithm to identify bonds, using all 3 maximizes that chance that at least one of them will identify a stabilizing disulfide.

The combined output of these three programs was 56 unique disulfide bonds. To reduce the number of potential residue pairs, a set of selection criteria was developed to prioritize the pairs with the highest likelihood of success. The first criterion involved identifying residue pairs that were too close together along the primary sequence as this might produce potential structural constraints that can modify the secondary structure, and thus lead to tertiary conflicts that can undermine functionality[134,135]. To apply this criterion, any residue pairs that were less than 10 amino acids apart in the primary sequence were eliminated. This reduced the number of candidates from 56 to 44 potential residue pairs. The second criterion involved preserving the integrity of the active site. Therefore, any residue pairs for which the Cα of either residue was located within 5Å of essential coordinating residues (Tyr162, Tyr196, His220, His222), the Fe(III), or the substrate in the crystal structure PDB ID: 3HHY were eliminated. This criterion further reduced the number of candidates from 44 to 30 potential residue pairs. The third criterion involved identifying the most flexible regions of the protein. Previous studies have shown that disulfide bonds were often introduced in flexible, less structurally defined regions while introducing disulfide bonds in parts of protein that were structurally well-defined did not contribute much to protein thermostability[133]. Any residue pairs in which either residue was located within the least flexible 25% of the enzyme (as determined by B-factors) were eliminated. This further reduced the number of candidates from 30 to 15 potential residue pairs.

The final criterion was using a program DynaMut2 to predict any mutations that would have a strongly destabilizing effect on the protein structure. This final criterion shrank the number of candidates from 15 to 9. The result of these selection criteria was that 84 % of the original designs predicted by the computer software's were eliminated.

2.4.2 Thermostability of wild type Rho1,2-CTD and the mutants

Protein stability comes in two flavours: i) thermodynamic stability, and ii) kinetic stability as defined above.[131] In order to gain a more complete overall picture of stability, both thermodynamic (T_{50}, $t_{1/2}$) and kinetic stability (T_m) were tested. In terms of kinetic stability, as expected there was a correlation between the results from T_{50} testing and $t_{1/2}$ testing, with a Pearson r value of 0.75 and a one-tailed p-value of 0.0031 suggesting the results obtained from these assays were reliable.

In both of these assays, the mutant K96C-D278C was shown to be the most stable, having a 4.7 °C increase in T_{50} and a 635 % increase in half-life at 40 °C compared to wild-type enzyme. Interestingly however, the second-best performing mutant in each of these assays differed. In terms of T_{50}, the second-best performing mutant was A35C/Q211C with an increase of 3.4 °C; while in terms of $t_{1/2}$, the second-best performing mutant was H85C-A252C, with an increase of 264 %. Interestingly, the mutant H85C-A252C performed quite modestly in the T_{50} assay, with the standard error of the WT T_{50} and H85C-A252C T_{50} having overlapping standard errors while the mutant A35C/Q211C performed quite modestly in the $t_{1/2}$ assay, having a 19 % increase in $t_{1/2}$ (see structural discussion below).

Moving onto thermodynamic stability, this type of stability was tested through melting-temperature testing. There was a moderate correlation between T_m and T_{50} (Pearson r = 0.60, One-tailed p-value = 0.0191) and between T_m and $t_{1/2}$ (Pearson r = 0.60, One-tailed p-value =

0.198). The best-performing mutants in tests of kinetic stability were also the best-performing in terms of thermodynamic stability; these being K96C-D278C with an increase in T_m of 5.5 °C, H85C-A252C with an increase of 8.8 °C, and A35C/Q211C with an increase of 10.5 °C.

Overall, based on the results from tests of both kinetic and thermodynamic stability, there were three mutants that clearly stood ahead of the rest of the pack: K96C-D278C, A35C/Q211C, and H85C-A252C. In an attempt to further improve stability, and to test whether the stabilizing effects of the disulfide bonds were additive, two different double disulfide mutants were subsequently designed by combining the best single disulfide mutant (K96C-D278C) with each one of the second-best performing mutants (A35C/Q211C or H85C-A252C).

For the double disulfide mutants, the effects were not additive as predicted. In terms of kinetic stability, each double disulfide mutant seemed to have stability in between that of the two single disulfide mutants it was made from. DB1 performed better than A35C/Q211C, but worse than K96C-D278C in terms of both T_{50} and $t_{1/2}$. DB2 performed better than H85C-A252C but worse than K96C-D278C in terms of $t_{1/2}$. In terms of T_{50}, DB2 performed better than H85C-A252C and was not substantially different than K96C-D278C. In terms of thermodynamic stability, both double disulfide mutants performed better than the single disulfide mutants they were made from based on T_m (see structural discussion below).

2.4.3 Kinetic activities of wild type Rho1,2-CTD and the mutants

In terms of protein engineering, there is a well-known limitation referred to as the activity-stability trade off. This idea implies that an increase in activity is accompanied by a concomitant decrease in protein stability and vice versa. Therefore, it was important to test the kinetic parameters of the disulfide mutants to determine the effect stabilization had on enzyme activity.

In terms of the single disulfide mutants, as expected there was an effect on the kinetic parameters of the enzyme. K96C-D278C and H85C-A252C showed decreases in catalytic efficiency of 26%, and 8% respectively compared to WT, while A35C/Q211C showed a negligible change (see Table 2.5 in results) though it should be noted there was some overlap between the standard error for catalytic efficiency of the WT and all three of these mutants. While it would be ideal to suffer no loss in catalytic activity, as of writing, it appears from the literature that no one has been able to totally overcome this trade-off. Nevertheless, the loss of catalytic activity in these mutants are small, especially in comparison to the increase in stability.

2.4.4 Total turnover number of wild type Rho1,2-CTD and the mutants

Finally, to get a sense of the balance between activity and stability and to identify mutants with a good mix of stability and activity, the static TTN was calculated. To calculate TTN an approximation was used; To calculate the TTN, the turnover number, k_{cat}, is divided by the deactivation constant, k_d (obtained from $t_{1/2}$ testing). Hence the value of TTN is dependent on a combination of activity and stability, so a higher TTN represents a higher combined activity and stability. In this measure, once again the mutant K96C-D278C stands above the rest, with over an 8-fold increase in TTN compared to the wild type (see Table 2.6 in results).

2.4.5 Structural description of engineered mutants

[145,146147]A recurring theme noted in disulfide engineering experiments is the observation that engineered disulfide bonds linking regions of relatively high mobility are those most likely to confer stability to the protein[148]. This idea appears to originate from a meta-analysis of previously engineered disulfide bonds where the effect on protein stability had been characterized and published. In this meta-analysis, the authors investigated the relationship between several structural features of each disulfide bond and the change in stability. In

summary, they reported: (1) stabilizing mutations were most often found in regions of medium to high mobility; (2) stabilizing mutations were more likely to be near the protein surface; (3) stabilizing mutations were associated with longer loop lengths (>25 residues); and (4) the introduced disulfide bond should not cause steric overlap.

The best-performing disulfides reported in this these can be compared to these criteria to see how well they fit. The overall best-performing single disulfide mutant was K96C-D278C. This bond is located near the C-terminus. Both cysteine residues forming this bond are in highly flexible regions of the protein Cα chain, as quantified by their B-factors in the structure file. This bond is also located close to the protein surface, and has a large loop length enclosed by the disulfide. It also does not cause steric overlap. Overall, it fits well with the criteria listed by Dani et al.[149].

Following K96C-D278C, the next two best-performing mutants were A35C/Q211C and H85C-A252C. Starting with A35C/Q211C, this bond connects the two homodimers of *Rho*1,2-CTD. In terms of flexibility, the residue Q211 is in a highly flexible region of the protein while the residue A35 is in a region of moderate flexibility. This bond is also located close to the protein surface, has a large loop length enclosed by the disulfide, and does not cause steric overlap. Overall, this bond also fits well with the criteria.

The final top performer, H85C-A252C connects the catalytic and dimerization domains of *Rho*1,2-CTD. Interestingly, in contrast to K96C-D278C and A35C/Q211C, the residue pair H85C-A252C, does not fit well with all of the 4 criteria reported by. While it does have a large loop length enclosed by the disulfide, and does not cause steric overlap, it is located within the core of the protein and is located in a region with low relative flexibility. In fact, this bond did not actually pass the third step in the design pipeline since both residues in the bond are among

the 25 % least flexible residues of *Rho*1,2-CTD based on their crystallographic B-factors (see Figure 2.14A). An exception was made as this bond is located at a domain interface and was predicted to be strongly stabilizing by the program DynaMut2[136]. Despite this location being in a rigid part of the protein, a disulfide bond formed between positions 85-252 was shown to convey a considerable increase in stability. Additionally, this mutant performed considerably better than the other residue pairs tested, which were located in regions of much higher flexibility. Take for example the bonds K230C-A280C, E231C-A280C, and S116C-D181C. These three bonds were among the worst performing, yet all of these bonds were located in highly flexible regions of the protein (see Figure 2.14B). While it is important to keep in mind that this is a limited sample, it suggests that perhaps it's unnecessary to specifically target highly mobile residues in disulfide bond engineering, and other factors such as ease unfolding (Chapter 4) may play a larger role.

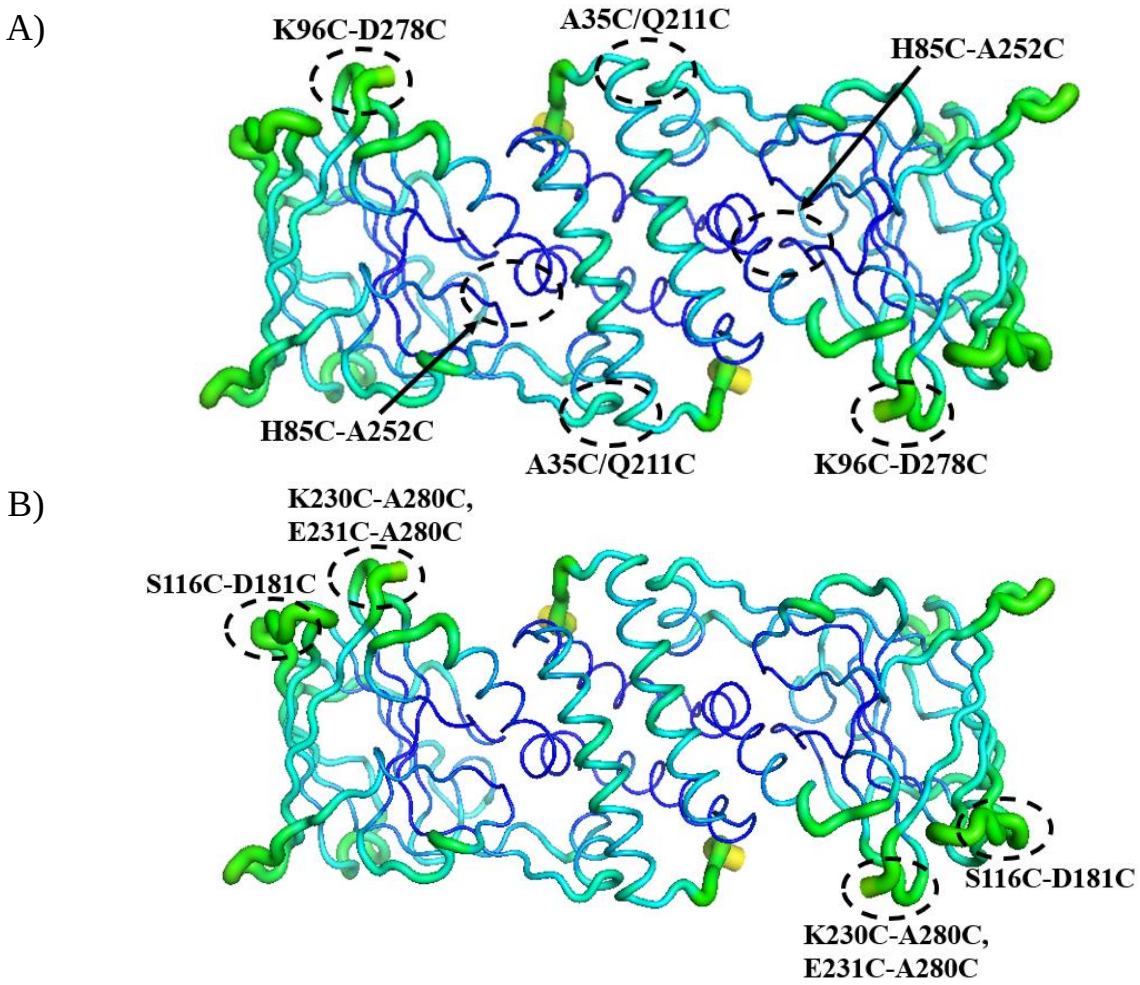

Figure 2.14: B-factor putty representation of *Rho*1,2-CTD biological assembly (PDB ID: 3HGI) showcasing locations of best and worst performing disulfide bonds. A) Locations of top-performing disulfide bonds, B) locations of lesser-performing disulfide bonds.

2.4.6 Conclusion

In conclusion, multiple thermostable *Rho*1,2-CTD mutants were engineered via the introduction of disulfide bonds. These disulfide mutants were designed using disulfide bond prediction software packages combined with a novel selective design pipeline. Several mutants displayed significantly improved thermostability properties. The design strategy used in this study combining computational prediction with relevant selection criteria demonstrated great success in selecting candidates with improved thermostability and has strong potential to be applied to engineer thermostability of other enzymes.

2.5 Materials and Methods

2.5.1 Identifying disulfide bond candidates and refining list of candidates.

Sequence and structural data (PDB ID # 3HGI[52]) for *Rho*1,2-CTD were fed into 3 computer programs with different predictive algorithms set to default parameters to generate lists of potential disulfide bond locations (Yosshi[115], SSBondPre[117], DisulfideByDesign2[116]). A set of selection criteria was subsequently applied to narrow down the list of potentially stabilizing disulfide bonds. Disulfide bonds were eliminated if: residues forming the bond were less than 10 residues apart in the primary structure, if bond was within 5Å of active site (Tyr162, Tyr196, His220, His222, Fe(III), or catechol), if residues forming bond were not part of 25 % most flexible residues (as determined by B-factors, note: one exception was made for H85C-A252C), if mutations were predicted to be strongly destabilizing (as determined by program DynaMut2[136]).

2.5.2 Standard recombinant expression of *Rho*1,2-CTD

Starter cultures were created by inoculating a small volume of Lysogeny broth (LB) (10 g Tryptone, 10 g NaCl, 5 g Yeast Extract per 1 L) containing appropriate antibiotic (100 µg/mL Ampicillin) with a frozen cell stock from the desired cell line; *E. coli* BL21DE3 cells containing pET31b+ plasmid with 1,2-CTD coding sequence (Rho1,2-CTD NCBI accession code: WLF46330, synthesized by GenScript, cloned into NdeI and XhoI restriction sites, containing C-terminal His Tag). Starter cultures were grown overnight at 37 ℃ with shaking at 275 rpm. Enzyme expression was achieved as follows; An Erlenmeyer flask containing 250 mL LB medium containing 100µg/mL ampicillin was inoculated using a 1:100 ratio of starter culture. The cells were grown at 37℃ at 275 rpm until OD reached 0.3-0.4, then the temperature was turned down to 18℃ and cells were grown until OD_{600} reached 0.6-0.7. Expression was induced

with 0.7mM of IPTG and cells were grown at 18℃ for 18-24 hours at 150rpm. Aliquots of culture were harvested by centrifugation at 3250g for 20-30mins at 4℃. Pellets were stored at -20℃.

2.5.3 Recombinant *Rho*1,2-CTD purification

Cell pellets were thawed for 15 minutes on ice and resuspend in Lysis buffer (10mM Tris-HCl, 150mM NaCl, 10mM Imidazole, 2mM $MgCl_2$, 5mM 1,4-Dithiothreitol, 1mM phenylmethylsulfonyl fluoride (protease inhibitor), 3U/mL Benzonase, 0.5mg/mL Lysozyme). Cells were lysed using a French Press. Lysate was centrifuged at 17,000g for 30mins at 4℃. *Rho*1,2-CTD was then purified by IMAC chromatography using a Bio-Rad column and Ni-NTA (Qiagen) resin. Ni-NTA slurry was added to a 25mL Bio-Rad column, and the column was equilibrated with 10 column volumes of lysis buffer. Supernatant from the cell lysate preparation was added to the column and the flowthrough was collected. The column was then washed with 1 column volume of wash buffer (10mM Tris-HCl, 150mM NaCl, 20mM Imidazole, adjusted to pH 8.0) 3 times to collect washes 1-3. The protein was then eluted: 1 column volume of elution buffer (10mM Tris-HCl, 150mM NaCl, 250mM Imidazole, adjusted to pH 8.0) was added, the column was inverted several times and incubated for 4 minutes before collecting the elution fraction. This step was repeated 3 times to collect elutions 1-3. The purified *Rho*1,2-CTD sample from elutions 1-3 was applied to a 120mL HiLoad™ 16/60 Superdex200 prep grade column (GE Healthcare Life Sciences) pre-equilibrated with gel filtration buffer (10mM Tris-HCl. 150mM NaCl, adjusted to pH 8.0) using an AKTA pure 25L FPLC system (Cytiva) with UNICORN™ 7.0 software, at a flow rate of 0.3mL/min.

2.5.4 General enzyme activity assay

Enzyme activity was followed by a continuous assay that detects 3-methyl cis, cis muconic acid formation over time from 3-methylcatechol cleavage at 260nm as previously described[52]. The assay was adapted to support a high throughput format in a UV Star 96 well microplate. The 300 µL reaction mixture contained 50mM Tris pH 7.2. After 2 minutes of pre-incubation, the reaction was started by adding the enzyme, and the increase in absorbance at 260nm ($\varepsilon = 16.0\ mM^{-1} \cdot cm^{-1}$) was monitored in a SpectraMax M5e spectrophotometer. Unless otherwise indicated, enzymatic activity was assayed at 40 °C.

2.5.5 Michaelis-Menten kinetic testing

Michaelis-Menten kinetics measurements for all the enzymes we determined by means of the assay described above using an enzyme concentration of 1.5 µg/mL, with 3-methylcatechol concentrations varying from 2.19 to 70 µM. Data points were obtained in triplicate. Data were fitted by to the Michaelis-Menten equation[150] $v = (V_{max}[S])/(K_m + [S])$ using GraphPad Prism version number 9.1.12. Apparent k_{cat} values were calculated on the basis of the molecular mass of the enzyme subunit (31740.73 Da) using the equation $k_{cat} = V_{max}/[ET]$, with ET representing total enzyme concentration in $mg \cdot ml^{-1}$.

2.5.6 Temperature of half-inactivation (T_{50})

T_{50} tests were performed using the assay described by above after heat treatment as previously described[151]. Heat treatment of purified protein was carried out by incubating the protein in 0.2 mL PCR tubes in a BioRad T100 Thermal Cycler. Proteins (30 µL of previously determined optimal concentration which varied between mutants) were heated at different temperatures ranging from 30 °C to 55 °C at intervals of 5 °C for 15 mins and cooled at 4 °C for 10 mins, followed by equilibration at the assay temperature of 25 °C for 5 mins. Samples were

centrifuged to remove any aggregated protein before assaying the enzymatic activity using a 3-methylcatechol concentration of 65µM. Activity of enzyme without undergoing heat treatment was considered to be 100%; the residual activities were quantitatively measured after heating at different temperatures for 15 mins. The T_{50} value is the temperature at which enzyme activity is reduced to 50% after a 15-min heat treatment. The precise value was obtained by determination of the inflection point of a fit of the residual activities at certain temperatures to a sigmoidal plot (sigmoidal Boltzmann fit using GraphPad Prism version number 9.1.12)

2.5.7 Thermal Shift Assay (T_m Testing)

Thermal shift assays were performed as originally described[139] using an iQ5 qPCR machine (BioRad). Enzyme and SYPRO Orange Dye (Ex/Em: 491/596 nm) (ThermoFisher) were added to appropriate wells in a Hard-Shell High Profile 96-well Semi-Skirted PCR plate (BioRad) at a final concentration of 400µg/mL and 5X, respectively in a 50mM Tris buffer, pH 7.2. The PCR plates were sealed using Microseal "B" seals (BioRad) and placed in the qPCR machine. The assay was performed using the melt curve function of the iQ5 qPCR; with a starting point at 20°C and an endpoint of 95°C. Measurements were taken at 0.5°C intervals with 15 seconds at each temperature. Melting points were found at the global minimum of the first derivative of RFU.

2.5.8 Half-Life ($t_{1/2}$) testing

The Rho1,2-CTD variants (0.2 mg/mL) were incubated [151]using a BioRad T100 Thermal Cycler at 40°C for time intervals ranging from 15-minute to 2-hour intervals. Their residual activities were assayed at 40°C with a 3-methylcatechol concentration of 65µM using the assay described in 2.5.4. The data were fitted to first-order plots and analyzed with the first-order rate constants (k_d) determined by linear regression of ln(residual activity) versus incubation time (t).

The time required for the residual activity to be reduced to half ($t_{1/2}$) was calculated using the following equation: $t_{1/2} = \ln 2/k_d$.

2.5.9 Verification of disulfide bonds formation (DTNB method)

The formation of disulfide bonds in *Rho*1,2-CTD mutants was examined using DTNB 5,5-dithio-bis-(2-nitrobenzoic acid (DTNB) (Millipore Sigma) which can quantitatively measure the number of free sulfhydryl groups in protein structure[152]. Two samples were prepared: a reduced and a non-reduced sample. The reduced sample was prepared by mixing the protein sample with 100 mM dithiothreitol (DTT), while the non-reduced was mixed with the same volume of reaction buffer (0.1 M Sodium Phosphate, 1 mM Ethylenediaminetetraacetic acid (EDTA), pH 8.0). Samples were mixed and incubated at 37 °C for two hours, and then underwent buffer exchange using Amicon Ultra Centrifugal Filters (Millipore) to remove DTT. Samples were then concentrated and mixed with reaction buffer and 0.2 mM DTNB. Samples were incubated for 15 minutes at room temperature and absorbance was measured at 412nm (extinction coefficient = 14.15 $mM^{-1} \cdot cm^{-1}$)

Chapter 3

Effects of incorporating Nδ-methylhistidine into the active site of catechol 1,2-dioxygenase on activity and kinetic properties

3.1 Abstract

There are only a limited number of genetically encoded amino acids with the ability to coordinate metals. Thus, to create metalloenzymes with diverse structures and functions, there is a reliance on diverse interaction networks between precisely positioned active site residues as a mechanism to compensate for this limited functionality. Expanding the range of metal coordination environments presents an opportunity to create metalloenzymes with improved properties. Here the introduction of a non-canonical Nδ-methylhistidine (NmH) into the active site of catechol 1,2-dioxygenase from *R. opacus* (*Rho*1,2-CTD) is investigated. This study shows that the substitution of native histidine 220 or 222 with NmH leads to a higher substrate binding affinity, but a dramatically reduced turnover number and catalytic efficiency. Thus, this study demonstrates the need to be able to accurately predict the effects of non-canonical amino acids on catalytic properties as altering the metal-coordinating environment can have unintended consequences.

3.2 Introduction

One of the key steps in the bacterial degradation of aromatic compounds in the environment is the oxidative cleavage of catechol and substituted catechols[153]. Cleavage of the ring component of these compounds is catalyzed by a family of non-heme iron-containing catechol dioxygenases (CTDs). These are divided into two structurally distinct sub-families that also differ mechanistically with respect to their mode of ring cleavage. Intradiol CTDs utilize non-heme Fe(III) to cleave the carbon–carbon bond between the phenolic hydroxyl groups by

inserting an oxygen molecule to yield muconic acid as the product. In contrast, extradiol CTDs employ non-heme Fe(II) or other divalent metal ions to cleave the aromatic nucleus *meta* to the hydroxyl-substituents[154,155].

Catechol 1,2-dioxygenase from *Rhodococcus Opacus* (*Rho*1,2-CTD) is an intradiol dioxygenase. Studies have shown that *Rho*1,2-CTD is highly active compared to 1,2-CTDs from other bacteria of the genus *Rhodococcus,* and has activity against a large range of phenolic compounds[52,74]. *Rho*1,2-CTD is a homodimeric enzyme containing two subunits of 280 residues. Each subunit is folded into two domains: a C-terminal catalytic domain and an N-terminal dimerization domain[1]. The N-terminal domain consists of roughly 100 residues that fold into five helices[1]. The catalytic domain consists of a non-heme iron(III) cofactor ligated by four amino acid side chains: two histidine (His220 and His222) and two tyrosine (Tyr196 and Tyr162) residues[51]. A fifth ligand, a water molecule/hydroxide ion completes a trigonal bipyramidal structure[52].

While *Rho*1,2-CTD is highly active, it is not highly stable. Chapter 2 demonstrated that the stability of this enzyme can be improved significantly through the rational design of disulfide bonds into the structure of the enzyme. However, there was a loss of catalytic activity. Now, Genetic Code Expansion (GCE) technologies are considered as a possible strategy to improve the catalytic activity of the stabilized mutants.

In recent years, there has been growing interest in the field of GCE, where the effects of site-specifically incorporating noncanonical amino acids (ncAAs) into enzymes with potential applications in industrial biocatalysis have been increasingly investigated[156]. GCE is a technique used to expand the repertoire of amino acids that can be incorporated into proteins during translation beyond the 20 canonical amino acids by reprogramming rarely used codons (typically

the amber (AMB) stop codon) to encode for non-canonical amino acids. This emerging technology offers the potential of unlocking new chemistries and improving catalytic properties. This study explores the potential of applying this new technology to improve the catalytic properties of *Rho*1,2-CTD.

In this context, the non-canonical amino acid Nδ-methylhistidine (NmH) presents interesting potential. This residue is an analog of histidine, where the delta nitrogen on the imidazole ring is bonded to a methyl group. It is proposed that this modification does several things[126]: (1) it fixes the tautomeric form of the imidazole ring, (2) it ensures the neutral charge state of the proximal histidine ligand throughout the catalytic cycle by preventing the possible formation of anionic ligand via deprotonation of the Nδ hydrogen and (3) it prevents the buildup of imidazolate character that can occur via polarization of the His Nδ-H bond. It is proposed that substituting one or both of the native histidine's in the active site of *Rho*1,2-CTD (H220 and H222) with the non-canonical amino acid Nδ-methylhistidine (NmH) could improve the catalytic properties of the enzyme. Indeed, several studies have demonstrated marked improvements in catalytic activity when replacing canonical histidine residues with NmH[123,124]. One study showed that the installation of NmH as the proximal heme ligand in myoglobin led to a 3.7-fold increase in catalytic efficiency[123]. Another determined that the introduction of a non-canonical Me-His23 nucleophile converted a promiscuous ester hydrolase activity of the MBHase BH32 into its predominant activity and substantially increased the rate of turnover[124].

The mechanisms underlying the benefits provided by substitution with NmH aren't fully understood at this time. Mutagenesis studies have shown that imparting imidazolate character onto the axial residue affects Fe-His bond strength, redox potential, and iron electronic properties. Regardless of the mechanism, the histidine's in *Rho*1,2-CTD play a similar role as in

APX2 and engineered myoglobin as axial iron-coordinating residues. Based on these observations, it is hypothesized that substitution of one or both native histidine's in *Rho*1,2-CTD with NmH could improve its' catalytic properties.

3.3 Results

3.3.1 Recombinant expression and purification of wild-type *Rho*1,2-CTD

WT protein was obtained as described in Chapter 2.

3.3.2 Recombinant expression with incorporation of Nδ-methylhistidine (NmH) into *Rho*1,2-CTD

Catalytic site His residues 220 and 222 were selected as targets for NmH substitution. To create each mutant, the codon encoding the targeted residue was mutated to an amber codon (Figure 3.1).

>WT *Rho*1,2-CTD

catATGACTACCACTGAATCTCCGACCGCTGCGGGTTCTGGTTCTGCGGCTACCGACAAATTCAAAGC
GGAACGTGCAACCGCTGACACCTCTCCGGAACGTCTGGCGGCGATCGCGAAAGACGCTCTGGGTGC
TCTGAACGATGTTATCCTGAAACACGGTGTTACCTACCCGGAATACCGTGTTTTCAAACAGTGGCTG
ATCGACGTTGGTGAAGGTGGTGAATGGCCGCTGTTCCTGGACGTTTTCATCGAACACTCTGTTGAAG
AAGTTCTGGCGCGTTCTCGTAAAGGTACCATGGGTAGCATCGAAGGTCCGTACTACATCGAAAACT
CTCCGGAACTGCCGTCTAAATGCACCCTGCCGATGCGTGAAGAAGATGAAAAGATCACCCCGCTGG
TTTTCTCTGGTCAGGTTACCGACCTGGACGGTAACGGTCTGGCAGGTGCTAAAGTTGAACTGTGGCA
CGCTGACAACGACGGTTACTACTCTCAGTTCGCGCCGCACCTGCCGGAATGGAACCTGCGTGGTACC
ATCATCGCTGACGAAGAAGGCCGCTACGAAATCACCACCATCCAGCCGGCGCCGTACCAGATTCCG
ACCGATGGTCCGACCGGTCAGTTCATCGAAGCGCAGAACGGTCACCCGTGGCGTCCGGCGCACCTG
CACCTGATCGTTTCTGCGCCGGGTAAAGAATCCGTTACCACCCAGCTGTATTTCAAAGGTGGTGAAT
GGATCGACTCTGACGTTGCGTCTGCGACCAAACCGGAACTGATCCTGGATCCGAAAACCGGTGACG
ACGGTAAAAACTACGTTACCTACAACTTCGTTCTGGACCCGGCGctcgag

>H220NmH *Rho*1,2-CTD

catATGACTACCACTGAATCTCCGACCGCTGCGGGTTCTGGTTCTGCGGCTACCGACAAATTCAAAGC
GGAACGTGCAACCGCTGACACCTCTCCGGAACGTCTGGCGGCGATCGCGAAAGACGCTCTGGGTGC
TCTGAACGATGTTATCCTGAAACACGGTGTTACCTACCCGGAATACCGTGTTTTCAAACAGTGGCTG
ATCGACGTTGGTGAAGGTGGTGAATGGCCGCTGTTCCTGGACGTTTTCATCGAACACTCTGTTGAAG
AAGTTCTGGCGCGTTCTCGTAAAGGTACCATGGGTAGCATCGAAGGTCCGTACTACATCGAAAACT
CTCCGGAACTGCCGTCTAAATGCACCCTGCCGATGCGTGAAGAAGATGAAAAGATCACCCCGCTGG
TTTTCTCTGGTCAGGTTACCGACCTGGACGGTAACGGTCTGGCAGGTGCTAAAGTTGAACTGTGGCA
CGCTGACAACGACGGTTACTACTCTCAGTTCGCGCCGCACCTGCCGGAATGGAACCTGCGTGGTACC
ATCATCGCTGACGAAGAAGGCCGCTACGAAATCACCACCATCCAGCCGGCGCCGTACCAGATTCCG
ACCGATGGTCCGACCGGTCAGTTCATCGAAGCGCAGAACGGTCACCCGTGGCGTCCGGCGUAGCTG
CACCTGATCGTTTCTGCGCCGGGTAAAGAATCCGTTACCACCCAGCTGTATTTCAAAGGTGGTGAAT
GGATCGACTCTGACGTTGCGTCTGCGACCAAACCGGAACTGATCCTGGATCCGAAAACCGGTGACG
ACGGTAAAAACTACGTTACCTACAACTTCGTTCTGGACCCGGCGctcgag

> H222NmH *Rho*1,2-CTD

catATGACTACCACTGAATCTCCGACCGCTGCGGGTTCTGGTTCTGCGGCTACCGACAAATTCAAAGC
GGAACGTGCAACCGCTGACACCTCTCCGGAACGTCTGGCGGCGATCGCGAAAGACGCTCTGGGTGC
TCTGAACGATGTTATCCTGAAACACGGTGTTACCTACCCGGAATACCGTGTTTTCAAACAGTGGCTG
ATCGACGTTGGTGAAGGTGGTGAATGGCCGCTGTTCCTGGACGTTTTCATCGAACACTCTGTTGAAG
AAGTTCTGGCGCGTTCTCGTAAAGGTACCATGGGTAGCATCGAAGGTCCGTACTACATCGAAAACT
CTCCGGAACTGCCGTCTAAATGCACCCTGCCGATGCGTGAAGAAGATGAAAAGATCACCCCGCTGG
TTTTCTCTGGTCAGGTTACCGACCTGGACGGTAACGGTCTGGCAGGTGCTAAAGTTGAACTGTGGCA
CGCTGACAACGACGGTTACTACTCTCAGTTCGCGCCGCACCTGCCGGAATGGAACCTGCGTGGTACC
ATCATCGCTGACGAAGAAGGCCGCTACGAAATCACCACCATCCAGCCGGCGCCGTACCAGATTCCG
ACCGATGGTCCGACCGGTCAGTTCATCGAAGCGCAGAACGGTCACCCGTGGCGTCCGGCGCACCTG
UAGCTGATCGTTTCTGCGCCGGGTAAAGAATCCGTTACCACCCAGCTGTATTTCAAAGGTGGTGAAT
GGATCGACTCTGACGTTGCGTCTGCGACCAAACCGGAACTGATCCTGGATCCGAAAACCGGTGACG
ACGGTAAAAACTACGTTACCTACAACTTCGTTCTGGACCCGGCGctcgag

Figure 3.1: DNA coding sequences for WT Rho1,2-CTD and variants H220NmH and H222NmH. The codons encoding His 220 and 222 are highlighted in cyan. Codons mutated to amber codon (UAG) are highlighted in yellow. Lowercase letters represent restriction sites that are not part of the coding sequence.

The selective incorporation of NmH into proteins using amber codons with an evolved orthogonal tRNA/tRNA synthetase pair (PylRS(NmH)/tRNA^{Pyl}) has been previously

described[157]. The otRNA/tRNA synthetase pair is encoded by the pEVOL plasmid which has been optimized for the production of proteins containing noncanonical residues[158]. This pEVOL construct was co-transformed with a pET31b+ plasmid encoding the modified *Rho*1,2-CTD sequence. Transformed cells were grown in LB media containing appropriate antibiotics. Expression of the otRNA/tRNA synthetase pair and the NmH-incorporated *Rho*1,2-CTD was induced with arabinose and IPTG respectively. NmH is also added to allow for incorporation (Figure 3.2). Following expression, the protein was purified by Ni-NTA IMAC purification, followed by SEC purification as described for WT.

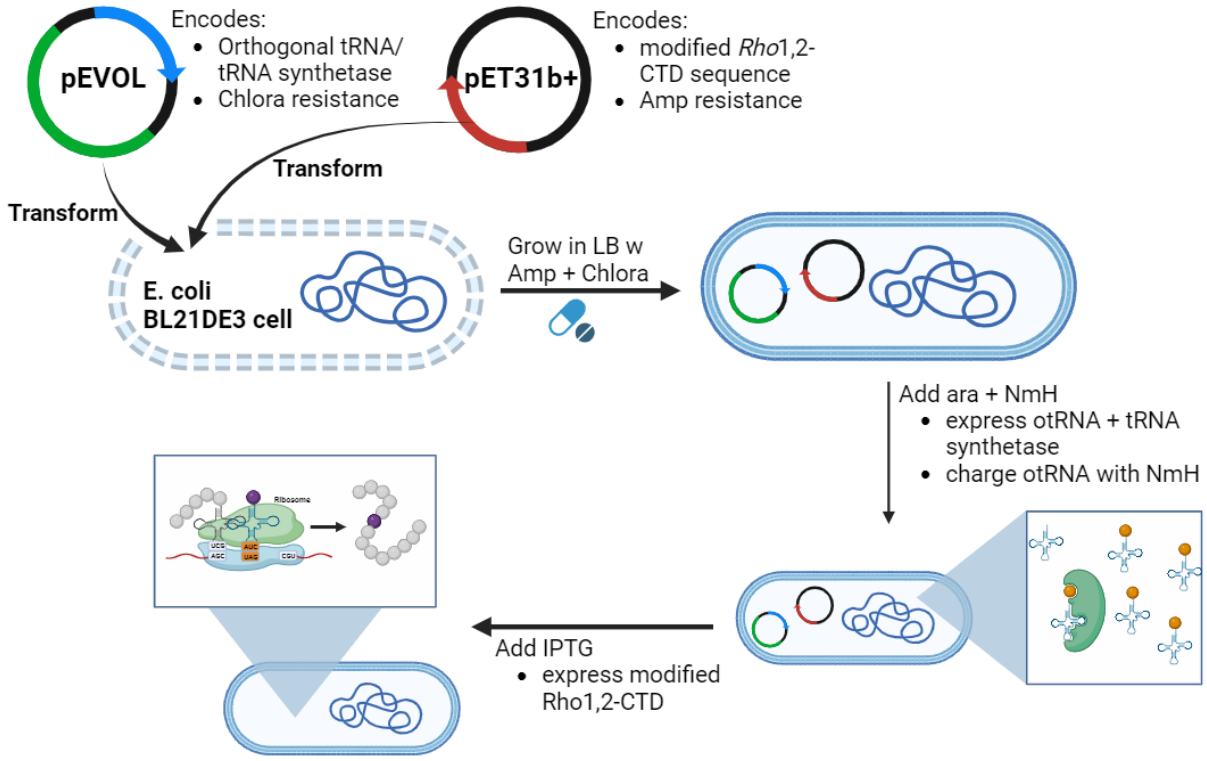

Figure 3.2: Diagram outlining the expression and purification process for NmH-incorporated *Rho*1,2-CTD. An E coli cell is transformed with two plasmids: pEVOL and pET31b+. The pEVOL plasmid contains the sequence for an orthogonal tRNA/tRNA synthetase pair that has been evolved to incorporate NmH and a chloramphenicol resistance gene. The pET31b+ plasmid contains the modified Rho1,2-CTD sequence and an ampicillin resistance gene. After transformation, cells are grown in LB media containing ampicillin and chloramphenicol so that only doubly transformed cells survive. Expression of the orthogonal tRNA/tRNA synthetase pair is induced with arabinose. NmH is added at the same time so the orthogonal tRNA synthetases can charge the orthogonal tRNAs with NmH. Expression of NmH-incorporated Rho1,2-CTD is induced with IPTG.

3.3.3. Purification of putative His220NmH-*Rho*1,2-CTD

Unlike in the purification of WT *Rho*1,2-CTD, there are 3 clearly visible bands of different molecular weights in the H220NmH variant elution fractions (Figure 3.3). The band around 34 kDa (indicated with the blue arrow) is likely full-length protein, which has an expected molecular weight of 31.5 kDa. The bands close to the 72 kDa molecular weight marker (indicated with the yellow arrow) can also be seen in the WT purification, though they appear more evident here due to there being relatively less NmH variant than WT *Rho*1,2-CTD present. These high molecular weight bands are likely dimers of *Rho*1,2-CTD, which have an expected molecular weight of 64 kDa. The lower molecular weight bands (indicated with the red arrow) could be truncated 1,2-CTD proteins. A *Rho*1,2-CTD protein truncated at the amber codon of position 220, instead of having NmH incorporated would have an approximate molecular weight of 24 kDa. A truncated version of the enzyme should be removed in the wash steps of purification as it would lack a His-Tag. Despite this, a band appearing to be approximately the same molecular weight as the truncated protein appears on the gel. This is considered below in the discussion.

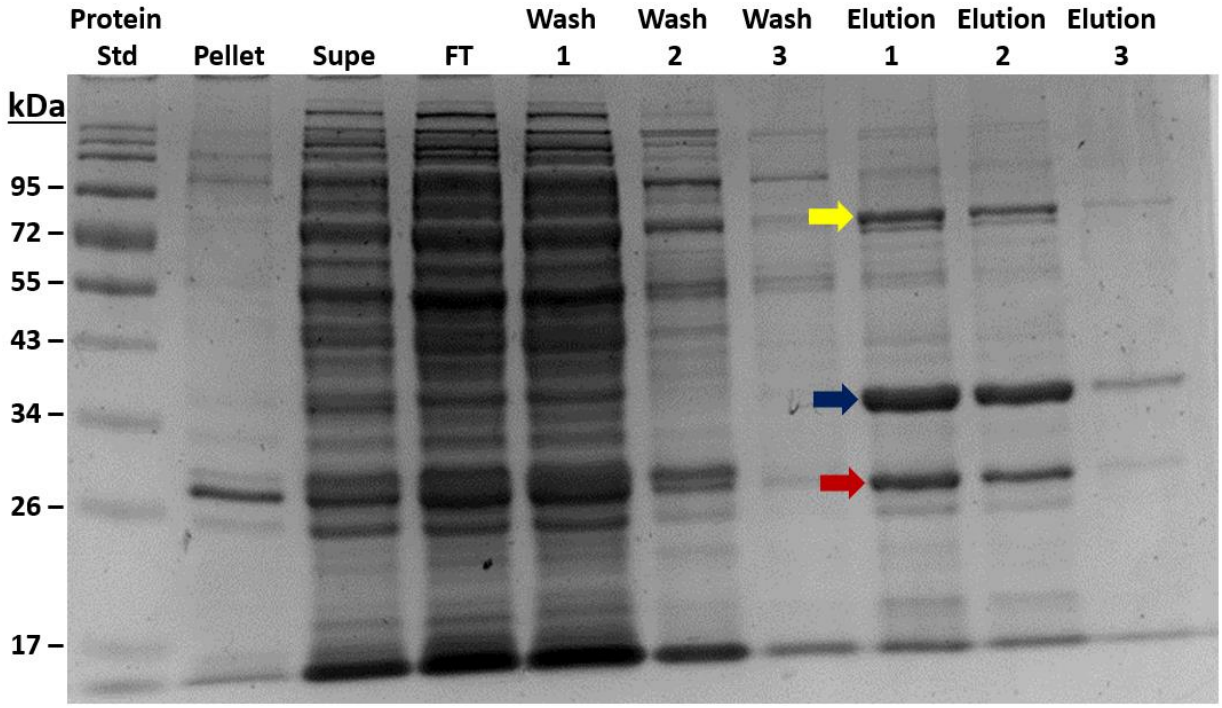

Figure 3.3: SDS-PAGE gel showing samples from the metal-affinity purification of H220NmH. Gel was stained with Coomassie blue and imaged using a Bio-Rad ChemiDoc™ MP imager. Pellet was obtained after lysed cells were centrifuged. Supernatant (Supe) was obtained from cell lysis. Flowthrough (FT) was obtained after running the supernatant through the nickel column. Washes were obtained by running wash buffer with an imidazole concentration of 20mM through the column. Elution's were obtained by running elution buffer with an imidazole concentration of 250mM through the column. Different coloured arrows identify bands of interest.

The size-exclusion purification profile shows a large, sharp peak in absorbance occurring between 63 and 70 minutes (Figure 3.4A). The steep peak demonstrated that the purification was likely successful as only one product appears to have eluted in this time frame. To confirm successful size-exclusion purification, fractions were run on an SDS-PAGE gel with the pooled eluates from Ni-NTA purification as a reference (Figure 3.4B). The gel shows a band in each fraction around the expected molecular weight of *Rho*1,2-CTD, confirming its presence in the purified samples. The gel also shows that the larger product of around 72 kDa seen after Ni-NTA purification was largely eliminated by size exclusion purification. Unfortunately, some of the presumed truncated 1,2-CTD still appears to be present in the purified fractions. To counter this,

only the fractions that appear the purest (fractions 29 and 30) were used for kinetic testing. In these fractions, the amount of truncated protein compared to full-length protein is quite small, so it should not have a significant impact on kinetic testing.

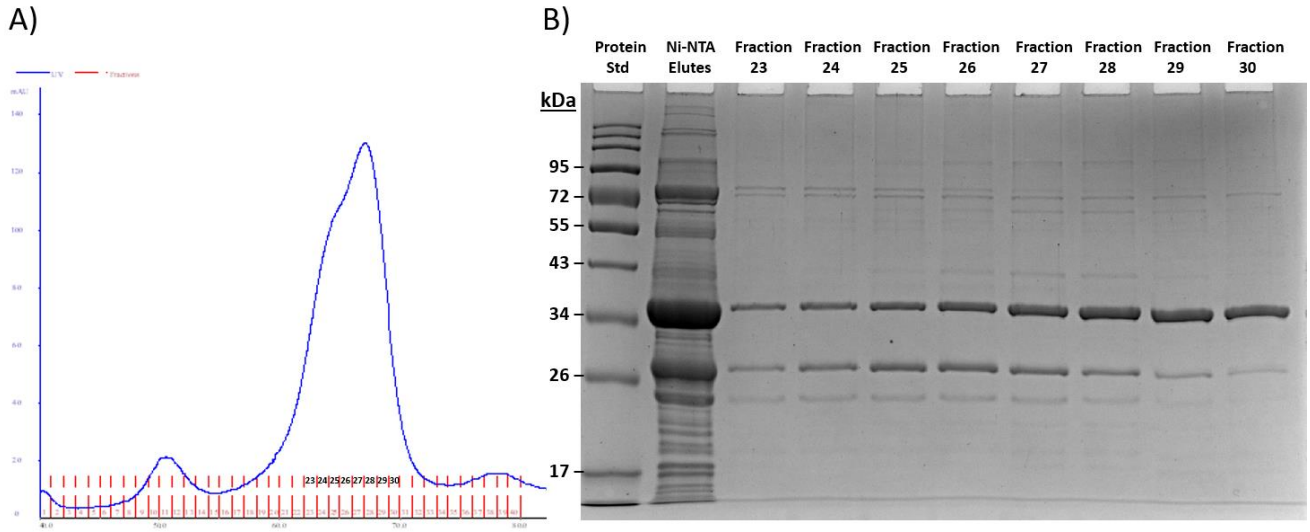

Figure 3.4: Size-exclusion purification of H220NmH. Size exclusion was performed using a HiLoad 16/600 Superdex 200 size exclusion column on AKTA Pure 25L FPLC. **A)** Size exclusion purification curve. Absorbance data was collected at 280 nm. Red numbers indicate fraction numbers. **B)** SDS-PAGE gel showing samples from fractions 23-30 of size-exclusion purification. Gel was stained with Coomassie blue and imaged using a Bio-Rad ChemiDoc™ MP imager. Ni-NTA Elutes refers to elutions 1-3 obtained from metal-affinity purification, which were combined and run on the gel as a reference.

3.3.4 Purification of putative His222NmH-*Rho*1,2-CTD

To address the issues with truncated protein persisting throughout the purification, which were noticed in the purification of H220NmH, the Ni-NTA IMAC enrichment protocol was modified slightly in an attempt to improve the separation of the putative full-length protein from the putative truncated protein. As truncated protein wouldn't have a His-Tag, it should bind with lower affinity to the nickel column. Therefore, the concentration of imidazole in the washes was increased from 20 mM to 50 mM in an attempt to elute the truncated protein but not the full-length protein. However, this modification was not very successful as much of the putative full-

length protein was eluted in the washes (Figure 3.5). To recover the full-length protein lost in the washes, the washes were re-combined, and purification was attempted again. In this second variation, two washes were performed using a concentration of 20 mM imidazole, followed by two washes using a concentration of 50 mM imidazole. In this case, much more of the full-length protein was recovered, and much of the impurity was eliminated (Figure 3.3). However, some impurities persisted including the putative dimer and putative AMB222 truncated version.

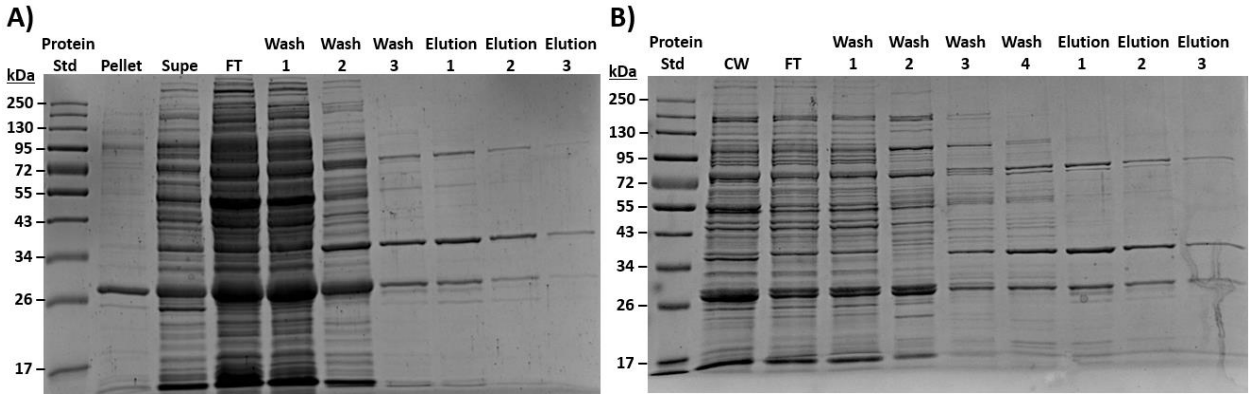

Figure 3.5: SDS-Page gels showing samples from the metal-affinity purification of H222NmH. Gel was stained with Coomassie blue and imaged using a Bio-Rad ChemiDoc^{TM} MP imager. **A)** Purification of cell lysate. Pellet was obtained after lysed cells were centrifuged. Supernatant (Supe) was obtained from cell lysis. Flowthrough (FT) was obtained after running the supernatant through the nickel column. Washes were obtained by running wash buffer with a low imidazole concentration (50mM) through the column. Elution's were obtained by running elution buffer with a high concentration of imidazole (250 mM) through the column. **B)** The washes shown in A) were recombined (CW) and re-purified. Flowthrough was obtained after running combined washes through column. Wash 1 and 2 were eluted using a wash buffer with a concentration of 20 mM. Wash 3 and 4 were eluted using a wash buffer with a concentration of 50 mM. Elutions were obtained using elution buffer with a 250 mM concentration of imidazole.

The size exclusion profile shows a large peak between 38 and 42 minutes, indicating elution of a product. A smaller peak was also present between 47-48 minutes and a large peak was present at the tail end of the purification (Figure 3.6A). None of these peaks seem to correspond to the proper elution time of *Rho*1,2-CTD, based on WT and His222NmH SEC profiles, but fractions were still collected for further scrutiny. Despite eluting earlier than

expected the product isolated from fractions corresponding to the first and second peaks appear to be the correct molecular weight for *Rho*1,2-CTD upon SDS-PAGE analysis (Figure 3.6B). This is quite unusual. A possible explanation could be that the FPLC column used for the size-exclusion purification was calibrated incorrectly. In addition to this problem, as was seen in the purification of H220, the truncated protein is still present in the purification fractions following size-exclusion purification. Fractions 1 and 2 were used for assays.

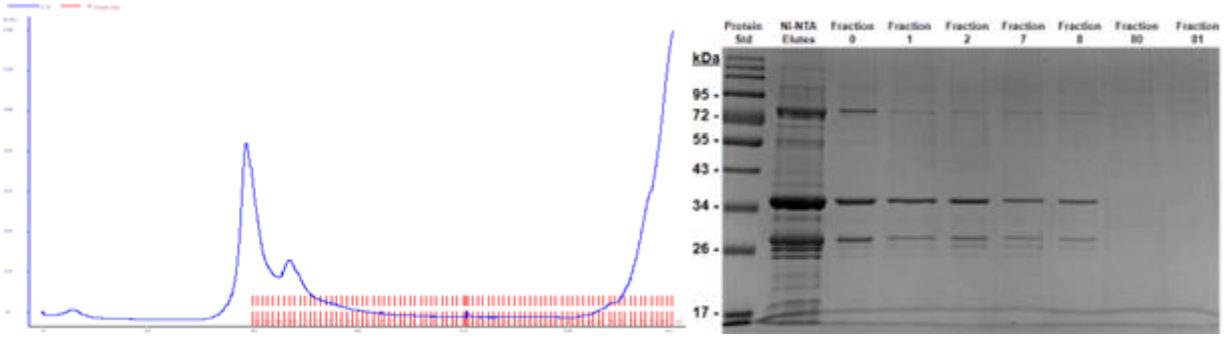

Figure 3.6: Size exclusion purification of the sample putatively containing H222NmH *Rho*1,2-CTD. Size exclusion was performed using a HiLoad 16/600 Superdex 200 size exclusion column on AKTA Pure 25L FPLC. **A)** Size exclusion purification curve. Absorbance data was collected at 280 nm. Red numbers indicate fraction numbers. **B)** SDS_PAGE gel showing samples from each of the peaks present in the FPLC profile. Gel was stained with Coomassie blue and imaged using a Bio-Rad ChemiDoc^{TM} MP imager. Ni-NTA Elutes refers to elutions 1-3 obtained from metal-affinity purification, which were combined and run on the gel as a reference.

3.3.5 Michaelis-Menten kinetic analyses

Nonetheless, the obtained purified version of WT-*Rho*1,2-CTD and putative variants H220NmH-*Rho*1,2-CTD and H222NmH-*Rho*1,2-CTD were used for kinetic testing to examine their respective catalytic activities against 3-methylcatechol (Figure 3.7). In the case of this assay, 3-methylcatechol was used as a substrate, which is converted by the enzymes into 3-carboxy-cis,cis muconic acid. This product absorbs at a wavelength of 260 nm which was used to detect product formation. The kinetic parameters were analyzed by fitting the data to a

Michaelis-Menten plot using GraphPad (Table 3.1). Both mutants appear to show improvement in K_m, suggesting they have increased productive substrate binding compared to the wild-type enzyme. However, both mutants have significantly lower catalytic efficiency owing to dramatically decreased k_{cat} values. While not the preferred outcome, the loss of catalytic functionality does suggest recognition of the amber codon and possibly incorporation of NmH in the catalytic site, as His incorporation would support WT functionality.

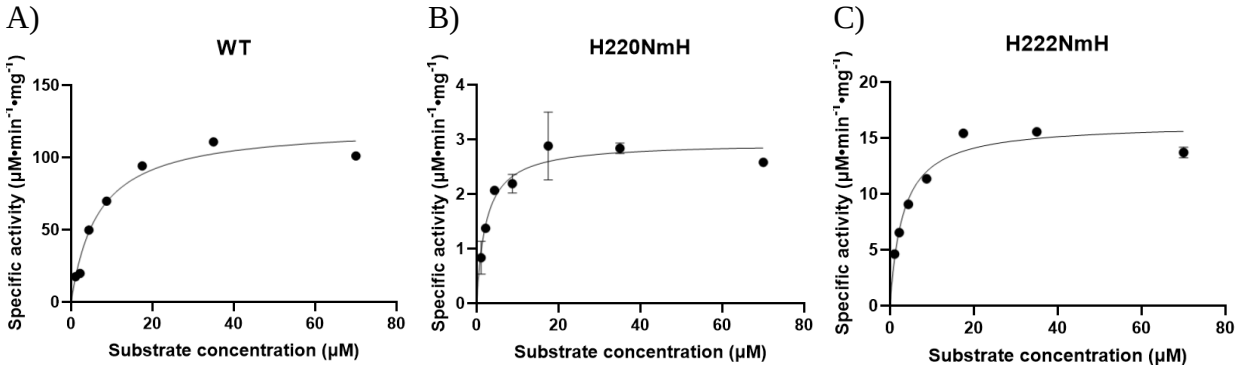

Figure 3.7: Michaelis-Menten kinetic curves of wild-type (WT), H220NmH, and H222NmH *Rho*1,2-CTD. Activity of Rho1,2-CTD was measured spectrophotometrically to measure the initial rate of product formation at differing substrate concentrations. These data were fit to the Michaelis-Menten equation using GraphPad Prism version 9.1.12 to extract the kinetic parameters. Specific activity was calculated by dividing the initial rate of product formation by the concentration of enzyme. Data plots specific activity vs substrate concentration and represents the mean of three technical replicates with error bars showing the standard error. A) Plot of WT, B) Plot of H220NmH, C) Plot of H222NmH

Table 3.1: Kinetic parameters for WT, H220NmH, and H222NmH Rho1,2-CTD.
Table shows the mean and the SEM for each parameter.

	WT*	**WT**	**H220NmH**	**H222NmH**
K_m μM	6.5	6.8 ± 0.8	2.3 ± 0.7	3.1 ± 0.4
k_{cat} (sec^{-1})	636.9	122 ± 4.4	2.9 ± 0.2	16.3 ± 0.5
k_{cat}/K_m ($sec^{-1}*\mu M^{-1}$)	98	18 ± 3	1.3 ± 0.4	5.2 ± 0.9

*Literature values from (Matera et al)[52]

3.4 Discussion

Chapter 2 showed that the thermostability of *Rho*1,2-CTD can be greatly improved via the rational design of disulfide bonds. However, this improvement in stability comes with a loss in catalytic activity. To compensate for this, GCE was used in an attempt to improve the catalytic activity of the stabilized enzyme by substituting native Iron coordinating histidine's with the noncanonical amino acid NmH. However, these variations resulted in decreased turnover numbers and reduced catalytic efficiency.

3.4.1 Purification of Nδ-methylhistidine (NmH) incorporated Rho1,2-CTD (H220 and H222)

Following the purification of both mutants it could be seen that there was a small, persistent impurity present in purified enzyme samples. Based on the size of this impurity, a likely explanation is that the impurity represents a truncated *Rho*1,2-CTD enzyme. Indeed, recombinant *Rho*1,2-CTD truncated at position 220 or 222 would have a molecular weight of 24 kDa. The location of the impurity on the SDS gels appears to support this explanation. The presence of truncated protein is an anticipated risk being that the amber codon naturally functions as a stop codon. Despite the inclusion of the reencoded PylRS/tRNAPyl, the amber codon can still act as a stop codon. In this system, the amber codon is recognized both by the orthogonal tRNA and Release Factor 1. If release factor 1 reaches the ribosome during translation before the orthogonal tRNA, it will terminate translation at the stop codon. To address this possibility, the His-Tag was placed at the C-terminal end of the protein so that in theory, only full-length protein should be purified in nickel metal-affinity chromatography. However, this does not appear to have happened as a small impurity of the same size as the truncated protein was purified with full-length protein.

A possible explanation for this is that the truncated protein was co-purified as a heterodimer dimer with full-length protein. The dimerization domain of the protein is made up of roughly the first 100 residues in the primary sequence, meaning that a truncation at position 220 or 222 would leave this domain fully intact. If bound to an intact dimer containing a His-Tag during purification, this construct would be eluted in the elution steps. This could also explain why the truncated version isn't entirely eliminated during size-exclusion purification as dimer composed of one full-length and one truncated subunit would have a similar molecular weight to a dimer between two full-length proteins (56 vs 63 kDa), see Figure 3.8.

>Wild Type

MTTTESPTAAGSGSAATDKFKAERATADTSPERLAAIAKDALGALNDVILKHGVTYPEYRVFK
QWLIDVGEGGEWPLFLDVFIEHSVEEVLARSRKGTMGSIEGPYYIENSPELPSKCTLPMREEDE
KITPLVFSGQVTDLDGNGLAGAKVELWHADNDGYYSQFAPHLPEWNLRGTIIADEEGRYEITT
IQPAPYQIPTDGPTGQFIEAQNGHPWRPAHLHLIVSAPGKESVTTQLYFKGGEWIDSDVASATK
PELILDPKTGDDGKNYVTYNFVLDPAHHHHHH

Number of residues: 286

MW: 31.52kDa

>H220NmH

MTTTESPTAAGSGSAATDKFKAERATADTSPERLAAIAKDALGALNDVILKHGVTYPEYRVFK
QWLIDVGEGGEWPLFLDVFIEHSVEEVLARSRKGTMGSIEGPYYIENSPELPSKCTLPMREEDE
KITPLVFSGQVTDLDGNGLAGAKVELWHADNDGYYSQFAPHLPEWNLRGTIIADEEGRYEITT
IQPAPYQIPTDGPTGQFIEAQNGHPWRPA(NmH)LHLIVSAPGKESVTTQLYFKGGEWIDSDVAS
ATKPELILDPKTGDDGKNYVTYNFVLDPAHHHHHH

Number of residues: 286

MW: 31.54kDa

>H222NmH

MTTTESPTAAGSGSAATDKFKAERATADTSPERLAAIAKDALGALNDVILKHGVTYPEYRVFK
QWLIDVGEGGEWPLFLDVFIEHSVEEVLARSRKGTMGSIEGPYYIENSPELPSKCTLPMREEDE
KITPLVFSGQVTDLDGNGLAGAKVELWHADNDGYYSQFAPHLPEWNLRGTIIADEEGRYEITT
IQPAPYQIPTDGPTGQFIEAQNGHPWRPAHL(NmH)LIVSAPGKESVTTQLYFKGGEWIDSDVAS
ATKPELILDPKTGDDGKNYVTYNFVLDPAHHHHHH

Number of residues: 286

MW: 31.54kDa

>Position 220 Truncation

MTTTESPTAAGSGSAATDKFKAERATADTSPERLAAIAKDALGALNDVILKHGVTYPEYRVFK
QWLIDVGEGGEWPLFLDVFIEHSVEEVLARSRKGTMGSIEGPYYIENSPELPSKCTLPMREEDE
KITPLVFSGQVTDLDGNGLAGAKVELWHADNDGYYSQFAPHLPEWNLRGTIIADEEGRYEITT
IQPAPYQIPTDGPTGQFIEAQNGHPWRPA

Number of residues: 219

MW: 24.05kDa

>Position 222 Truncation

MTTTESPTAAGSGSAATDKFKAERATADTSPERLAAIAKDALGALNDVILKHGVTYPEYRVFK
QWLIDVGEGGEWPLFLDVFIEHSVEEVLARSRKGTMGSIEGPYYIENSPELPSKCTLPMREEDE
KITPLVFSGQVTDLDGNGLAGAKVELWHADNDGYYSQFAPHLPEWNLRGTIIADEEGRYEITT
IQPAPYQIPTDGPTGQFIEAQNGHPWRPAHL

Number of residues: 221

MW: 24.30kDa

Figure 3.8: Amino acid sequences for Rho1,2-CTD variants and truncated versions

3.4.2 Functional characterization of the putatively NmH incorporated variants

Following successful recombinant expression and purification of *Rho*1,2-CTD, kinetic testing was performed to determine the kinetic parameters of the enzyme and compare them to previously published data. *Rho*1,2-CTD catalyzes the reaction of catechol (and its derivatives) to cis,cis-muconic acid (or its derivatives). In this study, 3-methylcatechol was used as it was shown that 1,2-CTD had the highest activity against this particular substrate[52]. Results from kinetic testing of the wild-type enzyme showed a K_m value almost identical to that of previously published studies, however the k_{cat} obtained in these experiments was smaller than previously published results[51]. This can likely be explained due to the fact that the enzyme used in this study was produced recombinantly and contained a His-tag while in the other study acquired protein from the native source[52]. Previous studies such as[159] have shown that His-tags can decrease catalytic activity. Since the standard errors obtained from analysis of the WT activity data were quite small, it provided confidence in the accuracy of the acquired results.

Confident that kinetic testing provided accurate kinetic values, testing was performed on NmH-incorporated enzymes H220NmH and H222NmH. Kinetic values obtained from the testing of both NmH-incorporated enzymes demonstrated significantly reduced catalytic activity compared to the wild-type enzyme (though it should be noted that quantification of purity might alter kinetic outcomes). Previous studies testing the fidelity and efficiency of PylHRS/tRNA pair confirmed its specificity for NmH over canonical His[157]. This would suggest that any activity observed must be coming from the NmH-incorporated enzyme (the truncated enzyme should have no activity).

There are other examples in the literature where substitution of NmH into the active site decreases catalytic properties. A previous study showed that substitution of catalytic site His 163

with NmH in an engineered ascorbate peroxidase (APX2) results in a 50% loss in k_{cat}. Similar to *Rho*1,2-CTD, this enzyme demonstrated an improvement in K_m value and a deterioration in k_{cat} value[126]. As His163 plays a similar role in APX2 as His220 and His222 play in *Rho*1,2-CTD as an iron-coordinating residue, there could be a similar underlying mechanism behind these changes in catalytic activity. However, it should be noted that the decrease in k_{cat} seen in *Rho*1,2-CTD is significantly greater than the decrease in k_{cat} seen in APX2.

In an attempt to investigate this further and elucidate possible reasons why NmH-incorporated enzymes displayed reduced activity against 3-methyl catechol, the active sites were modelled based on the crystal structure of *Rho*1,2-CTD (PDBID 3HGI) using PyMOL. While DFT or QM analyses would be needed to validate these proposed correlations, this analysis showed a few structural aspects that might contribute to the observed outcomes.

Modelling of H220NmH showed that the delta nitrogen of His220 forms a hydrogen bond with the carbonyl oxygen of Pro218 (Figure 3.9). Mutation of His to NmH would disrupt this interaction and the delta methyl group creates a potential steric clash with the carbonyl group. This clash could potentially alter the rotational angle of histidine, negatively affecting its ability to coordinate the iron(III). To validate the idea of a reduced amount of iron, the presence of iron could have been measured by absorbance[160]. Previous studies that substituted His for NmH also disrupted a hydrogen bond and created a potential steric clash between His and Asp in the case of APX2[126] and His and Ser in the case of engineered myoglobin[123]. The difference here could be due to the fact that in the case of *Rho*1,2-CTD, this steric clash occurs with the protein backbone, rather than a side chain, meaning the protein backbone has to shift to accommodate the extra methyl group which could interfere with the secondary structure of the protein.

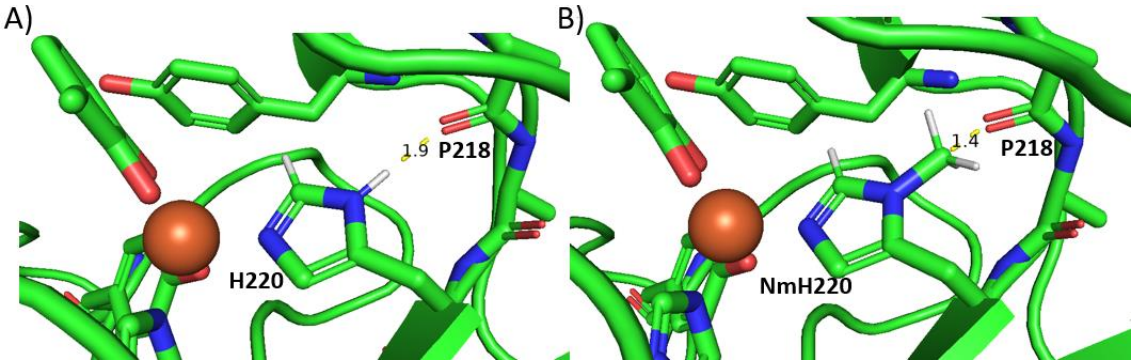

Figure 3.9: Models of *Rho*1,2-CTD active site, based on crystal structure PDB ID: 3HJQ showcasing interaction between residues 220 and 218. A) WT active site. B) H220NmH active site. The methyl group of NmH220 is within 1.4Å of the backbone carbonyl oxygen of P218, suggesting the possibility of a steric clash. Structures were visualized using PyMOL[69].

Modelling of H222NmH shows that the delta nitrogen of His222 appears to be involved in a hydrogen bonding network (Figure 3.10). In this network, the carbonyl oxygen of T235, the carbonyl oxygen of I102, and the delta nitrogen of His222 each form a hydrogen bond with a water molecule. The importance of this network is unknown, but perhaps the disruption of this network could have a destabilizing effect which would explain the decrease in activity displayed by H222NmH.

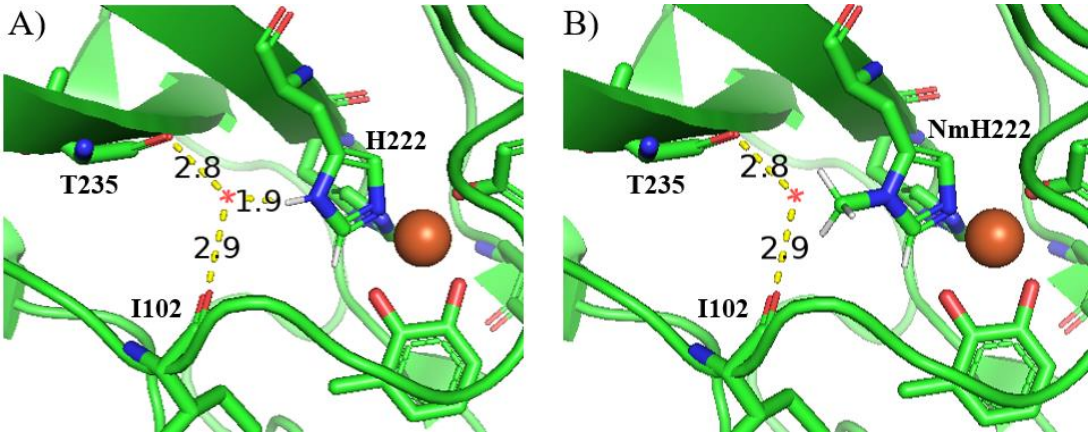

Figure 3.10: Models of *Rho*1,2-CTD active site based on crystal structure PDB ID: 3HJQ. Figure showcases potential hydrogen bonding network between residues 102, 222, 235 and a water molecule. Catalytic Fe(III) is represented by the orange sphere. Iron-coordinating residues are coloured by element; C is green, N is blue, O is red. A) WT active site. B) H222NmH active site. Since methyl groups can't participate in hydrogen bonding, replacing the hydrogen on the delta nitrogen with a methyl group may disrupt a hydrogen bonding network. Structures were visualized using PyMOL[69].

Another possible explanation is that these mutants have not incorporated NmH in the first place. It should be noted that codons encoding Gln, Tyr and Trp are all one base off from the amber codon (GAG, UAC/UAU, UGG vs UAG). Due to this, misreading of the amber codon is possible. Indeed, it has been noted that natural amino acids can be incorporated in response to the amber codon, apparently via misreading of Gln, Tyr, and Trp tRNAs[161].

3.4.3 Conclusions and future directions

In conclusion, an attempt was made to improve the catalytic properties of *Rho*1,2-CTD by incorporating the non-canonical amino acid Nδ-methylhistidine, into the active site of the enzyme. Kinetic testing showed an increase in productive binding, but a dramatically decreased k_{cat} value.

Ultimately, the decision was made to halt this project due to increasing complexity and effort required to continue and a lack of good results to show for the work. As such there are several loose ends on this project that were never tied up. The most important of these loose ends is the fact that incorporation was never verified. Since this wasn't done it is difficult to know for certain the true cause of the poor catalytic activity demonstrated by the mutants. If this project were to continue it would be important to verify the incorporation of Nδ-methylhistidine into the mutant enzymes using mass spectrometry or crystallization highlighted in the literature[126]. This would confirm whether the loss in activity is indeed caused by NmH incorporation or rather by issues with incorporation.

Overall, this negative result should not dissuade from the potential GCE offers in numerous applications. More so, it should serve as a cautionary tale that altering the metal-coordination site of enzymes can have unexpected effects, so it is important to thoroughly evaluate and study target enzymes to improve chances of success.

3.5 Materials and Methods

3.5.1 Materials

All chemicals were obtained from Sigma unless otherwise stated below. Wild-type and mutant H220NmH and H222NmH coding regions were synthesized and cloned into the pET31b+ vector (EMD Millipore) by Biobasics Inc. The genes were excised and amplified using PCR. The amplicons were then subcloned into empty pET31b+ vectors (EMD Millipore) at the XhoI and NdeI restriction sites in the Multiple Cloning Site. The pET31b+ vector confers resistance to ampicillin. Expression of *Rho*1,2-CTD in pET31b+ is under the control of a T7 Promoter induced by isopropyl-β-d-thiogalactopyranoside (IPTG) and encodes a C-terminal His-Tag. The pEVOL-PylRS(NmH)[157] plasmid encodes an orthogonal tRNA/tRNA synthetase enabling incorporation of the non-canonical amino acid Nδ-methylhistidine at the amber codon during protein translation. The orthogonal tRNA/tRNA synthetase pair in pEVOL is under the control of an Ara_{BAD} promoter induced by Arabinose, and pEVOL confers resistance to chloramphenicol.

3.5.2 Standard recombinant expression of wild-type *Rho*1,2-CTD

Starter cultures were created by inoculating 10 mL of LB media (10 g/L Peptone, 5 g/L Yeast Extract, 5 g/L NaCl) containing appropriate antibiotic (100 µg/mL Ampicillin) with a frozen cell stock from the desired cell line (*E. coli* BL21DE3 strain cells containing pET31b+ plasmid with *Rho*1,2-CTD coding sequence) and grown overnight at 37 °C with shaking at 275 rpm. Enzyme expression was achieved as follows: A 250 mL Erlenmeyer flask with 50-100 mL of LB medium containing 100 µg/mL ampicillin was inoculated using a 1:100 dilution of the over night starter culture. The cells were grown at 37 °C at 275 rpm until OD_{600} reached 0.3-0.4, then the temperature was turned down to 18 °C and cells were grown until OD_{600} reached 0.6-0.7.

Expression was induced with 0.7 mM of IPTG, and cells were grown at 18 ℃ for 18-24 hours at 150 rpm. Aliquots of culture were harvested by centrifugation at 3250 ×g for 20 mins at 4 ℃. Pellets were stored at -20 ℃.

3.5.3 Standard recombinant expression of Nδ-methylhistidine incorporated 1,2-CTD

The cell lines used for Nδ-methylhistidine (NmH) incorporation contained two plasmids, the pET31b+ encoding the gene of interest (H220NmH or H222NmH) and the pEVOL encoding the orthogonal tRNA/tRNA synthetase pair.

Starter cultures were created by inoculating 10mL of LB media containing appropriate antibiotics (100µg/mL Ampicillin, 33µg/mL Chloramphenicol) with a frozen cell stock from the desired cell line and grown overnight at 37℃ with shaking at 275rpm. Both Amp and Chlora were used to ensure only cells containing both pET31b+ and pEVOL survived.

Enzyme expression was achieved as follows; A 250mL Erlenmeyer flask with 50-100mL of LB medium containing 100µg/mL Ampicillin and 33µg/mL Chloramphenicol was inoculated using a 1:100 dilution of starter culture. The cells were grown at 37℃ at 275 rpm until OD reached 0.3-0.4, then the temperature was turned down to 18℃ and cells were grown until OD_{600} reached 0.6-0.7. Expression of the orthogonal tRNA/tRNA synthetase set was induced with 0.2% arabinose and 2mM Nδ-methylhistidine added at the same time. The culture was left 1 hour and then expression of NmH-incorporated 1,2-CTD was induced with 0.1mM IPTG. Cells were grown at 18℃ for 18-24 hours at 275rpm. 40mL aliquots of culture were harvested by centrifugation at 3250g for 20-30mins at 4℃. Pellets were stored at -20℃.

3.5.4 Wild-type and variant recombinant 1,2-CTD purification

Pellets were thawed for 15 minutes on ice and resuspend in 4 mL of Lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 10 mM Imidazole, 2 mM $MgCl_2$, 5 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride (protease inhibitor), 3 U/mL Benzonase (Millipore Sigma), 0.5 mg/mL Lysozyme). Aliquots were transferred into 15 mL ~~Falcon~~ Conical tubes. Cells were lysed on ice by sonication (Sonifier Cell Disruptor 350, SmithKline) for 12-24 cycles, until the sample clarified (25 seconds at 30% duty and 40 output power, with 30 second cooling periods). The obtained lysate was centrifuged at 17,000 g for 30 mins at 4 ℃. Resulting pellets were stored at -20 ℃. *Rho*1,2-CTD was then purified by immobilised metal affinity chromatography using an Econo-Pac® Chromatography column (Bio-Rad) and Ni-NTA (Qiagen) resin. Ni-NTA slurry was added to the column and the column was equilibrated with 10 column volumes of lysis buffer. The supernatant from the cell lysate preparation was added to the column and the flowthrough was collected. The column was then washed with 1 column volume of wash buffer (10 mM Tris-HCl, 150 mM NaCl, 20 mM Imidazole (or 50 mM Imidazole as indicated), adjusted to pH 8.0) 3 times to collect washes 1-3. The protein was then eluted: 1 column volume of elution buffer (10 mM Tris-HCl, 150 mM NaCl, 250 mM Imidazole, adjusted to pH 8.0) was added, the column was inverted several times and incubated for 4 minutes before collecting the elution fraction. This step was repeated 3 times to collection elutions 1-3. The purified sample from elution 1 was applied to a 120 mL HiLoad™ 16/60 Superdex200 prep grade column (GE Healthcare Life Sciences) pre-equilibrated with gel filtration buffer (10 mM Tris-HCl. 150 mM NaCl, 5 mM DTT, adjusted to pH 8.0) using an AKTA FPLC system (Amersham Biosciences) with UNICORN™ 4.0 software, at a flow rate of 1 mL/min.

3.5.5 Michaelis-Menten kinetic analyses

Activity of *Rho*1,2-CTD was measured spectrophotometrically using a SpectraMax M5e spectrophotometer in a UV Star 96 well microplate (Greiner bio-one). The reaction mixture contained 50mM Tris, pH 7.2 and 3-methylcatechol. Eight different 3-methylcatechol concentrations were tested; 70µM, 35µM, 17.5µM, 8.75µM, 4.38µM, 2.19µM, 1.09µM, 0.54µM. The reaction was started by adding enzyme (1.5µg/mL in reaction). Specific rates of *Rho*1,2-CTD with 3-methylcatechol were determined using absorbance readings at 260 nm, taken every 15 seconds. The raw data was exported to GraphPad version 9.1 for analysis of kinetic parameters. For calculation of each parameter the data of 3 replicates were used for each substrate concentration tested to obtain the mean and standard error. The data was analyzed by fitting to the Michaelis-Menten equation using GraphPad. Apparent k_{cat} values were calculated of the basis of the molecular mass of the enzyme subunit (31740.73 Da). Protein concentrations were determined using the Nanodrop 2000c Spectrophotometer using the purified protein analysis function.

3.5.6 *In silico* analyses

In silico analyses were performed in PyMOL v2.5 using the crystal structure of Rho1,2-CTD in complex with 3-methylcatechol (PDB ID: 3HJQ[52]). Potential polar contacts were determined using build-in PyMOL commands and verified by visual inspection of the distances and angles. Methyl groups were added in using the builder function of PyMOL and potential clashes were determined by visual inspection of angles and distances between atoms.

Chapter 4
General Discussion and Future Directions

4.1 Summary of Chapter 2 and 3

Through the studies described in Chapter 2, several mutant *Rho*1,2-CTD enzymes was generated through disulfide bond engineering with significantly improved thermostability. In Chapter 3, attempts were made to improve the activity of Rho1,2-CTD to see if it was possible to compensate for the activity lost when stabilizing the enzyme. These attempts were unsuccessful.

4.2 Possible improvements to the disulfide prediction pipeline

4.2.1 Why the prioritizing selection criteria were chosen

In Chapter 2, a design pipeline was created to generate potential locations for disulfide bond engineering and then prioritize predicted designs for higher likelihood of success. To prioritize the designs a set of selection criteria hypothesized to select those with a higher likelihood of success was applied.

The factors that determine if an engineered disulfide bond will increase or decrease stability of a protein are not well characterized, the criteria developed in the disulfide engineering pipeline in Chapter 2 were developed from guidelines drawn from disulfide engineering experiments. The first criteria involved eliminating bonds comprised of residues less than 10 amino acids apart in the primary sequence, as previous studies have shown that this had the potential to produce structural conflicts in native secondary structure[134,135]. The second criteria involved eliminating bonds located close to the active site because while substitutions near the active site can stabilize a protein, they are also likely to disrupt catalysis and substrate binding[171]. The third criteria involved eliminating bonds made up of residues in the least flexible 25% of the protein as stabilizing disulfide mutations are most often found in regions of medium

to high mobility[149,172]. The fourth criteria involved using a computer program DynaMut2[136,149] to check for energetically unfavourable contacts with the modeled disulfide since these can considerably offset the stabilizing effects of the cross-link[149].

4.2.2 A possible issue with criteria 3 in the design pipeline

The authors of this study couldn't find anything to question the validity of criteria's 1, 2 and 4 of the developed pipeline, however there is some evidence that third criteria related to prioritizing more flexible regions of the protein may not be appropriate. This idea that engineered disulfide bonds linking regions of relatively high mobility are those most likely to confer stability to the protein seems to have been popularized because of a study by Dani et al.[149]. A meta-analysis reported that stabilizing disulfide mutations were most often found in regions of medium to high mobility, an idea which substantiates what was reported previously [172]. However, the idea that using disulfide bonds to link areas of high mobility is most likely to increase stability may not be entirely accurate[148]**.** This idea reflects limited collective knowledge of the physicochemical determinants of the stabilizing effect of a novel disulfide bridge.

Indeed, the fact that this idea is not entirely accurate is consistent with outcomes in Chapter 2 with the disulphide bond H85C-A252C. The bond H85C-A252C is located in a region of low flexibility (see Figure 2.14 in Chapter 2) and technically should have been rejected according to the pipeline developed in this study. However, it was predicted by the program DynaMut2 to have a strong stabilizing effect. Because of this, an exception was made to the pipeline and a mutant Rho1,2-CTD enzyme with this bond was created and tested. This bond proved to convey considerable stability to the enzyme, despite being in an area of low flexibility, likely due to the fact that it connected the two domains of the protein (dimerization and catalytic).

All this to say, while there is certainly an association observed between flexible regions and stabilizing disulfide bonds in many studies, there are exceptions to this idea. Therefore, prioritizing or limiting designs entirely to regions of high flexibility risks eliminating bonds that could confer improved stability.

4.2.3 Unfolding as a different possible criterion

A different criterion that could be used instead of flexibility, which has demonstrated great success in previous studies, is targeting areas of the protein that unfold first. To provide support for this idea some studies involving rational disulfide bond design that used the idea of prioritizing flexible regions of the protein[111,162] can be compared to studies which used the idea of targeting areas of the protein that unfold first[145,163] (Table 4.1). While this data set is limited, from this comparison, it can be seen that disulfide bond design strategies that targeted regions of the protein that unfolded first, yielded greater increases in stability compared to strategies that targeted flexible regions of the protein.

Table 4.1: Comparison of stability gained through different disulfide bond design rationales

Method guiding disulfide design	**Enzyme**	**Change in T_{50}**	**Change in $t_{1/2}$**	**Reference**
Rigidify Flexible sites	Cellulase C	Improved by 4.1 °C	5.8-fold increase at 65 °C	Badieyan et al. (2012)[111]
	Lipase (LipB)	Improved by 8.5 °C	4.5-fold increase at 50 °C	Le et al. (2012)[162]
Target unfolding regions	Barnase	---	19-fold increase in urea solution	Clarke and Fersht (1993)[145]
	Thermolysin-like protease	Improved by 16.7 °C	120-fold increase at 92.5 °C	Mansfeld et al (1997)[163]

An interesting side note is that all the papers on disulfide engineering that were published after the study by Dani et al.[149] prioritized engineering of disulfide bonds in areas of the protein that were more flexible. The author of this thesis has been unable to find studies published after

this date which prioritized engineering in areas of the protein that unfold first. While a full meta-analysis would be needed to investigate this, it raises the interesting possibility that prioritizing engineering of disulfides in areas with higher flexibility has become overly pervasive in the field, perhaps at the expense of more effective approaches such as identifying and prioritizing areas of the protein that unfold first in the unfolding pathway.

This begs the question: which regions of *Rho*1,2-CTD unfold first? There are a few interesting ways reported in literature such as: 1) elucidating the unfolding pathways by protein engineering techniques to determine folding intermediate and transition states[166–169], 2) simulating denaturation by diluting the network of salt bridges and hydrogen bonds [170] and 3) MD simulations[165].

However, the aim of Chapter 2 was to create a pipeline that was a simple, high throughput way to prioritize candidates. Therefore, perhaps the easiest way could be to target regions that tend to unfold first during protein denaturation; subunit interfaces, domain interfaces, N-termini and C-termini. Interestingly, the three bonds that conveyed the greatest increases in thermostability (A35C/Q211C, H85C-A252C, and K96C-D278C) are all located in one of these types of regions. A35C/Q211C is located at the subunit interface, H85C-A252C is located at the domain interface between the dimerization and catalytic domains, and K96C-D278C is located at the C-terminus.

With all this in mind, to improve the pipeline for future studies, the third criterion of prioritizing flexible sites of the protein could be replaced with a criterion that entails prioritizing bonds located at subunit interfaces, domain interfaces, N-termini, and C-termini.

4.3 Difficulties with GCE

Finally, toward improving catalysis in stabilized mutants, I investigated the potential of GCE technologies. Though GCE is a promising technology, currently it still comes with many challenges and, frankly, is impractical. Some of the more prominent issues include the following: truncation at the amber site and misreading of the amber codon by Gln, Tyr, and Trp tRNAs. Both of these, along with the requirement to work in minimal media, contribute to lowering target protein yields. Additionally, this technology is expensive, and incorporation must be verified, further adding to costs.

The first big challenge with GCE, limiting its industrial application, is that there are low yields, due to truncation at the active site. Stop codon readthrough methodology at the amber site has been used to design proteins for diverse applications. However, this method suffers from low yields of the modified protein, as the suppressor tRNA that recognises the stop codon is unable to compete effectively with release factor 1 (RF1), which terminates translation[174]. Further affecting yields is the fact that codons encoding Gln, Tyr, and Trp are all one base off from the amber codon (GAG, UAC/UAU, UGG vs UAG). Due to this, misreading of the amber codon is quite possible. Indeed, it has been noted that natural amino acids can be incorporated in response to the amber codon, apparently via misreading of Gln, Tyr, and Trp tRNAs[174].

Another challenge is that GCE incorporation must be validated. This requires use of technology such as mass spectrometry. The ability to validate incorporation is particularly difficult when the ncAA and the canonical amino acid have similar masses. In the case of histidine and NmH, there is only a mass difference of about 15 Da. In the context of a protein with a single-subunit mass of 31,740.73 Da, an expected mass difference of 15 Da is very hard to detect significantly and is below the sensitivity of most MS instruments. Thus, very specialized

MS instruments are needed. Colleagues of mine failed to detect whether NmH was present in a tannase enzyme despite working with two different MS labs at the National Research Council of Canada. Not to say it is impossible, others[126] were able to successfully verify incorporation of NmH by performing trypsin digestion and using an ESI-Q-TOF system coupled with an Agilent 1200 system to perform a MALDI-MS.

Due to all these issues, GCE is currently difficult to work with, expensive and labour-intensive. Non-canonical amino acids are expensive and since the yields are low, significantly more reagents are needed to produce the same amount of protein as you would get with a protein build solely with canonical amino acids. With current technology, a lab that would like to do GCE requires a large number of resources, meaning only-well funded labs are able to take the risk of using GCE. Currently, some of these issues are being tackled by the University of Oregon's GCE4All Centre. This centre aims to optimize, develop, and broadly disseminate GCE technologies so they can be more widely used. GCE has promising potential of unlocking new chemistries within proteins, so if these technologies can be improved, it certainly would be a big breakthrough. Currently however, until the many current issues with GCE technology are dealt with, this technology will remain broadly inaccessible and impractical for any real-world application.

4.4 Industrial sources of catechol

Finally, we come back to the enzyme 1,2-CTD and its potential relevance to industrial applications with respect to the production of ccMA. Studies have shown that a one-step enzymatic conversion of catechol to ccMA using 1,2-CTD is a promising strategy for achieving high yields of ccMA without the added headaches of various pre-treatment strategies and difficult recovery processes[175,176]. While this strategy eliminates many problems, it comes with a

looming question; what would the industrial source of catechol be? Up to date numbers were difficult to find, but current demand for, and production of catechol is low, with the worldwide consumption in 1990 being 20,000 t/a[176]. Catechol is currently manufactured together with hydroquinone by direct hydroxylation of phenol with peroxides. Presently, there are three plants worldwide using this process: *Rhône-Poulenc* in France, *Enichem* in Italy, and *Ube Industries* in Japan.

Despite this current lack of production, there are many potential sources of catechol that as of yet go unexploited. One of these potential sources is pulp mill effluent. According to a survey made under section 71 of CEPA 1999, a majority of the 1,000,000 to 10,000,000 kg of catechol produced in Canada in 2006 was generated as a by-product of kraft pulp production[177].

Another very promising potential source of catechol is through the valorization of lignin[178]. Lignocellulosic biomass is globally abundant, renewable, cheap, and doesn't enter in competition with the food industry[179]. Lignin also shares structural similarity to catechols, making it a promising potential feedstock for the production of bio-catechols. In fact, several aromatic polyols, such as cresols, catechols or resorcinols, that preserve the lignin monomer structure, can be obtained by different techniques such as catalytic depolymerization[180], hydrodeoxygenation[181], thermal treatments, catalytic hydrogenation, hydrocracking, oxidation or hydrolysis[182]. An economic analysis showed that lignin valorization in a biorefinery scheme could be competitive with traditional production methods[183].

Finally, it is important to note the prices of catechol compared to ccMA. In particular the prices from AK Scientific are $15.00/10 g of catechol, and $249.00/10 g of ccMA. According to these quotes, the price of ccMA is more than 16X that of catechol, and this doesn't even factor in that the molecular weight of ccMA is greater than that of catechol. Thus, in conclusion, although

the current production of catechol is small, it is anticipated to grow based on the current emphasize on valorizing lignocellulosic biomass. This, it should still be noted that the thermostable enzymes developed in this study still have the potential to be employed profitably immediately.

Bibliography

1. Vaillancourt, F. H., Bolin, J. T. & Eltis, L. D. The Ins and Outs of Ring-Cleaving Dioxygenases. *Crit Rev Biochem Mol Biol* **41**, 241–267 (2006).

2. Pérez-Pantoja, D., González, B. & Pieper, D. H. Aerobic Degradation of Aromatic Hydrocarbons. in *Handbook of Hydrocarbon and Lipid Microbiology* (ed. Timmis, K. N.) 799–837 (Springer Berlin Heidelberg, 2010). doi:10.1007/978-3-540-77587-4_60.

3. Adler, E. Lignin chemistry—past, present and future. *Wood Sci Technol* **11**, 169–218 (1977).

4. PubChem Compound Summary for CID 289, Catechol. *National Center for Biotechnology Information* (2023).

5. Aggelis, G. *et al.* Evaluation of white-rot fungi for detoxification and decolorization of effluents from the green olive debittering process. *Appl Microbiol Biotechnol* **59**, 353–360 (2002).

6. Field, J. A. & Lettinga, G. Treatment and Detoxification of Aqueous Spruce Bark Extracts by Aspergillus niger. *Water Science and Technology* **24**, 127–137 (1991).

7. Mahadewswamy, M., Mall, I. D., Prasad, B. & Mishra, I. M. Removal of phenol by adsorption of coal fly ash and activated carbon. *Pollut Res* **16**, 107–175 (1997).

8. Mason, J. R. & Cammack, R. THE ELECTRON-TRANSPORT PROTEINS OF HYDROXYLATING BACTERIAL DIOXYGENASES. *Annu Rev Microbiol* **46**, 277–305 (1992).

9. Gibson, D. T., Koch, J. R. & Kallio, R. E. Oxidative degradation of aromatic hydrocarbons by microorganisms. I. Enzymic formation of catechol from benzene. *Biochemistry* **7**, 2653–2662 (1968).

10. Murray, K., Duggleby, C. J., Williams, P. A. & Sala-Trepat, J. M. The Metabolism of Benzoate and Methylbenzoates via the meta-Cleavage Pathway by Pseudomonas arvilla mt-2. *Eur J Biochem* **28**, 301–310 (1972).

11. Hughes, E. J. & Bayly, R. C. Control of catechol meta-cleavage pathway in Alcaligenes eutrophus. *J Bacteriol* **154**, 1363–1370 (1983).

12. Catelani, D., Colombi, A., Sorlini, C. & Treccani, V. Metabolism of biphenyl. 2-Hydroxy-6-oxo-6-phenylhexa-2,4-dienoate: the meta-cleavage product from 2,3-dihydroxybiphenyl by Pseudomonas putida. *Biochemical Journal* **134**, 1063–1066 (1973).

13. Jeffrey, A. M. *et al.* Initial reactions in the oxidation of naphthalene by Pseudomonas putida. *Biochemistry* **14**, 575–584 (1975).

14. Crawford, R. L. Pathways of 4-hydroxybenzoate degradation among species of Bacillus. *J Bacteriol* **127**, 204–210 (1976).

15. Mashetty, S. B., Manohar, S. & Karegoudar, T. B. Degradation of 3-hydroxybenzoic acid by a Bacillus species. *Indian journal of biochemistry & biophysics* **33**, 145–148 (1996).

16. Keyser, P., Pujar, B. G., Eaton, R. W. & Ribbons, D. W. Biodegradation of the phthalates and their esters by bacteria. *Environ Health Perspect* **18**, 159–166 (1976).

17. Mattevi, A. *et al.* Crystal structures and inhibitor binding in the octameric flavoenzyme vanillyl-alcohol oxidase: the shape of the active-site cavity controls substrate specificity. *Structure* **5**, 907–920 (1997).

18. Priefert, H., Rabenhorst, J. & Steinbüchel, A. Molecular characterization of genes of Pseudomonas sp. strain HR199 involved in bioconversion of vanillin to protocatechuate. *J Bacteriol* **179**, 2595–2607 (1997).

19. Grund, E., Denecke, B. & Eichenlaub, R. Naphthalene degradation via salicylate and gentisate by Rhodococcus sp. strain B4. *Appl Environ Microbiol* **58**, 1874–1877 (1992).

20. Hopper, D. J. & Taylor, D. G. Pathways for the degradation of m-cresol and p-cresol by Pseudomonas putida. *J Bacteriol* **122**, 1–6 (1975).

21. Cain, R. B. Anthranilic acid metabolism by microorganisms. Formation of 5-hydroxyanthranilate as an intermediate in anthranilate metabolism byNocardia opaca. *Antonie Van Leeuwenhoek* **34**, 417–432 (1968).

22. Jain, R. K., Dreisbach, J. H. & Spain, J. C. Biodegradation of p-nitrophenol via 1,2,4-benzenetriol by an Arthrobacter sp. *Appl Environ Microbiol* **60**, 3030–3032 (1994).

23. Gaal, A. & Neujahr, H. Y. Metabolism of phenol and resorcinol in Trichosporon cutaneum. *J Bacteriol* **137**, 13–21 (1979).

24. Karolina, N., Maria, U. & K, J. J. Novel 4-Chlorophenol Degradation Gene Cluster and Degradation Route via Hydroxyquinol in Arthrobacter chlorophenolicus A6. *Appl Environ Microbiol* **71**, 6538–6544 (2005).

25. Jean, A., N, T. K. & Rolf-Michael, W. A Functional 4-Hydroxysalicylate/Hydroxyquinol Degradative Pathway Gene Cluster Is Linked to the Initial Dibenzo-p-Dioxin Pathway Genes inSphingomonas sp. Strain RW1. *J Bacteriol* **181**, 3452–3461 (1999).

26. Zaborina, O., Latus, M., Eberspächer, J., Golovleva, L. A. & Lingens, F. Purification and characterization of 6-chlorohydroxyquinol 1,2-dioxygenase from Streptomyces rochei 303: comparison with an analogous enzyme from Azotobacter sp. strain GP1. *J Bacteriol* **177**, 229–234 (1995).

27. Sikkema, J., A, de B. J. & Poolman, B. Mechanisms of membrane toxicity of hydrocarbons. *Microbiol Rev* **59**, 201–222 (1995).

28. Subramanyam, R. & Mishra, I. M. Treatment of catechol bearing wastewater in an upflow anaerobic sludge blanket (UASB) reactor: Sludge characteristics. *Bioresour Technol* **99**, 8917–8925 (2008).

29. Subramanyam, R. & Mishra, I. M. Co-degradation of resorcinol and catechol in an UASB reactor. *Bioresour Technol* **99**, 4147–4157 (2008).

30. Subramanyam, R. & Mishra, I. M. Biodegradation of catechol (2-hydroxy phenol) bearing wastewater in an UASB reactor. *Chemosphere* **69**, 816–824 (2007).

31. Fedorak, P. M. & Hrudey, S. E. Anaerobic degradation of phenolic compounds with application to treatment of industrial waste waters. in *Biotreatment Systems* (ed. Wise, D. L.) 170–212 (CRC Press, 1988).

32. Calabrese, E. J. & Kenyon, E. M. *Air Toxics and Risk Assessments*. (Lewis Publishers, Inc., 1991).

33. Moussavi, G., Barikbin, B. & Mahmoudi, M. The removal of high concentrations of phenol from saline wastewater using aerobic granular SBR. *Chemical Engineering Journal* **158**, 498–504 (2010).

34. Lofrano, G., Rizzo, L., Grassi, M. & Belgiorno, V. Advanced oxidation of catechol: A comparison among photocatalysis, Fenton and photo-Fenton processes. *Desalination* **249**, 878–883 (2009).

35. Kumar, A., Kumar, S. & Kumar, S. Biodegradation kinetics of phenol and catechol using Pseudomonas putida MTCC 1194. *Biochem Eng J* **22**, 151–159 (2005).

36. Rigo, M., Alegre, R. M., Bezerra, J. R. M. V., Coelho, N. & Bastos, R. G. Catechol biodegradation kinetics using Candida parapsilopsis. *Brazilian Archives of Biology and Technology* **53**, 481–486 (2010).

37. Mehrizad, A. *et al.* Comparison of 4-chloro-2-nitrophenol adsorption on single-walled and multi-walled carbon nanotubes. *Iranian J Environ Health Sci Eng* **9**, 5 (2012).

38. Xie, N.-Z., Liang, H., Huang, R.-B. & Xu, P. Biotechnological production of muconic acid: current status and future prospects. *Biotechnol Adv* **32**, 615–622 (2014).

39. Harayama, S. & Rekik, M. Bacterial aromatic ring-cleavage enzymes are classified into two different gene families*. *Journal of Biological Chemistry* **264**, 15328–15333 (1989).

40. Guzik, U., Greń, I., Hupert-Kocurek, K. & Wojcieszyńska, D. Catechol 1,2-dioxygenase from the new aromatic compounds – Degrading Pseudomonas putida strain N6. *Int Biodeterior Biodegradation* **65**, 504–512 (2011).

41. Latus, M., Seitz, H., Eberspacher, J. & Lingens, F. Purification and Characterization of Hydroxyquinol 1,2-Dioxygenase from Azotobacter sp. Strain GP1. *Appl Environ Microbiol* **61**, 2453–2460 (1995).

42. Costas, M., Mehn, M. P., Jensen, M. P. & Que, L. Dioxygen Activation at Mononuclear Nonheme Iron Active Sites: Enzymes, Models, and Intermediates. *Chem Rev* **104**, 939–986 (2004).

43. Karegoudar, T. B. & Kim, C.-K. Microbial Degradation of Monohydroxybenzoic Acids. *The Journal of Microbiology* **38**, 53–61 (2000).

44. Mahiudddin, Md., Fakhruddin, A. N. M. & Abdullah-Al-Mahin. Degradation of Phenol via Meta Cleavage Pathway by *Pseudomonas fluorescens* PU1. *ISRN Microbiol* **2012**, 741820 (2012).

45. Vetting, M. W. & Ohlendorf, D. H. The 1.8 Å crystal structure of catechol 1,2-dioxygenase reveals a novel hydrophobic helical zipper as a subunit linker. *Structure* **8**, 429–440 (2000).

46. Daubaras, D. L., Hershberger, C. D., Kitano, K. & Chakrabarty, A. M. Sequence analysis of a gene cluster involved in metabolism of 2,4,5-trichlorophenoxyacetic acid by Burkholderia cepacia AC1100. *Appl Environ Microbiol* **61**, 1279–1289 (1995).

47. Eltis, L. D. & Bolin, J. T. Evolutionary relationships among extradiol dioxygenases. *J Bacteriol* **178**, 5930–5937 (1996).

48. Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R. & Goodenough, P. W. Evolution of functional diversity in the cupin superfamily. *Trends Biochem Sci* **26**, 740–746 (2001).

49. Armstrong, R. N. Mechanistic Diversity in a Metalloenzyme Superfamily. *Biochemistry* **39**, 13625–13632 (2000).

50. Gerlt, J. A. & Babbitt, P. C. Divergent Evolution of Enzymatic Function: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies. *Annu Rev Biochem* **70**, 209–246 (2001).

51. Ohlendorf, D. H., Lipscomb, J. D. & Weber, P. C. Structure and assembly of protocatechuate 3,4-dioxygenase. *Nature* **336**, 403–405 (1988).

52. Matera, I. *et al.* Catechol 1,2-dioxygenase from the Gram-positive Rhodococcus opacus 1CP: Quantitative structure/activity relationship and the crystal structures of native enzyme and catechols adducts. *J Struct Biol* **170**, 548–564 (2010).

53. Ferraroni, M. *et al.* Crystal Structure of 3-Chlorocatechol 1,2-dioxygenase Key Enzyme of a New Modified Ortho-pathway from the Gram-positive Rhodococcus opacus 1CP Grown on 2-chlorophenol. *J Mol Biol* **360**, 788–799 (2006).

54. Ferraroni, M. *et al.* Crystal Structure of the Hydroxyquinol 1,2-Dioxygenase from Nocardioides simplex 3E, a Key Enzyme Involved in Polychlorinated Aromatics Biodegradation*. *Journal of Biological Chemistry* **280**, 21144–21154 (2005).

55. Han, S., Eltis, L. D., Timmis, K. N., Muchmore, S. W. & Bolin, J. T. Crystal Structure of the Biphenyl-Cleaving Extradiol Dioxygenase from a PCB-Degrading Pseudomonad. *Science (1979)* **270**, 976–980 (1995).

56. Kita, A. *et al.* An archetypical extradiol-cleaving catecholic dioxygenase: the crystal structure of catechol 2,3-dioxygenase (metapyrocatechase) from Pseudomonas putida mt-2. *Structure* **7**, 25–34 (1999).

57. Vetting, M. W., Wackett, L. P., Lawrence, Q., D, L. J. & H, O. D. Crystallographic Comparison of Manganese- and Iron-Dependent Homoprotocatechuate 2,3-Dioxygenases. *J Bacteriol* **186**, 1945–1958 (2004).

58. Asturias, J. A., Eltis, L. D., Prucha, M. & Timmis, K. N. Analysis of three 2,3-dihydroxybiphenyl 1,2-dioxygenases found in Rhodococcus globerulus P6. Identification of a new family of extradiol dioxygenases. *Journal of Biological Chemistry* **269**, 7807–7815 (1994).

59. Bugg, T. D. H. Overproduction, purification and properties of 2,3-dihydroxyphenylpropionate 1,2-dioxygenase from Escherichiacoli. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1202**, 258–264 (1993).

60. Nogales, J., Canales, Á., Jiménez-Barbero, J., García, J. L. & Díaz, E. Molecular Characterization of the Gallate Dioxygenase from Pseudomonas putida KT2440: THE PROTOTYPE OF A NEW SUBGROUP OF EXTRADIOL DIOXYGENASES*. *Journal of Biological Chemistry* **280**, 35382–35390 (2005).

61. Sugimoto, K. *et al.* Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions. *Structure* **7**, 953–965 (1999).

62. Mampel, J., Providenti, M. A. & Cook, A. M. Protocatechuate 4,5-dioxygenase from Comamonas testosteroni T-2: biochemical and molecular properties of a new subgroup within class III of extradiol dioxygenases. *Arch Microbiol* **183**, 130–139 (2005).

63. Davis, J. K., He, Z., Somerville, C. C. & Spain, J. C. Genetic and biochemical comparison of 2-aminophenol 1,6-dioxygenase of Pseudomonas pseudoalcaligenes JS45 to meta-cleavage dioxygenases: divergent evolution of 2-aminophenol meta-cleavage pathway. *Arch Microbiol* **172**, 330–339 (1999).

64. Harpel, M. R. & Lipscomb, J. D. Gentisate 1,2-dioxygenase from pseudomonas. Purification, characterization, and comparison of the enzymes from Pseudomonas testosteroni and Pseudomonas acidovorans. *Journal of Biological Chemistry* **265**, 6301–6311 (1990).

65. Iwabuchi, T. & Harayama, S. Biochemical and Molecular Characterization of 1-Hydroxy-2-naphthoate Dioxygenase from Nocardioides sp. KP7*. *Journal of Biological Chemistry* **273**, 8332–8336 (1998).

66. Chatterjee, A., DeLorenzo, D. M., Carr, R. & Moon, T. S. Bioconversion of renewable feedstocks by Rhodococcus opacus. *Curr Opin Biotechnol* **64**, 10–16 (2020).

67. Jang, H. G., Cox, D. D. & Que, L. Jr. A highly reactive functional model for the catechol dioxygenases. Structure and properties of [Fe(TPA)DBC]BPh4. *J Am Chem Soc* **113**, 9200–9204 (1991).

68. Eley, K. L., Crowley, P. J. & Bugg, T. D. H. A Solvolytic C−C Cleavage Reaction of 6-Acetoxycyclohexa-2,4-dienones: Mechanistic Implications for the Intradiol Catechol Dioxygenases. *J Org Chem* **66**, 2091–2097 (2001).

69. Schrodinger LLC. The PyMOL Molecular Graphics System, Version 1.8. Preprint at (2015).

70. Earhart, C. A. *et al.* Structure of catechol 1,2-dioxygenase from Pseudomonas arvilla. *Biochem Biophys Res Commun* **338**, 198–205 (2005).

71. Ucun, H., Yildiz, E. & Nuhoglu, A. Phenol biodegradation in a batch jet loop bioreactor (JLB): Kinetics study and pH variation. *Bioresour Technol* **101**, 2965–2971 (2010).

72. Moussavi, G. & Heidarizad, M. The performance of SBR, SCR, and MSCR for simultaneous biodegradation of high concentrations of formaldehyde and ammonia. *Sep Purif Technol* **77**, 187–195 (2011).

73. Sarfaraz, S., Thomas, S., Tewari, U. K. & Iyengar, L. Anoxic treatment of phenolic wastewater in sequencing batch reactor. *Water Res* **38**, 965–971 (2004).

74. Shumkova, E. S., Solyanikova, I. P., Plotnikova, E. G. & Golovleva, L. A. Phenol degradation by Rhodococcus opacus strain 1G. *Appl Biochem Microbiol* **45**, 43–49 (2009).

75. Silva, A. S. *et al.* Enzymatic activity of catechol 1,2-dioxygenase and catechol 2,3-dioxygenase produced by Gordonia polyisoprenivorans. *Quim Nova* **35**, 1587–1592 (2012).

76. Suvilampi, J., Lepistö, R. & Rintala, J. Biological treatment of pulp and paper mill process and wastewaters under thermophilic conditions - A review. *Paperi ja Puu/Paper and Timber* **83**, 320–325 (2001).

77. Thiemens, M. H. & Trogler, W. C. Nylon Production: An Unknown Source of Atmospheric Nitrous Oxide. *Science (1979)* **251**, 932–934 (1991).

78. Li, M., Ruddy, T., Fahey, D., Busch, D. H. & Subramaniam, B. Terephthalic Acid Production via Greener Spray Process: Comparative Economic and Environmental Impact Assessments with Mid-Century Process. *ACS Sustain Chem Eng* **2**, 823–835 (2014).

79. Boustead, I. *Eco-profiles of the European plastics industry.* (2005).

80. *Market volume of purified terephthalic acid worldwide from 2015 to 2022, with a forecast for 2023 to 2030*. (2023).

81. *Canadian Environmental Sustainability Indicators: Greenhouse gas emissions*. (2023).

82. Warhurst, A. M., Clarke, K. F., Hill, R. A., Holt, R. A. & Fewson, C. A. Production of catechols and muconic acids from various aromatics by the styrene-degraderRhodococcus rhodochrous NCIMB 13259. *Biotechnol Lett* **16**, 513–516 (1994).

83. Han, L. *et al.* Engineering catechol 1, 2-dioxygenase by design for improving the performance of the cis, cis-muconic acid synthetic pathway in Escherichia coli. *Sci Rep* **5**, 13435 (2015).

84. Becker, J., Kuhl, M., Kohlstedt, M., Starck, S. & Wittmann, C. Metabolic engineering of Corynebacterium glutamicum for the production of cis, cis-muconic acid from lignin. *Microb Cell Fact* **17**, 115 (2018).

85. Salvachúa, D. *et al.* Bioprocess development for muconic acid production from aromatic compounds and lignin. *Green Chemistry* **20**, 5007–5019 (2018).

86. Xie, N.-Z. *et al.* OPTIMIZATION OF MEDIUM COMPOSITION FOR cis,cis-MUCONIC ACID PRODUCTION BY A Pseudomonas sp. MUTANT USING STATISTICAL METHODS. *Prep Biochem Biotechnol* **44**, 342–354 (2014).

87. Wang, S. *et al.* Development of a Plasmid-Free Biosynthetic Pathway for Enhanced Muconic Acid Production in Pseudomonas chlororaphis HT66. *ACS Synth Biol* **7**, 1131–1142 (2018).

88. Choi, S., Lee, H.-N., Park, E., Lee, S.-J. & Kim, E.-S. Recent Advances in Microbial Production of cis,cis-Muconic Acid. *Biomolecules* **10**, (2020).

89. Vardon, D. R. *et al.* cis,cis-Muconic acid: separation and catalysis to bio-adipic acid for nylon-6,6 polymerization. *Green Chem.* **18**, 3397–3413 (2016).

90. Jemli, S., Ayadi-Zouari, D., Hlima, H. Ben & Bejar, S. Biocatalysts: application and engineering for industrial purposes. *Crit Rev Biotechnol* **36**, 246–258 (2016).

91. Choi, J.-M., Han, S.-S. & Kim, H.-S. Industrial applications of enzyme biocatalysis: Current status and future aspects. *Biotechnol Adv* **33**, 1443–1454 (2015).

92. Gurung, N., Ray, S., Bose, S. & Rai, V. A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond. *Biomed Res Int* **2013**, 329121 (2013).

93. Singh, R., Kumar, M., Mittal, A. & Mehta, P. K. Microbial enzymes: industrial progress in 21st century. *3 Biotech* **6**, 174 (2016).

94. Kamini, N. R., Hemachander, C., Mala, J. G. S. & Puvanakrishnan, R. Microbial enzyme technology as an alternative to conventional chemicals in leather industry. *Curr Sci* **77**, 80–86 (1999).

95. Solano, F. Enzyme Engineering: Old and New Approaches. *Enzyme Engineering* **04**, (2014).

96. Basheer, S. M. & Chellappan, S. Enzyme Engineering. in *Bioresources and Bioprocess in Biotechnology: Volume 2 : Exploring Potential Biomolecules* (eds. Sugathan, S., Pradeep, N. S. & Abdulhameed, S.) 151–168 (Springer Singapore, 2017). doi:10.1007/978-981-10-4284-3_6.

97. Smith, M. Synthetic DNA and Biology. *Biosci Rep* **14**, 51–66 (1994).

98. Wohlgemuth, R. Industrial biotechnology – past, present and future. *N Biotechnol* **29**, 165 (2012).

99. Chen, K. & Arnold, F. H. Enzyme Engineering for Nonaqueous Solvents: Random Mutagenesis to Enhance Activity of Subtilisin E in Polar Organic Media. *Bio/Technology* **9**, 1073–1077 (1991).

100. Tracewell, C. A. & Arnold, F. H. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol* **13**, 3–9 (2009).

101. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr Opin Biotechnol* **16**, 378–384 (2005).

102. Reetz, M. T. Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *J Am Chem Soc* **135**, 12480–12496 (2013).

103. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).

104. Zhang, L., Chou, C. P. & Moo-Young, M. Disulfide bond formation and its impact on the biological activity and stability of recombinant therapeutic proteins produced by Escherichia coli expression system. *Biotechnol Adv* **29**, 923–929 (2011).

105. Landeta, C., Boyd, D. & Beckwith, J. Disulfide bond formation in prokaryotes. *Nat Microbiol* **3**, 270–280 (2018).

106. Karimi, M. *et al.* Reactivity of disulfide bonds is markedly affected by structure and environment: implications for protein modification and stability. *Sci Rep* **6**, 38572 (2016).

107. Roesler, K. R. & Rao, A. G. A single disulfide bond restores thermodynamic and proteolytic stability to an extensively mutated protein. *Protein Science* **9**, 1642–1650 (2000).

108. Hogg, P. J. Disulfide bonds as switches for protein function. *Trends Biochem Sci* **28**, 210–214 (2003).

109. Abkevich, V. I. & Shakhnovich, E. I. What can Disulfide Bonds Tell Us about Protein Energetics, Function and Folding: Simulations and Bioninformatics Analysis. *J Mol Biol* **300**, 975–985 (2000).

110. Long, L. *et al.* In Silico Rational Design and Systems Engineering of Disulfide Bridges in the Catalytic Domain of an Alkaline α-Amylase from Alkalimonas amylolytica To Improve Thermostability. *Appl Environ Microbiol* **80**, 798–807 (2014).

111. Badieyan, S., Bevan, D. R. & Zhang, C. Study and design of stability in GH5 cellulases. *Biotechnol Bioeng* **109**, 31–44 (2012).

112. Siadat, O. R., Lougarre, A., Lamouroux, L., Ladurantie, C. & Fournier, D. The effect of engineered disulfide bonds on the stability of Drosophila melanogaster acetylcholinesterase. *BMC Biochem* **7**, 12 (2006).

113. Wells, J. A. & Powers, D. B. In vivo formation and stability of engineered disulfide bonds in subtilisin. *Journal of Biological Chemistry* **261**, 6564–6570 (1986).

114. Wetzel, R., Perry, L. J., Baase, W. A. & Becktel, W. J. Disulfide bonds and thermal stability in T4 lysozyme. *Proceedings of the National Academy of Sciences* **85**, 401–405 (1988).

115. Suplatov, D., Timonina, D., Sharapova, Y. & Švedas, V. Yosshi: a web-server for disulfide engineering by bioinformatic analysis of diverse protein families. *Nucleic Acids Res* **47**, W308–W314 (2019).

116. Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* **14**, 346 (2013).

117. Gao, X., Dong, X., Li, X., Liu, Z. & Liu, H. Prediction of disulfide bond engineering sites using a machine learning method. *Sci Rep* **10**, 10330 (2020).

118. Kazlauskas, R. Engineering more stable proteins. *Chem Soc Rev* **47**, 9026–9045 (2018).

119. Fields, P. A. Review: Protein function at thermal extremes: balancing stability and flexibility. *Comp Biochem Physiol A Mol Integr Physiol* **129**, 417–431 (2001).

120. PW, H. & Somero, G. *Bio-Chemical Adaptation: Mechanism and Process in Physiological Evolution*. vol. 30 (2002).

121. Sterner, R. hard & Liebl, W. Thermophilic Adaptation of Proteins. *Crit Rev Biochem Mol Biol* **36**, 39–106 (2001).

122. Feller, G. Protein stability and enzyme activity at extreme biological temperatures. *Journal of Physics: Condensed Matter* **22**, 323101 (2010).

123. Pott, M. *et al.* A Noncanonical Proximal Heme Ligand Affords an Efficient Peroxidase in a Globin Fold. *J Am Chem Soc* **140**, 1535–1543 (2018).

124. Burke, A. J. *et al.* Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **570**, 219–223 (2019).

125. Guo, Q. & Cao, Y. J. Applications of genetic code expansion technology in eukaryotes. *Protein Cell* pwad051 (2023) doi:10.1093/procel/pwad051.

126. Green, A. P., Hayashi, T., Mittl, P. R. E. & Hilvert, D. A Chemically Programmed Proximal Ligand Enhances the Catalytic Properties of a Heme Enzyme. *J Am Chem Soc* **138**, 11344–11352 (2016).

127. Rorrer, N. A. *et al.* Renewable Unsaturated Polyesters from Muconic Acid. *ACS Sustain Chem Eng* **4**, 6867–6876 (2016).

128. Pandell, A. Enzymic-Like Aromatic Oxidations. Metal-Catalyzed Peracetic Acid Oxidation of Phenol and Catechol to cis,cis-Muconic Acid1. *J. Org. Chem.* **41**, 3992–3996 (1976).

129. Tsuji, J. & Takayanagi, H. Oxidative cleavage reactions of catechol and phenol to monoester of cis,cis,-muconic acid with the oxidizing systems of O2/CuCl, KOH/CuCl2. *Tetrahedron* **34**, 641–644 (1978).

130. Bolm, C., Beckmann, O. & Dabard, O. A. G. The Search for New Environmentally Friendly Chemical Processes. *Angewandte Chemie International Edition* **38**, 907–909 (1999).

131. Cavani, F. & Alini, S. Synthesis of adipic acid: on the way to more sustainable production. *Sustainable industrial chemistry* 367–425 (2009).

132. Van De Vyver, S. & Román-Leshkov, Y. Emerging catalytic processes for the production of adipic acid. *Catal Sci Technol* **3**, 1465–1479 (2013).

133. Niu, C., Zhu, L., Xu, X. & Li, Q. Rational Design of Disulfide Bonds Increases Thermostability of a Mesophilic 1,3-1,4-β-Glucanase from Bacillus terquilensis. *PLoS One* **11**, e0154036- (2016).

134. Betz, S. F. Disulfide bonds and the stability of globular proteins. *Protein Science* vol. 2 1551–1558 Preprint at https://doi.org/10.1002/pro.5560021002 (1993).

135. Petersen, M. T. N., Jonson, P. H. & Petersen, S. B. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein Engineering, Design and Selection* **12**, 535–548 (1999).

136. Rodrigues, C. H. M., Pires, D. E. V & Ascher, D. B. DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Science* **30**, 60–69 (2021).

137. Sanchez-Ruiz, J. M. Protein kinetic stability. *Biophys Chem* **148**, 1–15 (2010).

138. Colón, W. *et al.* Biological Roles of Protein Kinetic Stability. *Biochemistry* **56**, 6179–6186 (2017).

139. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212–2221 (2007).

140. Hanson, A. D. *et al.* The number of catalytic cycles in an enzyme's lifetime and why it matters to metabolic engineering. *Proceedings of the National Academy of Sciences* **118**, (2021).

141. Rogers, T. A. & Bommarius, A. S. Utilizing simple biochemical measurements to predict lifetime output of biocatalysts in continuous isothermal processes. *Chem Eng Sci* **65**, 2118–2124 (2010).

142. Han, Z., Han, S., Zheng, S. & Lin, Y. Enhancing thermostability of a Rhizomucor miehei lipase by engineering a disulfide bond and displaying on the yeast cell surface. *Appl Microbiol Biotechnol* **85**, 117–126 (2009).

143. Hwa, K.-Y., Subramani, B., Shen, S.-T. & Lee, Y.-M. An intermolecular disulfide bond is required for thermostability and thermoactivity of β-glycosidase from Thermococcus kodakarensis KOD1. *Appl Microbiol Biotechnol* **98**, 7825–7836 (2014).

144. Yin, X. *et al.* Contribution of Disulfide Bridges to the Thermostability of a Type A Feruloyl Esterase from Aspergillus usamii. *PLoS One* **10**, e0126864- (2015).

145. Clarke, J. & Fersht, A. R. Engineered disulfide bonds as probes of the folding pathway of barnase: Increasing the stability of proteins against the rate of denaturation. *Biochemistry* **32**, 4322–4329 (1993).

146. Eijsink, V. G. H., Vriend, G. & Van Den Burg, B. Engineering a Hyperstable Enzyme by Manipulation of Early Steps in the Unfolding Process. *Biocatal Biotransformation* **19**, 443–458 (2001).

147. Wedemeyer, W. J., Welker, E., Narayan, M. & Scheraga, H. A. Disulfide Bonds and Protein Folding. *Biochemistry* **39**, 4207–4216 (2000).

148. Dombkowski, A. A., Sultana, K. Z. & Craig, D. B. Protein disulfide engineering. *FEBS Lett* **588**, 206–212 (2014).

149. Dani, V. S., Ramakrishnan, C. & Varadarajan, R. MODIP revisited: re-evaluation and refinement of an automated procedure for modeling of disulfide bonds in proteins. *Protein Engineering, Design and Selection* **16**, 187–193 (2003).

150. Michaelis, L. & Menten, M. L. Die Kinetik der Invertinwirkung. *Biochem Z* **49**, 333–369 (1913).

151. Xie, Y. *et al.* Enhanced Enzyme Kinetic Stability by Increasing Rigidity within the Active Site*. *Journal of Biological Chemistry* **289**, 7994–8006 (2014).

152. Ellman, G. L. Tissue sulfhydryl groups. *Arch Biochem Biophys* **82**, 70–77 (1959).

153. NISHIZUKA, Y., ICHIYAMA, A., NAKAMURA, S. & HAYAISHI, O. A new metabolic pathway of catechol. *J Biol Chem* **237**, (1962).

154. Harayama, S., Kok, M. & Neidle, E. L. FUNCTIONAL AND EVOLUTIONARY RELATIONSHIPS AMONG DIVERSE OXYGENASES. *Annu Rev Microbiol* **46**, 565–601 (1992).

155. Que, L. Jr. & Ho, R. Y. N. Dioxygen Activation by Enzymes with Mononuclear Non-Heme Iron Active Sites. *Chem Rev* **96**, 2607–2624 (1996).

156. Agostini, F. *et al.* Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angewandte Chemie International Edition* **56**, 9680–9703 (2017).

157. Xiao, H. *et al.* Genetic Incorporation of Histidine Derivatives Using an Engineered Pyrrolysyl-tRNA Synthetase. *ACS Chem Biol* **9**, 1092–1096 (2014).

158. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J Mol Biol* **395**, 361–374 (2010).

159. Meng, L. *et al.* Effects of His-tag on Catalytic Activity and Enantioselectivity of Recombinant Transaminases. *Appl Biochem Biotechnol* **190**, 880–895 (2020).

160. Guo, S., Boyd, J., Sammynaiken, R. & Loewen, M. Identification and characterization of a unique cysteine residue proximal to the catalytic site of Arabidopsis thaliana carotenoid cleavage enzyme 1. *Biochem Cell Biol* **86**, 262–270 (2008).

161. Chin, J. W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu Rev Biochem* **83**, 379–408 (2014).

162. Le, Q. A. T., Joo, J. C., Yoo, Y. J. & Kim, Y. H. Development of thermostable Candida antarctica lipase B through novel in silico design of disulfide bridge. *Biotechnol Bioeng* **109**, 867–876 (2012).

163. Mansfeld, J. *et al.* Extreme Stabilization of a Thermolysin-like Protease by an Engineered Disulfide Bond*. *Journal of Biological Chemistry* **272**, 11152–11156 (1997).

164. Benson, N. C. & Daggett, V. Dynameomics: Large-scale assessment of native protein flexibility. *Protein Science* **17**, 2038–2050 (2008).

165. Pikkemaat, M. G., Linssen, A. B. M., Berendsen, H. J. C. & Janssen, D. B. Molecular dynamics simulations as a tool for improving protein stability. *Protein Engineering, Design and Selection* **15**, 185–192 (2002).

166. Fersht, A. R., Matouschek, A. & Serrano, L. The folding of an enzyme: I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* **224**, 771–782 (1992).

167. Serrano, L., Kellis, J. T., Cann, P., Matouschek, A. & Fersht, A. R. The folding of an enzyme: II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* **224**, 783–804 (1992).

168. Serrano, L., Matouschek, A. & Fersht, A. R. The folding of an enzyme: III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J Mol Biol* **224**, 805–818 (1992).

169. Matouschek, A., Serrano, L. & Fersht, A. R. The folding of an enzyme: IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J Mol Biol* **224**, 819–835 (1992).

170. Hespenheide, B. M., Rader, A. J., Thorpe, M. F. & Kuhn, L. A. Identifying protein folding cores from the evolution of flexible regions during unfolding. *J Mol Graph Model* **21**, 195–207 (2002).

171. Jones, B. J., Kan, C. N. E., Luo, C. & Kazlauskas, R. J. Chapter Six - Consensus Finder web tool to predict stabilizing substitutions in proteins. in *Methods in Enzymology* (ed. Tawfik, D. S.) vol. 643 129–148 (Academic Press, 2020).

172. Matsumura, M., Becktel, W. J., Levitt, M. & Matthews, B. W. Stabilization of phage T4 lysozyme by engineered disulfide bonds. *Proceedings of the National Academy of Sciences* **86**, 6562–6566 (1989).

173. Ferina, J. & Daggett, V. Visualizing Protein Folding and Unfolding. *J Mol Biol* **431**, 1540–1564 (2019).

174. Mala, P. & Saraogi, I. Enhanced Codon–Anticodon Interaction at In-Frame UAG Stop Codon Improves the Efficiency of Non-Natural Amino Acid Mutagenesis. *ACS Chem Biol* **17**, 1051–1060 (2022).

175. Aravind, M. K., Varalakshmi, P., John, S. A. & Ashokkumar, B. Catechol 1,2-Dioxygenase From Paracoccus sp. MKU1—A Greener and Cleaner Bio-Machinery for cis, cis-Muconic Acid Production by Recombinant E. coli. *Front Bioeng Biotechnol* **9**, (2021).

176. Fiege, H. *et al.* Phenol Derivatives. in *Ullmann's Encyclopedia of Industrial Chemistry* (2000). doi:https://doi.org/10.1002/14356007.a19_313.

177. *Final Screening Assessment for 1,2-Benzenediol (Catechol)*. (2008).

178. Kim, K. H. *et al.* Tandem conversion of lignin to catechols via demethylation and catalytic hydrogenolysis. *Ind Crops Prod* **159**, 113095 (2021).

179. Schmetz, Q. *et al.* Comprehension of an organosolv process for lignin extraction on Festuca arundinacea and monitoring of the cellulose degradation. *Ind Crops Prod* **94**, 308–317 (2016).

180. Zhang, X. *et al.* Phenolics Production through Catalytic Depolymerization of Alkali Lignin with Metal Chlorides. *Bioresources* **9**, (2014).

181. Liu, X. *et al.* Selective Hydrodeoxygenation of Lignin-Derived Phenols to Cyclohexanols or Cyclohexanes over Magnetic CoNx@NC Catalysts under Mild Conditions. *ACS Catal* **6**, 7611–7620 (2016).

182. Zakzeski, J., Bruijnincx, P. C. A., Jongerius, A. L. & Weckhuysen, B. M. The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. *Chem Rev* **110**, 3552–3599 (2010).

183. Mabrouk, A., Erdocia, X., Alriols, M. G. & Labidi, J. Economic analysis of a biorefinery process for catechol production from lignin. *J Clean Prod* **198**, 133–142 (2018).