University of Wisconsin Milwaukee

UWM Digital Commons

Theses and Dissertations

12-1-2017

Structure-Function Relationships in Bacterial Regulatory Proteins and an Enzyme Involved in Antibiotic Biosynthesis

Lanlan Han
University of Wisconsin-Milwaukee

Follow this and additional works at: https://dc.uwm.edu/etd

Part of the Biochemistry Commons

STRUCTURE-FUNCTION RELATIONSHIPS IN BACTERIAL REGULATORY PROTEINS AND AN

ENZYME INVOLVED IN ANTIBIOTIC BIOSYNTHESIS

by

Lanlan Han

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Biochemistry

at

The University of Wisconsin-Milwaukee

Dec 2017

ABSTRACT

STRUCTURE-FUNCTION RELATIONSHIPS IN BACTERIAL REGULATORY PROTEINS AND AN
ENZYME INVOLVED IN ANTIBIOTIC BIOSYNTHESIS

by

Lanlan Han

The University of Wisconsin-Milwaukee, 2017
Under the Supervision of Professor Nicholas R. Silvaggi

The first part of my thesis is focused on a new family of two-component response regulator proteins: Aspartate-Less Regulators (ALR). They lack the catalytic aspartate residue that mimics the phosphorylation mechanism of typical two component response regulators. We are using biophysical tools to determine two proteins with redox-sensitive ALR domains: repressor of iron transport regulator (RitR) from *Streptococcus pneumonia* R6 and diguanylate cyclase Q15Z91 from *Pseudoalteromonas atalantica*. The structure of inactive RitR(C128S) monomer showed that the ALR domain and the DNA-binding domain are linked by an α-helix that runs the length of the entire protein, with C128 near the C-terminal end. Bioinformatic analysis of all streptococcal RitR homologs showed that Cys128 is strictly conserved, suggesting that RitR may be a novel redox sensor. Hydrogen peroxide was used to oxidize the cysteine thiol group to determine the structure of the oxidized, dimeric form of RitR. Oxidation of C128 to the disulfide caused a conformational change that caused the DBD to release from the ALR domain. Surprisingly, the freed DBD was observed bound to the ALR domain of the other, disulfide-linked molecule of RitR, recapitulating almost exactly the structure of the inactive, monomeric protein. An extended dimeric conformation was found in the RitR(L86A/V93A) variant. It binds to the target DNA according to gel filtration and differential scanning fluorimetry. The crystal structure

of the RitR(L86A/V93A) ALR domain showed an unprecedented conformational change for a response regulator protein, where helix α4 is disordered and the two protomers swap their α5 helices to form the dimer. Combining with the C128D mutant in vivo studies, it seems that oxidation of C128 is part of the activation mechanism, but there must be an additional factor that leads to dimerization of the ALR domains. The second ALR protein Q15Z91 has R61 replacing the phosphorylatable aspartate residue in the ALR domain. According to the structure of Q15Z91 with GTP and c-di-GMP, Purified Q15Z91 is an activated but product-inhibited dimer. Substitution on C142 that at the same position with that in RitR demonstrated that this C142 residue is also a redox sensor that involved in Q15Z91 activity regulation.

The second part is a mechanistic enzymology project aimed at understanding the structure and mechanism of the novel pyridoxal-5'-phosphate (PLP)-dependent L-arginine hydroxylase/deaminase, MppP, from *Streptomyces wadayamensis* (SwMppP). SwMppP is predicted to be a type I/II aminotransferase based on primary sequence identity. However, NMR and ESI-MS results showed that SwMppP is not an aminotransferase, but rather a hydroxylase. The enzyme catalyzes the oxygen-dependent hydroxylation of L-arginine, forming 4-hydroxy-2-ketoarginine and the abortive side-product 2-ketoargine in a ratio of 1.7:1. This is exciting because SwMppP is the first PLP-dependent enzyme to react with oxygen in any context other than oxidative decarboxylation. The discovery of this new activity is especially surprising given that the tertiary structure of SwMppP is very similar to that of the prototypical aminotransferase, the *E. coli* aspartate aminotransferase (PDB entry 1ARS; RMSD of aligned Cα atoms is 3.7 Å). The major differences between the two enzymes are the disordered N terminus of SwMppP, and changes of a limited number of

amino acids near the PLP cofactor. The N-terminal helix transitions from a disordered, random-coil state to a helical conformation covering the active site only the substrate is bound. Specific roles of the un-conserved residues in the active site are being studied by mutagenesis. So far, most of the SwMppP mutants have lost the hydroxylase activity and only produce 2-ketoarginine. Our mechanistic studies have revealed that formation of the fully oxidized (hydroxylated) product requires 2 equivalents of dioxygen, while formation of 2-ketoarginine requires only one equivalent of dioxygen. Interestingly, the hydroxyl group of 4-hydroxy-2-ketoarginine comes from H_2O, not dioxygen. Mutagenesis, structural and kinetic studies were used to understand how the residues in the active site stabilize the quinonoid form of the L-arginine-PLP complex to promote the reaction with dioxygen. Our structural and kinetic characterization of the wild-type and variant forms of SwMppP have allowed us to propose a model where the oxygen incorporated in the hydroxy-arginine product is derived from water rather than from dioxygen. In addition, SwMppP exhibits very high substrate specificity. Either change on the substrate length or guanidine group would result in no binding affinity or little activity.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 4

LIST OF TABLES

Chapter 5

LIST OF SCHEMES

LIST OF ABBREVIATIONS

2KA	2-ketoarginine
2KE	2-ketoenduracididine
4HKA	4-hydroxy-ketoarginine
AATase	aspartate aminotransferase
ALR	aspartate-less receiver
APS	Advanced Photon Source
Atase	aminotransferase
c-di-GMP	bis-(3'-5')-cyclic dimeric guanosine monophosphate
CA	catalytic ATP/Mg^{2+} binding site
D-βhEnd	D-β-hydroxyenduracididine
DBD	DNA-binding domain
DGC	diguanylate cyclase
DHp	dimerization interface and the phosphorylatable histidine residue
DSF	Differential scanning fluorimetry
DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)
DTT	dithiothreitol
EcAAT	*E. coli* aspartate aminotransferase
EMSA	electrophoretic mobility shift assay
ESI-MS	mass spectrometry with electrospray ionization
FDASIA	Food and Drug Administration Safety and Innovation Act
GAIN	Generating Antibiotic Incentives Now
GDH	L-glutamate dehydrogenase

GTP	guanosine triphosphate
HAMP	histidine kinase, adenylyl cyclases, methyl-accepting chemotaxis proteins, and phosphatase
HK	histidine kinase
HPt	histidine-containing phosphoryl transmitter domain
HTH	helix-turn-helix
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermol titration Calorimetry
L-Cap	L-capreomycidine
L-DOPA	L-3,4-dihydroxyphenylalanine
L-End	L-enduracididine
L-βhEnd	L-β-hydroxyenduracididine
LAAO	L-amino acid oxidase
LCMS	Liquid Chromatograph Mass Spectrometer
LS-CAT	Life Science Collaborative Access Team
MCPs	methyl-accepting chemotaxis proteins
MESG	2-amino-6-mercapto-7-methylpurine ribonucleoside
MRSA	methicillin-resistant *Staphylococcus aureus*
NADH	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
NSDS	native SDS-PAGE
OCST	one-component signal transduction
PAS	Per, ARNT and SIM

PEG	polyethylene glycol
pGpG	5'-phosphoguanosylyl-(3'-5')-guanosine
PLP	pyridoxal-5'-phosphate
PMP	pyridoxamine phosphate
PNP	purine nucleoside phosphorylase
Q1	Quinonoid I
Q2	Quinonoid II
RD2LV	RitR(L86A/V93A) REC domain, residues 1-132
RDLV	RitR(L86A/V93A) REC domain, residues 1-124
REC domain	receiver domain
RitR	repressor of iron transport regulator
rmsd	root-mean-square deviation
RP-HPLC	reverse phase high performance liquid chromatography
RR	response regulator
SAD	single-wavelength anomalous diffraction
SAXS	small-angle X-ray scattering
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SEC-MALS	size exclusion chromatography with multi-angle light scattering
SSM	secondary structure matching
TCST	two-component signal transduction
TLS	translation-libration-screw
TNB-	2-nitro-5-thiobenzoate

VRE   vancomycin-resistant enterococci

ACKNOWLEDGMENTS

would also like to thank all my friends in Milwaukee, especially Guanguan Li, for all her

help in these years.

Chapter 1

Introduction to Bacterial Signal Transduction

All cells sense and respond to environmental stimuli using signal transduction systems. In prokaryotes, these systems typically take the form of either one-component signal transduction (OCST) systems or two-component signal transduction (TCST) systems. OCST systems consist of one protein, which contains both input (sensory) and output (effector) domains, but lacks phosphotransfer domains[15]. The quorum-sensing protein TraR [16] from *Agrobacterium tumefaciens* is a one-component signaling system. When the autoinducer N-(3-oxo-octanoyl)-L-homoserine lactone binds to the N-terminal domain, TraR dimerizes and binds to DNA to activate expression of plasmid conjugal transfer genes [16, 17]. In contrast, the more common TCST systems usually consist of five parts: (1) a sensory protein to detect the stimulus, (2) an autophosphorylated histidine kinase (HK; often part of the sensory protein), (3) a response regulator (RR) that accepts the phosphoryl group from the HK, (4) a target that the response regulator acts on, and (5) a phosphatase to restore the RR to its inactive state. The name "two-component" signal transduction system refers to the histidine kinase/response regulator pair. A typical TCST system is illustrated schematically in Figure 1-1. The sensor domain of the transmembrane histidine kinase protein triggers autophosphorylation of a specific histidine side chain of the HK domain in response to the stimulus, which can be the binding of some ligand [18], changes in pH [19], osmolality [20], temperature [21] and nutrients [22]. The phosphoryl group on the HK is then transferred onto the conserved aspartate residue on the receiver (REC) domain of a cognate response regulator protein. This phosphotransfer reaction is catalyzed by the

response regulator itself and requires Mg^{2+}.
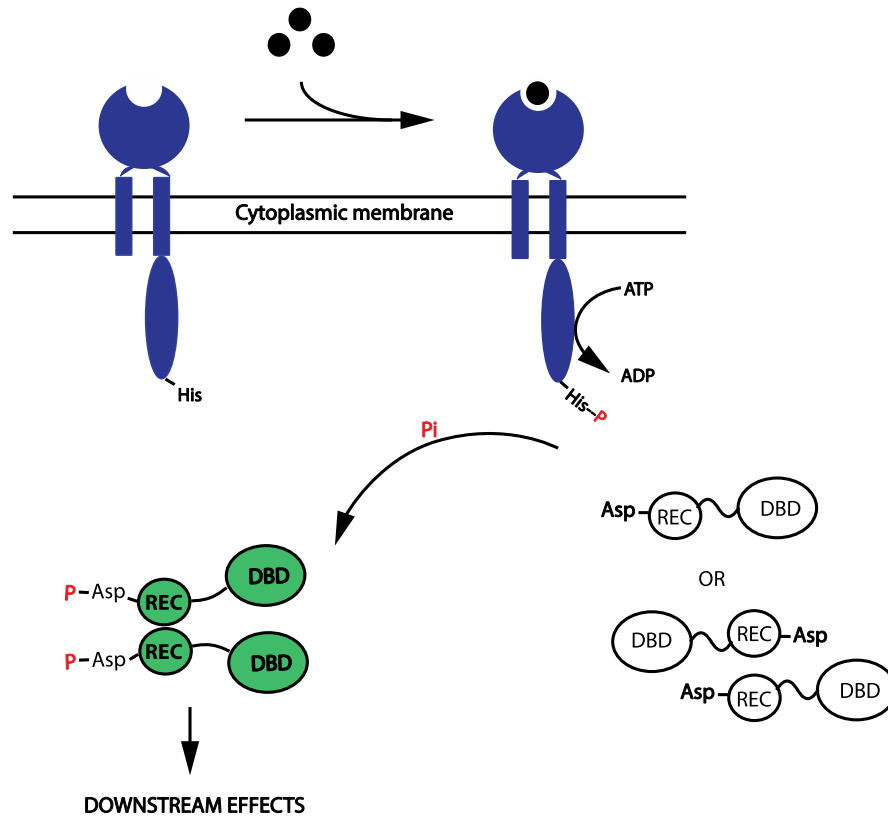
Figure 1-1: Schematic diagram of a typical two-component signal transduction system of the OmpR/PhoB family[5]. The stimulus (black solid circle) is represented as a small molecule that is recognized by the sensor-kinase (blue); ligand binding triggers autophosphorylation of the histidine kinase domain. The inactive response regulator (white) encounters the phosphorylation HK and transfers the phosphate group to an aspartate residue in the receiver domain (REC) of the response regulator. Inactive (unphosphorylated) response regulator proteins can exist as monomeric or inactive dimeric conformations. Phosphorylation of the REC domain causes conformational changes that result in formation of the active dimer conformation of the RR (green). The response regulator in this example has a DNA-binding domain (DBD) as the output domain, and activation would result in either up- or down-regulation of transcription from a specific gene locus.

Signal transduction systems in prokaryotes were recently found to be more complicated than originally thought due to the discovery of auxiliary factors [23] and diverse phosphotransmission pathways, such as the ArcB-ArcA anoxic redox control system in *E. coli* [24], the CheA-CheY/CheB chemotaxis system in *E. coli* [25], the KinA/KinB-Spo0F-Spo0B-Spo0A sporulation control system in *Bacillus subtilis* [7], and Ser/Thr and Tyr phosphorylation systems mediated by bacterial eukaryotic-like serine/threonine and tyrosine kinases and the cognate PP2C phosphatases[26-28].

1.1 Histindine Kinase

According to the domain compositions, HKs can be generally divided into two types: orthodox kinases and hybrid kinases. Orthodox histidine kinases typically contain a transmembrane domain, a four-helix coiled coil HAMP (histidine kinase, adenylyl cyclases, methyl-accepting chemotaxis proteins, and phosphatase) domain, a PAS (Per, ARNT and SIM) [29] domain, and a C-terminal transmitter domain. The transmitter domain contains the dimerization interface and the phosphorylatable histidine residue (DHp), and a catalytic ATP/Mg^{2+} binding site (CA). EnvZ is an orthodox HK from *E. coli* that located in the inner membrane and contains two transmembran regions and a large cytoplasmic domain. It senses environmental osmolarity changes [1] and then phosphorylates the cognate RR protein OmpR, which ultimately regulates the expression of the outer membrane porins OmpF and OmpC [30] (Figure 1-2A). At high osmolality , OmpF is repressed and OmpC is the major outmembrane porin. It was also found that higher osmolality induces intra-helical H bonds, which enhances the H243 autophosphrylation and phosphoryl transfer to OmpR[31].
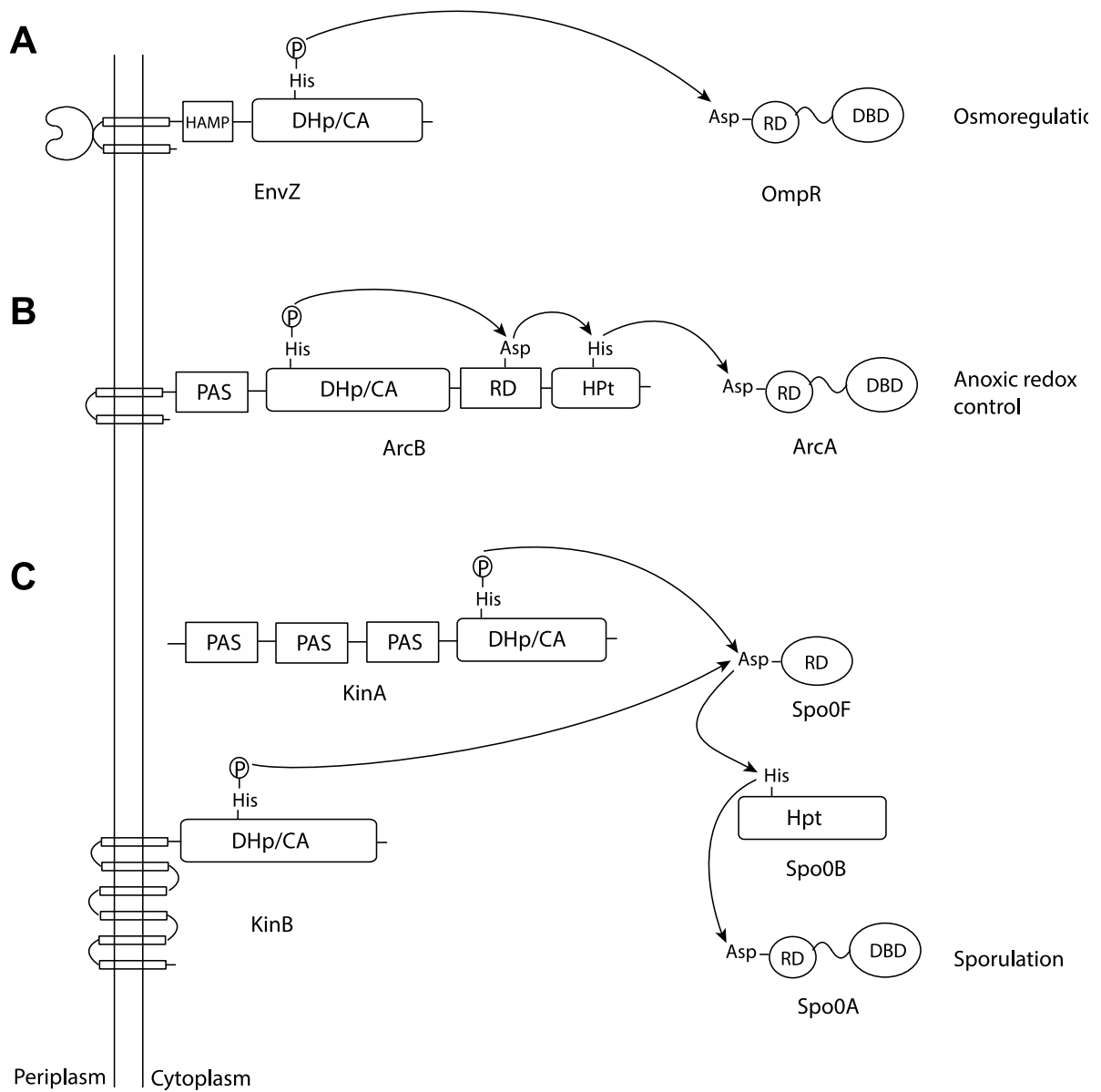
Figure 1-2: Prototypical histidine kinase with its cognate RR: Osmosis control (EnvZ [1]-OmpR) (A), a typical hybrid histidine kinase system, ArcB-ArcA [2] (B) and the KinA/KinB-Spo0F-Spo0B-Spo0A system [7] (C), which is involved in phosphor-relay control. HPt: Histidine-containing phosphoryl transmitter domain.

HKs usually have two or more transmembrane regions that separate the periplasmic

sensor region from the cytoplasmic transmitter domain. Some orthodox HKs do not have a transmembrane region, like the chemotaxis kinase CheA [32] and the nitrogen regulatory kinase NtrB [33]. PAS domains are common in sensor proteins and can sense diverse environmental or internal signals, including ligands, light, oxygen level and redox change[34]. The architecture of proteins containing PAS domains is diverse, but the overall structures of PAS domains are conserved[35]. Most PAS domains contain a central antiparallel β sheet with several α helices packed against one face[35], as exemplified by VicK[36] from *Streptococcus mutans* (Figure 1-3). The PAS domain contains a five-stranded central β sheet surrounded by five α helices. The uppermost HAMP domain consists four parallel α helices from two chains. Deletion of HAMP and PAS domains impaired the phosphatase activity of VicK[37]. The dimerization domain is usually composed of two long α helices. The phosphoryl receptor histidine locates in the helix (H217 in VicK). The loop connecting the two α helices determines the specificity of the receptor for one or more specific RR proteins [38]. The main body of kinases CA domains is conserved, consisting of an α/β-sandwich fold with a five-stranded mixed β sheet and three α helices, as observed in DesK[39] and PhoQ[40]. This domain is structurally homologous to the ATPase domains of DNA gyrase B[41], MutL[42], and Hsp90[43].
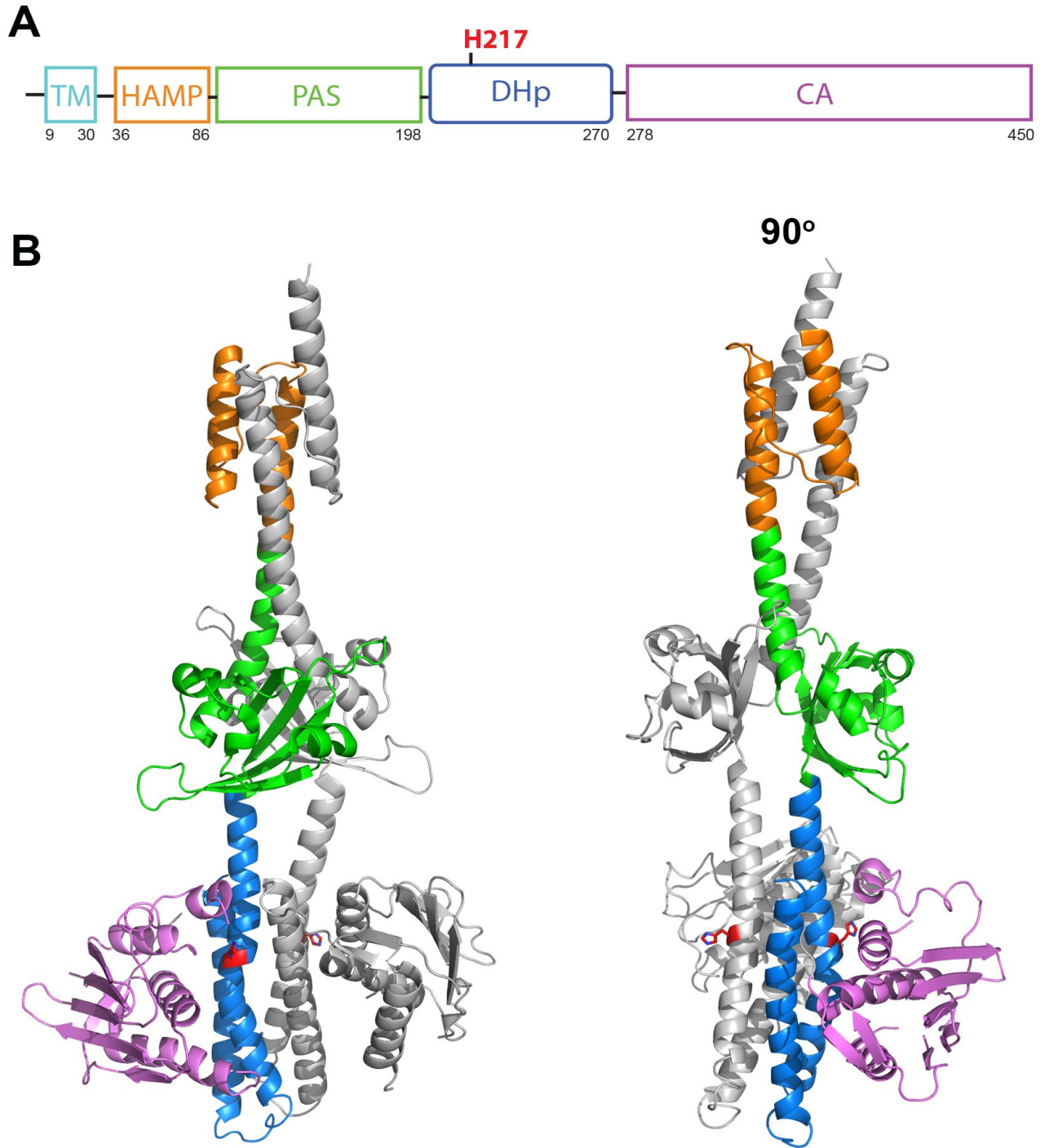
Figure 1-3: Schematic showing of VicK domain organization (A) and ribbon representation of VicK structure (PDB ID: 4I5S [36]) (the transmembrane domain is not included in the structure) (B). The colors of different domains in chain A are matching the schematic from A. Chain B is colored in grey. The conserved H217 residues in the DHp domain are shown in stick with red carbon atoms.

Hybrid kinases are found in some prokaryotes and almost all eukaryotes [44]. Hybrid kinases contain multiple phosphodonor and phosphoacceptor sites and undergo multi-step phospho transfer, such as ArcB [2, 24] (Figure 1-2B). ArcB from *E. coli* that is involved in anoxic redox control (Arc), is a typical HK of the hybrid kinase family. ArcB contains two N-terminal transmembrane regions, a kinase core, and a second His-containing phosphotransfer domain. Once the N-terminal transmembrane domain senses the stimulus, the signal is propagated into the transmitter domain by the HAMP domain, which triggers dimerization via the dimerization domain. The resulting four-helix bundle of the dimer is cross-phosphorylated on a conserved histidine residue of the dimerization domain using ATP and Mg^{2+}. The complexity of hybrid kinase structures allows different regulatory pathways.

1.2 Response regulators (RRs)

1.2.1 Activation mechanism of phosphorylation dependent RRs

In prokaryotic TCST systems, the most common and well-studied RRs are transcriptional regulators [15]. They have DNA-binding domains that bind specific DNA sequences, and usually function as homodimers. The largest subgroup of DNA binding RRs is OmpR/PhoB family. The RRs in OmpR/PhoB family are characterized with a REC domain and a winged helix-turn-helix domain connected by a linker [45].

X-ray crystal structures have been determined for a number of full-length RR proteins from the OmpR/PhoB family[10, 46-48]. In addition, the structures of some members of both

families have been determined in both the inactive, unphosphorylated and active, phosphorylated states. The latter is commonly approximated by BeF$_3$ adducts that mimic the actual phosphorylated protein, owing to the difficulty of producing fully phosphorylated (i.e. homogenous) protein for crystallization studies. The ability to compare the structures of unphosphorylated and "phosphorylated" RRs has allowed some insight into the phosphorylation-induced activation mechanisms in these proteins[14, 48].

The REC domains of OmpR/PhoB subfamily adopt a sandwich $(\beta/\alpha)_5$ topology with a central five-stranded parallel β sheet surrounded by 3 and 2 α helices on both sides (e.g. PhoB in Figure 1-4). The phosphorylatable aspartate usually located in the loop between $\beta3$ and $\alpha3$, and the other two acidic residues usually located in the loop between $\beta1$ and $\alpha1$. These three acidic triad residues are involved in the Mg^{2+} coordination and phosphoryl transfer from cognate HK. They dimerize through $\alpha4$-$\beta5$-$\alpha5$ interface upon phosphorylation.

One well studied RR of OmpR/PhoB family is PhoB [5], which controls the cellular response to the phosphate-limitation in PhoR-PhoB[5] (known as PhoP in some organisms) TCST system in *E. coli* and *Bacillus subtilis*. PhoR is an autophosphorylated transmembrane histindine kinase. It has a N-terminal transmembrane region and a cytoplasmic domain consisting of PAS domain, DHp domain and CA domain [49]. PhoB (PhoP) is a transcriptional regulator that modulates the expression of *pho* regulon genes by binding to the "*pho* box" sequences of the promoters [50]. Both the inactive monomeric structure [9] and the active dimeric structure [14] of PhoB were determined (Figure 1-4). Once the aspartate residue accepts the phosphoryl group from histidine kinase, a subtle structural rearrangement
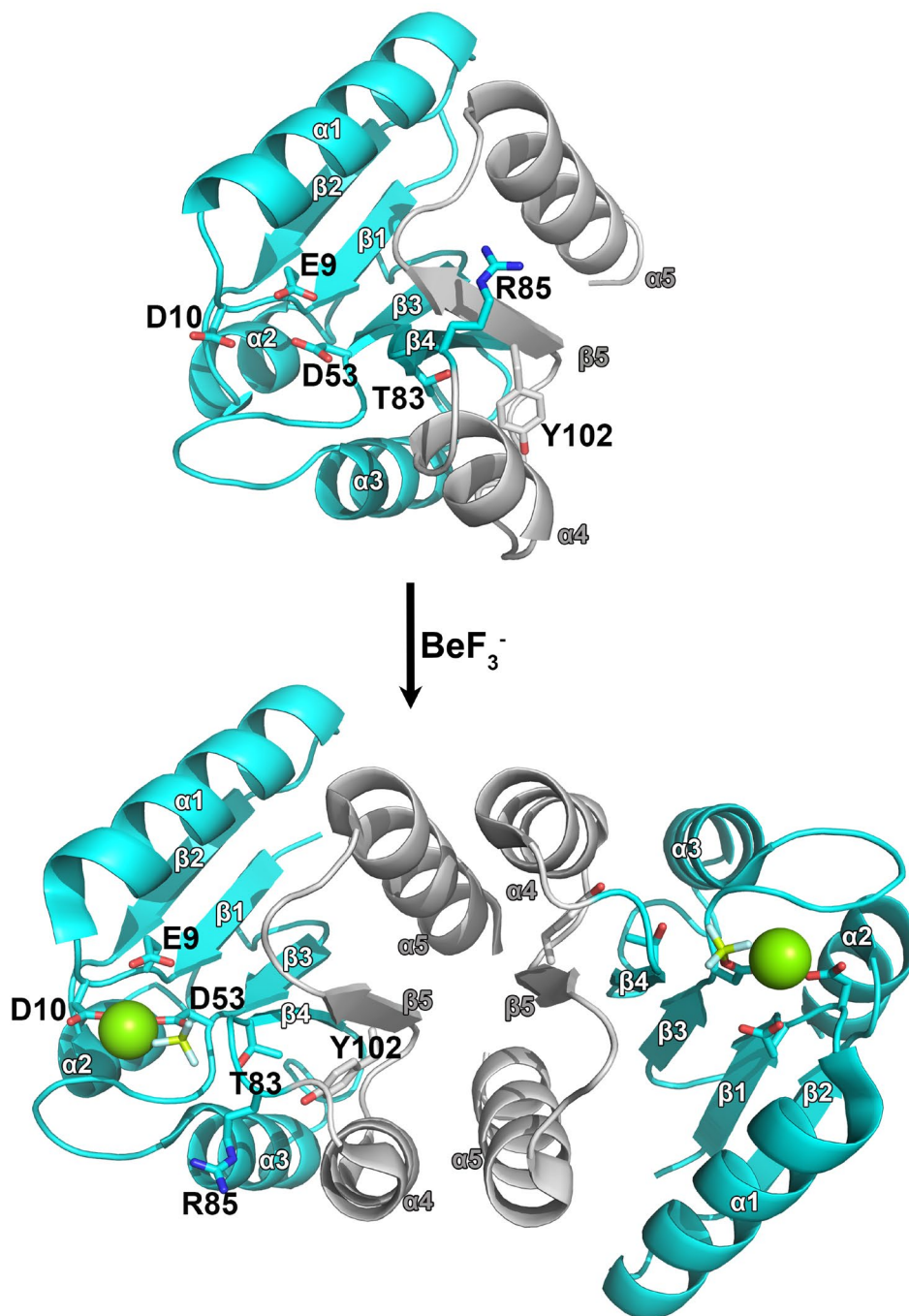
Figure 1-4: PhoB dimerization is triggered by beryllium trifluoride, a phosphoryl analog. The left side shows the structure of the inactive PhoB monomer (PDB ID: 1B00 [9]), and the right side shows the BeF$_3$-activated dimer (PDB ID: 1ZES [14]). The acidic triad (E9, D10 and D53) and "switch" residues (T83 and Y102) are shown as sticks. The yellow-green sphere represents the magnesium ion required for the phosphotransfer reaction.

brings the α3 helix closer to the active site. The α4 helix and β5 strand undergo a more pronounced conformational change. The side chain of T83 reorients and interacts with the phospho-Asp in the active site. In addition, the side chain of Y100 rotates around to fill the void left by the reorientation of T83. This reorientation of Y100 is thought to be an important factor in the formation of the α4-β5-α5 dimer interface [14]. Thus, T83 and T100 have come to be thought of as the "switch residues" that are largely responsible for triggering dimerization by modulating the shape of the interacting surface of the REC domain.

Comparisons of some full-length RRs structures in OmpR/PhoB family and the autophosphorylation rate, it was found that the interdomain interactions between REC domain and DBD of the inactive RRs affect the autophosphorylation rate[51]. According to the extent of interdomain interactions, OmpR/PhoB subfamily RRs are indeed divided into three subclasses (Figure 1-5) [12]. As shown in Figure 1-5A, the α4-β5-α5 dimer interface of DrrB [3] REC domain interacts extensively with the DBD, which represents the "closed" subclass. DrrD [10] (Figure 1-5B) has less interdomain interactions and the DNA recognition helices are solvent accessible, which belongs to the "open" subclass. The representative of third subclasses is HP1043 [11] (Figure 1-5C), which has little interdomain interactions and readily forms dimer through α4-β5-α5 dimer interface.
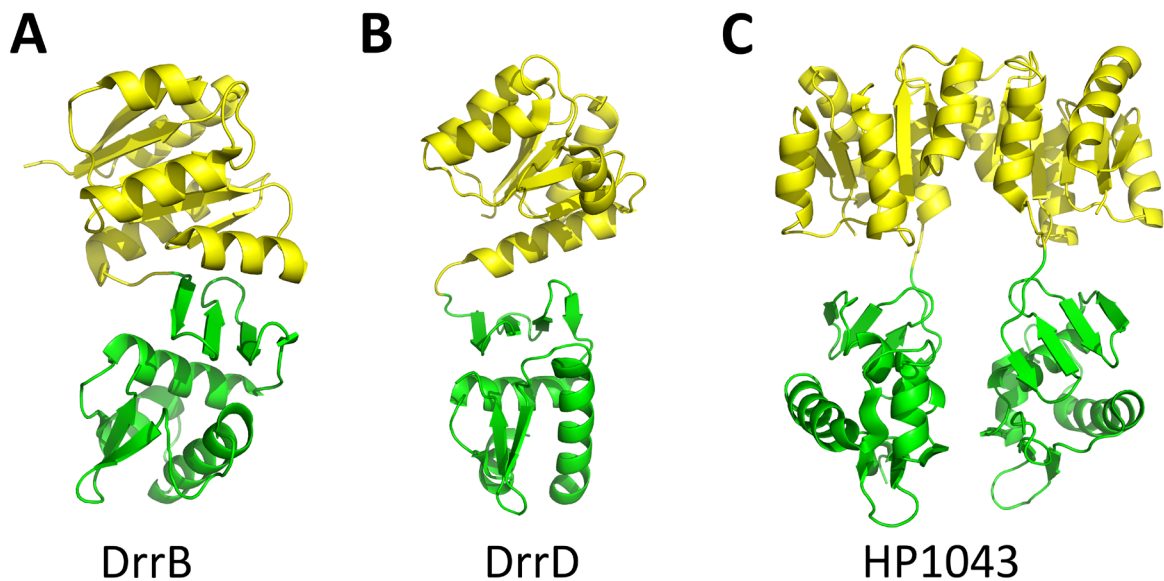
Figure 1-5: Structural comparison of various classes of full-length OmpR/PhoB subfamily structures. REC domains are colored in yellow, DBD colored in green. (A) The structure of DrrB (PDB ID: 1P2F [3]), representing the "closed" subclass that has extensively interdomain interactions. (B) The structure of DrrD (PDB ID: 1KGS [10]), representing the "open" subclass that has limited interdomain interactions (C) The structure of HP1043 (PDB ID: 2HQR [11]), representing the "free" subclass that completely lacks the

While DNA-binding effector domains are the most common, enzymatic domains (diguanylate cyclase, phosphatase, kinase etc), RNA binding domain and protein binding domain have also been observed in RRs [52]. Some RRs only have a REC domain and the regulatory function resides in the REC domain. For example, CheY is a chemotaxis-related regulator with a REC domain that directly regulates the direction of motor rotation of bacterial flagella after phosphorylation by binding to the rotor protein FliM and FliN[53, 54].

The chemotaxis system in *E. coli* mainly contains a transmembrane methyl-accepting chemotaxis proteins (MCPs), an adaptor CheW and the TCST system CheA-CheY/CheB.

11

MCPs, including Tar, Tsr, Tap, Trg, Aer [55], can bind various chemicals (e.g. sugars, amino acids, O_2) and control the autophosphorylation of CheA via the adaptor protein CheW. CheA is a histindine kinase that lacks the periplasmic and transmembrane domains typically found in HKs [56, 57]. CheY, a response regulator with REC domain only and CheB, a glutamate methylesterase, compete for the phosphoryl group on CheA. Once phosphorylated, CheY binds to the switch protein on the flagella motor to change the rotation direction, while CheB controls the adaptation of the MCPs by demethylation. Demethylated MCPs are poor kinase activators[58], which result in less activated CheA and less phosphorylated CheY.

Comparison of the inactive and BeF$_3$—activated mimic CheY structures[59-61] shows that CheY has a similar activation mechanism to that postulated for PhoB (Figure 1-4, Figure 1-
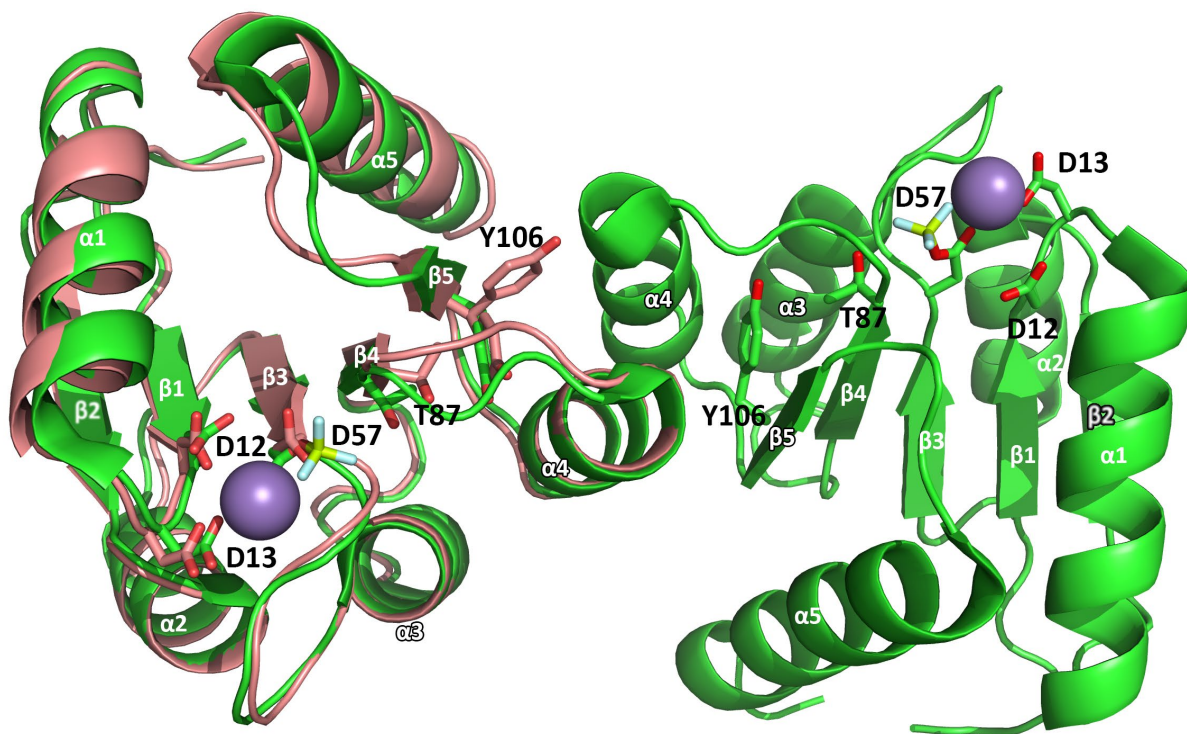
Figure 1-6: Overlay of the inactive CheY (light pink) with BeF$_3$ activated CheY dimer (green). Mn ions are represented in purple sphere. Acidic triad residues and switch residues are shown in stick.

6). Upon phosphorylation, the switch residue T87 of CheY rotates about to make a hydrogen bonding interaction with one of the fluoride atoms in the D57-BeF$_3$ adduct, analogous to one of the phosphate oxygen atoms. The side chain of Y106 also undergoes a conformational change, rotating inward to fill the void left by T87, where its phenolic oxygen makes a hydrogen bonding interaction with the main chain carbonyl O atom of E89 and the ring itself is buried in a hydrophobic pocket formed by residues on β4 and α4.

CheB has an N-terminal regulatory domain and a C-terminal methylesterase domain connected by a short linker. The regulatory domain of CheB is structurally homologous to CheY, but CheB contains a phosphorylation site and also an active site for auto-dephosphorylation on the REC domain [62]. The N-terminal REC domain inhibits the methylesterase activity in the inactive state, but after phosphorylation, the methylesterase activity will be transiently activated [62]. In addition, the inhibited full-length CheB structure is a "closed" conformation, because α4, α5 and β5 in the REC domain interact with the methylesterase domain and block the catalytic site [63]. Phosphorylation of D56 changes the conformations of α4 and α5, which disrupts the interdomain interactions and opens access to the catalytic site.

1.2.2 Conserved residues in phosphorylation-dependent RRs

Comparison of response regulator sequences (Figure 1-7) demonstrates that the residues mentioned above, Ser/Thr and Tyr/Phe switch residues, the acidic triad residues including phosphorylatable aspartic acid are all highly conserved in phosphorylation-

activated RRs. The acidic triad residues comprise a motif (Asp-Asp/Glu-Asp) that is essential for binding of the Mg^{2+} (or other divalent metal ions) required for phosphoryl group transfer. The conserved phosphorylatable aspartate residue plays a key role in activation. Substitution of this aspartate residue by glutamate, which has been used as a phospho-Asp mimic to examine activation in CheY [64], RcsB [65] and NtrC [66], resulted in constitutive activation. In contrast, substitution by asparagine or alanine in RcsB[65] and NtrC [66] eliminated activation in these proteins. The switch residues Thr/Ser and Tyr/Phe are essential for forming the active dimer conformation in many RRs. Substitutions of these two switch residues (*e.g.* T87 into isoleucine or Y106 into trptophan) impairs activation in CheY and affects chemotaxis [67]. Likewise, mutation of T87 to isoleucine prevents the side chain of Y106 from adopting the "inward-facing" conformation even when CheY is phosphorylated [67].

```
NarL    MSNQEPATILLIDDHPMLRTGVKQLISMAPDITVVGEASN-----------GEQGIELAESLDPDLILLDLN
FixJ    -----MQRLYIVDDDESVLDSLE--------FVLQGEGYQTQCFSSAEAFLQWSERHIECLTDTGCLILDCR
Spo0A   ---MEKIKVCVADDNRELVSLLSEYIEGQEDMEVIGVAYN-----------GQECLSLFKEKDPDVLVLDII
OmpR    --MQENYNILVVDDDMRLRALLERYL-TEQGFQ-VRSVAN-----------AEQMDRLLTRESFHLMVLDLM
PhoB    ----MARRILVVEDEAPIREMVCFVL-EQNGFQ-PVEAED-----------YDSAVNQLNEPWPDLILLDWM
CheY    -MADKELKFLVVDDFSTMRRIVRNLL-KELGFNNVEEAED-----------GVDALNKLQAGGYGFVISDWN
NtrC    --MQRGIV-WVVDDDSSIRWVLERAL-AGAGLT-CTTFEN-----------GAEVLEALASKTPDVLLSDIR
                    : :*     :     :                     :              .         :: *

NarL    MPGMNGLETLDKLREKSL--SGRIVVFSVSNHEEDVVTALKRGADGYLLKDMEPEDLLKALHQAAAGEMVLS
FixJ    MPGMGGQTLQTLLSQTRS--TLGIVFLTGYGDVPMAVDALKQGAVDFLQKPVDQKHLLRAVKNALI------
Spo0A   MPHLDGLAVLERLRESDLKKQPNVIMLTAFGQEDVTKKAVDLGASYFILKPFDMENLVGHIRQVSGNAS---
OmpR    LPGEDGLSICRRLRSQ--SNPMPIIMVTAKGEEVDRIVGLEIGADDYIPKPFNPRELLARIRAVLRRQA---
PhoB    LPGGSGIQFIKHLKRESMTRDIPVVMLTARGEEEDRVRGLETGADDYITKPFSPKELVARIKAVMRRIS---
CheY    MPNMDGLELLKTIRADGAMSALPVLMVTAEAKKENIIAAAQAGASGYVVKPFTAATLEEKLNKIFEKLG---
NtrC    MPGMDGLALLKQIKQRHP--MLPVIIMTAHSDLDAVVSAYQQGAFDYLPKPFDIDEAVALVERAISHYQ---
            :*   .*       :             ::  :    .             .         : *  .              :.
```

█ Acidic triad (phosphorylatable aspartate colored in red)   █ Swich residues   █ Conserved Lysine

Figure 1-7: Conserved residues in phosphorylation-activated response regulators.

In the REC domain of CheB [63], there is another important acidic residue E58. E58 is near the acidic triad and orients toward the metal-binding site. The E58K mutant of CheB exhibited a much slower rate of phosphoryl transfer from phospho-CheA and auto-dephosphorylation of CheB, which suggests that E58 contributes to the phosphoryl group transfer from CheA and autophosphotase activity of CheB. Phosphorylation-activated response regulators often have another conserved basic residue, a lysine, located near the phosphorylation site and interact with phospho-Asp.

However, these highly conserved residues may have different roles in the transition to the activated state. In CheY, D57 is the conserved phosphorylatable residue[68]. *In vivo*, the CheY mutant D57N was phosphorylated at an alternative residue, S56, and still regulated the rotation of flagella[69]. However, a CheY D57E mutant, which in other systems has been thought to mimic phospho-Asp, exhibited no activity in *vivo*, although it was still phosphorylated by CheA. In NtrC (nitrogen regulatory protein C), the variant D54E mimics the phospho-Asp and has an NtrC+ phenotype [70]. *In vitro* phosphorylation studies verify that the NtrC D54E protein can still be phosphorylated even though not stable [71]. Like CheY, NtrC also has a serine residue (S53) adjacent to the conserved phospho-acceptor residue D54. However, the NtrC D54N mutant does not have any active phenotype [70, 71]. The conserved basic residue K109 in CheY interacts with the phospho-Asp and contributes to the phosphorylation by providing a simple electrostatic on-off switch [72-74]. While mutating the analogous residue, K107 in CheB, to Arg does not affect the phosphorylation of D56, but it does slow down the dephosphorylation step [75]. The analogous residue in NtrC, K103, is also not required for phosphorylation [71].

1.2.3 Atypical, phosphorylation-independent RRs containing the conserved Asp residue

We talked about the phosphorylated RRs activation mechanism and the conserved residues above. For the remainder of this discussion, the phosphorylation-dependent response regulators will be referred to as "typical" or "canonical" RRs. Recently, a large group of RRs with atypical REC domains has been identified whose members respond to signals other than phosphorylation. For example, MarR from *E. coli* that represses the transcription of many antibiotic-resistance operons was found to be derepressed by copper induced oxidation of the redox sensor C80 on MarR [76].

The orphan regulator NblR from *Synechococcus* that lacks the cognate histidine kinase, has the conserved phosphorylatable Asp residue, but it is not required for NblR function [77, 78]. Experiments with yeast and bacterial two-hybrid systems demonstrated that NblR does not interact with NblS (a conserved histidine kinase in Cyanobacteria), but rather with NarB (nitrate reductase).

ChxR from *Chlamydia trachomatis* exists as a constitutively active homodimer without phosphorylation [79]. First, the structure of the ChxR REC differs from those of typical (phosphorylation-dependent) RRs in that ChxR lacks two helices (α2 and α3) found in the typical OmpR/PhoB-type regulators (Figure 1-8A) [8]. Second, ChxR does not have both of the Thr/Ser and Tyr/Phe switch residues. The tyrosine residue on β5 is present, but the Ser/Thr on β4 is missing. Furthermore, the two residues in ChxR equivalent to the switch residues (L72 and Y90) are oriented similarly to those of inactive OmpR/PhoB family regulators, which indicate ChxR has a different activation/dimerization mechanism. Based

on small-angle X-ray scattering (SAXS) analysis of full-length ChxR (Figure 1-8B), ChxR belongs to the "free" subclass that lacks interdomain interaction between the REC domain and DBD [12]. So ChxR can bind to DNA in a phosphorylation-independent manner.
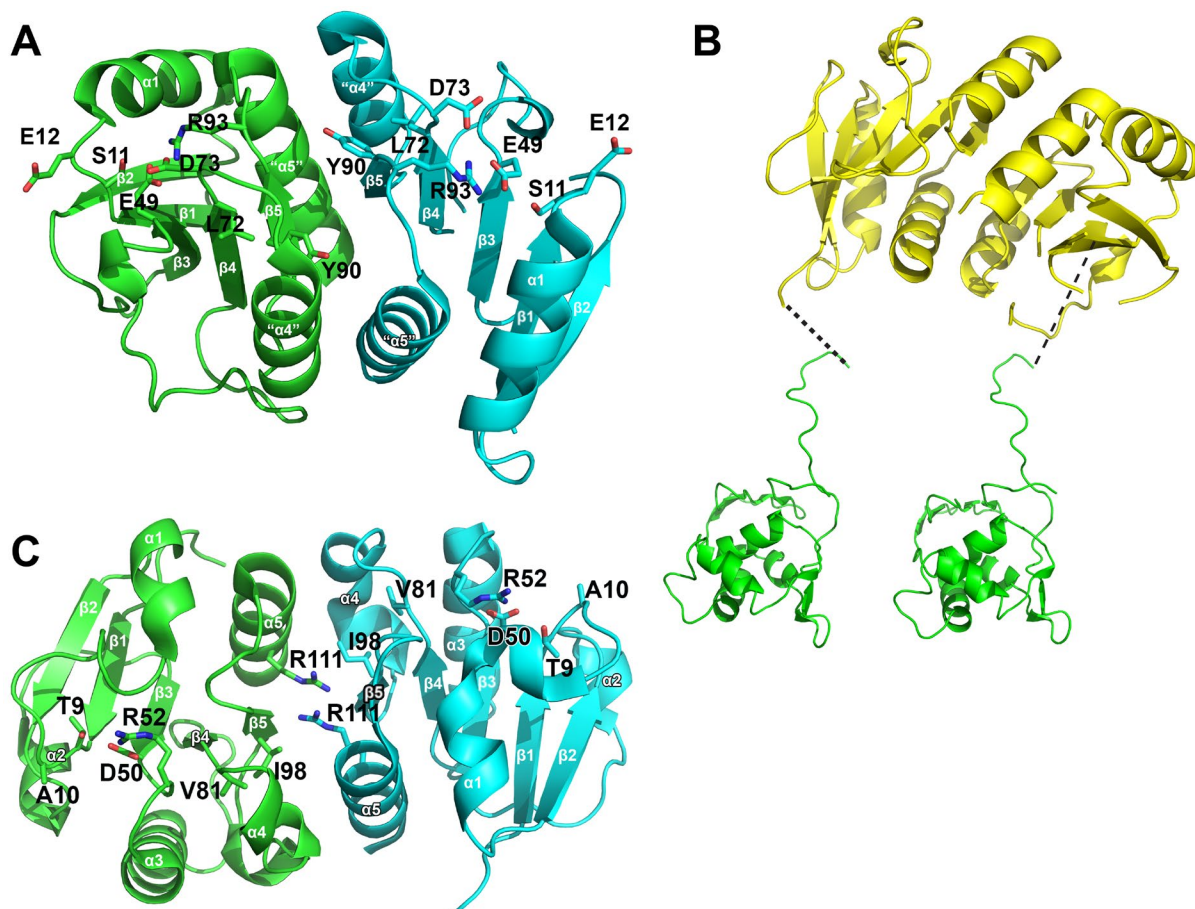
Figure 1-8: (A) The structure of ChxR REC domain (PDB ID: 3Q7R [8]). Chain A colored in green, chain B colored in cyan. In comparison with typical OmpR/PhoB RRs, ChxR lacks α2, α3 and the second and third helices are labeled as "α4" and "α5", respectively. (B) The model of full length ChxR dimer according to the best fit SAXS data [12] reflects a "free" conformation due to no interdomain interaction. (C) The structure of GlnR REC domain (PDB ID: 4O1H [13]). Chain A colored in green, chain B colored in cyan. The putative acidic triad (except D50) and switch residues are all missing in GlnR.

Another atypical regulator, GlnR from *Amycolatopsis mediterranei*, represses the expression of the glutamine synthetase gene and is thus involved in regulating nitrogen metabolism. GlnR is an orphan RR, because the cognate HK is not identified [80]. Like ChxR, GlnR also possesses the conserved phosphorylatable aspartate residue (D50), but functions as a homodimer without phosphorylation (Figure 1-8C). Interestingly, the other two acidic triad residues (E, D/E), the conserved Lys residue, and the switch residues (Ser/Thr, Tyr/Phe) are all not conserved in GlnR. In addition, Arg52 that located on the loop between β3 and α3 (Lβ3α3) is essential for maintaining dimerization by occupying the "phosphorylation pocket" and interacting with the conserved D50 [13].

1.2.4 Aspartate-Less Regulators (ALR)

Among atypical RRs, there are many proteins lacking the phosphorylatable Asp residue, such as the atypical response regulator HP1043 from *Helicobacter pylori*. HP1043 regulates the transcription of *tlpB* gene that encoding a methyl-accepting chemotaxis protein and is essential for cell growth [81]. In HP1043, a serine residue is substituted for the phosphorylatable Asp residue. The protein binds to DNA in a phosphorylation-independent way [11]. This raises the question of whether the loss of the phosphorylatable Asp residue is a common evolutionary theme in bacterial signal transduction. We recently defined a new subfamily of RR proteins with aspartate-less receiver (ALR) domains [6]. The ALR domains are common, accounting for ~4% of REC domain sequences in the Pfam database, and they are well distributed among bacteria, archaea, and eukaryotes. A large number (~28%) of ALR regulators have DNA-binding output domains. When compared to the entire class of

RRs, the ALR domain-containing response regulators are enriched for enzymatic output domains, such as diguanylate cyclase (GGDEF) domains, Ser-Thr kinase, and phosphatase domains. Although these atypical receiver domains have been known to exist for some time, the molecular details of their regulatory mechanisms have remained unclear. In the next two chapters, the structure-function relationships of two ALR proteins will be explored: RitR (Repressor of iron transport regulator, Chapter 2) and Q15Z91 (diguanylate cyclase, Chapter 3).

1.2.4.1 Repressor of iron transport regulator (RitR)

RitR is an orphan two-component signal transduction response regulator in *Streptococcus pneumoniae* R6. Analysis of the *S. pneumoniae* genome indicates that there are 13 pairs of TCST proteins (HK and RR) in this organism and a single orphan regulator, RitR [82-84]. Searching the genome sequence, there is no kinase gene close to the *ritR* gene locus. Instead, the adjacent gene *gnd* encodes a 6-phosphogluconate dehydrogenase and is co-transcribed with *ritR*, as evidenced by Northern Blot analysis [85]. RitR plays roles in iron uptake, the response to oxidative stress, and pathogenicity in a mouse lung model [85].

Analysis of REC domain sequences shows that RitR belongs to the OmpR/PhoB subfamily of RRs, and suggests that RitR likely has a similar tertiary structure to other OmpR/PhoB family members. A representative sequence alignment with the typical response regulator PhoB (Figure 1-9) shows that the conserved, phosphorylatable Asp residue in RitR is replaced by asparagine (N53; highlighted in yellow in Figure 1-9).

Furthermore, RitR does not have any other residue near this position that would be capable of accepting a phosphate group, such as Ser56 in CheY. The absence of any obvious candidates for phosphorylation in the sequence suggests that RitR does not have a phosphorylatable site in the "active site". Further underscoring the likely functional differences between RitR and other PhoB-type RRs is the substitution of a leucine residue in RitR for the Ser/Thr switch residue.

The three-dimensional structure of the RitR receiver domain (PDB ID: 4LZL [6]) was

```
RitR    MGKRILLLLEKERNLAHFLSLELQKEQYRVDLVEEGQKALSMALQTDYDLILLNVNLGDMM
PhoB    MARRILVVEDEAPIREMVCFVLEQNGFQPVEAEDYDSAVNQLNEPWPDLILLDWMLPGGS
        *.:***::*.*   :  .::.: *::: ::    .*: :.*:.    :    *****:  *

RitR    AQDFAEKLSRT---KPASVIMILDHWEDLQEELEVVQRFAVSYIYKPVLIENLVARISAI
PhoB    GIQFIKHLKRESMTRDIPVVM-LTARGEEEDRVRGLETGADDYITKPFSPKELVARIKAV
        . :* ::*.*     :   *:* *      : ::.:. ::  * .** **.  ::*****.*:

RitR    FRGRDFIDQHCSLMKVPRTYRNLRIDVEHHTVYRGEEMIALTRREYDLLATLMG-SKKVL
PhoB    MRRISPMAV-----EEVIEMQGLSLDPTSHRVMAGEEPLEMGPTEFKLLHFFMTHPERVY
        :*   . :      :      : * :*    * *  *** : :    *:.**  :*    ::*

RitR    TREQLLESVWKYESATETNIVDVYIRYLRSKLDVKGQKSYIKTVRGVGYTMQE--
PhoB    SREQLLNHVWGTNVYVEDRTVDVHIRRLRKALEPGGHDRMVQTVRGTGYRFSTRF
        :*****: **   :  .* . ***:** **. *:   *:.  ::*****.** :.!
```

Figure 1-9: Sequence Alignment of RitR and PhoB. The acidic triad residues are colored magenta (Mg^{2+}-binding) and red (phosphorylation site). The conserved Asp residue is replaced by Asn in RitR. The two switch residues are colored blue. The yellow boxes mark the differences in conserved residues between RitR and PhoB.

obtained by molecular replacement (Figure 1-10). The RitR REC domain is monomeric, which means that this structure represents an inactive (no DNA binding) conformation. It has the same $(\beta/\alpha)_5$ topology observed in other OmpR/PhoB family members, with only

subtle differences (see below).
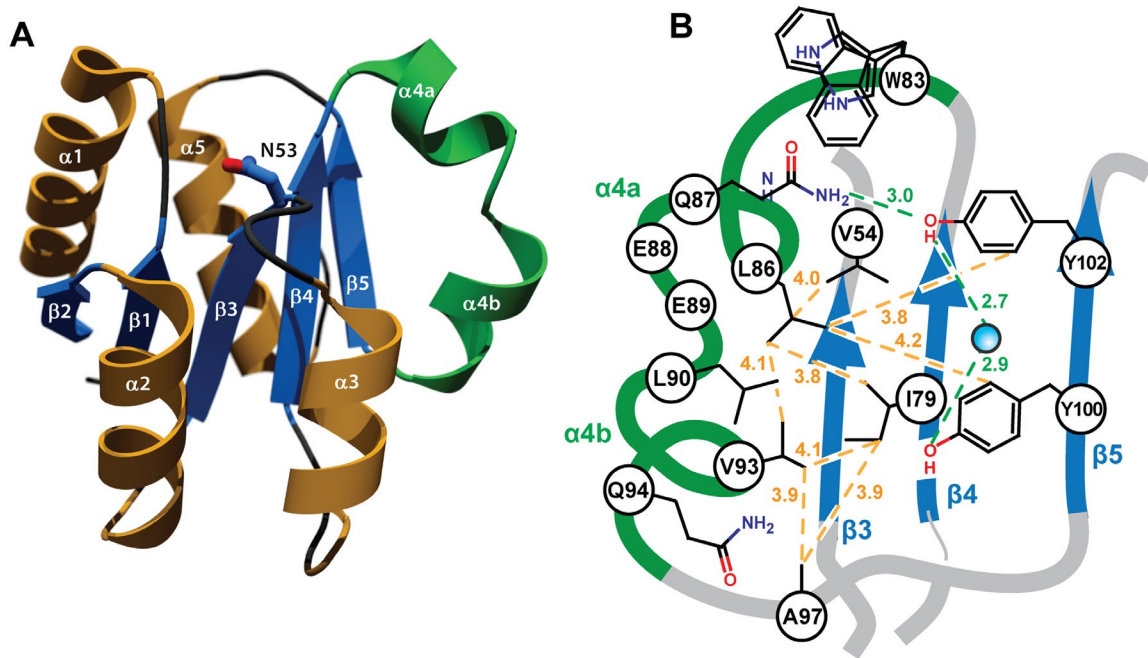
Figure 1-10: The structure of the RitR receiver domain (this figure is downloaded from PLOS Pathogens 11(4): e1004795 [6]). (A) Ribbon representation of the RitR REC domain with helices α1- α3 and α5 colored orange, the unusual α4 helix colored green, and the β-strands colored blue. N53 is shown in sticks. (B) Schematic view showing "hydrophobic interactions" (yellow dotted lines) and the hydrogen-bonding network (green dotted lines) in the "hydrophobic gate" region of the structure.

The most obvious difference is the broken α4 helix (colored green in Figure 1-10A), the pieces of which are labeled α4a and α4b, respectively. The hydrophobic residues on the α4 helix interact with residues on β3, β4 and β5, forming a hydrophobic core (as shown in Figure 1-10B). There are other RRs with unusual α4 helices, though none are "broken" like helix α4 of RitR. For example, DosR from *Mycobacterium tuberculosis* adopts an unusual (β/α)₄ fold, where the canonical α4 helix and β5 strand have been lost [86] (Figure 1-11).
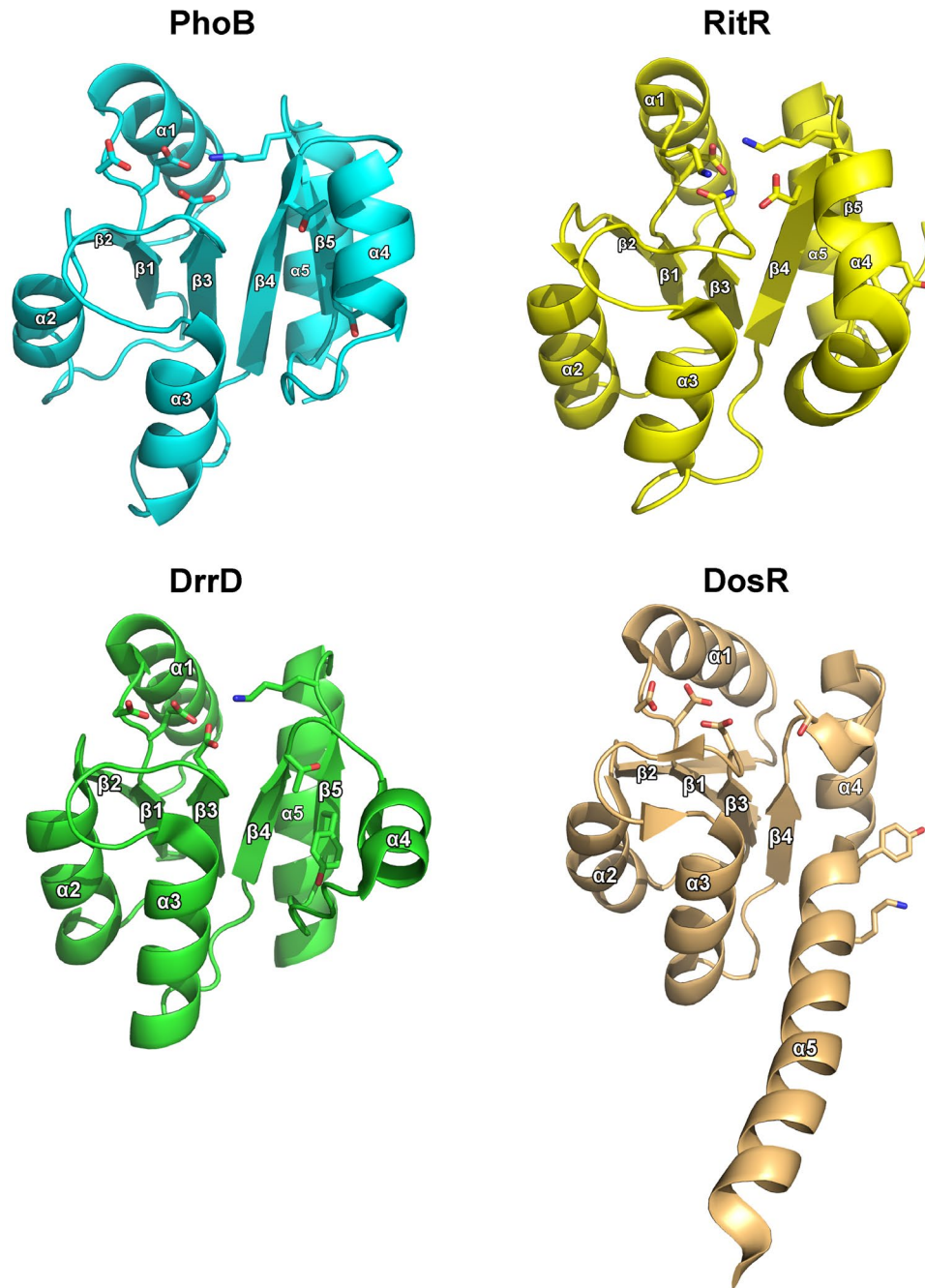
Figure 1-11: The α4 helix of response regulator REC domains appears to be particularly malleable. The conserved residues equivalents are shown in stick. The typical topology where α4 is parallel to β5 and unbroken is exemplified by PhoB (PDB ID 1B00 [9]), Other RR proteins have unusual α4 helices. For example, helix α4 is broken into two pieces in RitR (PDB ID: 4LZL [6]), while that of DrrD (PDB ID: 1KGS [10]) is roughly perpendicular to strand β5. Finally, DosR (PDB ID: 3C3W [86]) has a completely different topology where α4 is

missing entirely, as is strand β5.

Helix α4 in DosR is positioned where the canonical helix α5 is located in other NarL subfamily RR proteins. The conserved Y101 and K104 residues that are normally located on strand β5 and the loop Lβ5α5, are instead located in linker helix α5 and far away from the phosphorylation site. In DrrD from *Thermotoga maritima*, helix α4 is roughly perpendicular rather than parallel to the other helices [3], as shown in Figure 1-11.

In the so-called "active site," the phosphorylatable aspartate is replaced with an asparagine residue and there is no bound metal (Figure 1-12A). In what would be the metal-binding site, E9 interacts with N53 and the conserved residue K103. The basic residue K10, which is normally one of the residues of the acidic triad, sits near the top of the putative phosphorylation site and disrupts the acidic environment that promotes metal binding. As shown in Figure 1-12B, not only is the acidic environment is disrupted, but also the cleft for Mg^{2+} binding is eliminated in RitR. As mentioned above, RitR lacks a cognate histidine kinase. The lack of a cognate HK protein, together with the lack of key residues found in phosphorylation-dependent RRs makes it clear that RitR must be activated by a different mechanism from the typical, phosphorylation-dependent proteins.

Another feature of RitR is the only cysteine located in the linker region between REC domain and DBD. Bioinformatic results demonstrate that this cysteine128 is highly conserved in Streptococcus family. The specific role of this cysteine will be discussed in chapter 2.
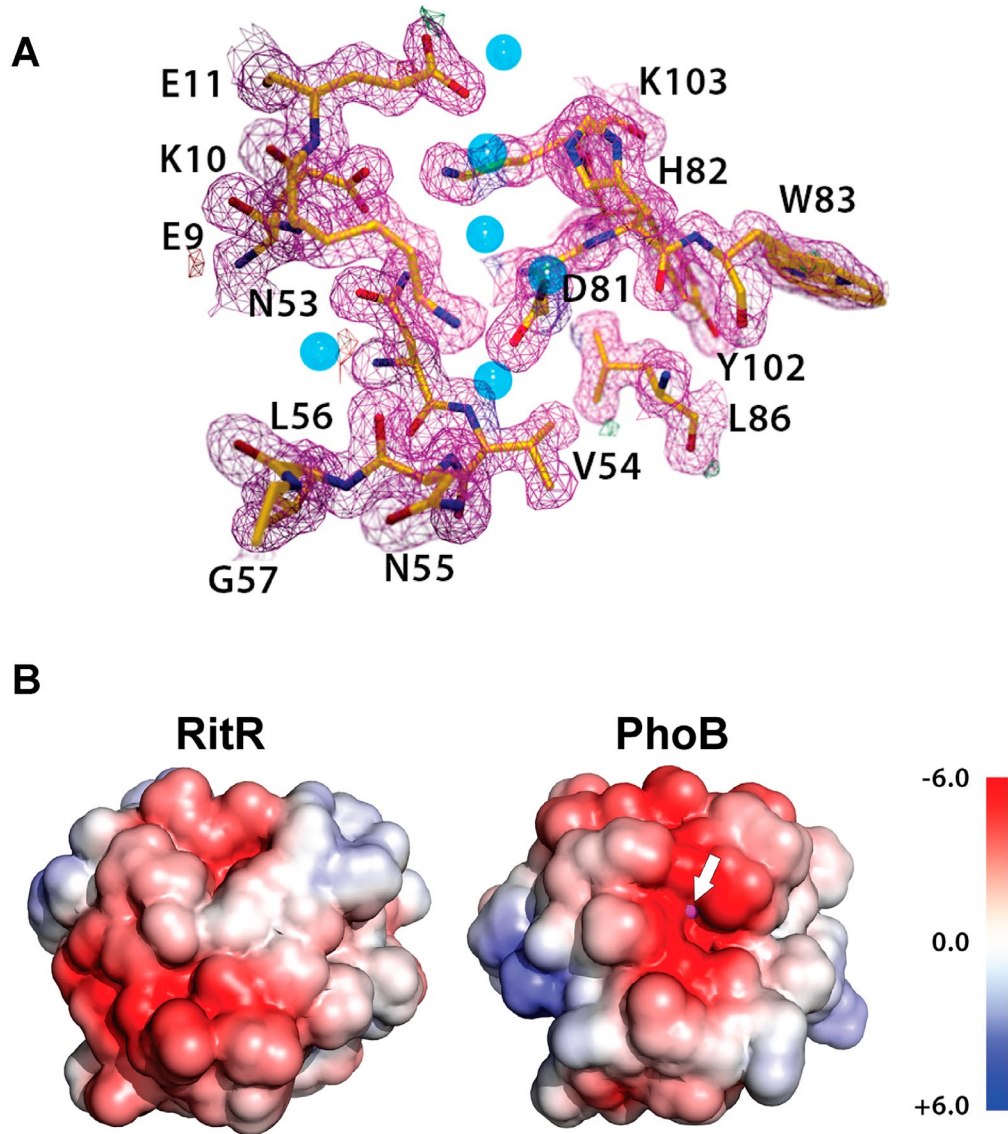
Figure 1-12: Structure of the RitR "active site" (this figure is downloaded from PLOS Pathogens 11(4): e1004795 [6]). The stereoview of the $2|F_o|-|F_c|$ electron density in the RitR putative phosphorylation site (magenta mesh) contoured at 1.5 σ (A) shows the network of interactions and the lack of a large electron density peak that would indicate bound metal. Water molecules are shown as blue spheres. A comparison of the vacuum electrostatic surface potentials of the RitR and PhoB REC domains (B) highlights the differences in the surface properties of RitR that account for the loss of metal binding. The Mg^{2+} in PhoB is shown as a magenta sphere and highlighted with a white arrow. Both the electronegative environment and the physical cleft for Mg^{2+} binding are lost in the RitR putative phosphorylation site.

1.2.4.2 Diguanylate cyclase Q15Z91

A search of response regulator sequences for ALR proteins that also possess a cysteine residue at the C-terminal end of the linker helix (α5) and an enzymatic output domain led to the identification of a protein known only by its Uniprot accession code, Q15Z91. This putative regulatory protein from *Pseudoalteromonas atlantica* Tc6 contains an ALR domain linked to a diguanylate cyclase effector domain. This protein was chosen for study in order to explore the mechanism of ALR activation in the context of enzymatic activity rather than DNA binding. An enzymatic activity was preferred due to the facts that (1) it is simpler to conceptualize the active and inactive states of an enzymatic activity and (2) it is significantly easier to quantify enzymatic activity than DNA-binding. A boon of the diguanylate cyclase domains in general is that the product, cyclic-di-guanosine monophosphate, is an important second messenger in biofilm formation. Thus, the biofilm phenotype is easily observed during heterologous expression in *E. coli* Origami™ 2 (DE3) cells, which will be discussed in detail in chapter 3.

The diguanylate cyclase (DGC) domain of Q15Z91 catalyzes the condensation of two molecules of GTP to yield bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) (Scheme 1-1). A separate phosphodiesterase enzyme cleaves c-di-GMP to give 5'-phosphoguanosylyl-(3'-5')-guanosine (pGpG). Cyclic di-GMP is an important second messenger that regulates biofilm formation, cell adhesion, and induction of virulence factors in some bacteria [87-89].
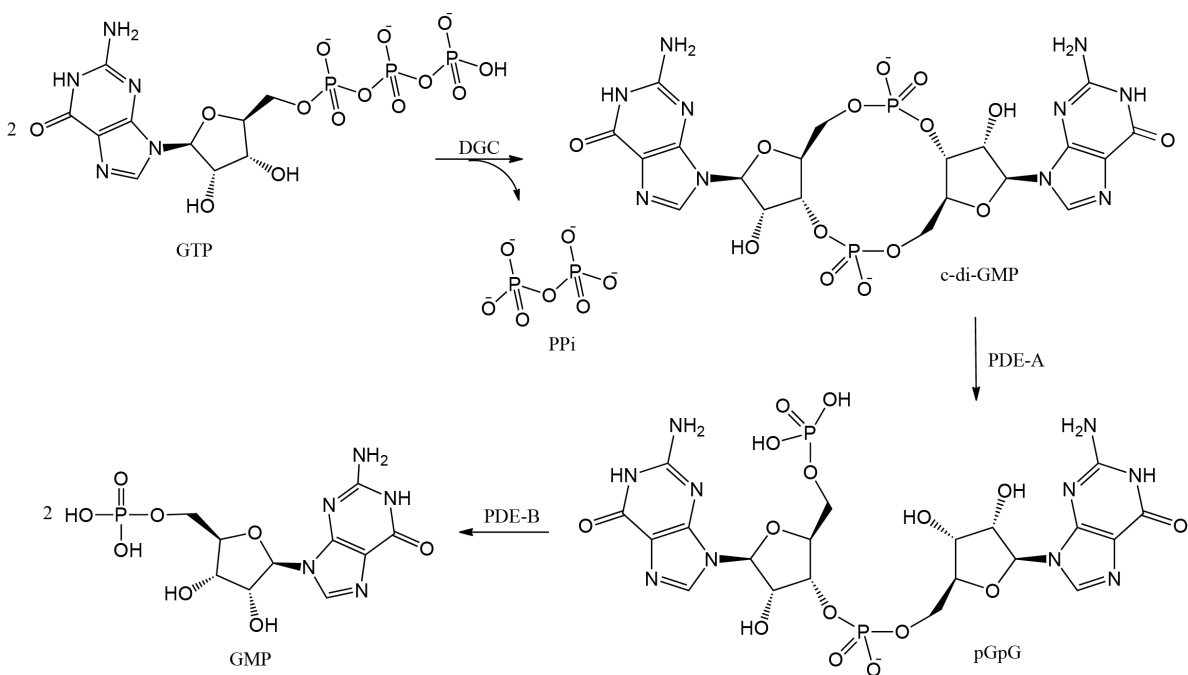
Scheme 1-1: Metabolism of cyclic-di-GMP. DGC: diguanylate cyclase; PDE-A: type A phosphodiesterase with EAL motif; PDE-B: type B phosphodiesterase with HD-GYP motif; pGpG: 5'-phosphoguanylyl-(3'-5')-guanosine.

The diguanylate cyclases usually contain a conserved GG(D/E)EF motif for catalysis and an RxxD motif for regulation of the DGC activity by product inhibition. A well-studied diguanylate cyclase, PleD from *Caulobacter crescentus*, is activated by phosphorylation [90]. As shown in Figure 1-13, PleD contains two CheY-like REC domains (a true receiver domain [D1] and an inactive "adapter" domain [D2]) and a DGC domain [91]. The receiver and adapter domains of PleD share the $(\beta/\alpha)_5$ topology common to typical OmpR/PhoB RRs. Upon phosphorylation of D53 in the acidic pocket of the REC domain (D1), the conserved switch residues T83 and F102 undergo a conformational change that triggers a rearrangement of the D1 and D2 domains relative to one another. This rearrangement

optimizes one face of the D1/D2 pair for dimerization between two PleD molecules [90, 92] (Figure 1-13A). Interestingly, the two REC domains (D1/D2) form an α4-β5-α5 dimer interface very similar to those seen in other OmpR/PhoB subfamily regulators, while the interactions between the PleD protomers includes residues from D1 domain (α1, β2, β3, α5) and D2 domain (α8 and α10).
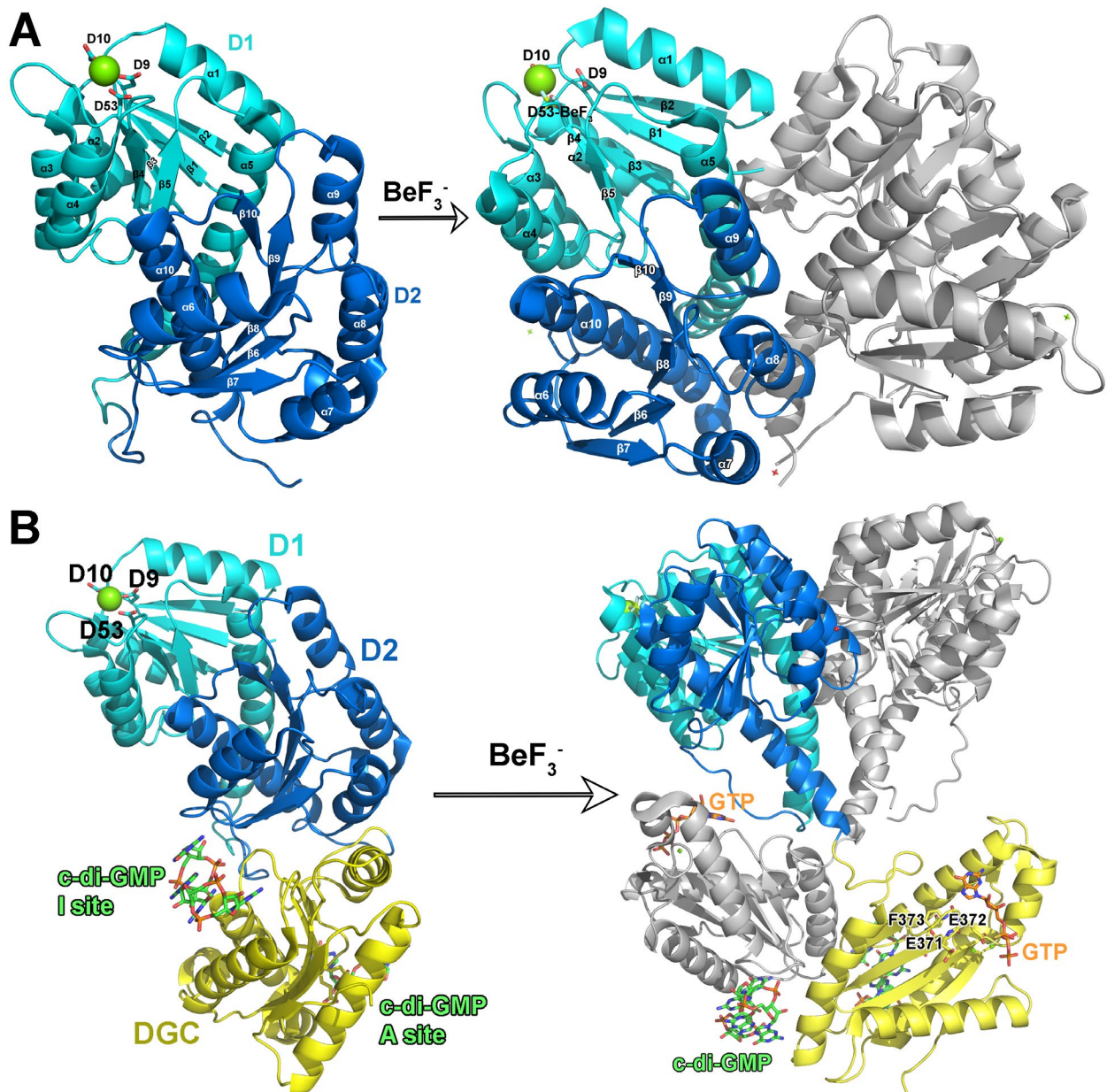
Figure 1-13: The receiver (D1) and adapter (D2) domains of PleD rearrange after BeF_3^- treatment (A). BeF_3^- was used to mimic phosphorylation of D53. The DGC domains were omitted for clarity. The acidic triad residues are shown as sticks. The Mg^{2+} ions are shown as green spheres. The inactive form of PleD is shown on the left, and the BeF_3 activated mimic structure is shown on the right. The D1 domain is colored cyan, and the D2 adapter domain is colored blue. In the activated PleD structure, the D1 and D2 domains of the second protomer in the dimer are colored grey. After D53 accepts the phosphoryl group, REC_1 and REC_2 form a tighter dimer interface. PleD dimerization (B) is promoted by phosphorylation and c-di-GMP binding immobilizes the DGC domain against the D1 and D2 domains. As above, the inactive PleD structure is shown on the left and the activated mimic on the right. The c-diGMP molecules are shown as sticks. In the inactive structure, c-di-GMP is bound to both the inhibitory (I) and active (A) sites. In the BeF_3 phospho-mimic structure of the PleD dimer, c-di-GMP is still bound in the inhibitory site, which restricts DGC domains from catalysis. Two GTP molecules (shown in stick) bind in each active site respectively.

The conserved DGC domain (GGEEF motif) consists of a five-stranded central β sheet surrounded by α helices, which is similar to the structure of adenylate cyclase and type I DNA polymerase. A single DGC domain binds only one GTP molecule, which means that a single domain contains only half of an active site. As shown in Scheme 1-2, for reactions to happen, two GTP molecules need to get closer and well aligned. Thus, activity requires the antiparallel association of two DGC domains, each with GTP bound. The regulatory logic is clear: antiparallel association of two DGC domains is much more likely in the dimeric form of PleD, where it is a unimolecular process, than it is in the case of free PleD monomers. In the active site of PleD, the conserved residues G368, G269 and E371 make key interactions for binding GTP, while E370 is essential for catalysis [91]. In PleD, inhibitory sites on the DGC domains, marked by an RxxD motif, bind intercalated c-di-GMP dimers that "pin" the DGC domains to the regulatory domains in the case of unactivated PleD, or to each other in an

unproductive back-to-back orientation in the case of activated (phosphorylated) PleD. This c-di-GMP dimer binding is the basis of the observed product inhibition that modulates DGC activity. Disruption of the inhibitory sites, for example by mutating the Arg or Gly residues of the RxxG motif, abrogates the product inhibition [93]. This allosteric feedback inhibition is not affected by phosphorylation status (Fig1-15B).
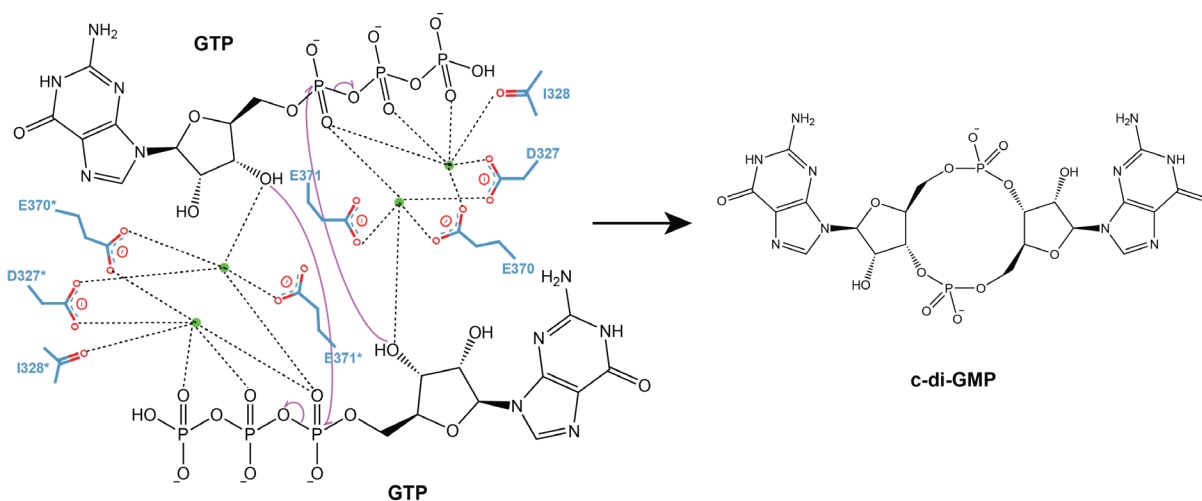
Scheme 1-2: Schematic representation of the diguanylate cyclase reaction where two GTP molecules are condensed to form c-di-GMP. The c-di-GMP contains two GMP moieties linked by O3'-P phosphodiester bonds. Four Mg^{2+} ions are shown as green spheres. The two GTP molecules are held in close proximity and with the proper alignment by four Mg^{2+} ions that are coordinated with the phosphate groups of GTP and the carboxyl groups of the conserved residues E370, E371, and D327. D327 is thought to promote deprotonation of the 3'-OH group of GTP to initiate the reaction. The deprotonated 3'-hydroxyl group would then attack the Pα-O bond, and release pyrophosphate (PPi). This is repeated in the other half of the active site so that a total of two molecules of PPi are released during each catalytic cycle.

The diguanylate cyclase WspR from *Pseudomonas aeruginosa* is very similar to PleD in that it is also regulated by feedback inhibition, though the exact mechanism differs [4]. WspR

contains a single CheY-like REC domain and a DGC domain (Figure 1-14A). These two domains are linked by a long helix α5 [4]. The REC domain has a similar structure to those of activated OmpR/PhoB RRs. The switch residue Y117 adopts an inward conformation, promoting association through the commonly observed α4-β5-α5 dimer interface (Figure 1-14B). However, no post-translational modification of WspR, such as phosphorylation, has been observed in the structure. *In vitro* activity study of Wsp and phosphorylated WspR showed that unphosphorylated WspR can catalyze c-di-GMP formation but phosphorylated WspR exhibited faster DGC activity [94]. WspR from *P. aeruginosa* functions as a compact, c-di-GMP-free dimer. However, other WspR homologs, such as that from *P. syringae,* require phosphorylation-induced tetramerization for high cyclase activity [95]. This activated cyclase activity was only obtained by phosphodiesterase treatment of WspR to remove c-di-GMP, followed by phosphorylation-induced tetramerization [95]. Both WspR exhibit a product immobilization inhibition. WspR from *P. aeruginosa* is purified as a mixture of dimer and tetramer [4, 95]. In the crystal lattice, WspR formed a distinct 222-symmetric tetramer, and the long helix α5 facilitates the antiparallel tetramer formation[4, 95], as shown in Figure 1-14A. Molecules of c-di-GMP dimers bind to the specific I sites (RxxD motif) that composed by two protomers from different dimers (e.g. chain A and chain C) in the tetramer and cross-link the DGC domains such that the active sites are forced apart for the cyclization of two GTP molecules.
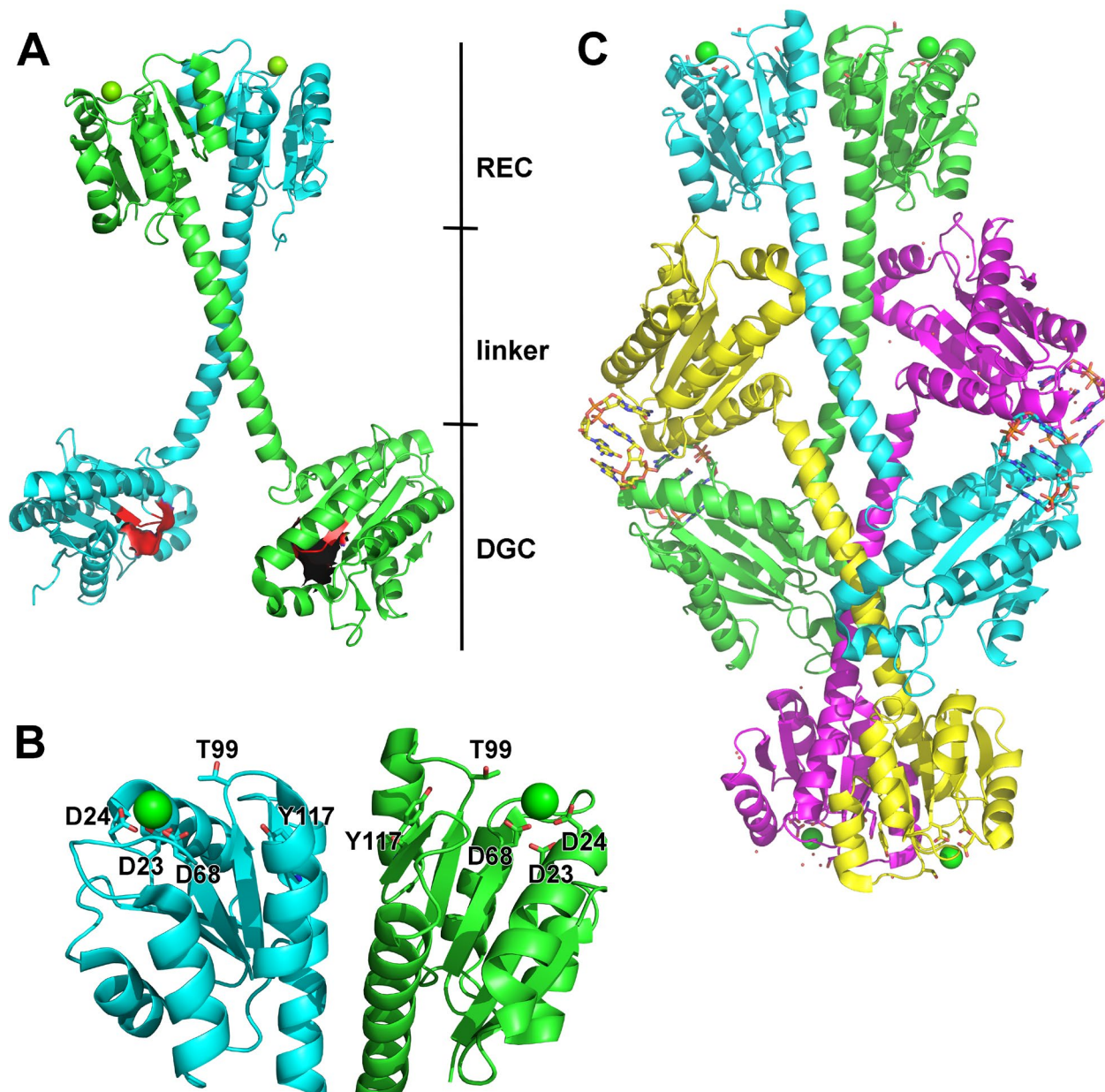
Figure 1-14: The Wsp dimer (A), REC domain of WspR dimer (B) and tetramer (C) structure from *Pseudomonas aeruginosa* (PDB ID: 3BRE [4]). Chain A is colored cyan, chain B is colored green, chain C is colored magenta, and chain D is colored yellow. Residues in phosphorylation site and switch residues are shown as sticks. The Mg^{2+} ions are represented with green spheres. The GTP binding active site are shown in red sphere in (A). Four c-di-GMP dimers in (C) are shown as sticks. In the crystal, two WspR dimers packed in an antiparallel way to form the inactive tetramer with c-di-GMP bound.

31

Both of these diguanylate cyclases, PleD and WspR, have the structural features required for phosphorylation-dependent activation. In the case of PleD, phosphorylation has been shown to be the primary regulatory factor, whereas with WspR, phosphorylation accelerates the DGC activity. In contrast, Q15Z91 has lost these structural features, such as the conserved, phosphorylatable aspartate residue and the Mg^{2+}-binding site. This prompted us to ask, "what is the mechanism of activation of Q15Z91?" Is it also controlled by the non-competitive allosteric product inhibition? These questions will be addressed in Chapter 3.

In summary, ALRs exists in most classes of bacteria (averaging ~2 per genome), and are also prevalent in archaea, fungi, plants and possibly even insects and primitive animals. The key "invariant" catalytic and signaling residues required in canonical REC domains are found to be changed in many ALR sequences. Although these atypical REC domains has been known to exist for some time, it remain unknown that how they are regulated in the absence of the conserved Asp. Combining structural data with biochemical and genetic examination of two founding members of the ALR family, RitR and Q15Z91, we demonstrate that ALRs might have evolved to accommodate more diverse environmental signals, while largely retaining their time-tested ancestral post-input signaling mechanisms. This work also provides a foundation for future experimentation to decipher the largely unknown functional roles of these proteins.

1.3 References

[1] Yoshida, T., Phadtare, S., and Inouye, M. (2007) Functional and structural characterization of EnvZ, an osmosensing histidine kinase of E. coli, *Methods in enzymology 423*, 184-202.

[2] Matsushika, A., and Mizuno, T. (1998) A dual-signaling mechanism mediated by the ArcB hybrid sensor kinase containing the histidine-containing phosphotransfer domain in Escherichia coli, *Journal of bacteriology 180*, 3973-3977.

[3] Robinson, V. L., Wu, T., and Stock, A. M. (2003) Structural analysis of the domain interface in DrrB, a response regulator of the OmpR/PhoB subfamily, *Journal of bacteriology 185*, 4186-4194.

[4] De, N., Pirruccello, M., Krasteva, P. V., Bae, N., Raghavan, R. V., and Sondermann, H. (2008) Phosphorylation-Independent Regulation of the Diguanylate Cyclase WspR, *PLoS biology 6*, e67.

[5] Bachhawat, P., Swapna, G. V. T., and Stock, A. M. (2005) Mechanism of Activation for Transcription Factor PhoB Suggested by Different Modes of Dimerization in the Inactive and Active States, *Structure (London, England : 1993) 13*, 1353-1363.

[6] Maule, A. F., Wright, D. P., Weiner, J. J., Han, L., Peterson, F. C., Volkman, B. F., Silvaggi, N. R., and Ulijasz, A. T. (2015) The Aspartate-Less Receiver (ALR) Domains: Distribution, Structure and Function, *PLOS Pathogens 11*, e1004795.

[7] Stephenson, K., and Hoch, J. A. (2002) Evolution of signalling in the sporulation phosphorelay, *Molecular microbiology 46*, 297-304.

[8] Hickey, J. M., Lovell, S., Battaile, K. P., Hu, L., Middaugh, C. R., and Hefty, P. S. (2011) The Atypical Response Regulator Protein ChxR Has Structural Characteristics and Dimer Interface Interactions That Are Unique within the OmpR/PhoB Subfamily, *The Journal of biological chemistry 286*, 32606-32616.

[9] Sola, M., Gomis-Ruth, F. X., Serrano, L., Gonzalez, A., and Coll, M. (1999) Three-dimensional crystal structure of the transcription factor PhoB receiver domain, *Journal of molecular biology 285*, 675-687.

[10] Buckler, D. R., Zhou, Y., and Stock, A. M. (2002) Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from Thermotoga maritima, *Structure 10*, 153-164.

[11] Hong, E., Lee, H. M., Ko, H., Kim, D. U., Jeon, B. Y., Jung, J., Shin, J., Lee, S. A., Kim, Y., Jeon, Y. H., Cheong, C., Cho, H. S., and Lee, W. (2007) Structure of an atypical orphan response regulator protein supports a new phosphorylation-independent regulatory mechanism, *The Journal of biological chemistry 282*, 20667-20675.

[12] Barta, M. L., Hickey, J. M., Anbanandam, A., Dyer, K., Hammel, M., and Hefty, P. S. (2014) Atypical Response Regulator ChxR from Chlamydia trachomatis Is Structurally Poised for DNA Binding, *PLOS ONE 9*, e91760.

[13] Lin, W., Wang, Y., Han, X., Zhang, Z., Wang, C., Wang, J., Yang, H., Lu, Y., Jiang, W., Zhao, G. P., and Zhang, P. (2014) Atypical OmpR/PhoB subfamily response regulator GlnR of actinomycetes functions as a homodimer, stabilized by the unphosphorylated conserved Asp-focused charge interactions, *The Journal of biological chemistry 289*, 15413-15425.

[14] Bachhawat, P., Swapna, G. V., Montelione, G. T., and Stock, A. M. (2005) Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states, *Structure 13*, 1353-1363.

[15] Ulrich, L. E., Koonin, E. V., and Zhulin, I. B. (2005) One-component systems dominate signal transduction in prokaryotes, *Trends in Microbiology 13*, 52-56.

[16] Vannini, A., Volpari, C., Gargioli, C., Muraglia, E., Cortese, R., De Francesco, R., Neddermann, P., and Di Marco, S. (2002) The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA, *The EMBO Journal 21*, 4393-4401.

[17] Zhu, J., Oger, P. M., Schrammeijer, B., Hooykaas, P. J., Farrand, S. K., and Winans, S. C. (2000) The bases of crown gall tumorigenesis, *Journal of bacteriology 182*, 3885-3895.

[18] Salvi, M., Schomburg, B., Giller, K., Graf, S., Unden, G., Becker, S., Lange, A., and Griesinger, C. (2017) Sensory domain contraction in histidine kinase CitA triggers transmembrane signaling in the membrane-bound sensor, *Proceedings of the National Academy of Sciences 114*, 3115-3120.

[19] Müller, S., Götz, M., and Beier, D. (2009) Histidine Residue 94 Is Involved in pH Sensing by Histidine Kinase ArsS of Helicobacter pylori, *PLoS ONE 4*, e6930.

[20] Moker, N., Reihlen P Fau - Kramer, R., Kramer R Fau - Morbach, S., and Morbach, S. Osmosensing properties of the histidine protein kinase MtrB from Corynebacterium glutamicum.

[21] Braun, Y., Smirnova Av Fau - Weingart, H., Weingart H Fau - Schenk, A., Schenk A Fau - Ullrich, M. S., and Ullrich, M. S. A temperature-sensing histidine kinase: function, genetics, and membrane topology.

[22] Zhang, X. X., Gauntlett, J. C., Oldenburg, D. G., Cook, G. M., and Rainey, P. B. Role of the Transporter-Like Sensor Kinase CbrA in Histidine Uptake and Signal Transduction.

[23] Buelow, D. R., and Raivio, T. L. (2010) Three (and more) component regulatory systems - auxiliary regulators of bacterial histidine kinases, *Molecular microbiology 75*, 547-566.

[24] Kwon, O., Georgellis D Fau - Lin, E. C., and Lin, E. C. Phosphorelay as the sole physiological route of signal transmission by the arc two-component system of Escherichia coli.

[25] Thakor, H., Nicholas, S., Porter, I. M., Hand, N., and Stewart, R. C. (2011) Identification of an anchor residue for CheA-CheY interactions in the chemotaxis system of Escherichia coli, *Journal of bacteriology 193*, 3894-3903.

[26] Dworkin, J. (2015) Ser/Thr phosphorylation as a regulatory mechanism in bacteria, *Current opinion in microbiology 24*, 47-52.

[27] Pereira, S. F. F., Goss, L., and Dworkin, J. (2011) Eukaryote-Like Serine/Threonine Kinases and Phosphatases in Bacteria, *Microbiology and Molecular Biology Reviews : MMBR 75*, 192-212.

[28] Kalantari, A., Derouiche, A., Shi, L., and Mijakovic, I. (2015) Serine/threonine/tyrosine phosphorylation regulates DNA binding of bacterial transcriptional regulators, *Microbiology 161*, 1720-1729.

[29] Nambu, J. R., Lewis, J. O., Wharton, K. A., and Crews, S. T. (1991) The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development, *Cell 67*, 1157-1167.

[30] Yoshida, T., Cai, S. j., and Inouye, M. Interaction of EnvZ, a sensory histidine kinase, with phosphorylated OmpR, the cognate response regulator.

[31] Wang, L. C., Morgan, L. K., Godakumbura, P., Kenney, L. J., and Anand, G. S. (2012) The inner membrane histidine kinase EnvZ senses osmolality via helix‐coil transitions in the cytoplasm, *The EMBO Journal 31*, 2648.

[32] Stock, A., Chen, T., Welsh, D., and Stock, J. (1988) CheA protein, a central regulator of bacterial chemotaxis, belongs to a family of proteins that control gene expression in response to changing environmental conditions, *Proceedings of the National Academy of Sciences of the United States of America 85*, 1403-1407.

[33] MacFarlane, S. A., and Merrick, M. (1985) The nucleotide sequence of the nitrogen regulation gene ntrB and the glnA-ntrBC intergenic region of Klebsiella pneumoniae, *Nucleic acids research 13*, 7591-7606.

[34] Taylor, B. L., and Zhulin, I. B. (1999) PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light, *Microbiology and Molecular Biology Reviews 63*, 479-506.

[35] Möglich, A., Ayers, R. A., and Moffat, K. (2009) Structure and Signaling Mechanism of Per-ARNT-Sim Domains, *Structure 17*, 1282-1294.

[36] Wang, C., Sang, J., Wang, J., Su, M., Downey, J. S., Wu, Q., Wang, S., Cai, Y., Xu, X., Wu, J., Senadheera, D. B., Cvitkovitch, D. G., Chen, L., Goodman, S. D., and Han, A. (2013) Mechanistic insights revealed by the crystal structure of a histidine kinase with signal transducer and sensor domains, *PLoS biology 11*, e1001493.

[37] Gutu, A. D., Wayne, K. J., Sham, L. T., and Winkler, M. E. (2010) Kinetic characterization of the WalRKSpn (VicRK) two-component system of Streptococcus pneumoniae: dependence of WalKSpn (VicK) phosphatase activity on its PAS domain, *Journal of bacteriology 192*, 2346-2358.

[38] Casino, P., Miguel-Romero, L., and Marina, A. (2014) Visualizing autophosphorylation in histidine kinases, *5*, 3258.

[39] Trajtenberg, F., Grana, M., Ruetalo, N., Botti, H., and Buschiazzo, A. (2010) Structural and enzymatic insights into the ATP binding and autophosphorylation mechanism of a sensor histidine kinase, *The Journal of biological chemistry 285*, 24892-24903.

[40] Marina, A., Waldburger, C. D., and Hendrickson, W. A. (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein, *Embo j 24*, 4247-4259.

[41] Wigley, D. B., Davies, G. J., Dodson, E. J., Maxwell, A., and Dodson, G. (1991) Crystal structure of an N-terminal fragment of the DNA gyrase B protein, *Nature 351*, 624-629.

[42] Ban, C., and Yang, W. (1998) Crystal Structure and ATPase Activity of MutL: Implications for DNA Repair and Mutagenesis, *Cell 95*, 541-552.

[43] Prodromou, C., Roe, S. M., O'Brien, R., Ladbury, J. E., Piper, P. W., and Pearl, L. H. (1997) Identification and structural characterization of the ATP/ADP-binding site in the Hsp90 molecular chaperone, *Cell 90*, 65-75.

[44] Wolanin, P. M., Thomason, P. A., and Stock, J. B. (2002) Histidine protein kinases: key signal transducers outside the animal kingdom, *Genome Biology 3*, reviews3013.3011-reviews3013.3018.

[45] Martínez-Hackert, E., and Stock, A. M. (1997) Structural relationships in the OmpR family of winged-helix transcription factors11Edited by M. Gottesman, *Journal of molecular biology 269*, 301-312.

[46] King-Scott, J., Nowak, E., Mylonas, E., Panjikar, S., Roessle, M., Svergun, D. I., and Tucker, P. A. (2007) The structure of a full-length response regulator from Mycobacterium tuberculosis in a stabilized three-dimensional domain-swapped, activated state, *The Journal of biological chemistry 282*, 37717-37729.

[47] Menon, S., and Wang, S. (2011) Structure of the response regulator PhoP from Mycobacterium tuberculosis reveals a dimer through the receiver domain, *Biochemistry 50*, 5948-5957.

[48] Park, A. K., Lee, J. H., Chi, Y. M., and Park, H. (2016) Structural characterization of the full-length response regulator spr1814 in complex with a phosphate analogue reveals a novel conformational plasticity of the linker region, *Biochemical and biophysical research communications 473*, 625-629.

[49] Gardner, S. G., Miller, J. B., Dean, T., Robinson, T., Erickson, M., Ridge, P. G., and McCleary, W. R. (2015) Genetic analysis, structural modeling, and direct coupling analysis suggest a mechanism for phosphate signaling in Escherichia coli, *BMC Genetics 16*, S2.

[50] Qi, Y., Kobayashi, Y., and Hulett, F. M. (1997) The pst operon of Bacillus subtilis has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the pho regulon, *Journal of bacteriology 179*, 2534-2539.

[51] Barbieri, C. M., Mack, T. R., Robinson, V. L., Miller, M. T., and Stock, A. M. (2010) Regulation of response regulator autophosphorylation through interdomain contacts, *The Journal of biological chemistry 285*, 32325-32335.

[52] Galperin, M. Y. (2010) Diversity of structure and function of response regulator output domains, *Current Opinion in Microbiology 13*, 150-159.

[53] Wadhams, G. H., and Armitage, J. P. (2004) Making sense of it all: bacterial chemotaxis, *Nature reviews. Molecular cell biology 5*, 1024-1037.

[54] Sarkar, M. K., Paul, K., and Blair, D. (2010) Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in Escherichia coli, *Proceedings of the National Academy of Sciences 107*, 9370-9375.

[55] Greer-Phillips, S. E., Alexandre, G., Taylor, B. L., and Zhulin, I. B. (2003) Aer and Tsr guide Escherichia coli in spatial gradients of oxidizable substrates, *Microbiology 149*, 2661-2667.

[56] Bourret, R. B., Davagnino, J., and Simon, M. I. (1993) The carboxy-terminal portion of the CheA kinase mediates regulation of autophosphorylation by transducer and CheW, *Journal of bacteriology 175*, 2097-2101.

[57] Griswold, I. J., Zhou, H., Matison, M., Swanson, R. V., McIntosh, L. P., Simon, M. I., and Dahlquist, F. W. (2002) The solution structure and interactions of CheW from Thermotoga maritima, *Nat Struct Mol Biol 9*, 121-125.

[58] Borkovich, K. A., Alex, L. A., and Simon, M. I. (1992) Attenuation of sensory receptor signaling by covalent modification, *Proc Natl Acad Sci U S A 89*, 6756-6760.

[59] Lee, S.-Y., Cho, H. S., Pelton, J. G., Yan, D., Henderson, R. K., King, D. S., Huang, L.-s., Kustu, S., Berry, E. A., and Wemmer, D. E. (2001) Crystal structure of an activated response regulator bound to its target, *Nat Struct Mol Biol 8*, 52-56.

[60] Lee, S. Y., Cho Hs Fau - Pelton, J. G., Pelton Jg Fau - Yan, D., Yan D Fau - Berry, E. A., Berry Ea Fau - Wemmer, D. E., and Wemmer, D. E. Crystal structure of activated CheY. Comparison with other activated receiver domains.

[61] Volz, K., and Matsumura, P. (1991) Crystal structure of Escherichia coli CheY refined at 1.7-A resolution, *The Journal of biological chemistry 266*, 15511-15519.

[62] Lupas, A., and Stock, J. (1989) Phosphorylation of an N-terminal regulatory domain activates the CheB methylesterase in bacterial chemotaxis, *The Journal of biological chemistry 264*, 17337-17342.

[63] Djordjevic, S., Goudreau, P. N., Xu, Q., Stock, A. M., and West, A. H. (1998) Structural basis for methylesterase CheB regulation by a phosphorylation-activated domain, *Proceedings of the National Academy of Sciences of the United States of America 95*, 1381-1386.

[64] Bourret, R. B., Hess, J. F., and Simon, M. I. (1990) Conserved aspartate residues and phosphorylation in signal transduction by the chemotaxis protein CheY, *Proc Natl Acad Sci U S A 87*, 41-45.

[65] Gupte, G., Woodward, C., and Stout, V. (1997) Isolation and characterization of rcsB mutations that affect colanic acid capsule synthesis in Escherichia coli K-12, *Journal of bacteriology 179*, 4328-4335.

[66] Klose, K. E., Weiss, D. S., and Kustu, S. (1993) Glutamate at the site of phosphorylation of nitrogen-regulatory protein NTRC mimics aspartyl-phosphate and activates the protein, *Journal of molecular biology 232*, 67-78.

[67] Zhu, X., Rebello, J., Matsumura, P., and Volz, K. (1997) Crystal structures of CheY mutants Y106W and T87I/Y106W. CheY activation correlates with movement of residue 106, *The Journal of biological chemistry 272*, 5000-5006.

[68] Sanders, D. A., Gillece-Castro, B. L., Stock, A. M., Burlingame, A. L., and Koshland, D. E., Jr. (1989) Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY, *The Journal of biological chemistry 264*, 21770-21778.

[69] Appleby, J. L., and Bourret, R. B. (1999) Activation of CheY mutant D57N by phosphorylation at an alternative site, Ser-56, *Molecular microbiology 34*, 915-925.

[70] Klose, K. E., Weiss, D. S., and Kustu, S. (1993) Glutamate at the Site of Phosphorylation of Nitrogen-regulatory Protein NTRC Mimics Aspartyl-Phosphate and Activates the Protein, *Journal of molecular biology 232*, 67-78.

[71] Moore, J. B., Shiau, S. P., and Reitzer, L. J. (1993) Alterations of highly conserved residues in the regulatory domain of nitrogen regulator I (NtrC) of Escherichia coli, *Journal of bacteriology 175*, 2692-2701.

[72] Lee, S. Y., Cho, H. S., Pelton, J. G., Yan, D., Berry, E. A., and Wemmer, D. E. (2001) Crystal structure of activated CheY. Comparison with other activated receiver domains, *The Journal of biological chemistry 276*, 16425-16431.

[73] Drake, S. K., Bourret, R. B., Luck, L. A., Simon, M. I., and Falke, J. J. (1993) Activation of the phosphosignaling protein CheY. I. Analysis of the phosphorylated conformation by 19F NMR and protein engineering, *The Journal of biological chemistry 268*, 13081-13088.

[74] Lukat, G. S., Lee, B. H., Mottonen, J. M., Stock, A. M., and Stock, J. B. (1991) Roles of the highly conserved aspartate and lysine residues in the response regulator of bacterial chemotaxis, *The Journal of biological chemistry 266*, 8348-8354.

[75] Stewart, R. C. (1993) Activating and inhibitory mutations in the regulatory domain of CheB, the methylesterase in bacterial chemotaxis, *The Journal of biological chemistry 268*, 1921-1930.

[76] Hao, Z., Lou, H., Zhu, R., Zhu, J., Zhang, D., Zhao, B. S., Zeng, S., Chen, X., Chan, J., He, C., and Chen, P. R. (2013) The multiple antibiotic resistance regulator MarR is a copper sensor in Escherichia coli, *Nature Chemical Biology 10*, 21.

[77] Ruiz, D., Salinas, P., Lopez-Redondo, M. L., Cayuela, M. L., Marina, A., and Contreras, A. (2008) Phosphorylation-independent activation of the atypical response regulator NblR, *Microbiology 154*, 3002-3015.

[78] Kato, H., Chibazakura, T., and Yoshikawa, H. (2008) NblR Is a Novel One-Component Response Regulator in the Cyanobacterium <I>Synechococcus elongatus</I> PCC 7942, *Bioscience, Biotechnology, and Biochemistry 72*, 1072-1079.

[79] Hickey, J. M., Weldon, L., and Hefty, P. S. (2011) The atypical OmpR/PhoB response regulator ChxR from Chlamydia trachomatis forms homodimers in vivo and binds a direct repeat of nucleotide sequences, *Journal of bacteriology 193*, 389-398.

[80] Malm, S., Tiffert, Y., Micklinghoff, J., Schultze, S., Joost, I., Weber, I., Horst, S., Ackermann, B., Schmidt, M., Wohlleben, W., Ehlers, S., Geffers, R., Reuther, J., and Bange, F. C. (2009) The roles of the nitrate reductase NarGHJI, the nitrite reductase NirBD and the response regulator GlnR in nitrate assimilation of Mycobacterium tuberculosis, *Microbiology 155*, 1332-1339.

[81] Schar, J., Sickmann, A., and Beier, D. (2005) Phosphorylation-independent activity of atypical response regulators of Helicobacter pylori, *Journal of bacteriology 187*, 3100-3109.

[82] Lange, R., Wagner, C., de Saizieu, A., Flint, N., Molnos, J., Stieger, M., Caspers, P., Kamber, M., Keck, W., and Amrein, K. E. (1999) Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of Streptococcus pneumoniae, *Gene 237*, 223-234.

[83] Tettelin, H., Nelson, K. E., Paulsen, I. T., Eisen, J. A., Read, T. D., Peterson, S., Heidelberg, J., DeBoy, R. T., Haft, D. H., Dodson, R. J., Durkin, A. S., Gwinn, M., Kolonay, J. F., Nelson, W. C., Peterson, J. D., Umayam, L. A., White, O., Salzberg, S. L., Lewis, M. R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A. M., Utterback, T. R., Hansen, C. L., McDonald, L. A., Feldblyum, T. V., Angiuoli, S., Dickinson, T., Hickey, E. K., Holt, I. E., Loftus, B. J., Yang, F., Smith, H. O., Venter, J. C., Dougherty, B. A., Morrison, D. A., Hollingshead, S. K., and Fraser, C. M.

[84] Throup, J. P., Koretke, K. K., Bryant, A. P., Ingraham, K. A., Chalker, A. F., Ge, Y., Marra, A., Wallis, N. G., Brown, J. R., Holmes, D. J., Rosenberg, M., and Burnham, M. K. (2000) A genomic analysis of two-component signal transduction in Streptococcus pneumoniae, *Molecular microbiology 35*, 566-576.

[85] Ulijasz, A. T., Andes, D. R., Glasner, J. D., and Weisblum, B. (2004) Regulation of iron transport in Streptococcus pneumoniae by RitR, an orphan response regulator, *Journal of bacteriology 186*, 8123-8136.

[86] Wisedchaisri, G., Wu, M., Sherman, D. R., and Hol, W. G. J. (2008) Crystal Structures of the Response Regulator DosR from Mycobacterium tuberculosis Suggest a Helix Rearrangement Mechanism for Phosphorylation Activation, *Journal of molecular biology 378*, 227-242.

[87] Jenal, U., and Malone, J. (2006) Mechanisms of cyclic-di-GMP signaling in bacteria, *Annual review of genetics 40*, 385-407.

[88] Tischler, A. D., and Camilli, A. (2004) Cyclic diguanylate (c-di-GMP) regulates Vibrio cholerae biofilm formation, *Molecular microbiology 53*, 857-869.

[89] Ryan, R. P. (2013) Cyclic di-GMP signalling and the regulation of bacterial virulence, *Microbiology 159*, 1286-1297.

[90] Paul, R., Abel, S., Wassmann, P., Beck, A., Heerklotz, H., and Jenal, U. (2007) Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization, *The Journal of biological chemistry 282*, 29170-29177.

[91] Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U., and Schirmer, T. (2004) Structural basis of activity and allosteric control of diguanylate cyclase, *Proc Natl Acad Sci U S A 101*, 17084-17089.

[92] Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U., and Schirmer, T. (2007) Structure of BeF3−-Modified Response Regulator PleD: Implications for Diguanylate Cyclase Activation, Catalysis, and Feedback Inhibition, *Structure 15*, 915-927.

[93] Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M., and Jenal, U. (2006) Allosteric control of cyclic di-GMP signaling, *The Journal of biological chemistry 281*, 32015-32024.

[94] Hickman, J. W., Tifrea, D. F., and Harwood, C. S. (2005) A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels, *Proc Natl Acad Sci U S A 102*, 14422-14427.

[95] De, N., Navarro, M. V., Raghavan, R. V., and Sondermann, H. (2009) Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR, *Journal of molecular biology 393*, 619-633.

Chapter 2

The Atypical Response Regulator Protein Repressor of Iron Transport Regulator (RitR) Exhibits an Alternative Mechanism of Activation

2.1 Background

RitR is an orphan response regulator in *S. pneumonia* R6 that has been shown to down-regulate the expression of iron-transport proteins to avoid oxidative damage during aerobic growth [1, 2]. RitR differs from typical two-component systems in that it lacks a cognate HK protein, the conserved phosphorylatable aspartate residue (D53) is substituted with asparagine, and the binding site for Mg^{2+}, which is required for the phosphorylation reaction, is apparently disrupted in RitR. Together, these differences suggest that RitR cannot be phosphorylated and, thus, must use a different mechanism of activation. So My work is using biophysical tools to determine the regulatory mechanisms of proteins with redox-sensitive ALR domains.

2.2 Methods

2.2.1 Cloning, expression, and purification of wild-type and mutant RitR

The gene encoding wild-type RitR from *Streptococcus pneumoniae* strain R800 was cloned into the pE-SUMO$_{kan}$ expression vector (LifeSensors Inc, Malvern, PA) using primers containing BsaI and XbaI restriction sites. The His$_6$-tagged SUMO-RitR fusion protein was expressed from *E. coli* BL21 Star (DE3) cells (Invitrogen Inc, Carlsbad, CA) carrying the pE-

SUMO-RitR plasmid. Cultures were grown in Luria-Bertani medium with 50 μg/mL kanamycin at 37 °C with agitation at 250 rpm. When the cultures reached an OD_{600} of ~1.0, protein expression was induced with 0.4 mM IPTG. The temperature was then reduced to 16 °C and the cultures were grown overnight. Cells were harvested by centrifugation, resuspended in 5 mL/g of buffer A (25 mM TRIS pH 8.0, 300 mM NaCl, 10 mM imidazole) supplemented with 0.1 mg/mL DNAse I (Worthington Biochemical Corp., Lakewood, NJ). Cells were lysed using a Branson Sonifier S-450 cell disruptor (Branson Ultrasonics Corp., Danbury, CT) for a total of 5 min of sonication at 60% amplitude with 30 s pulses separated by 50 s rest periods. The temperature was maintained at or below 4 °C by suspending the steel beaker in an ice bath directly over a spinning stir bar. The lysate was clarified by centrifugation at 39,000 x g for 45 minutes and then applied to a 5 mL HisTrap column (GE Lifesciences, Piscataway, NJ) at a flow rate of 5 mL/min to isolate the His_6-SUMO-RitR fusion protein. The protein was eluted by a 4-step gradient of buffer B (25 mM TRIS pH 8.0, 300 mM NaCl, and 250 mM imidazole; 5, 15, 50, and 100% B). The His_6-SUMO-RitR fusion protein eluted in the third and fourth steps and was ~90% pure, as judged on coomassie-stained SDS-PAGE gels. Peak fractions were pooled and dialyzed overnight against 4 L of 25 mM TRIS pH 8.0, 150 mM NaCl, and ~3 μM SUMO protease (LifeSensors Inc). The dialysate was passed through the HisTrap column a second time to remove the cleaved His_6-SUMO tag as well as the protease. The resulting RitR preparation was >95% pure according to SDS-PAGE results. The protein was desalted using a HiTrap Desalting column (GE Lifesciences) into storage buffer (pH 6.5, 25 mM TRIS, 150 mM NaCl, 100 mM Malic acid) and stored at -80 °C. All mutants mentioned in the chapter 2 were expressed and purified using the same protocol as the wild type RitR.

2.2.2 Preparation of purification of the oxidized RitR dimer (RitR$_{ox}$)

RitR$_{ox}$ was prepared by treating purified, wild-type RitR with 1 mM hydrogen peroxide for 3 h, followed by size exclusion chromatography to obtain the oxidized dimer. A HiPrep 26/60 Sephacryl S100 HR column was equilibrated with 2 column volumes of buffer C (pH 6.5, 20 mM TRIS, 150 mM NaCl, 100 mM malic acid, 1 mM H$_2$O$_2$). The first peak to elute from the column, which was consistent in size with the dimer (~60 kDa), contained the dimeric RitR$_{ox}$ protein.

2.2.3 Preparation of protein for structure determination

The mutant SeMet labeled RitR(C128S) from T7 Express Crystal cells (New England BioLabs, Ipswich, MA) carrying the pE-SUMO-C128S plasmid. Cells were grown in SelenoMethionine Medium Complete (Molecular Dimensions, Newmarket, Suffolk, UK) with 50 µg/mL kanamycin. This selenomethionine-labelled RitR(C128S) mutant was purified using the same protocol described for RitR. But it was stored in 25 mM TRIS pH 8.0, 5 mM DTT for crystallization. The mutant RitR(C128D) was purified using the same protocol described for RitR. But it was stored in 10 mM ADA pH 7.0 buffer.

Two forms of a RitR(L86A/V93A) ALR domain were devised to probe the structural consequences of this double mutation: one corresponding to the original RitR ALR domain structure, which ends well before the putative redox-active C128 residue (RitR(L86A/V93A) residues 1-124; RDLV) and a longer version that included the entire α5 helix (residues 1-132; RD2LV). Selenomethionine-labelled protein was prepared for both

double mutants from T7 Express Crystal cells (New England BioLabs, Ipswich, MA) carrying the pE-SUMO-RDLV or pE-SUMO-RD2LV plasmid. Cells were grown in SelenoMethionine Medium Complete (Molecular Dimensions, Newmarket, Suffolk, UK) with 50 µg/mL kanamycin. Both selenomethionine-labelled double mutant RitR$_{L86A/V93A}$ REC domains were purified using the same protocol described for RitR.

2.2.4 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) was used to search for buffer conditions where the relatively unstable WT RitR might be stable enough to crystallize. Wild-type RitR was diluted to 20 µM with water (1.1ml final volume) and 2µl of SYPRO Orange dye was added. This protein solution (10 µl) was mixed 1:1 with the Slice pH buffer screen (Hampton Research) in a 96-well PCR plate. The plate was heated from 20°C to 95°C in 1 °C steps in an Eppendorf RealPlex real-time PCR machine. At each temperature step, the fluorescence signal from the SYPRO Orange dye (ThermoFisher Scientific) in each well of the plate was monitored. Plotting the second derivative of the fluorescence intensity vs temperature plot gives a peak corresponding to the T_m of the protein in that buffer. Buffers showing at least a 2-3 °C upward shift in the T_m were investigated as storage buffers for RitR.

DSF was also used to assess the ability of RitR and various mutants to bind to DNA by monitoring the ability of DNA to stabilize the proteins. To test the buffer and pH effects on the T_m, RitR and RitR(L86A/V93A) were incubated with a dsDNA oligo corresponding to the putative primary RitR binding site in *S. pneumoniae* R6 (BS3-3: 5'-

ATGAAACTTTCTCATTATTTAATTGGCTTACTTCT-3') in different buffers at different pH. Firstly, 1500 µl solution containing 200 µM protein and 100 µM BS3-3 was incubated at room temperature for 2 hours. Then 2 µl of the protein/BS3-3 mixture was mixed with 18 µl of different buffers at different pH containing the SYPRO Orange dye. Protein-only negative controls were prepared by mixing 200 µM RitR or RitR(L86A/V93A) protein with different buffers only containing SYPRO Orange dye.

Finally, DSF was used to double check the gel shift results by looking at the shifts in T_m when different forms of RitR were incubated with BS3-3 in the binding buffer[1] (20 mM HEPES pH 7.2, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1 mM EDTA) and with the optimal protein:DNA ratio for binding (protein:DNA = 4:1). The protein variants investigated included $RitR_{ox}$, $RitR(L86A/V93A)_{ox}$ and $RitR(L86A/V93A/C128S)_{ox}$. Oxidized proteins were prepared by incubating with 1mM H_2O_2 at 4 °C overnight. Each RitR variant (RitR, RitRox, $RitR_{L86A/V93A}$, $RitR(L86A/V93A)_{ox}$, RitR(L86A/V93A/C128S), and $RitR(L86A/V93A/C128S)_{ox}$) was mixed with BS3-3 at final concentrations of 400 µM protein and 100µM DNA in 25 mM TRIS pH 8.0, 150 mM NaCl buffer for 2 hours at room temperature. Then 2 µl of each protein/BS3-3 mixture was mixed with 18 µl of different buffers (buffer 1: 20 mM TRIS pH 6.5, 150 mM NaCl, 100 mM Malic acid; buffer 2: binding buffer; or buffer 3: binding buffer supplemented with 150 mM NaCl) containing SYPRO Orange dye. The protein-only controls were prepared by mixing 400 µM protein with buffer with SYPRO™ orange fluorescent dye.

2.2.5 Site-directed mutagenesis

The K10A, N53A, N53D, D81A, L86A, L90A, V93A, L86A/V93A, F96G, Y100A, Y102A, R113A, C128A, C128S, C128Dand Y163F mutants were prepared by the Ulijasz lab. Other mutants, L86A/V93A/C128S, RDLV and RD2LV, were constructed as part of this work. The full-length mutants were obtained by the Quick-Change™ mutagenesis method using the primers listed in Table 2-1. The Quick-Change™ protocol used the Pfusion DNA polymerase and the high fidelity (HF) master mix (New England Biolabs, Ipswich, MA). The PCR program included an initial denaturing step at 98 °C for 30s, followed by the cyclic portion of the program (denature at 98°C for 10 s, anneal at (T_m of primer pair – 2) °C for 25 s, extend for 210 s at 72 °C; 17 cycles), and a final extension step at 72 °C for 10 min. After PCR, DpnI (20 U) was added to the 50 µl PCR reaction to digest the methylated template DNA at 37 °C for 1 h. The DpnI-treated PCR reaction (5 µl) was used to transform NEB5α competent *E. coli* (New England Biolabs, Ipswich, MA). Transformants were sequenced to confirm the presence of the mutation. The gene of two ALR domains of RitR(L86A/V93A) were amplified from pSUMO-RitR(L86A/V93A) plasmid. After BsaI and XbaI digestion, the genes were ligated into XbaI digested pSUMO kan plasmid. The success of ligation was checked by sequencing.

Table 2-1: Primers used for mutagenesis

Names	Primer sequence
sumo-RitRL86A_RD_Fwd	GCGGTCTCAAGGTATGGGGAAACGGATTTTAT
sumo-RitRL86A_RD_Rev	GCTCTAGATTAAATGAAGTCCCGACCTCGG

sumo-RitR(C128S)_Fwd	GACTTCATTGATCAACACAGCAGTCTGATGAAAGTTC
sumo-RitR(C128S)_Rev	GAACTTTCATCAGACTGCTGTGTTGATCAATGAAGTC
sumo-RDLV_Rev	GCTCTAGATTAAATGAAGTCCCGACCTCGG
sumo-RD2LV_Rev	GCTCTAGATTATTTCATCAGACTGCAGTGTTGATC
sumo-RitR(C128D)_Fwd	GACTTCATTGATCAACACGACAGTCTGATGAAAGTTC
sumo-RitR(C128D)_Rev	GAACTTTCATCAGACTGTCGTGTTGATCAATGAAGTC

2.2.6 Determination of C128 pK$_a$ by DTNB assay [6, 7]

The thiol-reactive reporter compound 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) can be used to determine the number of free, solvent-exposed thiol groups in a protein or to estimate the pK$_a$ in cases where a single thiol group is present. The assay is based on the formation of 2-nitro-5 thiobenzoate (TNB$^-$) upon reaction of DTNB with thiol groups. The ionized reporter group, TNB^{2-} absorbs UV light at 412 nm (Scheme 2-1).
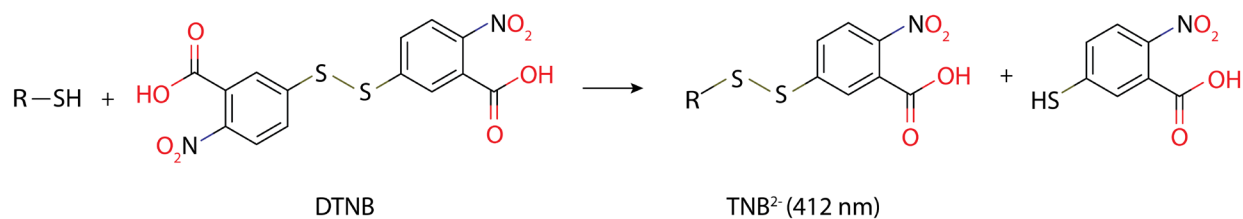
Scheme 2-1: The mechanism of DTNB assay

Reactions contained 5 μM RitR and 20 μM DTNB in 50 mM CHB buffer (50 mM citric acid, 50 mM HEPES, 50 mM Bicine, 150 mM NaCl, with pH adjusted by titration with NaOH

or HCl as appropriate) at pH values ranging from 4 to 9. The reactions were conducted at room temperature (~21 °C) in an Evolution™ 300 UV-Vis spectrophotometer (Thermo Scientific). The absorption at 412 nm was recorded until it plateaued. Then the pH-dependent rate of each reaction $k_{(pH)}$ was determined by fitting the time course to the following equation:

$$A_{412}(t) = A_{412,0} + \Delta A_{412}(1 - e^{-k(pH)t}) \qquad \text{Equation 1}$$

Since $k_{(pH)}$ is directly dependent on the proportion of the cysteine thiolate group, the pK_a of the C128 thiol was obtained by fitting the plot of $k_{(pH)}$ vs pH to the following equation:

$$k_{(pH)} = \frac{k}{1 + 10^{pK_{a,thiol} - pH}} \qquad \text{Equation 2}$$

where $k_{(pH)}$ is the observed rate constant at a particular pH value, and k is the maximal rate of TNB^{2-} formation at high pH.

2.2.7 Crystallization, structure determination, and model refinement

Crystals of selenomethionine (SeMet)-substituted RitR(C128S) were grown by the vapor diffusion method with hanging drop geometry. Drops were comprised of equal parts protein solution (20 mg/ml RitR(C128S) in 25 mM TRIS, pH 8.0) and crystallization solution (25% (w/v) polyethylene glycol (PEG) 3350, 0.1 M ammonium acetate). Partially-hollow rods appeared after several days at 16°C. Crystals were prepared for flash-cooling by sequential soaks in solutions containing 25% PEG (w/v) 3350, 0.2 M ammonium acetate, and 5, 10, or 20% glycerol (v/v).

Crystals of RitR(C128D) were grown by the same method, except with hanging drops composed of 1 μl of protein solution (15.3 mg/mL RitR(C128D) in 10 mM ADA pH 7.0) and 1 μl of crystallization solution (0.1 M HEPES pH 7.5, 22.5% Jeffamine ED-2001 pH 7.0). Long rod-shaped crystals grew after 3 days at 16 °C. Crystals were prepared for flash-cooling by sequential soaks in solutions containing 0.1 M HEPES pH 7.5, 22.5% Jeffamine ED-2001 pH 7.0, and 5, 10, or 20% glycerol (v/v).

Crystals of RitR$_{ox}$ were grown by the hanging-drop, vapor diffusion method with drops consisting of 2 μl of protein solution (6 mg/ml RitR$_{ox}$ in pH 6.5, 20 mM TRIS, 150 mM NaCl, 100 mM malic acid, 1 mM H_2O_2) and 1 μl of crystallization solution (10% (w/v) PEG 3350, 0.3 M magnesium sulfate). Prism-shaped crystals appeared after several days at 16 °C. Crystals were prepared for flash-cooling by soaking in a solution containing 10% (w/v) PEG 3350, 0.3 M magnesium sulfate, and 20% glycerol (v/v).

Crystals of RitR oxidized by treatment with $CuCl_2$ rather than H_2O_2 were grown from drops containing 1 μl of protein solution (5 mg/mL RitR$_{CuOx}$ in pH 6.5, 25 mM TRIS, 150 mM NaCl, 100 mM Malic acid, 1 mM $CuCl_2$) and 1 μl of microseeding solution (4% Tacsimate pH 6.0, 7.5% PEG 3350 containing microseeds derived from its own crystals). Single prism-shaped crystals grew after several days at 16 °C. Crystals were prepared for flash-cooling by sequential soaking in solutions containing 4% Tacsimate pH 6.0, 7.5% PEG 3350, and 5, 10, or 20% glycerol (v/v).

Crystals of selenomethionine-substituted RitR(L86A/V93A) REC domain, residues 1-124 (RDLV), were grown in drops consisting of equal parts of protein solution (4.4 mg/ml RDLV in 10 mM TRIS pH 8.5, 50 mM Succinic acid) and crystallization solution (20% (w/v)

PEG 3350, 0.2 M magnesium sulfate, 0.1 M TRIS pH 8.5, 1 mM DTT). Small rods appeared after several days at 16 °C. Crystals were prepared for flash-cooling by sequential soaks in solutions containing 20% PEG (w/v) 3350, 0.2 M magnesium sulfate, 0.1 M TRIS pH 8.5, and 5, 10, or 20% glycerol (v/v).

Crystals of SeMet-substituted RitR(L86A/V93A) REC domain, residues 1-132 (RD2LV), were grown by the hanging-drop, vapor diffusion method in drops comprised of equal parts of protein solution (19 mg/ml RD2LV in 10 mM glycine pH 8.8, 25 mM L-Arg) and crystallization solution (22.5% (w/v) polyethylene glycol (PEG) 3350, 12% Tacsimate pH 8.2). Small plate-shaped crystals appeared after several days at 16 °C. Crystals were prepared for flash-cooling by sequential soaks in solutions containing 22.5% (w/v) PEG 3350, 12% Tacsimate pH 8.2, and 5, 10, or 20% glycerol (v/v).

X-ray diffraction data for SeMet RitR(C128S) were collected at beamline 21-ID-F of the Life Science Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS), Argonne National Laboratory. The RitR(C128D) data set was collected at LS-CAT beamline 21-ID-G. The RitR$_{ox}$ data set was collected at LS-CAT beamline 21-ID-F. The SeMet RDLV data was collected at LS-CAT beamline 21-ID-G. The SeMet RD2LV data were collected at LS-CAT beamline 21-ID-D. All data sets were processed with HKL2000 [8].

The structure of RitR(C128S) was determined by the single-wavelength anomalous diffraction (SAD) method using 1.5 Å-resolution data collected from a single crystal of SeMet-substituted RitR(C128S) at 0.97889 Å, 61.0 eV below the tabulated K-edge wavelength for Se (0.97950Å). The program autoSHARP[9] was used to solve the Se substructure, which contained 16 of the 18 Se atoms in the asymmetric unit, and to

calculate density-modified electron density maps. An initial model comprising ~90 % of the asymmetric unit contents was built using BUCCANEER[10].

The structure of RDLV was determined by the single-wavelength anomalous diffraction (SAD) method using 1.9 Å-resolution data collected from a single crystal of SeMet-substituted proteins at 0.97857 Å, 61.0 eV below the tabulated K-edge wavelength for Se (0.97950Å). The program autoSHARP[9] was used to solve the Se substructure, which contained 16 of the 18 Se atoms in the asymmetric unit, and to calculate density-modified electron density maps. An initial model comprising ~90 % of the asymmetric unit contents was built using BUCCANEER[10]. The structure of RD2LV was determined by the single-wavelength anomalous diffraction (SAD) method using 2.6 Å-resolution data collected from a single crystal of SeMet-substituted proteins at 0.97872 Å, 61.0 eV below the tabulated K-edge wavelength for Se (0.97950Å). The program autoSHARP[9] was used to solve the Se substructure, which contained 16 of the 18 Se atoms in the asymmetric unit, and to calculate density-modified electron density maps. An initial model comprising ~90 % of the asymmetric unit contents was built using BUCCANEER[10].

Chain A of this model was used as the search model for molecular replacement in PHASER[11] to phase the RitR$_{ox}$ structure. Both models were subjected to iterative cycles of manual model building in COOT[12] and maximum likelihood-based refinement using the PHENIX package (phenix.refine[13]). Ordered solvent molecules were added automatically in phenix.refine and culled manually in COOT. Hydrogen atoms were added to the model using phenix.reduce[14] and were included in the later stages of refinement to improve the stereochemistry of the model. Positions of H atoms were refined using the riding model

with a global B-factor. Regions of the model for translation-libration-screw (TLS) refinement were identified using phenix.find_tls_groups (P.V. Afonine, unpublished work) and the TLS parameters were refined in phenix.refine. Once the refinement converged, the model was validated using the tools implemented in COOT and PHENIX [15, 16]. Sections of the backbone with missing or uninterpretable electron density were not included in the final model. Data collection and model refinement statistics are listed in Appendix A. Coordinates and structure factors for both RitR(C128S) and RitR$_{ox}$ models have been deposited in the Protein Data Bank (www.rcsb.org) with accession codes 5U8K and 5U8M.

2.2.8 Synthesis of binding site 3 (BS3) dsDNA oligonucleotides

Early footprinting experiments detected 3 binding sites for RitR(BS1-BS3)[1]. Since RitR was deemed to have the highest affinity for BS3, this sequence was used to investigate the DNA-binding activity of RitR and its variants. To facilitate co-crystallization of RitR bound to DNA, different lengths of BS3 were designed (Table 2-2), since the length of the oligonucleotide is a key parameter in the crystallization of protein:DNA complexes. Both strands of BS3-1, BS3-2, BS3-3, BS3-4, and BS3-5 were synthesized by Integrated DNA Technologies (Coralville, IA). Both strands of BS3-6 were synthesized via standard automated DNA synthesis techniques in a 1.0 μM scale using commercial 1000 Å CPG-succinyl-nucleoside supports. Deprotection of the nucleobases and phosphate moieties as well as cleavage of the linker were carried out under mild deprotection conditions using a mixture of 40% aq. MeNH$_2$ and 28% aq. NH$_3$ (1:1) at room temperature for 2 h. Oligonucleotides were purified by 20% denaturing polyacrylamide gel electrophoresis.

Quantification of oligonucleotides was carried out using a Phosphorimager equipped with ImageQuant Version 5.2 software. Double-stranded DNA was obtained by mixing the two single-stranded oligonucleotides in 250 mM TRIS pH 6.5, 150 mM NaCl buffer and heating to 90 °C for 5 min, followed by slow cooling to room temperature.

Table 2-2: Sequences of the BS3 variants used in RitR binding assays.

Name	Sequence	Length (bp)
BS3-1	5'-GGTATGAAACTTTCTCATTATTTAATTGGCTTACTTCT-3' 3'-CCATACTTTGAAAGAGTAATAAATTAACCGAATGAAGA-5'	38
BS3-2	5'-GGTATGAAACTTTCTCATTATTTAATTGGCTTAC-3' 3'-CCATACTTTGAAAGAGTAATAAATTAACCGAATG-5'	34
BS3-3	5'-ATGAAACTTTCTCATTATTTAATTGGCTTACTTCT-3' 3'-TACTTTGAAAGAGTAATAAATTAACCGAATGAAGA-5'	35
BS3-4	5'-AAACTTTCTCATTATTTAATTGGCTTACTTCT-3' 3'-TTTGAAAGAGTAATAAATTAACCGAATGAAGA-5'	32
BS3-5	5'-CTTTCTCATTATTTAATTGGCTTACTTCT-3' 5'-GAAAGAGTAATAAATTAACCGAATGAAGA-3'	29
BS3-6	5'-CTTTCTCATTATTTAATTGGCTTA-3' 3'-GAAAGAGTAATAAATTAACCGAAT-5'	24

2.2.9 Size exclusion chromatography (SEC)

The oligomeric states of wild-type and mutant forms of RitR were assessed by size exclusion chromatography (SEC). All separations were carried out using an Agilent 1220

Compact HPLC equipped with a 250 x 4.6 mm BioBasic SEC-300 column equilibrated with pH 6.5, 10 mM TRIS, 150 mM NaCl, and 100 mM malic acid. The column was calibrated with the Gel Filtration Molecular Weight Marker kit from Sigma-Aldrich (Cytochrome C (12.4 kDa), Carbonic Anhydrase (29 kDa), Bovine Serum Albumin (66 kDa), and Sweet Potato Amylase (200 kDa)). The H_2O_2- or $CuCl_2$- treated RitR proteins were injected (5 μl of 4 mg/mL proteins) onto the column and separated at a flow rate of 0.5 ml/min at ambient temperature. Wild-type and mutant forms of RitR was assayed for DNA-binding activity by mixing 260 μM protein (RitR, RitR(L86A/V93A), RitR(L86A/V93A/C128S) and other mutants) with 100 μM each of the BS3 variant oligonucleotides (BS3-1, BS3-2, BS3-3, BS3-4, BS3-5, BS3-6). SEC experiments were carried out under the same conditions used for the RitR proteins in the absence of DNA.

2.2.10 Non-reducing SDS-PAGE and Native SDS-PAGE (NSDS)

Proteins were incubated with different metals for 2 hours at room temperature. These samples were then mixed with SDS loading dye lacking reducing agent, boiled for 5 min, and loaded on 12% SDS polyacrylamide gel. Proteins were revealed by staining with GelCode™ Blue Safe Protein Stain (Thermo SCIENTIFIC, Rockford, IL) for 1 h and destained in MilliQ H_2O. The slightly modified protocol for running native SDS-PAGE is followed the protocol from Dr. Petering Lab[17].

2.3 Results

2.3.1 Determining the optimal RitR storage buffer for crystallization

Like many DNA-binding proteins, such as the γδ resolvase[18], MarA[19], and MarR[20], the full-length RitR protein is not stable in low ionic strength buffers. In low ionic strength buffers like 25 mM TRIS pH 8.0, RitR precipitated out as a flocculent white solid, and this precipitate could be redissolved by addition of NaCl. Since it is generally preferred to search for crystallization conditions with the protein in a dilute, low ionic strength buffer, we performed experiments to find the lowest NaCl concentration where RitR would be stable. To begin, purified RitR was desalted into 25 mM TRIS pH 8.0, with concentrations of NaCl ranging from 50 to 200 mM in 50 mM steps. Samples were incubated overnight at 4 °C and the amount precipitate was assessed qualitatively. The lowest NaCl concentration tested that did not give visible precipitate was 150 mM. With the ionic strength optimized, we set about exploring the effect of buffer identity and pH on RitR stability. This was done using differential scanning fluorimetry together with the Solubility & Stability and Slice pH screens (Hampton research). The Solubility and Stability Screen, which is a collection of 96 small-molecule additives, was run in the 25 mM TRIS pH 8.0, 150 mM NaCl buffer. The top 10 conditions from the Solubility & Stability and the Slice pH screens, as judged by the degree of increase in the T_m above the average T_m, are shown in Table 2-3 and Table 2-4, respectively.

Since malonate is used as a cryo-protecting agent for protein crystals, 100mM malonate was added to the protein storage buffer. The agents with similar structures, such as DL-

Table 2-3: Compounds from the Solubility & Stability Screen (Hampton Research) giving the highest T_m values in the DSF assay

T_m - $T_{m,averaged}$ (°C)	Additive
6.2	Spermine tetrahydrochloride
4.8	GSH, GSSG
4.7	Choline acetate
4.7	Ammonium sulfate
4.6	Succinic acid pH 7.0
4.4	Sodium malonate pH 7.0
4.4	Sodium sulfate decahydrate
4.3	L-Arginine, L-Glutamic acid
4.1	1-Ethyl-3-methylimidazolium acetate
4.0	Sodium chloride

Table 2-4: Buffers from the Slice pH Screen (Hampton Research) giving the highest T_m values in the Thermofluor assay

T_m - $T_{m,averaged}$ (°C)	Additive
4.6	DL-malic acid pH 5.6
4.4	Succinic acid pH 6.0
4.2	DL-malic acid pH 5.3
4.0	Succinic acid pH 5.7
3.5	BIS-TRIS propane pH 7.3
3.4	BIS-TRIS pH 6.6
3.3	ADA pH 7.0
3.3	HEPES sodium pH 6.6
3.2	MES monohydrate pH 6.5
3.1	Sodium cacodylate trihydrate pH 6.1

malic acid (ACROS Organics) and citric acid (ACROS Organics), were also used for wild-type RitR protein storage and crystallization.

According to the Slice pH screen results, the optimal pH is ~6 to 6.5. We chose pH 6.5, 25 mM Tris, 150 mM NaCl, 100 mM malonic acid/DL-malic acid/citric acid buffers to store wild-type RitR. RitR in these three buffers were crystallized. Crystals were obtained at all these buffers. But better crystals were obtained in DL-malic buffer, so the final RitR storage buffer I used is pH 6.5, 25 mM Tris, 150 mM NaCl, 100 mM DL-malic acid buffer.

2.3.2 Monomeric structure of RitR(C128S)

Gel filtration chromatography suggests that inactive RitR is predominantly monomeric, but a second species, consistent in size with a dimer, is also present and accounts for ~10-20% of the sample. Likely because of this heterogeneity, the full-length, wild-type RitR proved impossible to crystallize. To facilitate crystallization, the only cysteine residue (C128) was mutated to serine (RitR(C128S)). This variant proved to be 100% monomeric as judged by gel filtration chromatography and native-PAGE.

The structure of RitR(C128S) was determined to 1.7 Å resolution by using phases derived from SeMet single-wavelength anomalous diffraction (Se-SAD)(Figure 2-1). It is a monomer, which is in agreement with biochemical data mentioned above. There are two monomers in the asymmetric unit. But these two monomers do not form a dimer, nor with their symmetry mates. RitR(C128S) contains regulatory ALR domain and DNA binding domain (DBD). The only cysteine residue in RitR, C128, is located at the C-terminal end of

the long α5 linker helix, which connects the ALR domain with the DNA-binding domain (DBD). The ALR domain structure is same with the earlier solved RitR ALR domain only structure (PDB ID: 4LZL[3], Figure 1-). DBD has the helix-turn-helix (HTH) motif, which is the simplest and most common motif in prokaryotes and eukaryotes[21]. This HTH motif contains two fixed angled helices connected by a small loop. The more C-terminal helix (colored in red in Figure 2-1) is the recognition helix that used for contacting with specific DNA major groove.
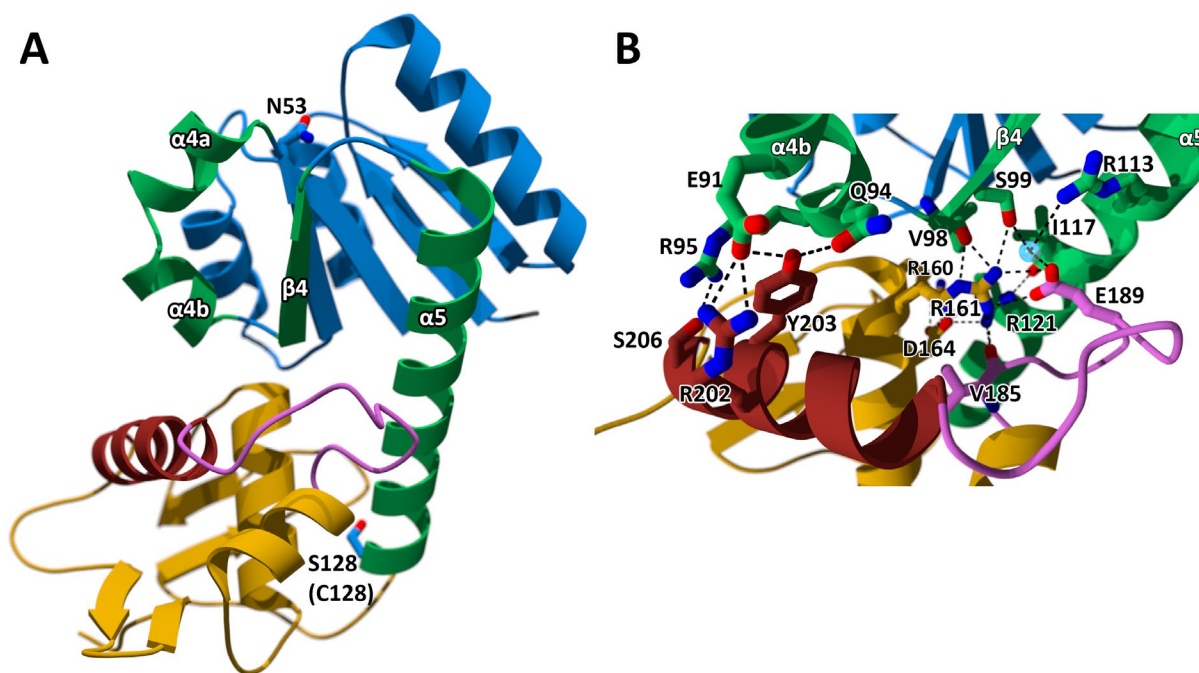
Figure 2-1: Ribbon representation of the structure of full-length RitR(C128S) (A) and a close-up on the inter-domain interface (B). Important amino acid side chains are shown as sticks. The regulatory domain is shown in blue, except for the putative α4-β5-α5 dimer interface, which is colored green. The DNA-binding domain is colored yellow, except for the DNA-contacting helix and trans-activation loop, which are colored red and magenta, respectively.

There is an extensive network of salt bridges and hydrogen bonds between the regulatory ALR domain and DBD, with each domain contributing a surface area of approximately ~850 Å^2 (Figure 2-1B). E91 makes salt bridges with R202 (3.1 Å). E91 and Q94 make hydrogen bonds to Tyr203 with distances of 2.0 Å and 2.2 Å, respectively). R95 interacts with S206 (2.5 Å), residues that are found on the DNA-contacting helix of the DBD, helix α4b, and helix α5 of the ALR domain. Both α4 and α5 are contributed to the dimerization interface in other OmpR/PhoB type response regulator proteins. The DNA-contacting helix interacts with helix α4b and is thus not free to interact with DNA. Additionally, the trans-activation loop is involved in the interdomain interface and would thus be unavailable to interact with RNA polymerase. All of these facts demonstrate that this monomeric conformation is not the active, DNA-binding conformation of RitR.

2.3.3 The thiol group of the conserved C128 residue has a perturbed pKa

We assumed that, like most DNA-binding RR proteins, the active, DNA-binding form of RitR is a homodimer. Comparing the sequences of all RitR homologs in the *Streptococci* (Figure 2-2A) shows that this Cys residue is strictly conserved. This raises the possibility that RitR could be a redox-responsive response regulator protein. In other cysteine-based redox sensor proteins, the side chain thiol groups of the key cysteine residues have been found to have significantly depressed pK_a values [22, 23]. In order to ascertain whether or not the C128 thiol group also has a perturbed pK_a, and could thus be expected to be redox active under physiological conditions, the pK_a of C128 was determined using the DTNB assay. The rate of reaction of DTNB with thiol groups depends directly on the proportion of

thiols that are present as the thiolate anion. The reaction (Scheme 2-1) releases TNB^-, which readily deprotonates to give TNB^{2-}, a species that absorbs at 412 nm. By monitoring the rate of the DTNB reaction at different pH values, it is possible to derive an estimate of the pK_a of the thiol group. Reactions containing 20 μM DTNB in a buffer system consisting of citric acid, HEPES, and bincine (50mM each) were initiated by the addition of 5 μM RitR, and the absorbance at 412 nm was monitored (Figure 2-2 B). Nonlinear fitting of these progress curves to Equation 1 gave estimates of the pH-dependent rate constant of the reaction, $k_{(pH)}$. These values were plotted as a function of pH, and then fit to Equation 2 (Figure 2-2C) to gives an estimate of the pK_a of C128 of 7.2 ± 0.2. This is slightly decreased from the thiol pK_a of cysteine in solution, which is generally quoted above 8.3. Different buffer systems were tested for this assay, giving a pK_a range for C128 of 6.9-7.5. The perturbed pK_a of C128 suggests that this residue is more reactive than a typical surface-exposed cysteine residue and thus may be part of a redox-sensing switch mechanism in RitR.
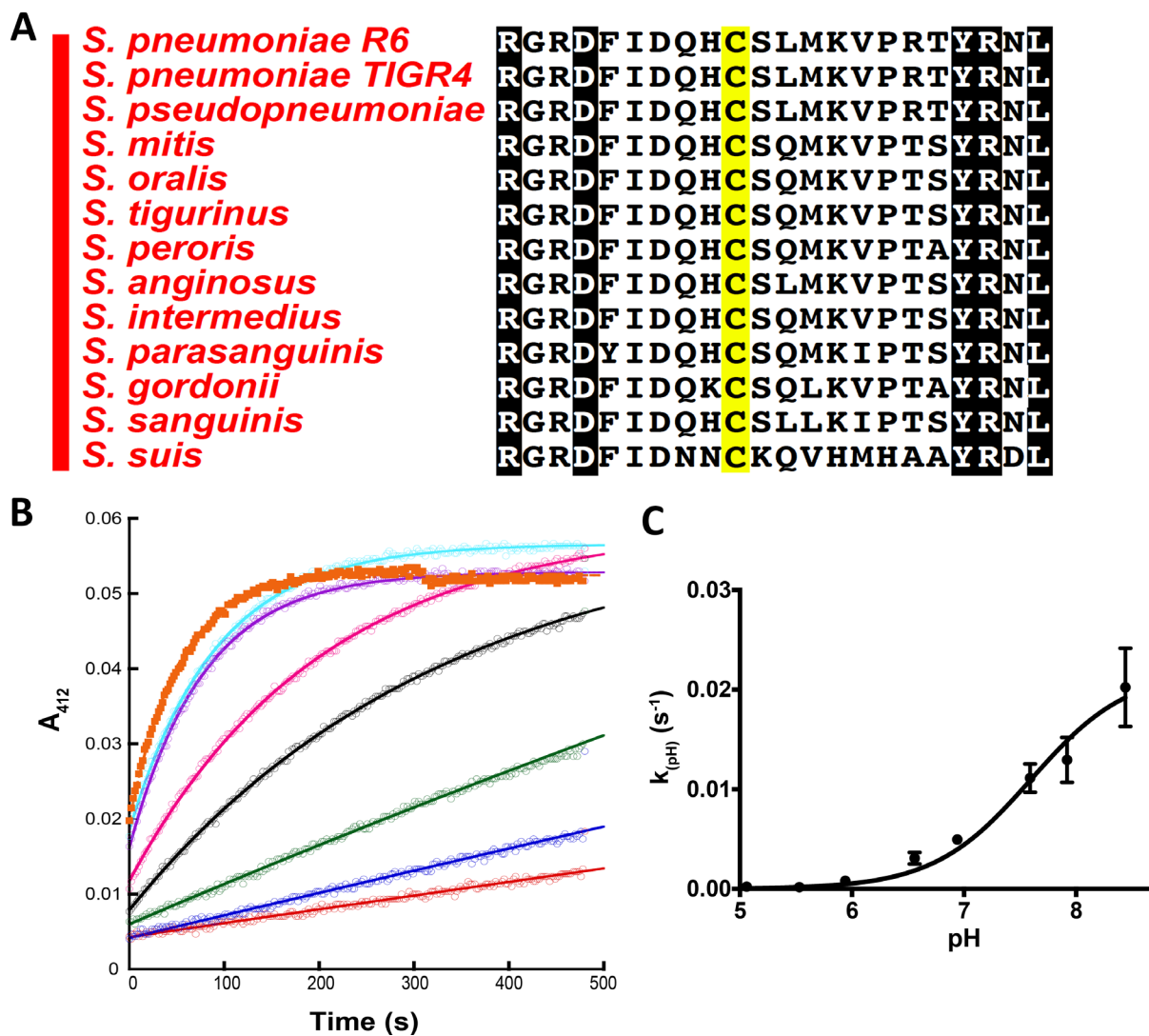
Figure 2-2: (A) Sequence alignment of RitR homolog in a non Group A/B subset of Streptococcal species demonstrated a conservation of redox sensor C128. (B) The absorbance at 412 nm (dots) and fitting lines at pH 5.1 (red), 5.5 (blue), 6.0 (green), 6.6 (black), 6.9 (magenta), 7.6 (cyan), 7.9 (purple) and 8.4 (orange). (C) The pH dependence of the rate constants of C128 with DTNB. The plot of rate constants ($k_{(pH)}$) vs pH gives the C128 pKa of 7.2 ± 0.2.

2.3.4 The structure of RitR$_{ox}$

RitR is a transcriptional regulator with a DNA-binding effector domain, which suggests that RitR probably functions as a homodimer, since oligomerization is an extremely common feature of DNA-binding proteins, and of response-regulators in particular. The depressed pK$_a$ of C128 suggested the possibility of a redox-based activation mechanism in RitR. Accordingly, the consequences of H$_2$O$_2$-mediated oxidation of RitR with respect to the oligomeric state and structure were investigated by size exclusion chromatography and X-ray crystallography.

Initial experiments showed that brief incubations with RitR failed to produce any change in the oligomeric state of the protein (Figure 2-3A). Using different buffers for the oxidation, such as citric acid, malic acid, or malonic acid, all at 50mM and pH 6.5, likewise did not alter the oligomeric state when RitR was treated with 1 mM H$_2$O$_2$ (Figure 2-3B). Extending the oxidation time to 19h did increase the proportion of dimer present, the maximum proportion of dimer ($\sim$90%) was only obtained when the SEC running buffer contained 1 mM H$_2$O$_2$. Once the optimal oxidation conditions were established, a preparative scale Sephacryl S-200 10/300 column was used to prepare samples of RitR dimer (*i.e.* RitR$_{ox}$) for crystallization. The dimer peak from the S-200 column was collected and stored in pH 6.5, 25 mM TRIS, 150 mM NaCl, 100 mM malic acid, and 1 mM H$_2$O$_2$. This material was concentrated to 10 mg/ml and subjected to screening for crystallization conditions.
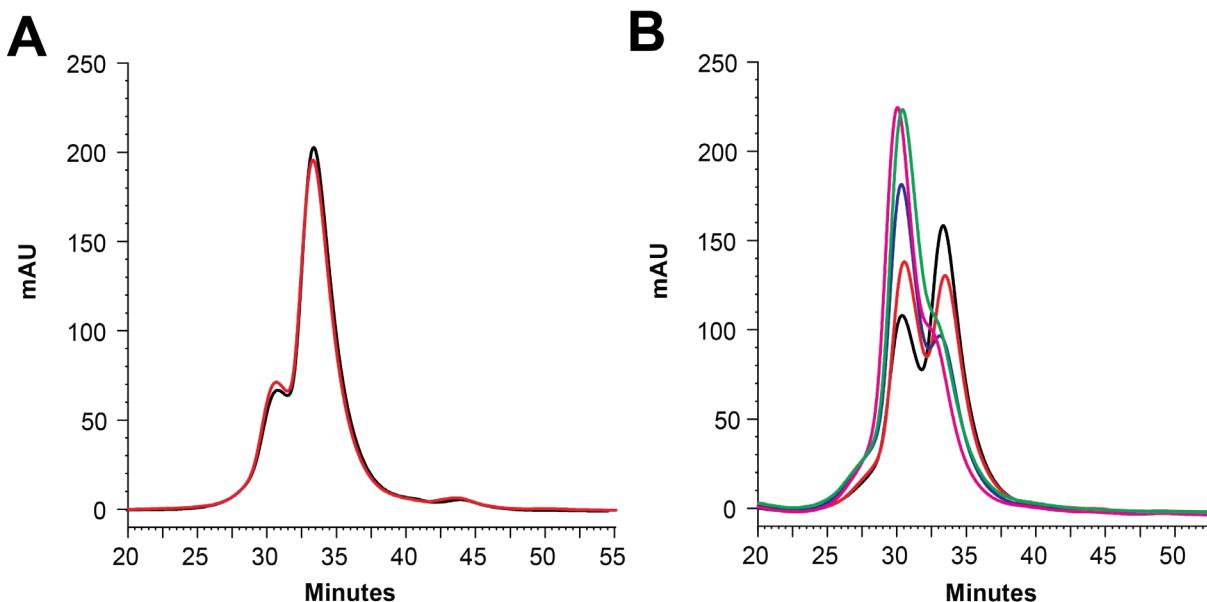
Figure 2-3: Different amounts of oxidized RitR dimer in different buffer conditions were determined by gel filtration on a Sephacryl S-200 10/300 column. (A) The gel filtration results of RitR in pH8, 25 mM Tris, 150 mM NaCl buffer. Black line represents no H_2O_2 treatment, while red line represents 1 mM H_2O_2 treatment. (B) The gel filtration results of RitR in pH 8, 25 mM Tris, 150 mM NaCl, 100 mM malonic acid buffer with different H_2O_2 treatment conditions. Black line represents no H_2O_2 treatment. Red line represents RitR was treated with 1 mM H_2O_2 for 3 h. Blue line represents RitR was treated with 1 mM H_2O_2 for 19 h. Green represents RitR was treated with 1 mM H_2O_2 for 3 h and containing 1mM H_2O_2 in the running buffer. Magenta represents RitR was treated with 10 mM H_2O_2 for 3 h and containing 1mM H_2O_2 in the running buffer. The maximum amounts of $RitR_{ox}$ were got with 1 mM H_2O_2 in the running buffer.

Diffraction-quality crystals of $RitR_{ox}$ were obtained from a hanging drop containing 2 μl of protein solution (6 mg/ml $RitR_{ox}$ in pH 6.5, 20 mM TRIS, 150 mM NaCl, 100 mM malic acid, 1 mM H_2O_2) and 1 μl of crystallization solution (10% (w/v) PEG 3350, 0.3 M magnesium sulfate. The structure was determined to 1.7 Å resolution by using RitR(C128S) as the

model. The structure of RitR$_{ox}$ shows that the tertiary structures of the individual domains do not change significantly. However, the two chains present in the asymmetric unit form a domain-swapped dimer (Figure 2-4A). Upon oxidation, the DNA-binding domain must be released from the REC domain. The structural change that triggers release of the DNA-binding domain may be a rearrangement of the C-terminal end of the long α5 linker helix, which appears to unwind upon oxidation of C128 (Figure 2-4B) to form an inter-chain disulfide bond. Since there is no DNA to bind, and the restricted conformation imposed by the inter-chain disulfide bond, the DNA-binding domain of one protomer interacts with the REC domain from the other protomer to almost perfectly recapitulate the inactive, monomeric structure (Figure 2-4D). Overlaying the REC domain of chain A together with the DBD of chain B onto the RitR(C128S) monomer gives a root-mean-square deviation for aligned Cα atoms of 0.7 Å. This inactive "cross-dimer" conformation may be an intermediate that forms before the oxidized dimer encounters its cognate DNA sequence, or it may indicate that oxidation of C128 is not the entire signal.
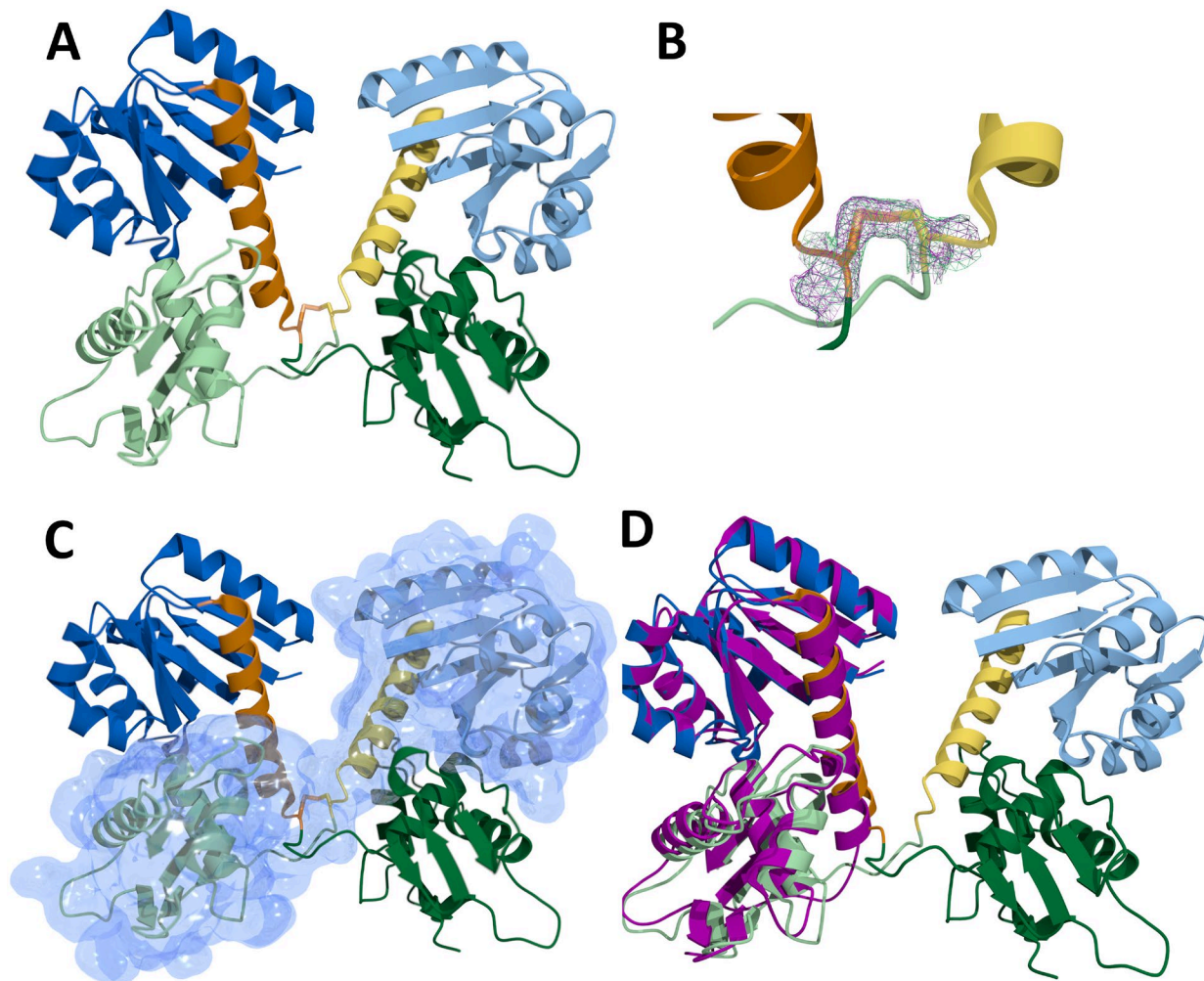
Figure 2-4: The structure of RitR_{ox}. (A) Ribbon representation of the RitR_{ox} structure, Rec domain colored in blue (dark blue in chain A and light blue in chain B); $\alpha 5$ helix colored in orange (dark orange in chain A and light yellow in chain B); DNA binding domain colored in green (dark green in chain A and light green in chain B). (B) Close up view of disulfide bond. The $2|F_o|-|F_c|$ electron density map contoured at 1.0σ is shown as magenta mesh, and the $2|F_o|-|F_c|$ composite omit map, also contoured at 1.0σ is shown as green mesh. (C) RitRox structure with transparent blue surface in chain B. (D) Superimpose RitRox with inactive monomeric RitR(C128S) (purple). RMSD = 0.7 Å

2.3.5 RitR incubated with different metals

Because RitR is the repressor of iron transport, we hypothesized that iron or another divalent metal might be involved in the activation mechanism of RitR. To study the effect of metals on RitR dimerization, RitR was treated with $Fe^{III}Cl_3$, $Fe^{II}Cl_2$, $MgCl_2$, $MnCl_2$, $CaCl_2$, $CuCl_2$, or $ZnCl_2$, all in pH 6.5, 20 mM TRIS, 150 mM NaCl, 100 mM malic acid buffer. Size exclusion chromatography was used to check the oligomeric state, and the results are shown in Figure 2-5.

Figure 2-5: Oligomer state of RitR after incubated with different metals. Blue line represents RitR only. Cyan line represents RitR(L86A/V93A) only. This mutant represented an extended dimer, which will be discussed below. Magenta line represents RitR with 1 mM metals. Green represents RitR with 10 mM different metals. Brown and green represent 1 mM and 10 mM different metals, respectively. (A) RitR were incubated with $CuCl_2$. (B) RitR were incubated with $Fe^{III}Cl_3$ with DTT. (C) RitR were incubated with $Fe^{III}Cl_3$ (D) RitR were incubated with $Fe^{II}SO_4$ with 0.1 N HCl. (E) RitR were incubated with $MgCl_2$. (F) RitR were incubated with $ZnCl_2$. (G) RitR were incubated with $CaCl_2$. (H) RitR were incubated with $MnCl_2$.
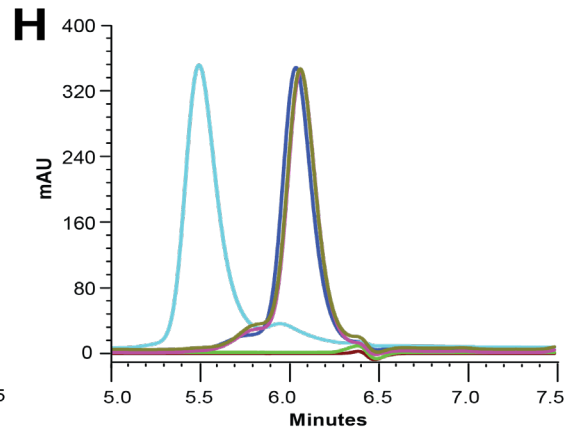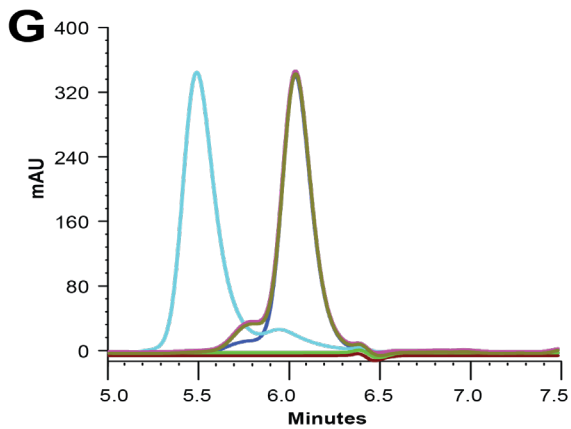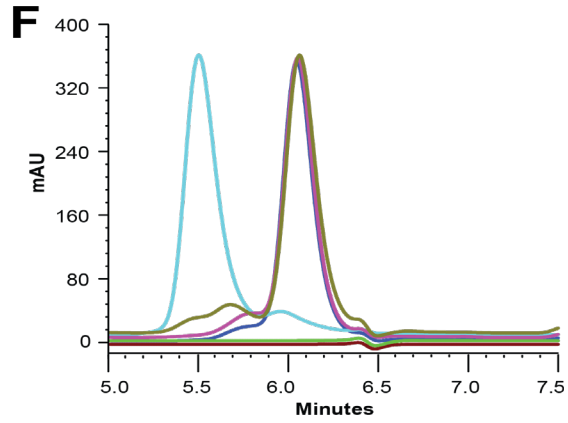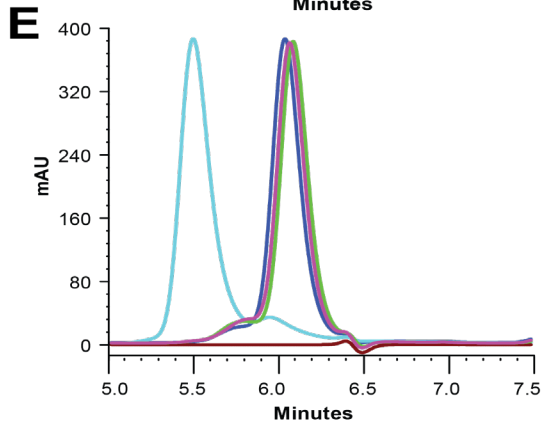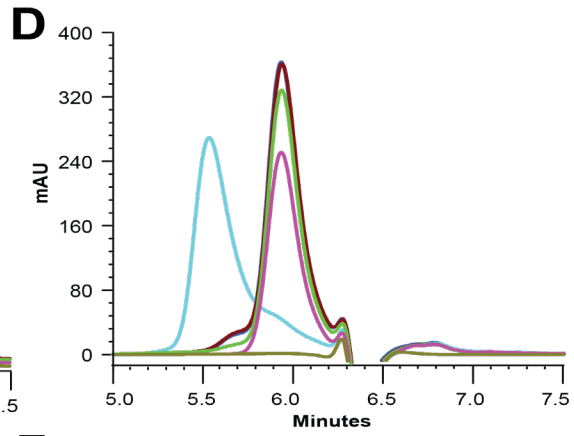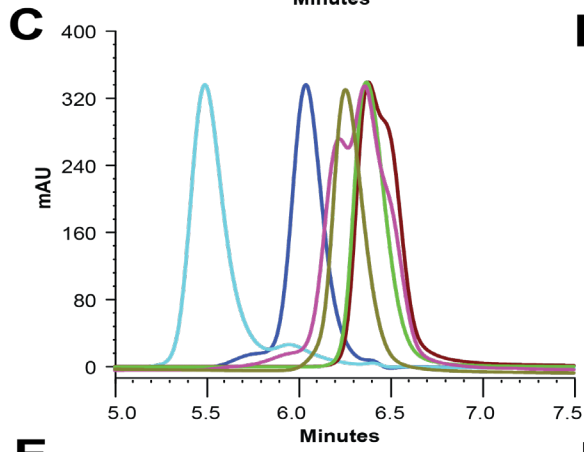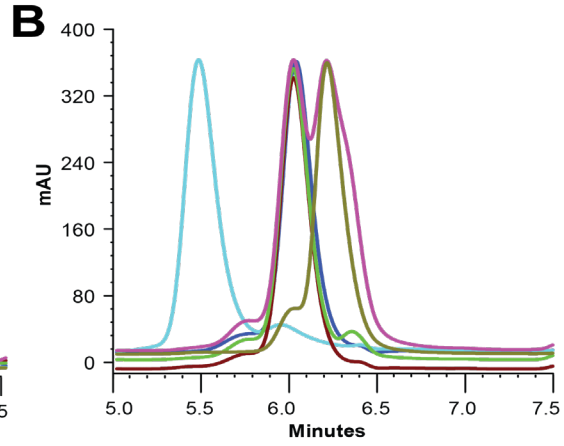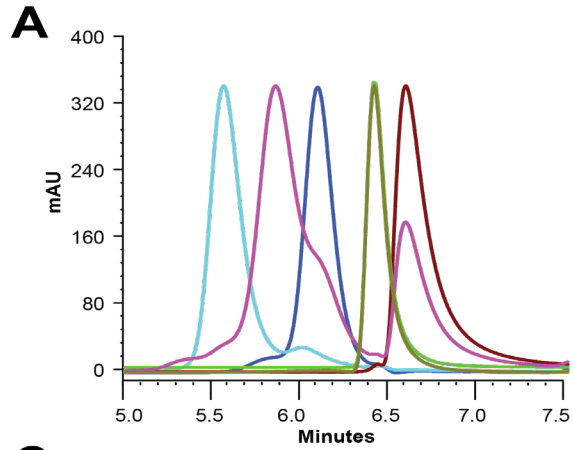
SDS-PAGE, non-reducing-PAGE, and native SDS-PAGE were also used to check the oligomeric state after treatment with the various metals. The results of these experiments are shown in Figure 2-6.
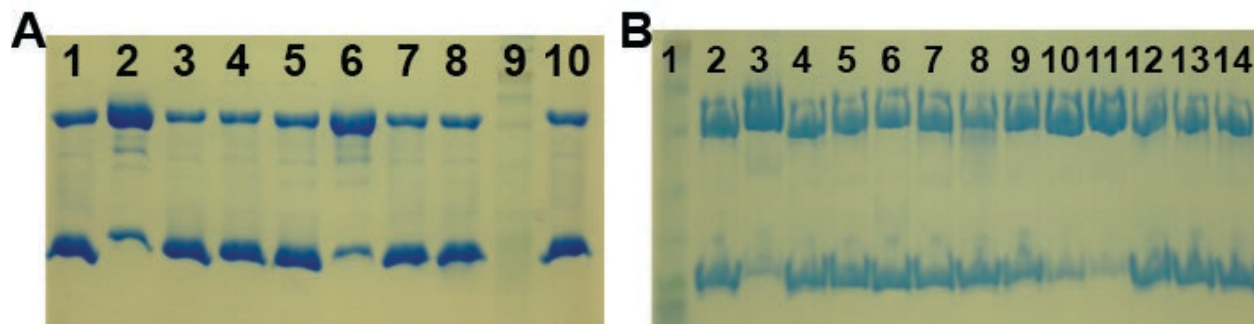
Figure 2-6: (A) Non-reducing-PAGE results of RitR treated with different metals. 1, wild-type RitR; 2, RitR(L86A/V93A); 3, RitR with 1 mM $Fe^{III}Cl_3$; 4, RitR with 10 mM $MgCl_2$; 5, RitR with 10 mM $MnCl_2$; 6, RitR with 1 mM $CuCl_2$; 7, RitR with 10 mM $CuCl_2$; 8, RitR with 1 mM $ZnCl_2$; 9, Marker; 10, RitR with 10 mM $ZnCl_2$. (B) Native SDS-PAGE results of RitR treated with different metals. 1, Marker; 2, wild-type RitR; 3, RitR(L86A/V93A); 4, RitR with 1 mM $Fe^{III}Cl_3$; 5, RitR with 10 mM $MgCl_2$; 6, RitR with 10 mM $MnCl_2$; 7, RitR with 1 mM $CaCl_2$; 8, RitR with 10 mM $ZnCl_2$; 9, RitR with 10 μM $CuCl_2$; 10, RitR with 100 μM $CuCl_2$; 11, RitR with 1 mM $CuCl_2$; 12, RitR with 10 μM $Fe^{III}Cl_3$ with 0.1 mM DTT; 13, RitR with 100 μM $Fe^{III}Cl_3$ with 0.1 mM DTT; 14, RitR with 1 mM $Fe^{III}Cl_3$ with 0.1 mM DTT. Top bands located around the 50 kD marker, corresponding to dimer, while the bottom bands located around the 25 kD marker, corresponding to monomer.

When the metal-treated samples of RitR were run on a non-reducing gel (Figure 2-6A) to preserve any disulfide bonds, wild-type RitR runs mainly as a monomer, while the double mutant RitR(L86A/V93A) runs mainly as a dimer. This constitutively dimeric double mutant will be discussed in detail below. Treatment of the wild-type RitR with Fe^{3+}, Fe^{2+},

Mg^{2+}, Mn^{2+}, Zn^{2+}, did not alter the dimer/monomer ratio. However, treatment with Cu^{2+} promoted dimerization of RitR as judged from the non-denaturing SDS-PAGE gel. This result was confirmed by native SDS-PAGE (nSDS-PAGE). Increasing the $CuCl_2$ concentration increased the proportion of dimer obtained. These results indicated that Cu^{2+} induced the dimerization of RitR.

The structure of RitR dimer by $CuCl_2$ treatment (data not shown) is same with $RitR_{ox}$ by H_2O_2 treatment, because copper also can oxidize the cysteine residue to form disulfide bond [24-26]. So the structure of $RitR_{ox,CuCl2}$ is also an inactive conformation.

2.3.6 Structural features important for RitR dimerization

The active, DNA-binding form of RitR is hypothesized to be a homodimer, which binds to specific sequences in the *piu* promotor and represses the expression of genes associated with iron uptake and up-regulates the expression of genes involved in the response to oxidative stress. We saw above that the DBD of the RitR dimer obtained by oxidation alone (*i.e.* in the absence of DNA) is characterized by a conformational change at the C-terminal end of α5 and a domain-swapping arrangement that matches almost perfectly with the inactive monomeric structure. There is virtually no contact between the ALR domains in this so-called "cross dimer." By analogy to other RR proteins whose active structures are known, we hypothesize that the ALR domains will dimerize in the active, DNA-binding state, likely through the same α4-β5-α5 interface that interacts with the DBD in the inactive structure. Again, by analogy to other RRs, there is likely to be a conformational change that

makes formation of this ALR dimer more favorable in the activated state. To investigate what residues are involved in the disruption of the ALR-DBD interface, and thus facilitate dimerization of the ALR domains, site-directed mutagenesis was employed. Specific point mutants were chosen based on the RitR(C128S) and RitR$_{ox}$ structures and are concentrated in regions of the structure known to be important in other RRs, or in regions that are particularly unusual in RitR. For example, residues in the active site that differentiate RitR from typical REC domains were investigated (K10D, N53D and N53A, D81A). Likewise, residues on the unusual α4 helix that were hypothesized to be part of a "hydrophobic gate" that could influence dimerization were also mutated (L86A, L90A, V93A, L86A/V93A, F96G). Residues on β5 that are involved in dimerization in other RRs were targeted (Y100 and Y102), since these residues make extensive interactions with the DBD in both structures. Finally, we tested the effects of mutating residues in the α5 linker helix that make key interactions in the ALR-DBD interface (R133A). C128 was targeted due to its central hypothesized role in redox-based sensing (C128A, C128S, C128D) and Y163 was mutated to Phe to break the water-mediated interaction with C128. The locations of the mutagenesis sites are shown in Figure 2-7, and the rationale for each is given in Table 2-5.
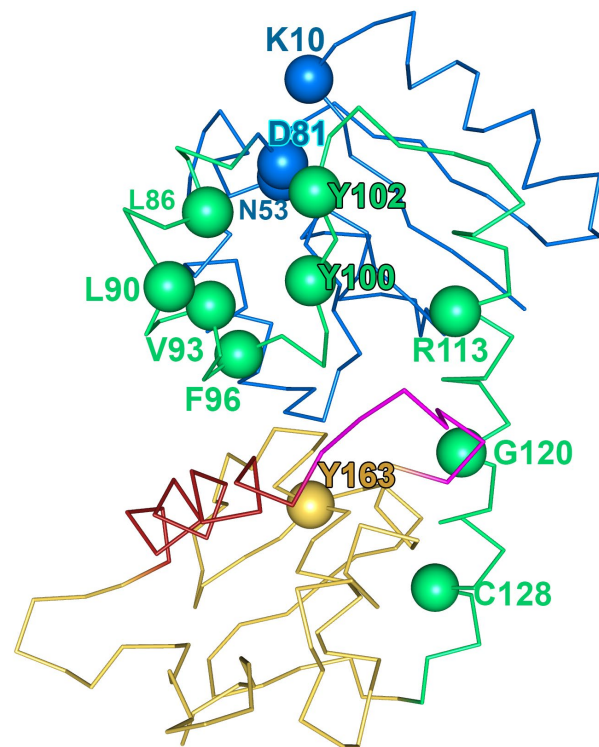
Figure 2-7: Cα trace of full-length RitR(C128S) structure with the sites chosen for mutagenesis marked as colored spheres. The ALR domain is colored blue, except for the α4-β5-α5 face, which is colored green. The DBD is colored pale yellow, except for the DND-contacting helix (red) and the trans-activation loop (magenta).

Table 2-5: Summary of the rationale of each mutation

Mutation Site	Reason for mutation
K10	Basic residue in the putative phosphorylation site, usually acidic residue in typical RRs.
N53	Phosphorylatable site
D81	Switch residue
Y100	Switch residue
F96	Hydrophobic core and impede the Y100 reorientation for dimerization
Y102	Point outward and affect the dimer interface formation
R113	Interact with Y100 and impede the Y100 reorientation for dimerization
C128	Redox sensing
Y163	Interact with S128(C) through a water
L86	Hydrophobic core and impede the Y100 reorientation for dimerization
L90	Hydrophobic core and impede the Y100 reorientation for dimerization
V93	Hydrophobic core and impede the Y100 reorientation for dimerization

70

The oligomeric state of each of the mutants was assessed by size exclusion chromatography in the hope of finding at least one variant that was dimeric even in the fully reduced state. Accordingly, each mutant was analyzed on the BioBasic SEC-300 column in the presence or absence of dithiothreitol (DTT). Wild-type RitR mainly exists as monomer, with a small portion of dimer (green trace in Figure 2-8A), which disappears when DTT is added (red trace). According to the sequence alignment in Figure 1-3, one of the acidic triad residues of typical response regulators is replaced with a basic residue, lysine, in RitR(K10; Figure 2-8B). The side chain of K10 disrupts both the acidic environment and the cleft for Mg^{2+} binding. The mutant K10D, which partially reconstitutes the Mg^{2+}-binding site, like the wild-type protein, exists mainly as monomer (Figure 2-8C). Interestingly, changing N53, which in typical RRs is the phosphorylatable aspartic acid residue, leads to a shift in the oligomeric state in the N53A variant (Figure 2-8D). This protein has a greater proportion of dimer than the wild-type, but monomer is still the major species. Changing N53 back to Asp (N53D) leads to an unstable and likely mostly unfolded protein (Figure 2-8E).
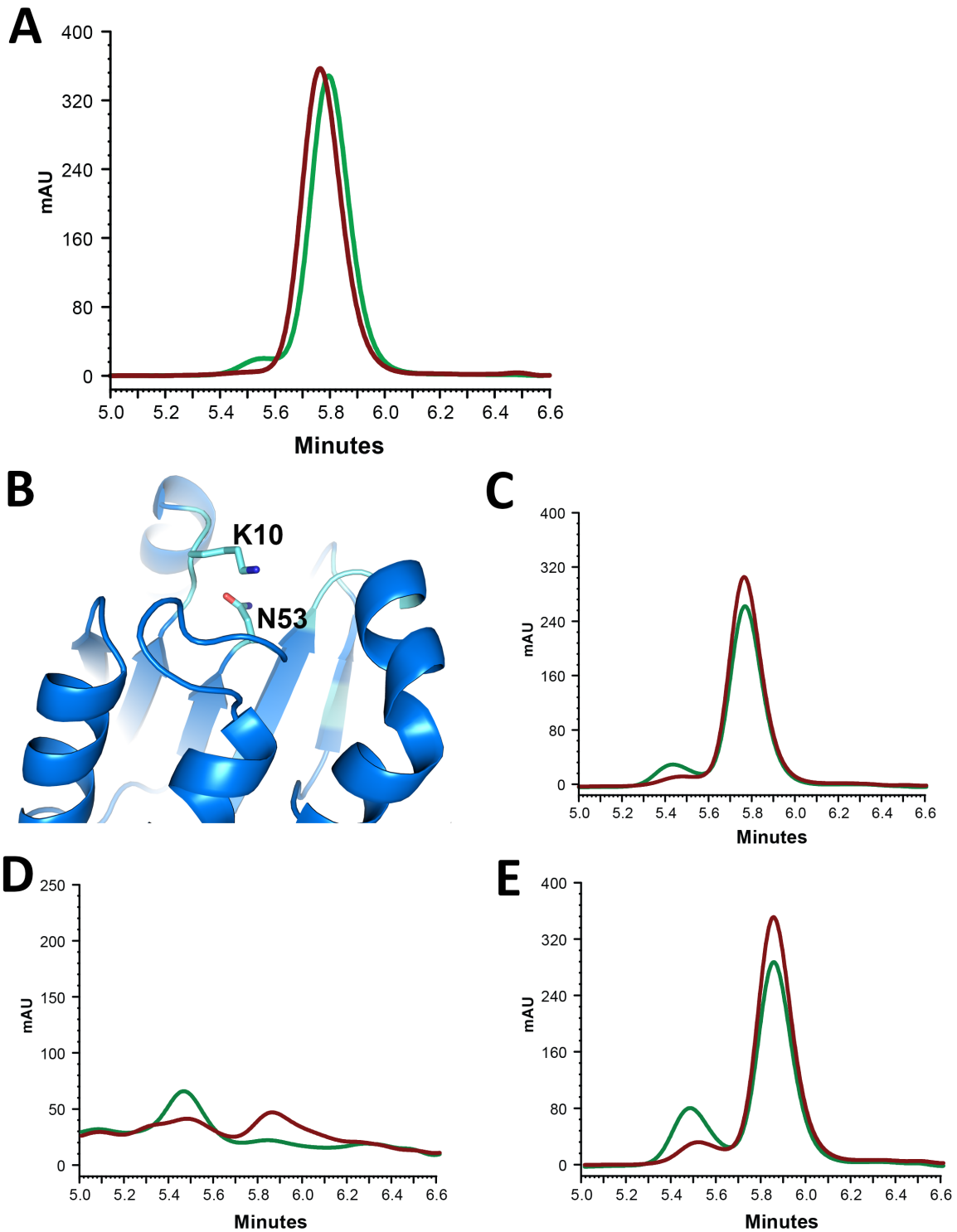
Figure 2-8: Oligomer states of wild-type RitR (A), RitR(K10D) (C), RitR(N53D) (D) and RitR(N53A) (E) without any treatment (green) or with 10 mM DTT treatment (red). (B) The putative active site of RitR. K10 and N53 are shown in stick.

In RitR, the residue that is equivalent to the threonine of the Y/T switch found in typical RRs is D81. This residue is oriented toward N53 and is held there by interactions with a number of active site residues (Figure 2-9). Compared with the corresponding residue in atypical regulator ChxR (L72) and typical regulator PhoP (T79), D81 in RitR represents the active conformer (Figure 2-9). This conformation of D81 is reminiscent of the interaction between T83 and the phospho-Asp side chain in activated PhoB[27] (Figure 1-4). In the inactive state, L72 in ChxR and T79 in PhoP forms part of gate residues, but in RitR, D81 points toward the putative active site and leaves L86 replacing the corresponding position. Some other atypical RRs that have been studied, such as ChxR, have a leucine residue at this position. It also shifts away from the active site.

The hydroxyl group of tyrosine of inactive ChxR and PhoB point outward from the domain, which interrupts the dimer interface formation. In the structure of non-phosphorylated PhoP (Figure 2-9), even it forms a similar "active" dimer with the switch residue Y98 pointing outward, but this dimer was obtain at high crystallization concentration and the helix α4 exhibits very high flexibility, which suggests this high concentration induced dimer is not a stable dimer. The phosphorylation-induced dimerization of PhoP (PDB ID: 2PL1[5]) is stable with Y98 side chain pointing inward. So in RitR structure, Y/T switch is effectively "uncoupled", where Asp81 is in the active form but Tyr100 is in the inactive form.
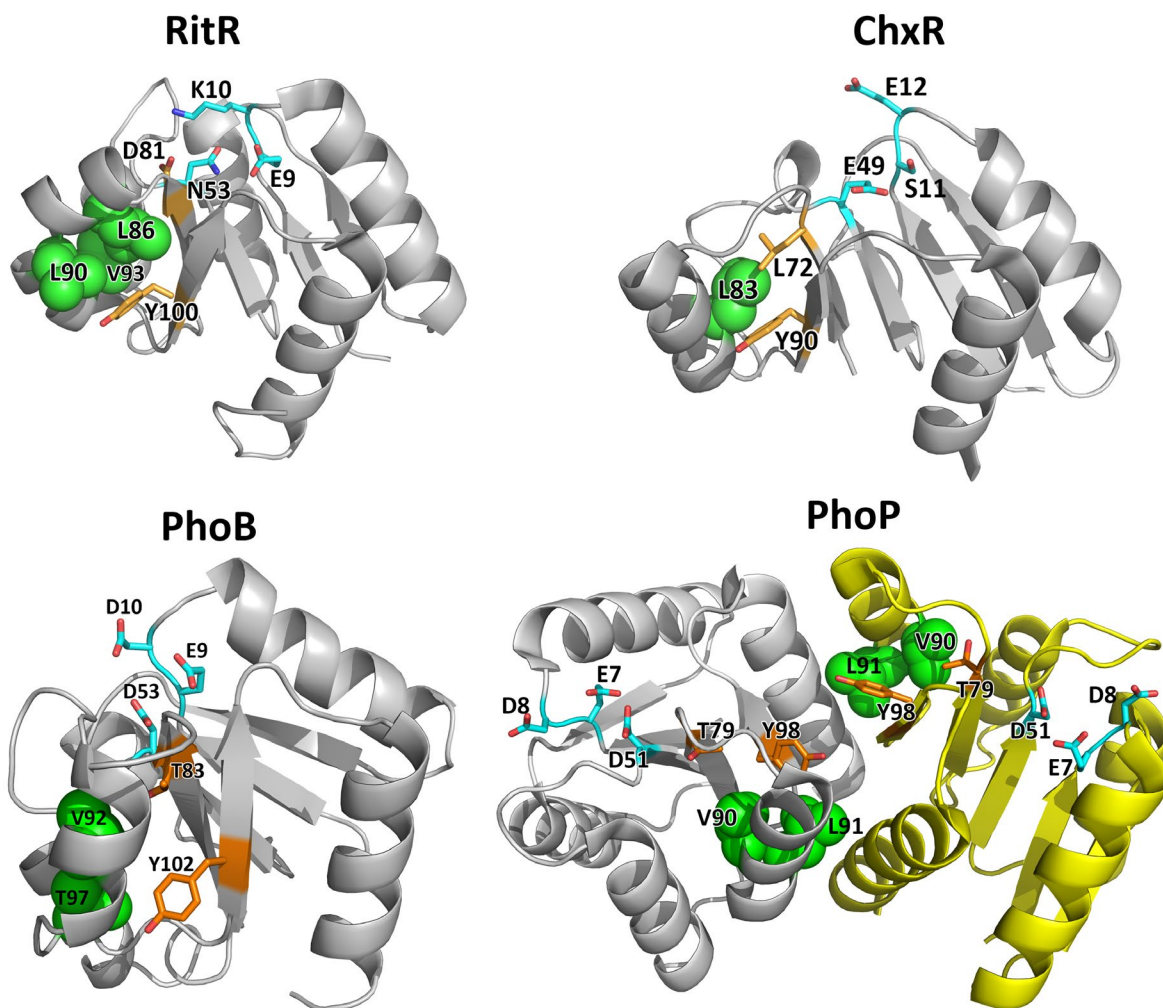
Figure 2-9: Comparison of gate regions of two ALR REC domains, RitR (PDB ID 4LZL[3]) and ChxR (PDB ID 3Q7T[4]), and the canonical REC domain of PhoB (PDB ID 1B00) and PhoP (PDB ID 2PKX[5]). These structures are all in an inactive state according to the positions of their Y/T switch residues. Even PhoP forms a dimer (Chain A colored in grey, chain B colored in yellow), but the dimer is a high protein concentration induced, unstable dimer. The activated dimer of PhoP is a more stable dimer with the switch residues shift inward. Gate residues are represented in green spheres, Y/T switch residues are shown as sticks with orange carbon atoms, and important catalytic residues are shown as sticks with cyan carbon atoms. Note that the Y/T switch residues L72 in ChxR, T83 in PhoB and T79 in PhoP form part of the gate, whereas D81 in RitR does not and is turned toward N53, resembling the active conformation of PhoB. In RitR, the switch residue side chain position is substituted by gate residue L86, which is strategically placed in front of tyrosine 100, allowed by the unusual formation of the kinked α4 helix in RitR.

74

In order to determine whether or not D81 and Y100 act similarly to the switch residues in other RR proteins, both residues were mutated (singly) to alanine and the effects on the oligomeric state assessed by SEC. Substitution of Y100 with Ala does not significantly increase the proportion of RitR in the dimeric conformation (Figure 2-10A). After treatment with DTT, there was no evidence of the dimer form of the Y100A variant. In contrast, the D81A variant exists as 65% dimer and 35% monomer (Figure 2-10B). After treatment with DTT, the proportion of dimer is less (~50%), but still represents a significant portion of the sample.
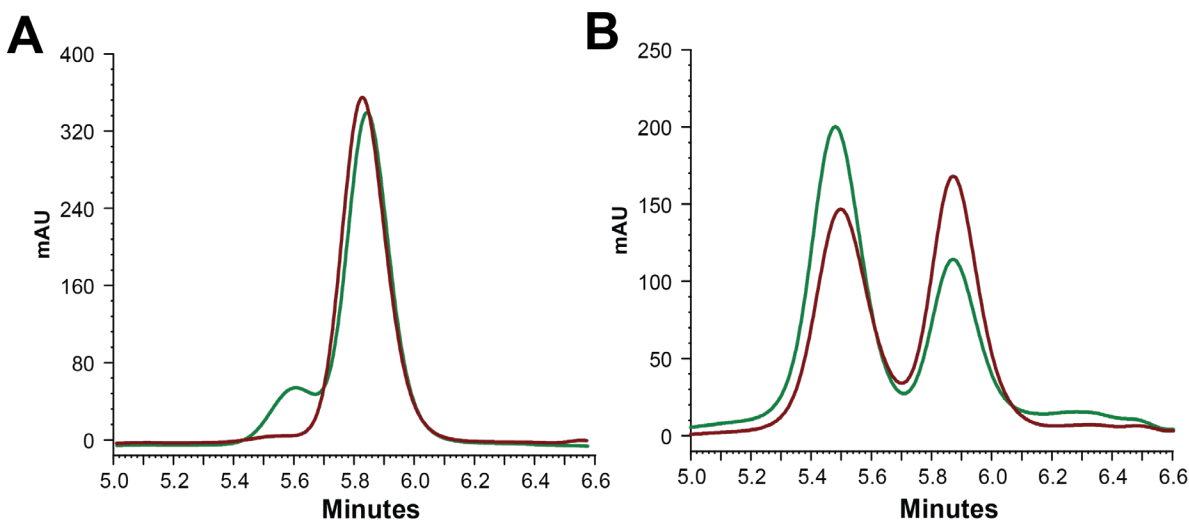
Figure 2-10: Results of SEC experiments on the Y100A (A) and D81A (B) variants. The green traces show the results in the absence of DTT treatment, and the red traces show the data after treatment with 10 mM DTT.

As shown in Figure 2-11, F96, which is located at the C-terminal end of helix α4b, is involved in the hydrophobic interactions within the hydrophobic core (β3, α3, β4). Substituting Gly for Phe at position 96 does increase the proportion of dimer compared to the wild type, but after DTT treatment, there was no dimer detectable. Also, the F96G

variant was not stable in solution.
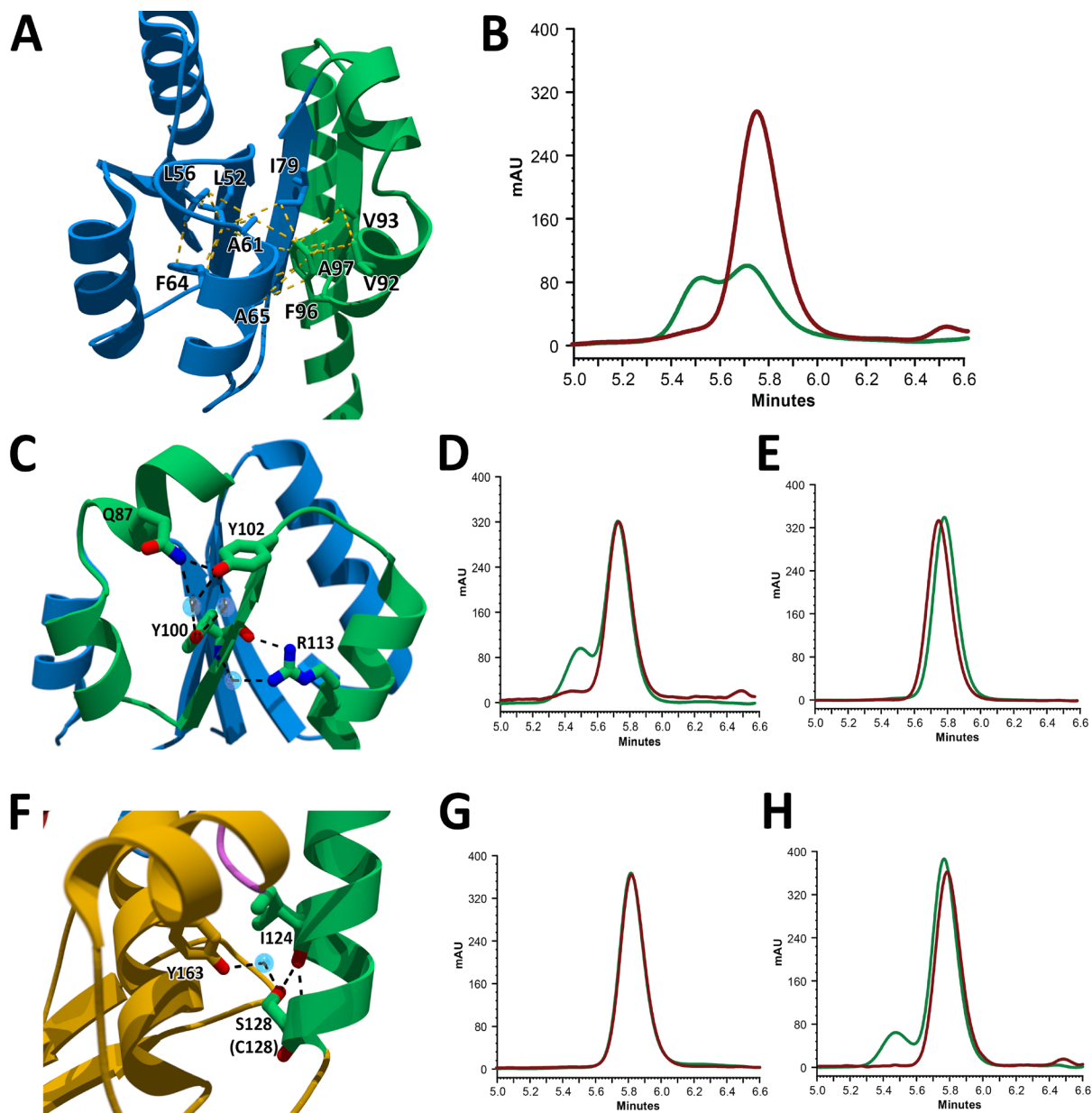
Figure 2-11: (A) F96 residue is involved in the hydrophobic interactions within hydrophobic core. (B) Oligomer states of RitR(F96G) (B), RitR(Y102A) (D), RitR(R113A) (E), RitR(C128A) (G) and RitR(Y163F) (H) with or without DTT treatment. Green represents no DTT treatment; brown red represents proteins were treated with 10 mM DTT. (C) Y102 and R113 both interact with the switch residue Y100 through water molecules. (F) Y163 interacts with serine in redox sensor C128 position through a water molecule.

Like Y100 discussed above, Y102 also points outward, away from the core of the ALR domain and projects out from the α4-β5-α5 face of the protein (Figure 2-11C), which may interfere with formation of the dimer interface. In addition, Y102 forms hydrogen bonds with Q87 and Y100 (through a water molecule), which likely impedes the reorientation of the Y100 side chain. When Y102 is mutated to Ala, there is little effect on the monomer-dimer equilibrium. The slight increase in the proportion of dimer becomes undetectable after treatment with 10 mM DTT by analytical SEC (Figure 2-11D). R113, which is located on the α5 linker helix, is also involved in interactions with Y100 and also makes a key interaction with the DBD in the inactive structures. Substitution of R113 with Ala results in a protein that is purely monomeric, whether or not it is treated with DTT (Figure 2-11E).

As expected, the C128 mutants (C128A) are purely monomeric (Figure 2-11G). The hydroxyl group of Y163 forms a water-mediated hydrogen bonding interaction with the hydroxyl group of S128 in the C128S variant (Figure 2-11F). This interaction could potentially be involved in the pK_a perturbation of C128. To test this idea, Y163 was mutated to Phe to remove the hydroxyl group. This change has very little effect on the monomer-dimer equilibrium. There is a slight increase in the proportion of dimer compared to the wild-type (Figure 2-11H), and treatment with DTT pushes the protein fully toward the monomeric form.

One feature of the RitR ALR domain that distinguishes it from typical OmpR/PhoB-type RRs is the kinked α4 helix. Helix α4 is broken in two, with an approximately 100° angle between the helical axes of the two pieces (α4a and α4b). No REC domain has been shown to have a similar conformation of helix α4. Three hydrophobic residues located in α4a

(L86), the short stretch of helix between α4a and α4b (L90), and in α4b (V93), pack against the central β sheet and make van der Waals interactions with V54 (Lβ3α3), I79 (β4), A97 (Lα4β5), Y100 (β5) and Y102 (β5) to form the hydrophobic core of this part of the protein (Figure 2-12A). This solidly packed core would seem to be a mechanical barrier impeding the reorientation of Tyr100 as seen in the structures of activated REC domains: a sort of "hydrophobic gate." If reorientation of Y100 is required for RitR dimerization as seen in other RRs, and if the hydrophobic gate residues do influence the conformation of Y100, then mutating these residues should have an impact on dimerization.

To test this idea, a series of mutants was constructed, including L86A, L90A, and V93A. The oligomeric state of each mutant was analyzed by size exclusion chromatography. The mutant L90A exists as a roughly equal mixture of monomer and dimer, and treatment with DTT shifts the equilibrium toward the monomeric form (~30% dimer; Figure 2-12B). Interestingly, RitR(L86A) and RitR(V93A) both exist as predominantly dimer in solution, and the response of each of these mutants to DTT is quite different (Figure 2-121C, D). The RitR(L86A) variant is nearly insensitive to DTT, with only slightly more monomer present after treatment with 10mM of the reductant. On the other hand, the V93A variant shifted dramatically, becoming predominantly monomeric after DTT treatment. Since they had the greatest abount of dimer of any mutants tested, the RitR(L86A) and RitR(V93A) mutants were combined (RitR(L86A/V93A); Figure 2-12E). The double mutant was slightly more stable to DTT treatment than the RitR(L86A) variant. In all, Ala substitution on any of the three hydrophobic gate residues dramatically increases the proportion of RitR dimers. The RitR(L86A/V93A) double mutant was purified in large quantities for crystallographic structure determination.
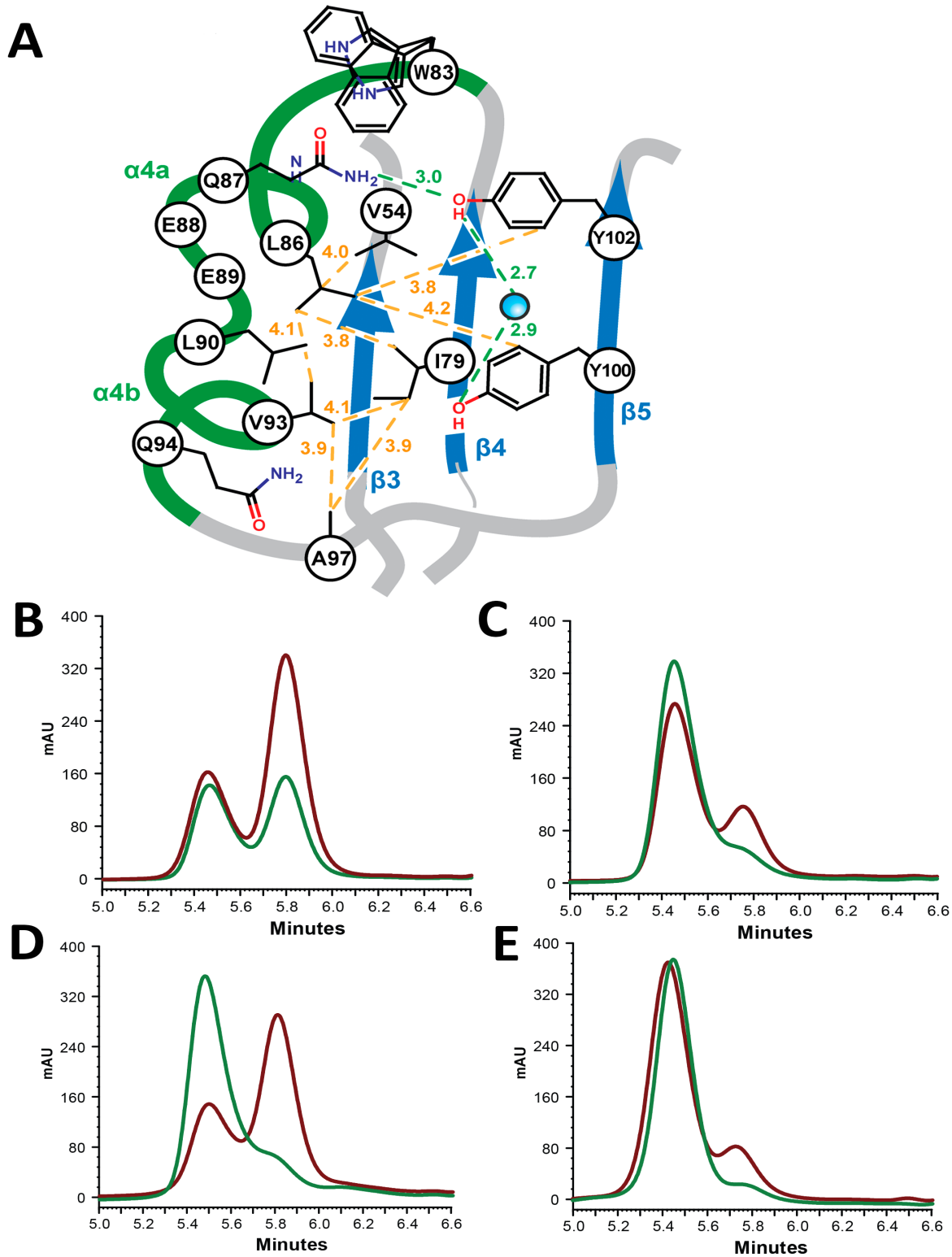
Figure 2-12: The effects of gate residues on RitR dimerization. (A) Schematic view of RitR gate residues van

der Waals interactions (yellow dotted lines) and hydrogen-bonding network (green dotted lines) in the

dimer/gate region. Oligomer states of wild-type RitR(L90A) (B), RitR(L86A) (C), RitR(V93A) (D) and RitR(L86A/V93A) (E) without any treatment (green) or with 10 mM DTT treatment (red).

2.3.7 Mutations to the gate residues L86 and V93 result in an "extended" dimer

We were surprised to learn that the retention time of the RitR(L86A/V93A) dimer was significantly earlier (5.4 min) than that of the "cross dimer" conformation (5.6 min) observed for the wild-type protein oxidized with H_2O_2 (Figure 2-13). In addition, the tiny monomer peak of the RitR(L86A/V93A) variant eluted earlier than that of the wild-type RitR monomer. Both of these observations suggest that the conformation of RitR(L86A/V93A) is more extended, making it behave like a larger protein on the SEC column. This led us to believe that the inter-domain interactions between the RitR(L86A/V93A) ALR domain and the DNA-binding domain are disrupted. This is supported by the small difference in thermal stability between the double mutant and the wild-type protein measured by differential scanning fluorimetry. The average T_m value (46.6°C) measured for RitR(L86A/V93A) is lower than that of wild-type RitR (52.2 °C), which suggests that RitR(L86A/V93A) is less stable than RitR. The D,L-malic acid buffer used for storage of the wild-type protein is still a good buffer to store RitR(L86A/V93A) (Table 2-6 and Table 2-7).
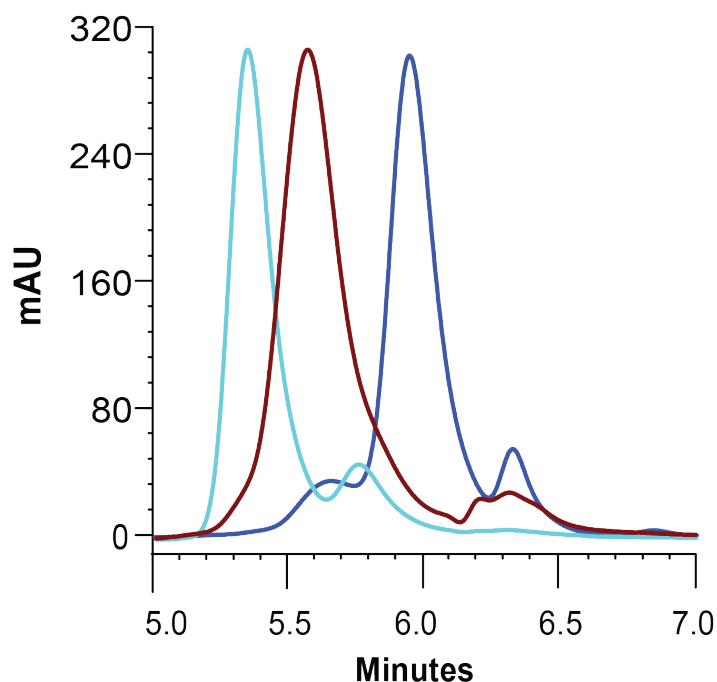
Figure 2-13: SEC retention times of wild-type RitR (as purified, colored in blue), wild-type treated with H$_2$O$_2$ (colored in brown), and the RitR(L86A/V93A) variant (colored in cyan).

Table 2-6: Buffers from the Solubility & Stability Screen (Hampton Research) giving the highest T$_m$ values in the Thermofluor assay.

N.O.	T$_m$ - T$_{m,averaged}$ (°C)	Buffer
1	4.3	Sodium phosphate monobasic, Potassium phosphate dibasic
2	3.5	Methyl α-D-glucopyranoside
3	3.3	Lithium citrate tribasic tetrahydrate
4	3.2	Sodium sulfate decahydrate
5	3.1	Sucrose
6	2.9	Succinic acid pH 7.0
7	2.7	Spermine tetrahydrochloride
8	2.6	Sodium malonate pH 7.0
9	2.5	DL-Malic acid pH 7.0

Table 2-7: Buffers from the lice pH Screen (Hampton Research) giving the highest T_m values in the Thermofluor assay.

N.O.	T_m (°C)	Buffer
1	3.0	DL-malic acid pH 5.3
2	2.7	DL-malic acid pH 5.9
3	2.7	Succinic acid pH 6.0
4	2.5	Succinic acid pH 5.7

2.3.8 Size exclusion chromatography of RitR : BS3 and RitR(L86A/V93A) : BS3

Since the two domains in the double mutant appeared to be dissociated from one another, while still allowing the protein to dimerize, it is possible that the RitR(L86A/V93A) mutant dimer conformation is the active conformation. To ascertain whether or not this variant is active, the DNA-binding ability of RitR(L86A/V93A) to bind to DNA was assessed by SEC. RitR is known to directly bind and repress *piu* transcription, and three binding sites (BS1-3) were identified by electrophoretic mobility shift assay (EMSA)[1]. Due to the much higher affinity of RitR for BS3, we concentrated on this target sequence. In this thesis, I tested the affinity of six lengths of BS3 for crystallization (Table 2-1). And BS3-1 is BS3.

As shown above, and in Figure 2-14A below, RitR exists mainly as a monomer in the absence of any oxidant. Incubating the wild-type protein with the BS3-1 olignucleotide did

not result in formation of a stable RitR:BS3-1 complex. Oxidation of RitR with 1 mM H_2O_2 for several hours prior to the addition of BS3-1 DNA, resulted in about 60% RitR forming the "cross dimer" $RitR_{ox}$ conformation. There is no peak corresponding to the $RitR_{ox}$:BS3-1 complex. The RitR(L86A/V93A) variant, on the other hand, is present almost exclusively as the extended dimer, and when DNA was added, the majority of the protein elutes as a complex with the BS3-1 oligonucletide (retention time ~5.1 min). Oxidation of RitR(L86A/V93A) with H_2O_2 did not significantly alter the oligomeric state of the double mutant or its ability to bind to the BS3-1 DNA. To verify the binding of RitR(L86A/V93A) to BS3-3 while the wildtype RitR protein does not, differential scanning fluorimetry was used. When the protein bound to BS3-3, the protein would get stabilized and would not be easier to be denatured (higher T_m). But if the protein cannot bind to the DNA, T_m value would not be affected. In all three buffers, RitR T_m value does not change significantly, while the T_m value of mutant RitR(L86A/V93A) significantly increases (Figure 2-15). These results show that wild-type RitR in both its reduced and oxidized forms is not able to bind to the dsDNA oligonucleotide BS3-1. However, there is a complex peak formed for the sample RitR(L86A/V93A) with dsDNA BS3-1, while RitR(L86A/V93A) with its presumed altered conformation, is able to do so, whether or not it is treated with H_2O_2. This suggests that the L86A/V93A substitutions bypass the C128-mediated redox-sensing trigger, resulting in a constitutively active form of RitR.
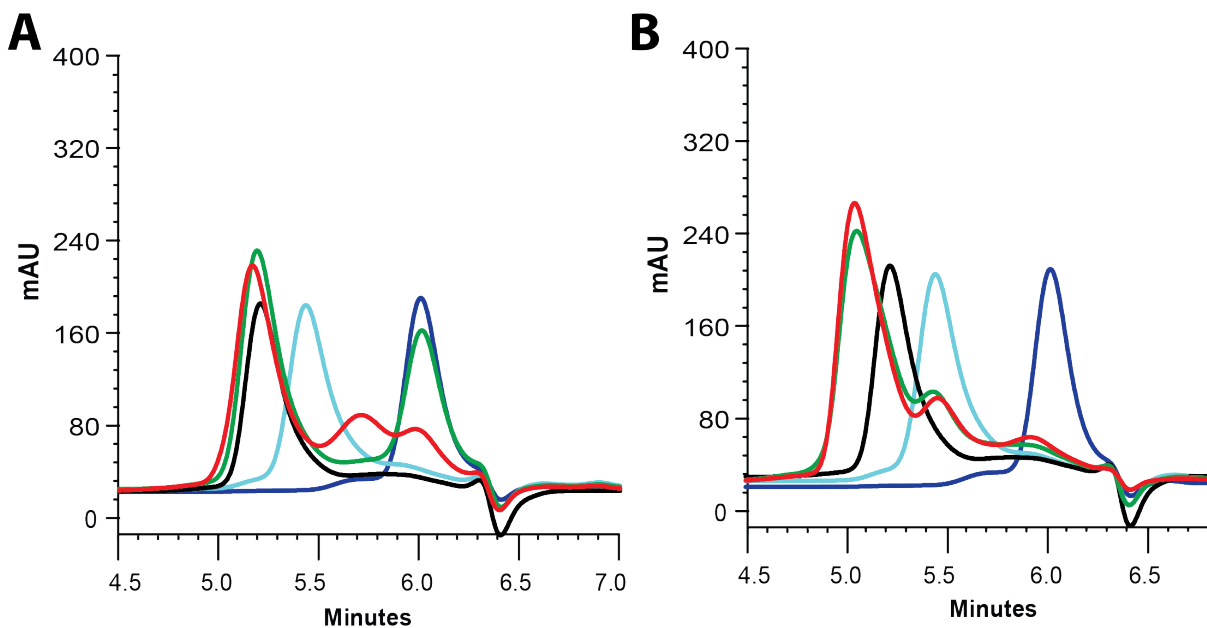
Figure 2-14: Size exclusion chromatography results for RitR with and without the BS3-1 dsDNA oligonucleotide (A) and RitR(L86A/V93A) with and without BS3-1 (B). In both panels, the traces for the wild-type RitR protein alone are blue, the RitR(L86A/V93A) protein alone cyan, BS3-1 alone black, RitR or RitR(L86A/V93A) with BS3-1 without H_2O_2 green, and with 1 mM H_2O_2 red. For the mixture of RitR or $RitR_{ox}$ with BS3-1, dsDNA BS3-1 and proteins were eluted separately. For the mixture of RitR(L86A/V93A) with BS3-1, the RitR(L86A/V93A):BS3-1 complex was eluted, as indicated by the earlier elution time when compared to the BS3-1-only and protein-only peaks.
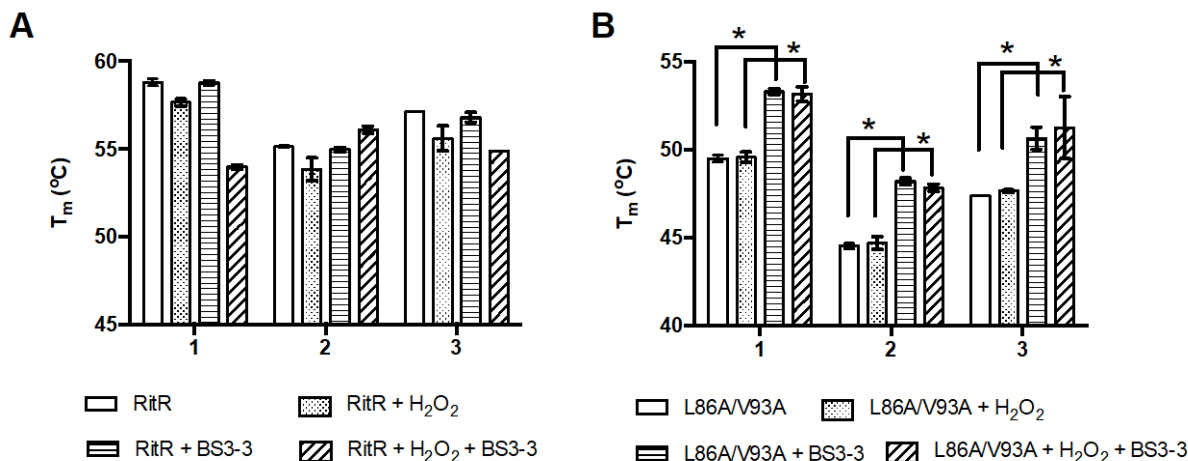
Figure 2-15: T_m value change of (A) RitR, RitR$_{ox}$ and (B) RitR(L86A/V93A) after incubated with BS3-3 dsDNA. **1**: pH 6.5, 25 mM TRIS, 150 mM NaCl, 100 mM DL-malic acid; **2**: 20 mM HEPES pH 7.2, 5 mM MgCl$_2$, 1 mM CaCl$_2$, 0.1 mM EDTA; **3**: 20 mM HEPES pH 7.2, 5 mM MgCl$_2$, 1 mM CaCl$_2$, 0.1 mM EDTA, 150 mM NaCl. After incubated with BS3-3, RitR T_m did not change significantly in three different buffers. Similar results were obtained for RitR$_{ox}$. But the mutant RitR(L86A/V93A) T_m value increased significantly in all three buffers.

2.3.9 Single mutants in helix α4 also bind to BS3-1

The studies above verified that mutagenesis on both L86 and V93 that located in the kinked α4 helix bind to the target DNA BS3-1. To investigate the binding ability of three single mutants (L86A, L90A, V93A) in α4 helix, the oligomer states of them with BS3-1 were determined by SEC. The results were shown in Figure 2-16. These three single mutants on α4 helix all bind to BS3-1, because a complex peak eluted earlier than BS3-1 only peak.
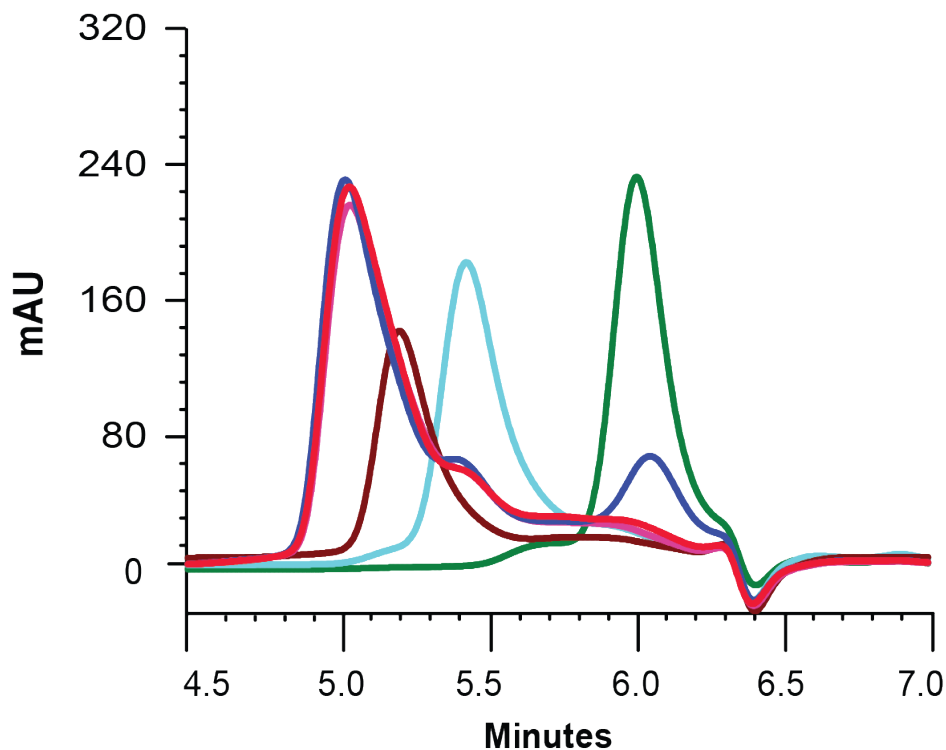
Figure 2-16: Gate residue variants bind to the dsDNA BS3-1 due to a complex peak eluted. In the panel, the traces for the wild-type RitR protein alone are green, the RitR(L86A/V93A) protein alone cyan, BS3-1 alone brown, RitR(L86A) with BS3-1 magenta, RitR(L90A) with BS3-1 blue and RitR(V93A) with BS3-1 red.

2.3.10 Crystallization of RitR(L86A/V93A)

RitR(L86A/V93A) is almost entirely dimeric as observed by SEC, but there are detectable amounts of monomer. Since this heterogeneity would make the protein difficult to crystallize, the dimer form was purified by preparative SEC with the same protocol used to purify the RitR$_{ox}$ dimer (with the exception of the H$_2$O$_2$ treatment). The fractions containing the RitR(L86A/V93A) dimer were concentrated to 7,2, 10.1, or 34.0 mg/ml and

used to set up sparse matrix screens for the identification of crystallization conditions (Index, PEG/Ion, SaltRx, and PEGRx from Hampton Research). These screens resulted in only clear drops, phase separation, or amorphous precipitate; no conditions resulted in crystalline protein. Examination of the conditions that gave phase separation (super-concentrated protein) showed that the protein favors neutral pH (HEPES or BIS-TRIS at pH 6.5-7.5), PEG3350, and a variety of salts having monovalent cations (ammonium sulfate, lithium sulfate, sodium formate, sodium malonate). These conditions were optimized with different RitR(L86A/V93A) concentrations on both sitting drop and hanging drop methods at different temperatures (room temperature, 16°C, 22°C). Considering the protein crystallization diagram, increase the protein concentration (supersaturation) is necessary for nucleation. So the highest protein concentration I tried is 88 mg/mL. But only stronger phase separation was obtained. There is no crystal formed. For those phase separation conditions, additive screen (Hampton) was also tried. But still no crystal formed. I also tried to use some other proteins crystals as microseeds (sbADC, RitR$_{ox}$, RitR(C128D)) to crystallize the RitR(L86A/V93A) protein, but none succeeded.

Some proteins only crystallize with the cognate DNA, for example, AmrZ in *Pseudomonas aeruginosa*[28]. Based on the shorter retention time on size exclusion column, RitR(L86A/V93A) is assumed to be an extended dimer. Because it's so hard to crystallize RitR(L86A/V93A) protein, RitR(L86A/V93A) probably is the protein that only crystallizes with cognate DNA. Then I tried to co-crystallize RitR(L86A/V93A) with dsDNA.

2.3.11 Shorten BS3-1 length for crystallization

One variable of protein-DNA complex crystallization is the length of DNA. The general principle is "smallest functional". So different lengths of oligonucleotides were designed to provide blunt-ended dsDNA fragments. In addition, the ratio of protein to DNA is also important for successful crystallization. This ratio relates directly to the shape of the protein component, rather than the oligonucleotide, because the overall shape of the B-DNA is regular and cylindrical.
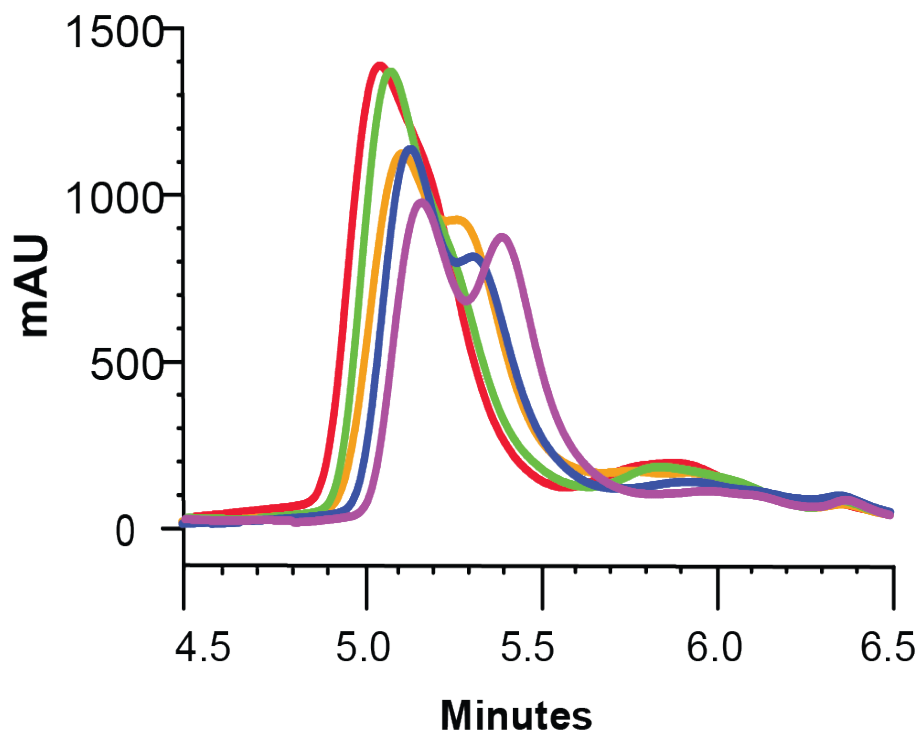
Figure 2-17: Complex peak area comparison of 216 µM RitR(L86A/V93A) with 100 µM of BS3-1 (red), BS3-2 (orange), BS3-3 (green), BS3-4 (blue) and BS3-5 (purple).

To find out which blunt-ended dsDNA at what ratio to RitR(L86A/V93A) protein gives the most quantity of protein-DNA complex for crystallization, the complex peak area of different ratios of RitR(L86A/V93A) to dsDNA (BS3-1 to BS3-5) was determined by SEC. RitR(L86A/V93A) protein (216 μM) was mixed with 100 μM of dsDNA (BS3-1 to BS3-5). The results were shown in Figure 2-17. Comparing the complex absorbance at 280 nm of RitR(L86A/V93A) with different lengths of DNA, BS3-3 with 35 bp nucleotides gives the most quantity of protein-DNA complex.

2.3.12 The ratio of RitR(L86A/V93A) : BS3-3 and RitR(L86A/V93A) : BS3-6

Based on the above results, BS3-3 gives the most quantity of protein-DNA complex, so BS3-3 was chosen to cocrystallize with RitR(L86A/V93A). Another shorter dsDNA (BS3-6) was also used in cocrystallization with RitR(L86A/V93A) protein. Because the dsDNA is limiting, to get the maximum protein-DNA complex, the "best" ratio of RitR(L86A/V93A) over BS3-3/BS3-6 was determined by comparing the absorbance of complex peak at different protein : DNA conditions. According to the gel filtration results shown in Figure 2-18, the absorbance of protein-DNA complex increases as the protein to DNA ratio increases. For RitR(L86A/V93A) with BS3-3 (Figure 2-18A), the maximum complex was obtained at the ratio of 4 : 1 with the excess protein. Increasing the ratio to 10:1 does not increase the protein-DNA complex peak absorbance. For RitR(L86A/V93A) with BS3-6 (Figure 2-18B), the BS3-6 DNA peak overlaps with the RitR(L86A/V93A) protein only peak. The binding affinity is lower than that of BS3-3, because even treated with 10 fold more protein, the complex absorbance still not get maximum. So the maximum ratio of 10 : 1 with the excess

protein was used. The two conditions were used for incubating protein with DNA, and then the complex was separated on a HiPrep™ 26/60 Sephacryl® S100 HR column. Before
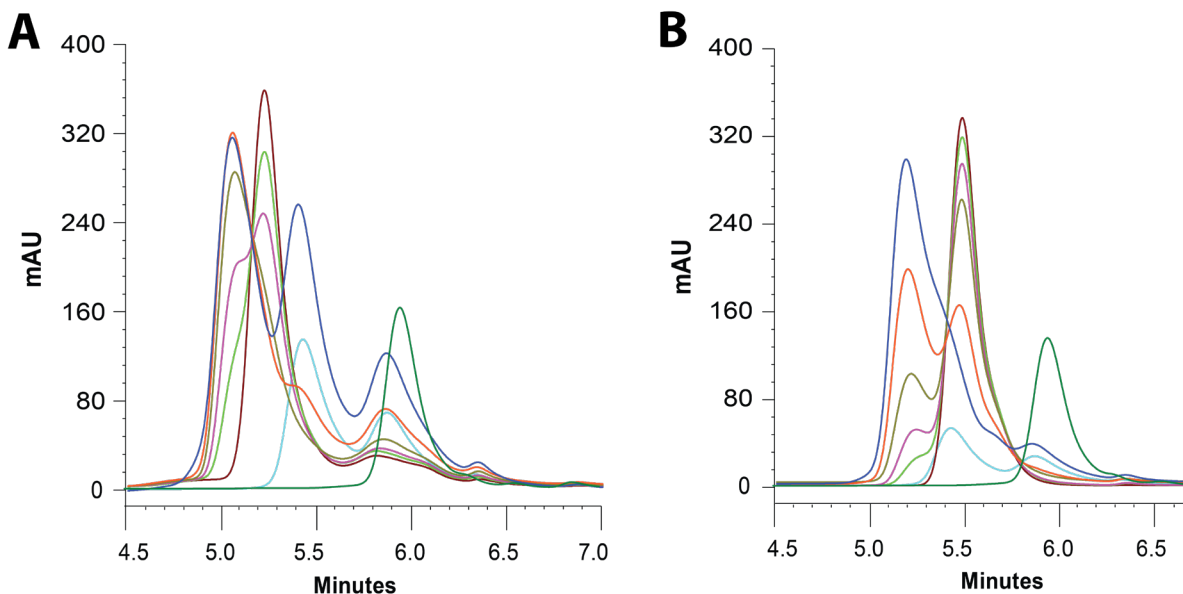
Figure 2-18: Gel filtration results of different ratios of RitR(L86A/V93A) : BS3-3 (A) and RitR(L86A/V93A) : BS3-6 (B). In both panels, the traces for the RitR(L86A/V93A) protein alone are dark green, the RitR(L86A/V93A) protein alone cyan, BS3-3 or BS3-6 alone brown, 100 μM BS3-3 or BS3-6 were incubated with 25 μM (green), 50 μM(purple), 100 μM (dark yellow), 200 μM (orange) and 500 μM (blue) of RitR(L86A/V93A).

making the crystallization tray, extra BS3-3/BS3-6 DNA was added into the complex solution. But there is no crystal was obtained but salt crystals.

2.3.13 Structure of the RitR(L86A/V93A) REC domain

Likely owing to the higher degree of flexibility in the "extended dimer" of the RitR(L86A/V93A), we were not successful in obtaining diffraction-quality crystals of the

full-length protein. Attempts to "lock down" the DNA-binding domains by co-crystallizing a number of BS3 variants with the double mutant. This was also unsuccessful. In order to determine whether or not the ALR domains of RitR(L86A/V93A) formed a stable dimer, an expression construct was prepared that included only residues 1-124 (RDLV; the same residues used in the original wild-type ALR domain structure). Diffraction-quality crystals were obtained from hanging drops containing 1 μl protein solution (4.4 mg/ml RDLV in 10 mM TRIS pH 8.5, 50 mM Succinic acid) and 1 μl crystallization solution (20% (w/v) PEG 3350, 0.2 M magnesium sulfate, 0.1 M TRIS pH 8.5), which diffracted to 1.8 Å resolution. The crystals belonged to space group $P222_1$ (a = 50.1 Å, b = 62.9 Å, and c = 80.3 Å) and contained 2 chains per asymmetric unit. We expected that the structure would be solved easily by molecular replacement, since the structure of the wild-type ALR domain was known. However, the molecular replacement calculation consistently failed, no matter which REC domain structure was used as the search model. SeMet-substituted RDLV was prepared, and it crystallized in the same condition used for the native protein, except 1 mM DTT was added to prevent heterogeneity due to oxidation of the SeMet residues. The data extended to 1.9 Å resolution and the anomalous signal from the 14 Se atoms provided excellent initial phase estimates that allowed Autobuild [29] to build an approximately 74% complete model. The two chains in the asymmetric unit form a dimer, which is consistent with the size exclusion chromatography results (data not shown). There are numerous hydrogen bonds between two chains, with an interface area of approximately 2,261 $Å^2$.

Shockingly, the topology of the RitR(L86A/V93A) ALR dimer is drastically altered from the $(\beta/\alpha)_5$ topology seen in all previous REC domain structures (Figure 2-19, Figure 2-10). There are two primary differences between the RDLV and all other REC domain structures.

First, the kinked α4 helix rearranges to form an extension of strand β4 and a long loop leading into β5. Owing to the loss of α4, β5 is no longer parallel to β4. Thus the central sheet in the RDLV ALR domain is mixed rather than parallel as it is in the wild-type ALR. Second, "α5," which is technically the 4th α-helix in the structure, is no longer docked with β5 and the rest of the domain, but rather participates in a domain-swapping interaction with the other protomer of the dimer. The two protomers literally swap their "α5" helices, and other than coming from the opposite dimer, there is little difference in the positions of "α5" of the RDLV and α5 of the monomeric ALR with respect to the rest of the tertiary structure.

Figure 2-19: Ribbon representation of the RDLV structure (A) showing the arrangement of the secondary structure elements. Chain A of the RDLV dimer is colored dark blue and dark orange (the rearranged β4, β5 and α5); chain B is colored light blue.and yellow (the rearranged β4, β5 and α5). The Y100 residues of each chain (shown as sticks) stack with each other in the dimer interface. Note the loss of the α4 helix, which becomes part of β4 and a long loop leading into β5. This reverses the direction of strand β5 and results in a mixed β sheet structure, rather than the parallel sheet typical of REC domains. The "α5" helices are swapped between the two chains, but occupy a position in the structure very similar to that of α5 in the monomeric RitR REC domain. (B) The overlay of the RDLV structure (chain A, colored in dark blue) onto the wild-type RitR REC domain structure (colored in grey) highlights the extent of the structural rearrangements leading to this dimeric form of the ALR domain.

92
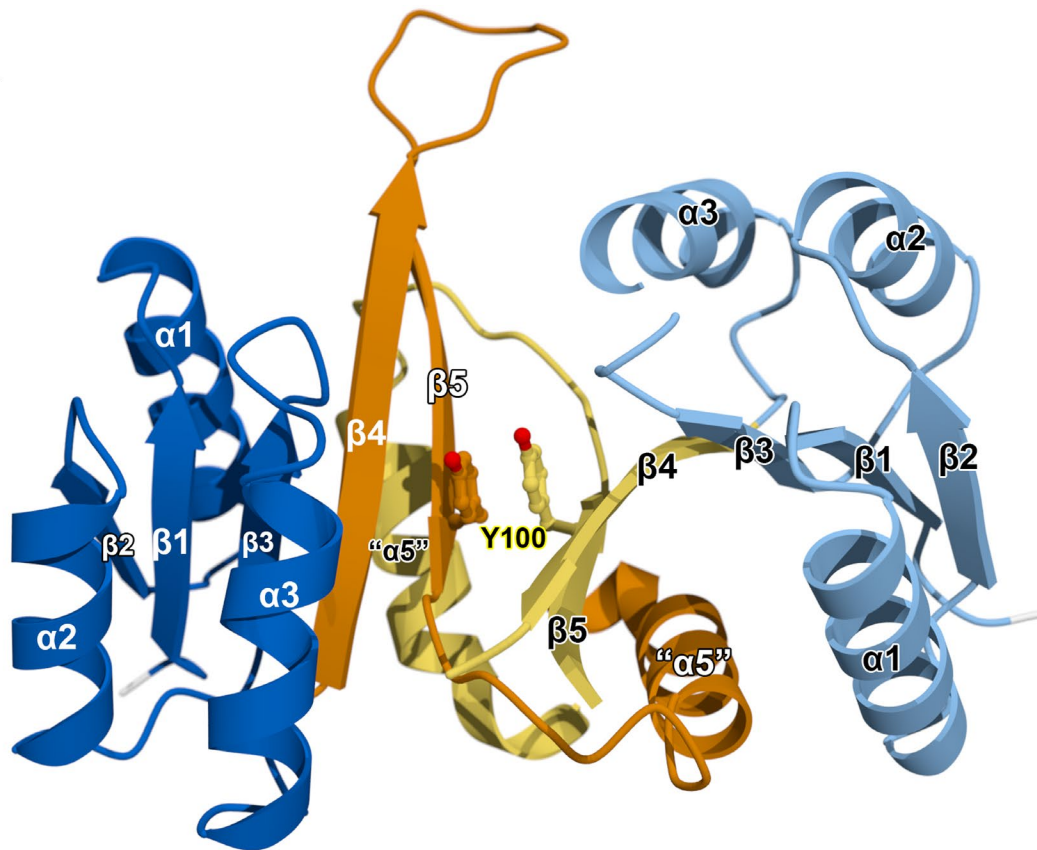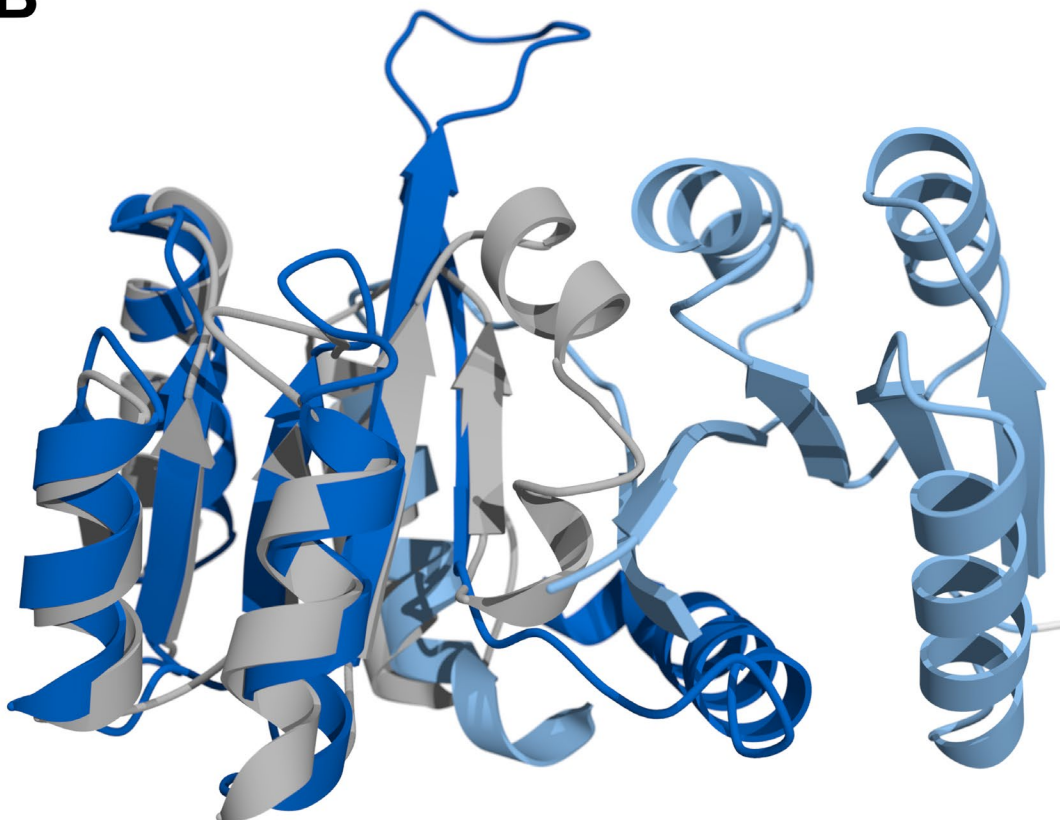
Of course, the pronounced changes in the topology of the RDLV REC domain lead to serious changes in the active site and switch residues (D81 and Y100; Figure 2-20). With the rearrangement of the hydrophobic core in the β4-α4 region of the protein, the side chains of N53 and N55, which point away from each other in the wild-type structure, both rotate ~180° to participate in a hydrogen bonding network that includes the amide nitrogen of V54. The side chain of the switch residue D81, which makes a hydrogen bonding interaction with the V54 amide nitrogen atom in the wild-type REC domain moves slightly in the RDLV REC domain so that it no longer interacts with V54. The other switch residue, Y100, also has an ~180° change in the orientation of its side chain, though in this case it is due to the "flipping" of the β5 strand rather than a simple rotation about the Cα-Cβ bond. In this orientation, the rings of Y100 from each protomer stack against one another at the heart of the dimer interface.

Figure 2-20: Comparison of the active site and switch residues in the wild-type RitR REC domain (A) and the RDLV variant (B). The wild-type REC domain is colored grey; RDLV dimer is colored in dark blue (chain A) and light blue (chain B). Because the kinked α4 helix forms the loop between β4 and β5 and changes the orientation of β5, the loop between β3 and α3 then moves due to the changed interactions. The N55 side chain flips and occupies the space occupied by K10 in the wild-type RitR REC domain.
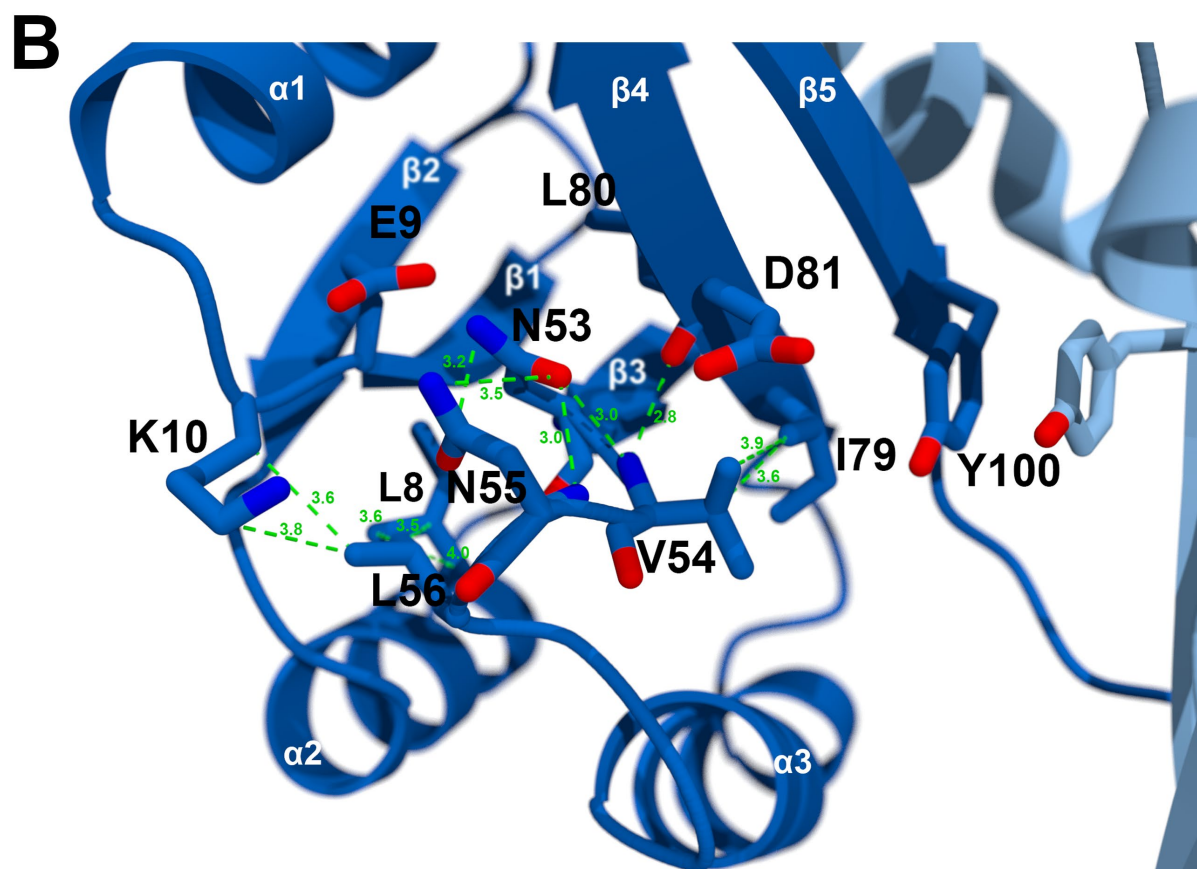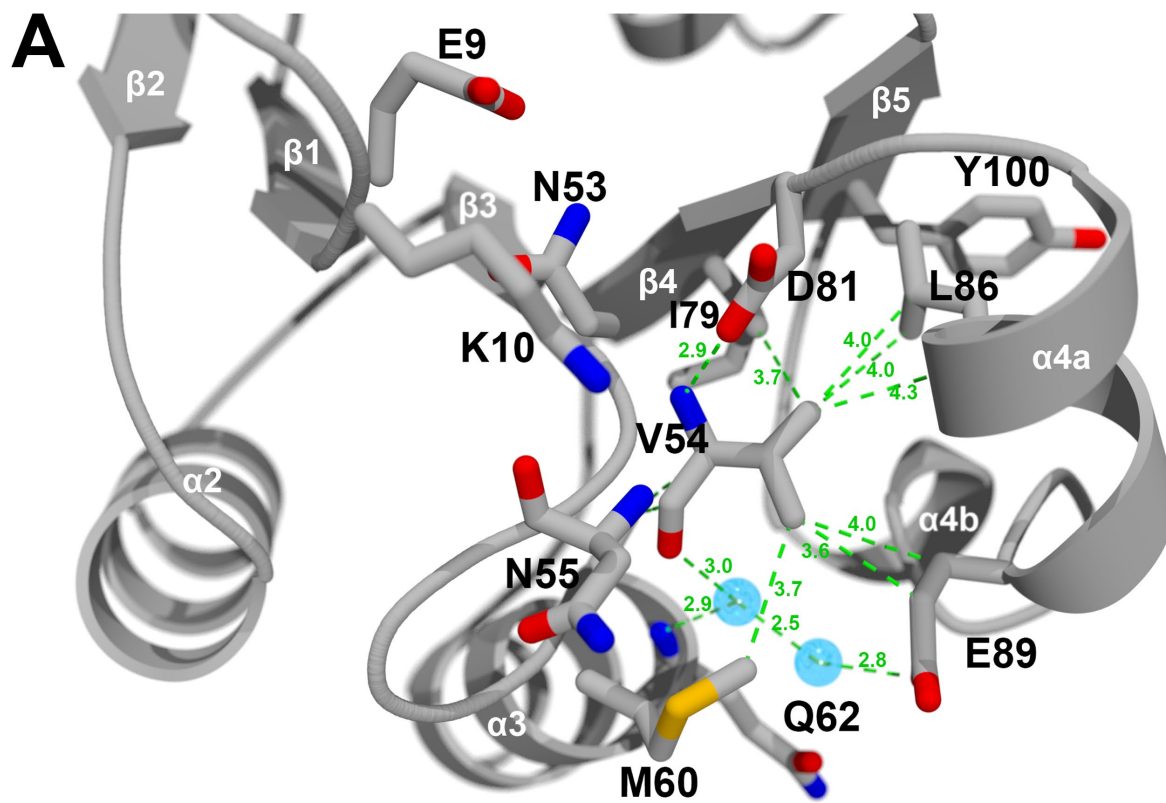
2.3.14 Structure of the RD2LV REC domain construct

Because of the radical nature of the structural changes observed in the RDLV variant, we thought it important to confirm that the structure is biologically relevant. To this end, we constructed a slightly longer form of the RDLV REC domain that includes the putative redox-sensing C128. This longer construct was dubbed RD2LV. SeMet-substituted protein was produced and diffraction-quality crystals were obtained from Hanging drops containing 1 µl of protein solution (19 mg/ml RD2LV in 10 mM glycine pH 8.8, 25 mM L-Arg) and 1 µl of crystallization solution (22.5% (w/v) polyethylene glycol (PEG) 3350, 12% Tacsimate pH 8.2).. These crystals diffracted to 2.6 Å resolution. They belonged to space group $P2_12_12$ with a = 115.3 Å, b = 125.4 Å, and c = 40.2 Å. The asymmetric unit contained four molecules arranged as two homodimers (Figure 2-21B).
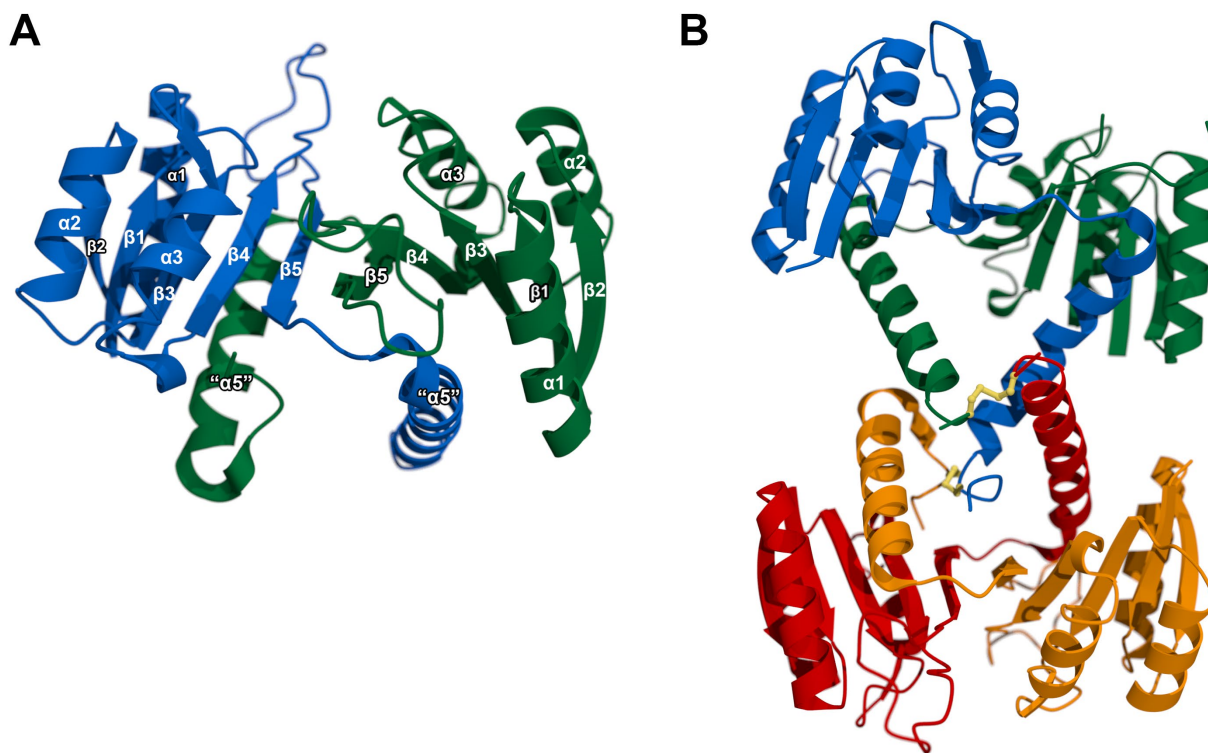
Figure 2-21: Ribbon representations of the RD2LV REC domain structure (A) and the RD2LV asymmetric unit (two dimers). The presence of C128 does not alter the domain swapped dimer, which is nearly identical in both structures. There are inter-dimer disulfide bonds that technically turn this into a tetramer. This appears to be an artifact of crystallization, as no tetramer is detectable in SEC traces of RD2LV (data not shown).

The structure of RD2LV is nearly identical to that of the shorter RDLV protein, both in terms of the tertiary structure of the protomer as well as the quaternary structure of the dimer (Figure 2-21A). The largest difference between the two structures is the fact that the "α5" helix is longer in RD2LV, which is not unexpected. This leads both to a modest increase in the number of hydrogen bonds and salt bridges between two chains, while the buried surface area of the interface is nearly unchanged (2,272 Å^2 vs 2,261 Å^2). More importantly, there are two disulfide bonds between the C128 residues of the two dimers (Figure 2-21B). The inter-dimer disulfide bonds are likely an artifact of the high concentrations required for crystallization, since there is not clear evidence of tetramerization in the analytical SEC data for the RD2LV variant (data not shown). Presumably, the α5 helices of the two protomers in a single dimer could also become disulfide linked, and this would hold the two DBDs in the proper orientation for DNA binding. The arrangement of the α5 helices in the RD2LV dimer suggests that the highly unusual REC domain structure observed here may be biologically relevant.

2.3.15 RitR with GADDY motif (RitR_GADDY)

A comprehensive investigation of OmpR family proteins[30] shows that RitR homologs have lost not only the phosphorylatable Asp residue, but also several highly conserved acidic residues that comprise the REC "GADDY motif" (see residues highlighted with red in Fig. 2-22), which is begins in the loop between α4 and β5 (Lα4β5) and ends near the middle of strand β5. The GADDY motif contributes to hydrogen bonding in the dimer interface of canonical REC domains [3, 31]. The terminal Tyr residue of the GADDY motif is the Y/T couple residue. We hypothesized that the loss of the GADDY motif in RitR (the motif becomes (F/Y)AVSY in RitR homologs) may explain why it only forms dimers in response to Cys128 oxidation.
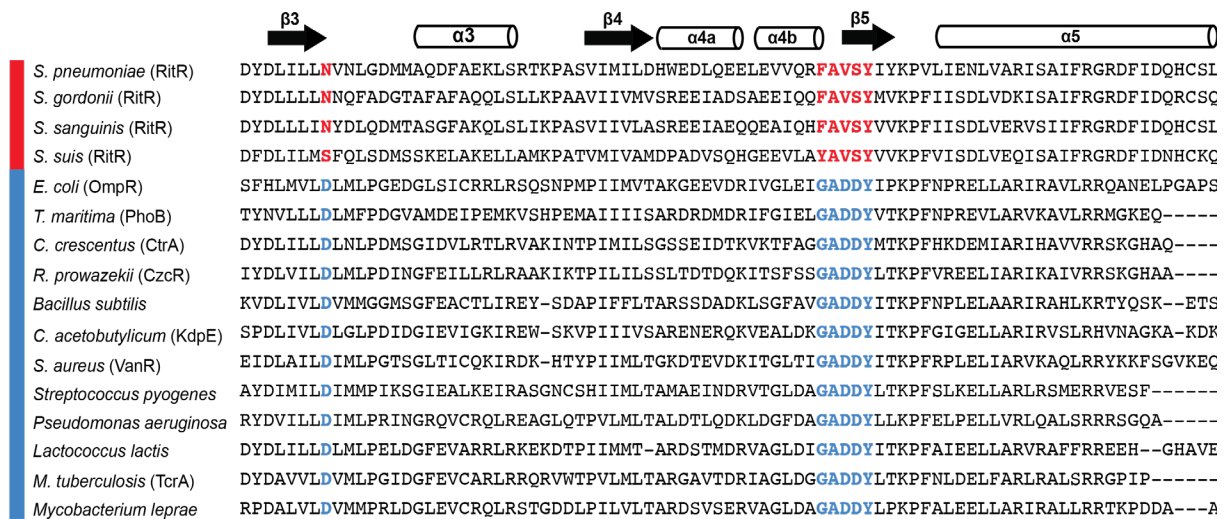
Figure 2-22: Sequence alignment of RitR homologs with other DNA binding proteins in the OmpR family. The major differences in the Streptococcal RitR homologs and the typical REC domain proteins are labeled in red. Asn/Ser substitutes for the phosphorylatable Asp residue, and the conserved GADDY motif is replaced with (F/Y)AVSY in RitR proteins.

To explore the hypothesis, we constructed a triple mutant of RitR where the equivalent "FAVSY" sequence is changed to GADDY (RitR_GADDY). The oligomeric state of RitR_GADDY was investigated by SEC and the results are shown in Figure 2-23. The RitR_GADDY variant has significantly more dimer than wild-type RitR protein in the absence of added oxidant. The retention time of the RitR_GADDY dimer is the same as that of RDLV, but earlier than the RitR_ox cross dimer. The retention time of the RitR_GADDY monomer is also earlier than that of the wild-type RitR monomer. These observations suggest that both the dimeric and monomeric forms of RitR_GADDY are more "extended" than the wild-type, likely due to dissociation of the ALR and DBD domains. The dimer/monomer ratio in RitR_GADDY is still dependent on the redox state of C128. After treatment with 1 mM H_2O_2, RitR_GADDY forms pure extended dimer, while it forms pure extended monomer after treatment with DTT (Figure 2-23).
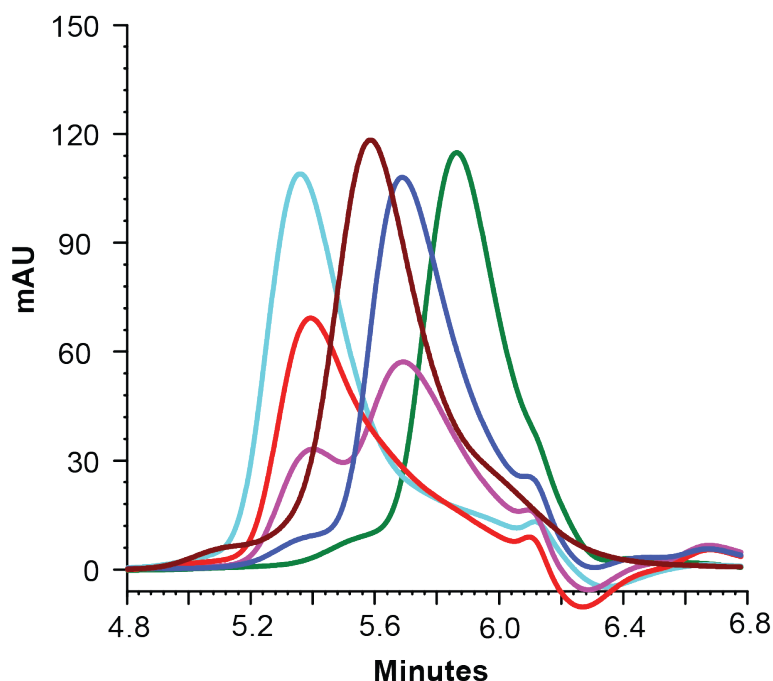
Figure 2-23: Size exclusion chromatography results of the RitR_GADDY mutant with or without H_2O_2 or DTT treatments. The mutant RitR_GADDY protein (magenta) exists as a mixture of extended dimer and extended

99

monomer. After treatment with H_2O_2 (red), it forms almost pure dimer. And after DTT treatment (dark blue), it forms almost pure monomer. The RitR$_{ox}$ "cross dimer" is dark red, the RitR(L86A/V93A) ("extended dimer") is cyan, wild-type RitR is green.

2.3.16 The role of C128 in the RitR(L86A/V93A) variant

C128 in RitR is highly conserved among a non-Group A/B subset of streptococcal species. To check if C128 has a role in the DNA-binding activity of RitR, we mutated C128 to Ser in the RitR(L86A/V93A) variant. The RitR(L86A/V93A) variant was used for these experiments, since it is the only form of RitR that we are reliably able to bind to DNA. The affinity of this RitR(L86A/V93A/C128S) triple mutant (LVC) for the BS3 target sequence was qualitatively assessed by SEC and differential scanning fluorimetry. As is evident in figure 2-24, the LVC mutant binds to the BS3 forming protein-DNA complex. There is a peak corresponding to the complex with a retention time of 5.1 min, and stability of the protein significantly increased in the presence of DNA. As expected from the results above, it seems that oxidation of C128 is only the beginning of the signal, and some other factor is required to trigger the structural rearrangement of the α4-β5-α5 face of the REC domain that leads to dimerization, and it would seem, to DNA binding.
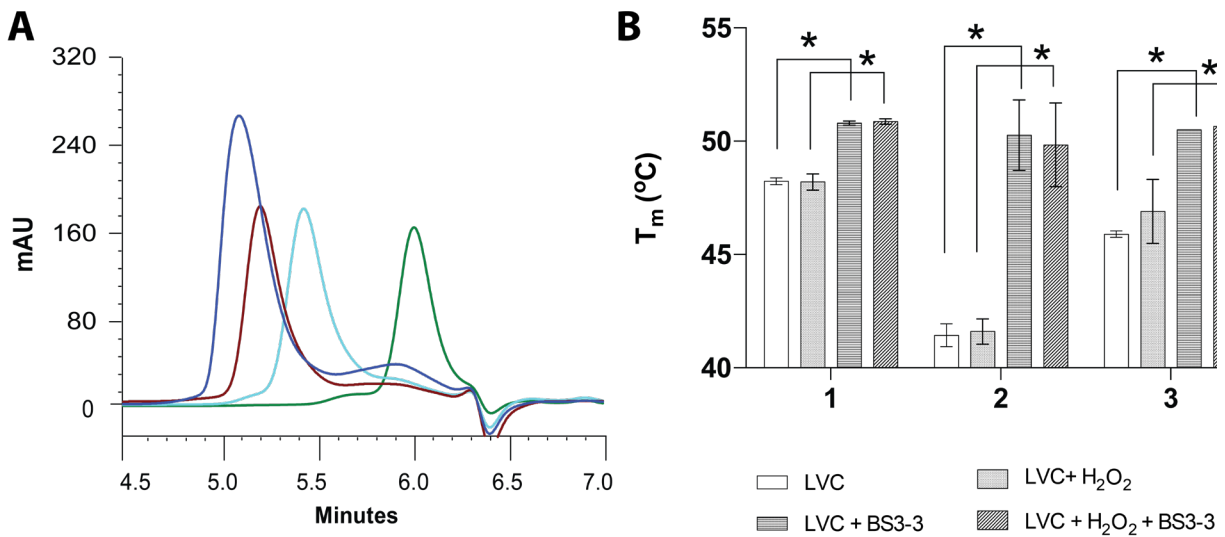
Figure 2-24: The LVC triple mutant still binds to the BS3-3 target dsDNA. Size exclusion chromatography results of LVC in the presence of BS3-3 (A) show that a peak eluted at 5.1 min that is not present in traces of either the protein or DNA alone and likely corresponds to the complex between the two. The presence of BS3-3 also significantly increased the stability of LVC as judged by differential scanning fluorimetry (B). **1**: pH 6.5, 25 mM TRIS, 150 mM NaCl, 100 mM DL-malic acid; **2**: 20 mM HEPES pH 7.2, 5 mM MgCl$_2$, 1 mM CaCl$_2$, 0.1 mM EDTA; **3**: 20 mM HEPES pH 7.2, 5 mM MgCl$_2$, 1 mM CaCl$_2$, 0.1 mM EDTA, 150 mM NaCl.

Our collaborators in the Ulijasz laboratory at Loyola University Chicago constructed three C128 mutants in the natural chromosomal position in *Streptococcus pneumoniae* strain R800 and D39: C128A (sensory dead), C128S (reduced mimetic) and C128D (oxidized mimetic). They found that the C128A mutant strain exhibited significantly impaired growth under aerobic conditions, whereas the C128S exhibited an intermediate growth defect. Conversely, the C128D mutant grew faster than the wild type. The results of the streptonigrin killing assay, which is an indirect indicator of cellular iron concentrations,

suggest that these effects are due to RitR-mediated repression of iron transport (to be published).
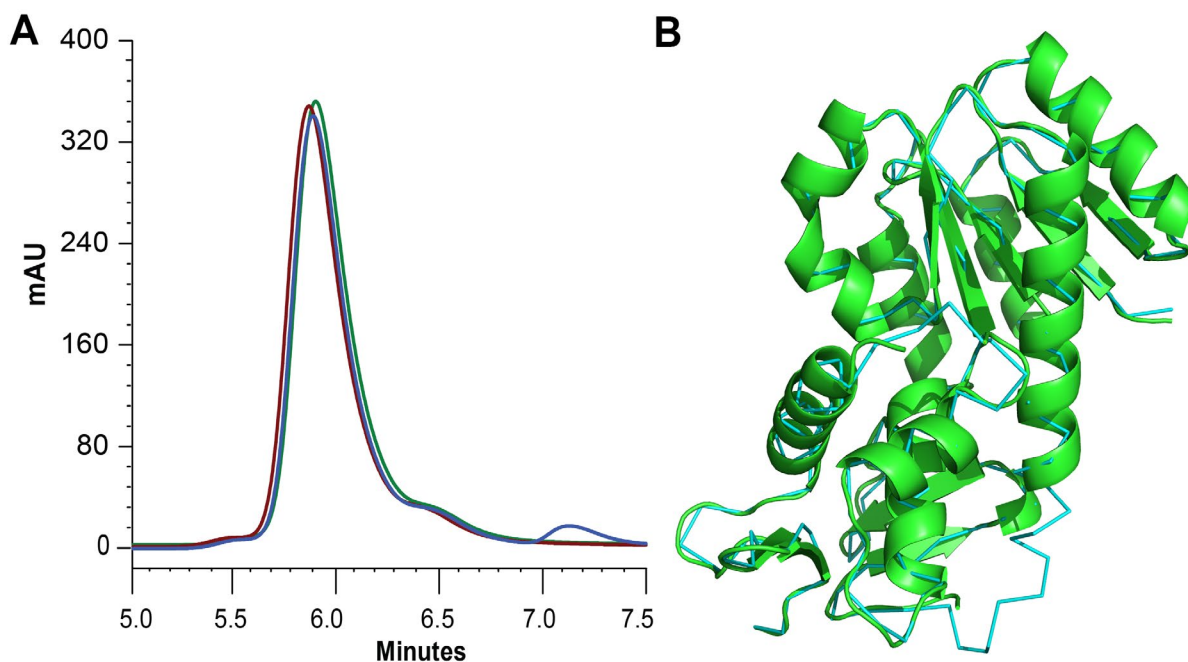
Figure 2-25: (A) Size exclusion chromatography results of RitR(C128D) with 1 mM H_2O_2 or 1mM DTT. The traces of RitR(C128D) only colored in blue, H_2O_2 treatment colored in red and DTT treatment colored in green. (B) Overlay of RitR(C128D) and RitR(C128S) structures. RitR(C128D) is colored green, while the RitR(C128S) structure is colored cyan. The rmsd for the matched $C\alpha$ atoms in the two structures is 0.30 Å.

To investigate the structure and function relationship of the RitR(C128D) mutant, the oligomeric state and the protein structure were determined. The results are shown in Figure 2-25. The RitR(C128D) mutant exists as monomer, even after treatment with 1 mM H_2O_2 for several days. The structure of RitR(C128D) is very similar to that of the RitR(C128S) structure. Superposition of the RitR(C128S) and RitR(C128D) structures using secondary structure matching (SSM)[32] gives a root-mean-square deviation (rmsd) of 0.3 Å

for 220 of the 229 Cα atoms in the RitR(C128S) chain A. This indicates that the two structures are essentially identical. The long α5 linker helix still connects the REC and DBD domains of the protein and runs almost the entire length of the protein. However, the C-terminal end of this helix is disordered in the RitR(C128D) variant. This suggests that the Cys > Asp substitution mimics the oxidized form of RitR at least enough to trigger the disordering of the C-terminal end of α5 that was observed in the $RitR_{ox}$ structure, even though it was apparently not enough to trigger dissociation of the domains.

2.4 Conclusions

To propagate a signaling response, the typical TCST system employs a sensory histidine kinase that phosphorylates a receiver domain on a conserved aspartate residue. However, RitR is an orphan TCST response regulator in *Streptococcus pneumonia* R6 that lacks a cognate histidine kinase. In addition, there are some other characteristics suggesting that RitR has a different activation mechanism: the phosphorylatable aspartate is replaced by asparagine in RitR; D10 in the acidic triad is substituted by basic lysine; one switch residue T83 is substituted with D81; and there is a conserved cysteine residue near the DBD, C128, that is not present in other response regulators. This cysteine residue appears to be redox active and to influence the quaternary structure of RitR.

Analogy to other DNA-binding response regulators suggests that the active form of RitR binds to DNA as a dimer. As purified, RitR exists as a monomer in equilibrium with a very small amount of homodimer. To get homogenous protein for crystallization, C128 was

mutated to serine; the resulting C128S variant is purely monomeric. The structure of RitR(C128S) showed that the ALR domain and the DNA-binding domain are linked by an unusually long α-helix that runs the length of the entire protein, with C128 near the C-terminal end. Bioinformatic analysis of all streptococcal RitR homologs showed that C128 is strictly conserved, suggesting that RitR may be a novel redox sensor. The environment of the cysteine thiol group, as inferred from the position of S128 Oγ is quite polar (Figure 2-11F). Thus there is no clear indication why the C128 thiol group has a pK_a of 7.2 (Figure 2-2), which is significantly lower than the pK_a of free cysteine in aqueous solution. It is possible that other hydrogen bonding interactions, *e.g.* the water-mediated hydrogen bond to Y163, as well as the concentration of basic amino acids toward the C-terminus of α5 (H127, H147, and R160, as well as R135, which is located on the loop after α5) might mitigate the unfavorable interaction between a cysteine thiolate and the helix dipole. However, what is evident is that the C-terminus of helix α5, despite an extensive hydrogen bonding network around C128 that includes H147, D115, R160 and D164, is one of the more dynamic regions of the structure. The last few turns of helix α5 become disordered in the H_2O_2-oxidized protein, as well as the oxidized (sulfenate/sulfinate)-mimetic mutant RitR(C128D) (Figure 2-4 and Figure 2-25). Cysteine sulfenic acid (Cys-SOH) is a not-uncommon feature in redox-active proteins. For example, sulfenic acid was found to be an intermediate for disulfide bond formation *in vitro*[33]. The dramatically faster growth phenotype of the C128D mutant strain of *S. pneumoniae* verifies that the disordering of the C-terminal end of α5 is one step of RitR activation.

Iron is an important micronutrient in bacteria, which influences cell growth, virulence and metabolism. To stop the repressor signal, the *gnd* gene that co-transcribed with RitR,

may be involved in the shutdown of oxidization-induced RitR activation. The *gnd* encodes a 6-phosphogluconate dehydrogenase that catalyzes NADPH formation. NADPH is, in turn, the cosubstrate of glutathione reductase, which may be involved in repressing RitR activation by reducing the sulfenic acid or disulfide bond into the reduced thiol group (Figure 2-26).

$$Fe^{3+} + H_2O_2 \longrightarrow HOO^{\cdot}$$

$$Fe^{2+} + H_2O_2 \qquad\qquad O_2^{\cdot -} + H^+$$

$$\text{RitR-S}^- \xrightarrow[\text{Or } H_2O_2]{HO^{\cdot}} \text{RitR-SOH} \xrightarrow{H_2O_2} \text{RitR-S}$$

Glutathione reductase

GSSG   GSH

NADPH

gnd ⟶ 6-phosphogluconate dehydrogenase
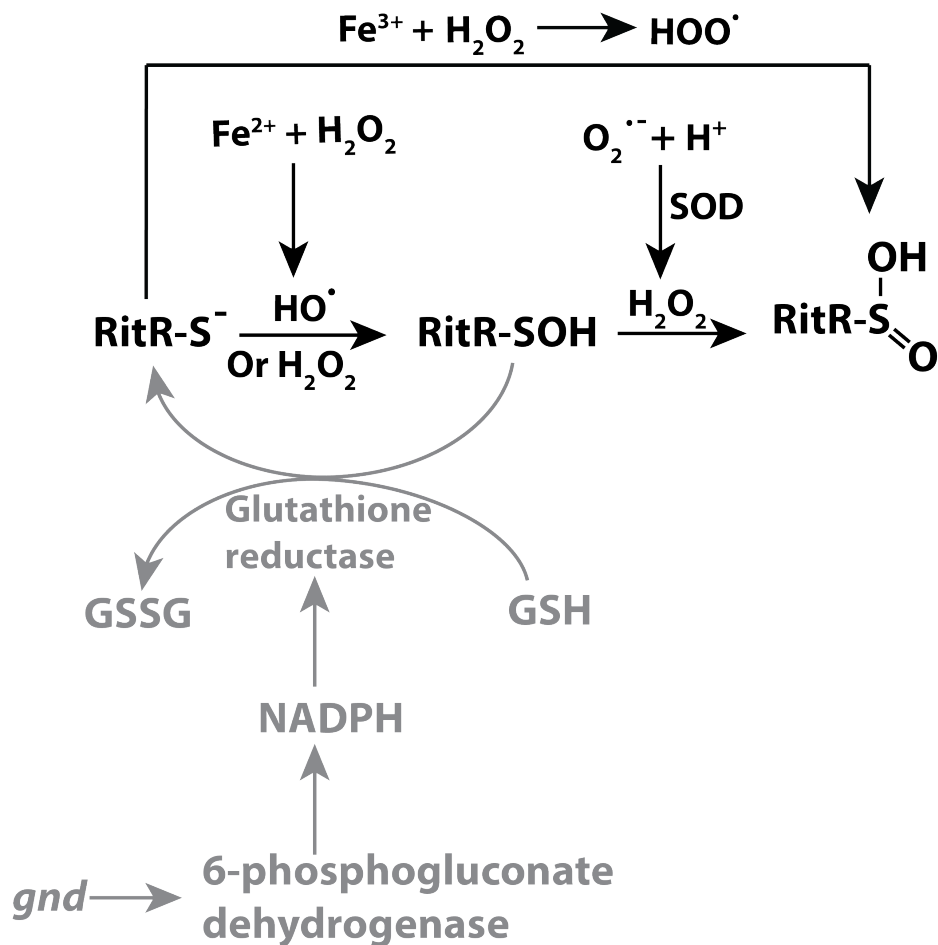
Figure 2-26: Proposed thiol group oxidation and reduction metabolism of RitR.

The structure of an oxidized, dimeric form of RitR was determined by X-ray crystallography. Oxidation of C128 to the disulfide caused a conformational change that

caused the DBD to release from the ALR domain. Surprisingly, the freed DBD was observed bound to the ALR domain of the other, disulfide-linked molecule of RitR, recapitulating almost exactly the structure of the inactive, monomeric protein.

A second dimeric conformation was found in the RitR(L86A/V93A) variant. L86 and V93 were both mutated to Ala. These residues are found in the kinked α4 helix that interacts with the DBD and form part of the hydrophobic core of this region of the protein. RitR(L86A/V93A) elutes significantly earlier than the "cross-dimer" conformation of the oxidized, inactive wild-type RitR, suggesting that it has a different, likely more extended, overall shape. This "extended-dimer" binds to the target DNA according to gel filtration and differential scanning fluorimetry.

Attempts to crystallize the full-length RitR(L86A/V93A) protein alone, or in complex with oligonucleotides of various lengths were unsuccessful. Only blunt-ended dsDNA was tested in these experiments; it is possible that sticky-end dsDNA might promote crystallization of the protein:DNA complex by forming long, repeating assemblies of oligonucleotides[34]. Because the protein component is usually globular, packing of the bound protein within the lattice formed by neighboring DNA oligonucleotide assemblies can be important for growth of a three-dimensional crystal.

Since the full-length protein did not crystallize readily, we targeted the isolated RitR(L86A/V93A) REC domain for crystallization. The crystal structures of two constructs were determined: a 1-124 version that matches the original wild-type REC domain construct (RDLV) and a longer 1-132 version that includes the putative redox-active C128 (RD2LV). Both RDLV and RD2LV exhibited the same unprecedented conformational change

for a response regulator protein, where helix α4 is disordered and the two protomers swap their "α5" helices to form the dimer. The α5 helices point in opposite directions, which would place the DNA-binding domains far apart and give the extended shape suggested by gel filtration.

The fact that strains harboring a reduced mimetic RitR mutants (e.g. C128S) have a growth defect phenotype, it seems that oxidation of C128 is part of the activation mechanism. The $RitR_{ox}$ structure demonstrates that the dissociation of the REC domain from the DBD upon oxidation, but the dissociated DBD re-binds to the REC domain of the other protomer. So there must be an additional factor that leads to dimerization of the ALR domains, the active form of which likely has a similar structure to the RitR(L86A/V93A) REC domain.

Based on these studies, we propose a model for RitR activation, as shown in Figure 2-27. Under reducing conditions, RitR mainly exists as inactive monomer. Upon oxidation, as might happen during the oxidative burst phase of the *pneumococcus* growth where millimolar amounts of H_2O_2 are produced [35], RitR favors the dissociated and extended conformation. In the absence of target DNA, RitR would eventually form the inactive cross-dimer. In the presence of DNA, it may bind to DNA at low affinity. Under the influence of some still-unknown factor, the ALR domain of RitR would undergo a conformational rearrangement of the α4-β5-α5 region to promote dimerization and place the two DBDs far from the ALR domains. In that way, RitR can keep the extended, active conformation to bind to DNA with high affinity. Alternatively, the ALR domains might be triggered to

dimerize before binding to DNA, and only bind to target DNA when the active conformation has been obtained.
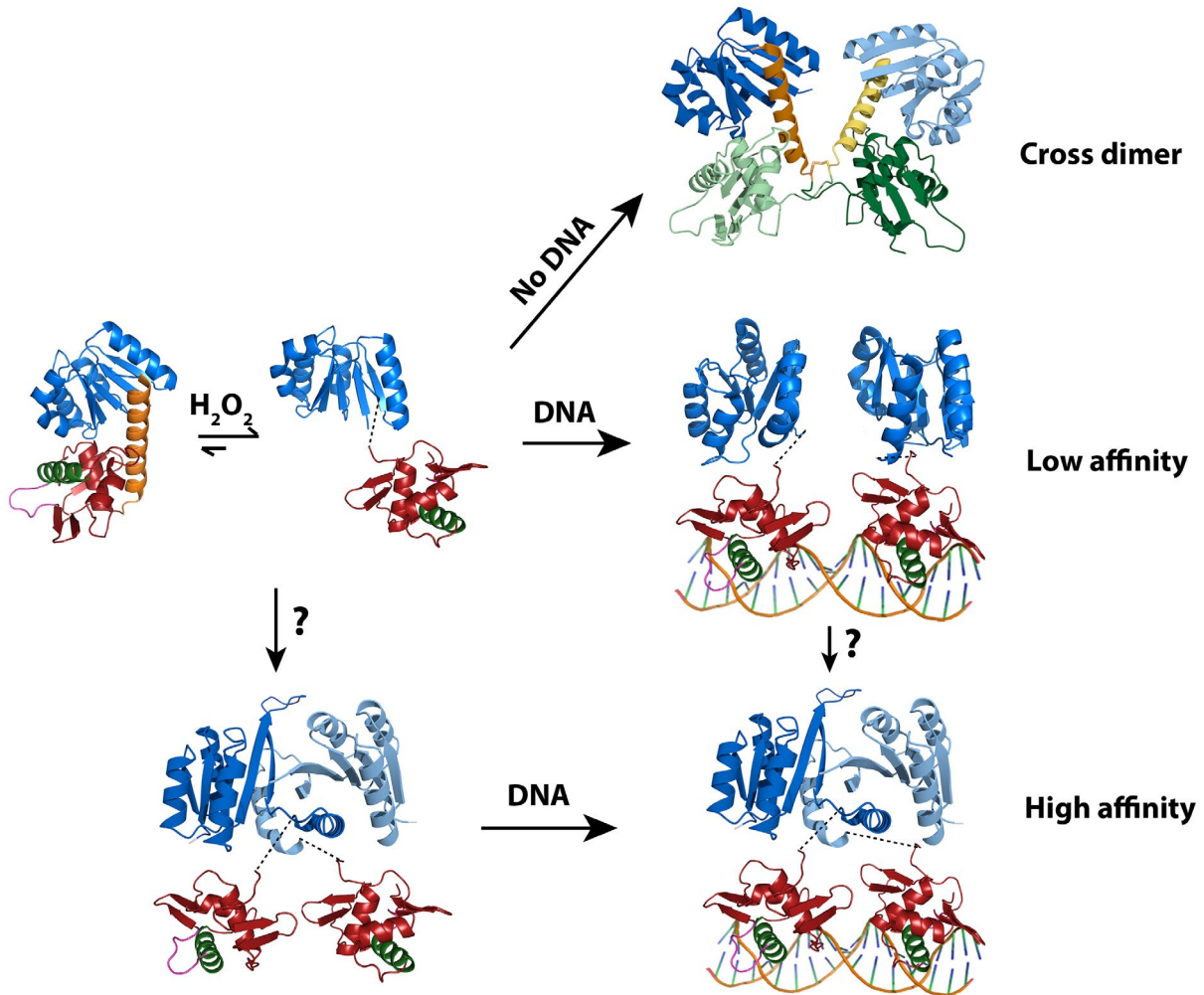
Figure 2-27: The proposed RitR activation mechanism.

Comparative transcriptome analysis demonstrated that, in addition to repressing expression of the *piu* iron transport genes, RitR also activates expression of bacterial ferritin, several genes involved in sugar uptake and other genes involved in DNA repair [1]. To check whether up-regulated sugars might contribute to RitR activation (dimerization),

co-incubation of RitR with several carbohydrates (sucrose, sorbitol, mannitol, glucose,

isoleucine) was conducted at room temperature for several hours. Size exclusion

chromatography results showed that no oligomer state change was observed. So future

studies on finding the trigger for RitR ALR domain dimerization are still needed to explore.

2.5 References

[1] Ulijasz, A. T., Andes, D. R., Glasner, J. D., and Weisblum, B. (2004) Regulation of iron transport in Streptococcus pneumoniae by RitR, an orphan response regulator, *Journal of bacteriology 186*, 8123-8136.

[2] Ong, C. L., Potter Aj Fau - Trappetti, C., Trappetti C Fau - Walker, M. J., Walker Mj Fau - Jennings, M. P., Jennings Mp Fau - Paton, J. C., Paton Jc Fau - McEwan, A. G., and McEwan, A. G. Interplay between manganese and iron in pneumococcal pathogenesis: role of the orphan response regulator RitR.

[3] Maule, A. F., Wright, D. P., Weiner, J. J., Han, L., Peterson, F. C., Volkman, B. F., Silvaggi, N. R., and Ulijasz, A. T. (2015) The Aspartate-Less Receiver (ALR) Domains: Distribution, Structure and Function, *PLOS Pathogens 11*, e1004795.

[4] Hickey, J. M., Lovell S Fau - Battaile, K. P., Battaile Kp Fau - Hu, L., Hu L Fau - Middaugh, C. R., Middaugh Cr Fau - Hefty, P. S., and Hefty, P. S. The atypical response regulator protein ChxR has structural characteristics and dimer interface interactions that are unique within the OmpR/PhoB subfamily.

[5] Bachhawat, P., and Stock, A. M. Crystal structures of the receiver domain of the response regulator PhoP from Escherichia coli in the absence and presence of the phosphoryl analog beryllofluoride.

[6] Palm, G. J., Khanh Chi, B., Waack, P., Gronau, K., Becher, D., Albrecht, D., Hinrichs, W., Read, R. J., and Antelmann, H. (2012) Structural insights into the redox-switch mechanism of the MarR/DUF24-type regulator HypR, *Nucleic acids research 40*, 4178-4192.

[7] Winther, J. R., and Thorpe, C. (2014) Quantification of Thiols and Disulfides, *Biochimica et biophysica acta 1840*, 10.1016/j.bbagen.2013.1003.1031.

[8] Otwinowski Z Fau - Minor, W., and Minor, W. Processing of X-ray diffraction data collected in oscillation mode.

[9] Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP, *Methods in molecular biology (Clifton, N.J.) 364*, 215-230.

[10] Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains, *Acta crystallographica. Section D, Biological crystallography 62*, 1002-1011.

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

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

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

[14] Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation11Edited by J. Thornton, *Journal of molecular biology 285*, 1735-1747.

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

[16] Urzhumtseva, L., Afonine Pv Fau - Adams, P. D., Adams Pd Fau - Urzhumtsev, A., and Urzhumtsev, A. Crystallographic model quality at a glance.

[17] Nowakowski, A. B., Wobig Wj Fau - Petering, D. H., and Petering, D. H. Native SDS-PAGE: high resolution electrophoretic separation of proteins with retention of native properties including bound metal ions.

[18] Rimphanitchayakit, V., Hatfull, G. F., and Grindley, N. D. (1989) The 43 residue DNA binding domain of gamma delta resolvase binds adjacent major and minor grooves of DNA, *Nucleic acids research 17*, 1035-1050.

[19] Rhee, S., Martin, R. G., Rosner, J. L., and Davies, D. R. (1998) A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator, *Proc Natl Acad Sci U S A 95*, 10413-10418.

[20] Alekshun, M. N., Levy, S. B., Mealy, T. R., Seaton, B. A., and Head, J. F. (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution, *Nat Struct Mol Biol 8*, 710-714.

[21] Chaffey, N. (2003) Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. Molecular biology of the cell. 4th edn, *Annals of Botany 91*, 401-401.

[22] Paulsen, C. E., and Carroll, K. S. (2013) Cysteine-Mediated Redox Signaling: Chemistry, Biology, and Tools for Discovery, *Chemical Reviews 113*, 4633-4679.

[23] Roos, G., Foloppe N Fau - Messens, J., and Messens, J. Understanding the pK(a) of redox cysteines: the key role of hydrogen bonding.

[24] Hiniker, A., Collet, J. F., and Bardwell, J. C. (2005) Copper stress causes an in vivo requirement for the Escherichia coli disulfide isomerase DsbC, *The Journal of biological chemistry 280*, 33785-33791.

[25] Matsui Lee, I. S., Suzuki, M., Hayashi, N., Hu, J., Van Eldik, L. J., Titani, K., and Nishikimi, M. (2000) Copper-Dependent Formation of Disulfide-Linked Dimer of S100B Protein, *Archives of Biochemistry and Biophysics 374*, 137-141.

[26] Pecci, L., Montefoschi, G., Musci, G., and Cavallini, D. (1997) Novel findings on the copper catalysed oxidation of cysteine, *Amino Acids 13*, 355-367.

[27] Bachhawat, P., Swapna, G. V., Montelione, G. T., and Stock, A. M. (2005) Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states, *Structure 13*, 1353-1363.

[28] Pryor, E. E., Jr, Wozniak, D. J., and Hollis, T. (2012) Crystallization of Pseudomonas aeruginosa AmrZ protein: development of a comprehensive method for obtaining and optimization of protein-DNA crystals, *Acta Crystallographica Section F 68*, 985-993.

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

[30] Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., and Bryant, S. H. (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, *Nucleic acids research 45*, D200-d203.

[31] Toro-Roman, A., Mack Tr Fau - Stock, A. M., and Stock, A. M. Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face.

[32] Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta crystallographica. Section D, Biological crystallography 60*, 2256-2268.

[33] Rehder, D. S., and Borges, C. R. (2010) Cysteine sulfenic acid as an intermediate in disulfide bond formation and non-enzymatic protein folding, *Biochemistry 49*, 7748-7755.

[34] Pio, F., Ni, C.-Z., Mitchell, R. S., Knight, J., McKercher, S., Klemsz, M., Lombardo, A., Maki, R. A., and Ely, K. R. (1995) Co-crystallization of an ETS Domain (PU.1) in Complex with DNA: ENGINEERING THE LENGTH OF BOTH PROTEIN AND OLIGONUCLEOTIDE, *Journal of Biological Chemistry 270*, 24258-24263.

[35] Pericone, C. D., Bae, D., Shchepetov, M., McCool, T., and Weiser, J. N. (2002) Short-sequence tandem and nontandem DNA repeats and endogenous hydrogen peroxide production contribute to genetic instability of Streptococcus pneumoniae, *Journal of bacteriology 184*, 4392-4399.

Chapter 3

Uncharacterized Protein Q15Z91 from *Pseudoalteromonas atlantica* is an ALR-Regulated Diguanylate Cyclase

3.1 Introduction

Aspartate-less receiver (ALR) proteins are a subclass of response regulator proteins in bacterial signal transduction systems. In typical two-component signaling systems, the receiver domain of the regulator would be phosphorylated in the conserved aspartate residue by the cognate histidine kinase. In TCSTs with ALR-containing response regulators, there is no conserved Asp residue for phosphorylation, which means that ALRs must necessarily respond to different regulatory stimuli. Chapter 2 described my work aimed at elucidating the activation mechanism of an ALR protein, RitR. I showed that it functions as an "extended dimer" there the DNA-binding domains are released from the ALR domain, partly in response to oxidation of C128. Owing to its activity as a DNA-binding protein, it was difficult to make quantitative studies of RitR activation. The assays used to detect DNA binding often gave contradictory results and could be highly dependent on the experimental system (*e.g.* choice of buffer, salt concentration, and temperature). For this reason, we chose to study an ALR-containing response regulator whose effector domain has an enzymatic activity, which would allow a much more nuanced and quantitative assessment of activation. We identified just such a protein through BLAST [2, 3] searches

using residues 1-132 of RitR, which comprise the ALR domain and the extended α5 helix that contains the redox-active cysteine (C128). By manually filtering the results of the BLAST search for sequences that contained the same "HCS" motif found in RitR, we found an uncharacterized protein from *Pseudoalteromonas atlantica* TC6, which is known only by its UNIPROT accession code, Q15Z91, which catalyzes the formation of the bacterial second messenger molecule bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) from two molecules of GTP. In this chapter I describe my structural and enzyme kinetics studies of Q15Z91.

3.2 Methods

3.2.1 Expression and purification of Q15Z91

The coding sequence of Q15Z91 (Genbank locus ID Patl_0265) was optimized for expression in *E. coli* and synthesized by GenScript Inc (Piscataway, NJ). The synthetic gene was sub-cloned into the pE-SUMO kan expression vector (LifeSensors Inc, Malvern, PA) using primers containing BsaI and XbaI restriction sites (Q15Z91 forward: 5'-GGTCTCAAGGTATGCAAAATCCGGAAAGTGC-3', Q15Z91 reverse: GCTCTAGATCATTATGACACCAGACGGTTACGA-3'). The His_6-tagged SUMO-Q15Z91 fusion protein was expressed from Origami2 *E. coli* cells (Invitrogen Inc, Carlsbad, CA) carrying the pE-SUMO-Q15Z91 plasmid. Cultures were grown in Luria-Bertani medium with 50 µg/mL kanamycin at 37 °C. When the cultures reached an OD_{600nm} of ~0.7, protein expression was induced with 0.4 mM IPTG. The temperature was reduced to 16 °C and the

113

cultures were grown overnight with shaking at 250 rpm. Cells were harvested by centrifugation, resuspended in 5 mL of buffer A (25 mM TRIS pH 8.0, 300 mM NaCl, 10 mM imidazole) supplemented with 1 mg/mL hen egg lysozyme (Hampton Research) and 0.1 mg/mL DNAse I (Worthington Biochemical Corp., Lakewood, NJ) per gram wet weight of cells. Cells were lysed using a Branson Sonifier S-450 cell disruptor (Branson Ultrasonics Corp., Danbury, CT) for a total of 10 min of sonication at 60 % amplitude with 30 s pulses separated by 50 s rest periods. The temperature was maintained at or below 4 °C by suspending the steel beaker in an ice bath directly over a spinning stir bar. The lysate was clarified by centrifugation at 39,000 x g for 45 minutes and then applied to a 5 mL HisTrap column (GE Lifesciences, Piscataway, NJ) at a flow rate of 5 mL/min to isolate the His_6-SUMO-Q15Z91 fusion protein. The protein was eluted by a 4-step gradient of buffer B (25 mM TRIS pH 8.0, 300 mM NaCl, and 250 mM imidazole; 5, 15, 50, and 100%). The His_6-SUMO-Q15Z91 fusion protein eluted in the third step and was ~90 % pure, as judged on coomassie-stained SDS-PAGE gels. Fractions corresponding to A_{280nm} peaks on the chromatogram were pooled ~3 μM SUMO protease (LifeSensors Inc) were added. The sample was loaded into SnakeSkin dialysis tubing and dialyzed overnight against 4 L of SUMO protease digestion buffer (25 mM TRIS pH 8.0, 150 mM NaCl). The dialysate was passed through the HisTrap column a second time to remove the cleaved His_6-SUMO tag as well as the protease. The resulting Q15Z91 preparation was > 95% pure. The protein was desalted using a 2x5 mL HiTrap Desalting column (GE Lifesciences) equilibrated with 25 mM TRIS pH 7.5, 150 mM NaCl and stored at -80 °C. The mutants described below were purified using the same protocol as the wild type protein.

Q15Z91 co-purified with c-di-GMP. To get c-di-GMP-free Q15Z91 protein, SadR, a phosphodiesterase from *Pseudomonas aeruginosa*, was used to linearize the c-di-GMP. The SadR expression construct was a kind gift of professor Holger Sondermann in the Department of Molecular Medicine, College of Veterinary Medicine, Cornell University. The His$_6$-tagged SadR protein was purified as described by De, et al[4], omitting the precision protease cleavage step. SadR treatment results in the degradation of c-di-GMP to free pGpG, which is easily removed by gel filtration. The His$_6$-tagged SadR protein was added to the purified Q15Z91 and dialyzed in 4 L of 50 mM TRIS pH 8.0, 150 mM NaCl, 2 mM MgCl$_2$ buffer at 4 °C overnight. The dialysate was passed through the HisTrap column a third time to separate c-di-GMP free Q15Z91 (Q15Z91$_{free}$) from the His$_6$-tagged SadR. The resulting Q15Z91 protein was desalted using a HiTrap Desalting column equilibrated with 25 mM TRIS pH 7.5, 150 mM NaCl and stored at -80 °C. In the desalt step, the free pGpG would be separated from protein. The c-di-GMP free mutants described below were purified using the same protocol as Q15Z91$_{free}$.

3.2.2 Mutagenesis of Q15Z91

The full-length mutants were obtained using the Q5 site-directed mutagenesis kit (New England Biolabs, Ipswich, MA) as described for RitR in Chapter 2. The sequences of the mutagenic PCR primers are listed in Table 3-1.

Table 3-1: Primers used for mutagenesis

Name	Primer sequence
sumo-Q15Z91_C142S_Fwd	CCTGATTTCTTCCACCGATGGCC
sumo-Q15Z91_C142S_Rev	GCCAGGGTGCGCAGTTTC
sumo-Q15Z91_R212N/D215A_Fwd	TCTGCTGCGATGGGCCGCTACGGC
sumo-Q15Z91_R212N/D215A _Rev	TTCATTCTGCAGGCTAGCTTCAATCGTGGTTG
sumo-Q15Z91_I99A _Fwd	GAAATCACTGGCTCTGTCGGGCC
sumo-Q15Z91_I99A _Rev	GGGGTGTCTTTGCCATAC
sumo-Q15Z91_G102A _Fwd	GATTCTGTCGGCCCTGGAAGCGG
sumo-Q15Z91_G102A _Rev	AGTGATTTCGGGGTGTCTTTG

3.2.3 Crystallization, structure determination, and model refinement

Crystals of Q15Z91 alone were grown by the vapor diffusion method with sitting-drop geometry. Crystals formed in many different conditions. The highest quality crystals formed in drops comprised of equal parts protein solution (8 mg/ml Q15Z91 in 10 mM TRIS/100mM citric acid pH 4.5, 150 mM NaCl, and crystallization solution (20% (w/v) PEG 3350, and 0.2 M sodium citrate tribasic dehydrate). Cube-shaped crystals appeared after several months at 4°C and were prepared for flash-cooling by soaks in solutions containing 2 μl of well solution and 0.7 μl of glycerol (~26%). Crystals of Q15Z91 without GTP bound were grown by mixing 1 μl of 8 mg/mL Q15Z91 that pre-incubated with 5mM $MgCl_2$, 200 μM $ZnCl_2$, 3 mM GTP with crystallization solution (8% v/v Tacsimate pH 7.0, 20% w/v Polyethylene glycol (PEG) 3,350). Cube-shaped crystals appeared after several months at 4 °C. Crystals were prepared for flash-cooling in the same way as Q15Z91 alone crystals.

X-ray diffraction data were collected from crystals of Q15Z91 at beamline 21-ID-D of the Life Science Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS). Data were collected from crystals of the protein with GTP at LS-CAT beamline 21-ID-G. Data were processed with HKL2000 [5]. The structure of Q15Z91 was determined by molecular replacement using 3BRE REC domain and effector domain as two ensembles in Phaser[5]. The initial model comprising 88% of the asymmetric unit contents was built by Autobuild[6]. Chain A of this model was used as the search model for molecular replacement in PHASER[7] to phase the Q15Z91 with GTP bound structure. Both models were subjected to iterative cycles of manual model building in COOT[8] and maximum likelihood-based refinement using the PHENIX package (phenix.refine[9]). Ordered solvent molecules were added automatically in phenix.refine and culled manually in COOT. Hydrogen atoms were added to the model using phenix.reduce[10] and were included in the later stages of refinement to improve the stereochemistry of the model. Positions of H atoms were refined using the riding model with a global B-factor. Regions of the model for translation-libration-screw (TLS) refinement were identified using phenix.find_tls_groups (P.V. Afonine, unpublished work) and the TLS parameters were refined in phenix.refine. Once the refinement converged, the model was validated using the tools implemented in COOT and PHENIX [9, 11]. Sections of the backbone with missing or uninterpretable electron density were not included in the final model. Data collection and model refinement statistics are listed in Table 3-2. Coordinates and structure factors for both Q15Z91 and Q15Z91 with GTP models will be deposited in the Protein Data Bank (www.rcsb.org).

3.2.4 Quaternary structure characterization by size exclusion chromatography

SEC experiments were carried out using an Agilent 1220 Compact HPLC equipped with a 250 x 4.6 mm BioBasic SEC-300A column equilibrated with 25 mM TRIS pH 7.5, 150 mM NaCl. The column was calibrated with the Gel Filtration Molecular Weight Marker kit from Sigma-Aldrich (Cytochrome C (12.4 kDa), Carbonic Anhydrase (29 kDa), Bovine Serum Albumin (66 kDa), and Sweet Potato Amylase (200 kDa)). Q15Z91 was pre-incubated with 1 mM H_2O_2 or 1mM DTT and 5 μl of the treated proteins were injected onto the column and separated at a flow rate of 0.5 ml/min at ambient temperature. The separation was monitored using a variable wavelength UV-visible detector set to 280 nm.

3.2.5 Size exclusion chromatography with multi-angle light scattering (SEC-MALS)

To confirm the molecular weights estimated from the calibrated SEC runs of Q15Z91 on the BioBasic SEC-300A column, the separations were repeated on the SEC-MALS system in the laboratory of Dr. Blake Hill at the Medical college of Wisconsin. The multiple-angle light scattering detector provides a direct estimate of the molecular weights of particles eluting from the column that is independent of the retention time. The same 250 x 4.6 mm BioBasic SEC-300A column was used with 24 hour equilibration in 25 mM TRIS pH 7.5, 150 mM NaCl at 0.5 ml/min. The running buffer is same as used above, except that it was filtered through a 0.022 μm filter, rather than a 0.22 μm filter. Due to the relatively low sensitivity of the MALS detector the separation was done with 100-200μg of Q15Z91, rather than the ~15 μg used for the UV/vis-monitored separations. The concentrated

proteins were centrifuged at 16,900 x g for 20 min at 4°C before loading.

3.2.6 Biofilm assay on borosilicate tubes

Assays to quantify the degree of biofilm formation in *E. coli* expressing wild-type or mutant forms of Q15Z91 were performed as a way to assess the *in vivo* activity of the proteins. Overnight cultures of Origami2 *E. coli* carrying the pSUMO-Q15Z91 (or mutant) plasmids were adjusted to the same inoculum amounts in 3 ml of fresh LB broth with 50 μg/ml kanamycin in 14 ml borosilicate glass tubes. After growth with rotating at 37 °C for ~5 hours, the cultures were equalized by adding extra fresh LB$_{kan}$ medium and then protein expression was induced by the addition of 0.4 mM IPTG. The cultures were allowed to grow overnight at 25°C. The next day, the cultures were removed from their respective test tubes, which were washed three times with tap water. After washing, crystal violet was used to stain any biofilm that had been deposited on the walls of the tubes. After staining, the culture tubes were washed with tap water five times. The culture tubes were then dried and the residual stain – that bound in the biofilm – was dissolved in 2 mL of 80% (v/v) ethanol/20% (v/v) acetone. The absorbance at 540 nm was measured on a spectrometer and taken to be a qualitative indicator of the degree of biofilm formation induced by wild-type Q15Z91 or the various mutants tested.

3.2.7 Enzymatic assay

3.2.7.1 RP-HPLC based end-point assay

As a convenient check for catalytic activity of the various forms of Q15Z91 studied here, protein was incubated for a set period of time with GTP and $MgCl_2$ and the reaction mixtures were characterized by reverse phase high performance liquid chromatography (RP-HPLC). Each 200 μl reaction contained 20 μM enzyme, 25 mM TRIS pH 7.5, 150 mM NaCl, 2mM $MgCl_2$, and 0.5 mM GTP. The GTP was prepared fresh as a 100 mM stock in mqH$_2$O. Reactions were incubated at room temperature for 30 min, and then quenched by boiling for 5 min. After centrifugation at 30,000 g for 10 min, the supernatants were analyzed on a 150 x 3 mm Aquasil C18 column (Thermo Scientific). Buffer A was 10 mM NaH_2PO_4/K_2HPO_4 pH 6.3, and buffer B was 30% (v/v) methanol, 70% (v/v) buffer A. The gradient from 5% to 100% B over 28 min, then held at 100% B for 5 min, 5% buffer B for 5 min. All separations were run at 50 °C and were monitored at 254 nm.

3.2.7.2 Biomol Green microplate assay

Reactions were prepared in a volume of 500 μl and contained 10 μM enzyme, 25 mM TRIS pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, and 0.5 mM GTP. Reactions were initiated with the addition of GTP. Aliquots were taken at set intervals and quenched by boiling for 5 min. After centrifugation at 30,000 g for 10 min, the supernatants were kept at 4 °C. The negative control was the reaction buffer containing GTP. The amount of pyrophosphate released was quantified by adding 0.005 U pyrophosphatase (Thermo Scientific) to each

reaction, followed by a 1.5 h incubation room temperature. Then 50 µl of the pyrophosphatase-treated reaction mixture were mixed with 100 µl Biomol Green reagent and incubated at room temperature for 20 min. The absorbance of the mixtures at 620 nm was recorded and converted to concentration of phosphate via a standard curve prepared with 0.03125 nM-2 nM inorganic phosphate.

3.2.7.3 EnzCheck® Pyrophosphate Assay Kit (Sigma)

This assay were conducted by following the optimized procedure published by Webb[12, 13]. Pyrophosphatase can catalyze pyrophosphate into 2 molecules of inorganic phosphate. In the presence of inorganic phosphate, purine nucleoside phosphorylase (PNP) catalyzes 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG) into ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine that absorbs maximally at 360 nm. Each 1 mL reaction contained 1 µM enzyme with 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG), purine nucleoside phosphorylase in the reaction buffer was incubated at 22 °C for 10 min. The reactions were initiated with the addition of GTP and the absorbance at 360 nm was monitored. The concentrations of GTP tested range from 50 µM to 1000 µM.

3.3 Results

3.3.1 Q15Z91 is an ALR-regulated diguanylate cyclase

Q15Z91 is 310 amino acids long (34.6 kDa) and consists of an N-terminal ALR domain

fused to a diguanylate cyclase domain. Alignment of the RitR and Q15Z91 sequences shows that the phosphorylatable aspartate residue is replaced by arginine in Q15Z91 (Figure 3-1). The conserved C128 that has been shown to play an important role in RitR activation is retained in Q15Z91, though the residues flanking it are different between the two proteins. The effector domain contains the GGEEF motif characteristic of diguanylate cyclase enzymes.

```
Q15Z91      MQNPESANVLKQTLLVVDNPGEMLKALTRQF-STSFHLLVCNDNEEVLKLALNAPIDLVL
RitR        ---------MGKRILLLEKERNLAHFLSLELQKEQYRVDLVEEGQKALSMALQTDYDLIL
               :  :  :*::::   ::  :  *:  ::  .   .::: :  ::.::.*.:**::   **:*

Q15Z91      IRVSPRDSDWLLICTRLKQHPLTIDIPLILYGKDTPKSLIL--SGLE----------AGA
RitR        LNVNLGDMMAQDFAEKL--------------SRTKPASVIMILDHWEDLQEELEVVQRFA
            :.*.  *      :.  :*                  .:  .*  *:*:   .   *            *

Q15Z91      LDFVELPTDIAMLNIKIRNHMQVSAKLRTLALIS--CTDGLTGVPNRMQLDTTYNRFWYA
RitR        VSYIYKPVFIENLVA------RISAIFRGRDFIDQHCS--LMKVPR------TY------
            :.::  *.  *   *           ::** :*    :*.         *   **.        **

Q15Z91      AIRGQHELSVLMIDIDFFKGFNDTYGHVAGDECLKKVATTIEASLQRESDAMGRYGGEEF
RitR        ------------------------------------------------------------

Q15Z91      LVLLPFTDKVGAQKIAEMILRNIEALDIKNKASTVNGKVTVSVGAATLKLHDINDHTFNH
RitR        ------------------------------------------------------------

Q15Z91      PEFLVEQADRCLYQAKKQGRNRLVS
RitR        -------------------------
```

Figure 3-1: Sequence alignment of Q15Z91 with the RitR REC domain. Arginine (colored in red) substitutes the conserved aspartate residue of the typical response regulator proteins (N53 in RitR). The conserved Cys128 of RitR is colored in orange; the GGEEF motif of Q15Z91 is colored blue.

3.3.2 Purification of Q15Z91

The pSUMO-Q15Z91 plasmid was transformed into BL21 Star(DE3) cells, T7 LysY/Iq cells (New England Biolabs, Ipswich, MA), and Origami2 cells (Novagen) Single colonies were used to inoculate large-scale expression cultures, which were grown at 37 °C to an OD_{600nm} of ~0.7, at which point protein expression was induced by the addition of 0.4 mM IPTG. Three strains were attempted to express Q15Z91, but only the Origami2 cells expressed soluble proteins and exhibited an obvious biofilm phenotype. Induction at higher or lower OD_{600nm} resulted in lower yields of soluble protein.

Q15Z91 was not stable in the purification buffers. To find the optimal storage buffer and pH, the Solubility & Stability Screen (Hampton research) and Slice pH (Hampton research) were used in differential scanning fluorimetry assays. The results showed that Q15Z91 is not stable, its average T_m being 33°C; this is not surprising given that *P. atlantica*

Table 3-2: Buffers from the Solubility & Stability Screen giving the highest T_m values in the DSF assay.

T_m - $T_{m,averaged}$ (°C)	Additive
8.4	Adipic acid
6.7	Ethylenediaminetetra acetic acid disodium
5.9	TCEP hydrochloride
5.9	Sodium phosphate monobasic, potassium phosphate dibasic
5.6	Ammonium sulfate
5.5	L-Argininamide dihydrochloride
5.1	Ethylenediamine dihydrochloride
5.1	Magnesium sulfate hydrate

is a psychrophilic bacterium. The top 10 conditions as judged by their increase in the T_m above the average, are shown in Table 3-2 and Table 3-3. Based on these data, 25mM TRIS/100mM citric acid pH 4.5, 150 mM NaCl was chosen as the storage buffer for Q15Z91.

Table 3-3: Buffers from the Slice pH Screen giving the highest T_m values in the DSF assay.

T_m - $T_{m,averaged}$ (ºC)	Additive
5.5	Sodium citrate tribasic dihydrate pH 4.5
4.9	Citric acid pH 4.4
4.7	DL-malic acid pH 4.7
4.6	Sodium citrate tribasic dihydrate pH 4.2
4.6	DL-malic acid pH 5.0
4.3	DL-malic acid pH 5.3
4.3	Succinic acid pH 5.1
4.1	Citric acid pH 4.1
4.1	Succinic acid pH 4.8
4.1	Succinic acid pH 5.4

3.3.3 The structure of Q15Z91 with c-di-GMP and citrate bound

The structure of Q15Z91 was determined by molecular replacement using 2.0 Å resolution data. The model contains residues 11-310, as well as two citrate molecules and two c-di-GMP dimers. As shown in Figure 3-2A, Q15Z91 is a dimer, with citrate bound at the active sites (A site) of the diguanylate cyclase (DGC) domains, and c-di-GMP dimers at

124

the inhibitory sites (I site). Chain A (light green) has no inter-domain interface, while chain B (light blue) has a weak interface. And there is a S-S bond between two chains. We can see clear electron density for the disulfide bond (Figure 3-2B). This Q15Z91 dimer structure is an inactive conformation, because the c-di-GMP bound in the I sites immobilizes the two DGC domains in an unproductive back-to-back orientation.
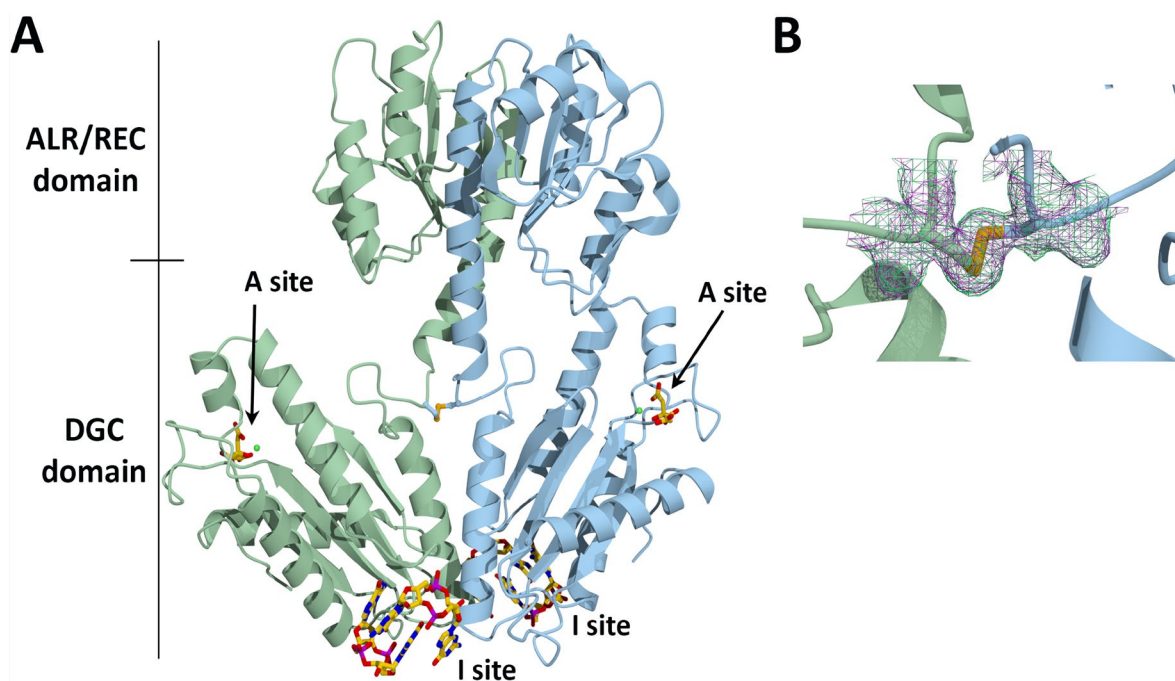
Figure 3-2: The structure of Q15Z91 dimer with c-di-GMP and citrate bound (A). The two protomers are colored light green and light blue. The ALR domain of Q15Z91 is at the top of the image, and the DGC catalytic effector domain is at the bottom. Citrate and c-di-GMP are shown as sticks with yellow carbon atoms. Mg^{2+} ions are shown as small green spheres associated with the tri-phosphate moieties of the citrate molecules. The citrate molecules from crystallization buffer are bound in the active sites (A site) and the c-di-GMP dimers are bound in the inhibitory sites (I site) of the DGC domain. The two polypeptides comprising the dimer are linked by a disulfide bond formed between C142 on both chains (B). The $2|F_o|-|F_c|$ electron density map contoured at 1.0σ is shown as a magenta mesh, and the $2|F_o|-|F_c|$ composite omit map, also contoured at 1.0σ, is shown as a green mesh.

The ALR domain of Q15Z91 has the same $(\beta/\alpha)_5$ topology and $\alpha 4$-$\beta 5$-$\alpha 5$ dimer interface (Figure 3-3) characteristic of typical OmpR/PhoB RRs. As shown in Figure 3-3B, the switch residue F110 (equivalent to Y100 of RitR) is oriented inward, which is the conformation associated with the activated forms of typical OmpR/PhoB RRs and is thought to facilitate dimer formation. However, the other Ser/Thr switch residue is replaced by G91 in Q15Z91. In the phosphorylation-dependent RRs, phosphorylation induces a conformational change of the Ser/Thr switch residue that results in the reorientation of the Tyr/Phe switch residue to facilitate dimerization. In Q15Z91, the residue that is equivalent to the Ser/Thr switch residue in typical RRs is G91, so obviously side chain reorientation cannot be an integral part of the activation mechanism in Q15Z91. This is similar to what was seen in RitR, where the Ser/Thr switch equivalent, D81, had a minimal impact on dimerization. In addition, R61 in Q15Z91 replaces the phosphorylatable aspartate residue of the typical RRs. Comparison of the Q15Z91 and RitR ALR domains shows that the side chain of R61 takes the place of the "lid residue" K10 of RitR and that the loop between $\beta 1$ and $\alpha 1$ shifts away from the "phosphorylation site" comparison with that in RitR. It seems that the ALR domain of Q15Z91 has evolved to be an obligate dimer, and that activation may be more subtle than a change from monomer to dimer as has been seen for the DNA-binding RR proteins.

Figure 3-3: ALR domains of the Q15Z91 dimer (A) and the overlay of the Q15Z91 and RitR ALR domains (PDB ID 4LZL[1]) (B). Chain A of Q15Z91 ALR domain colored in light green, chain B colored in blue. RitR ALR domain is colored cyan. The active site residues and switch residues equivalent are shown in stick. The Cα atom of G91 (Ser/Thr switch residue equivalent) in Q15Z91 is shown in sphere.
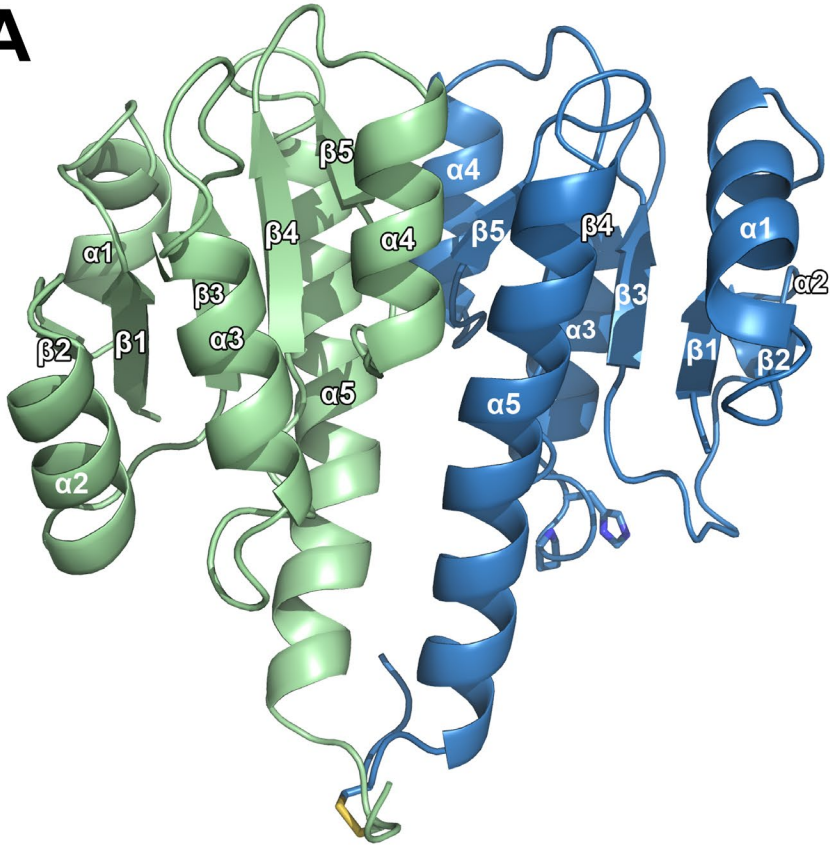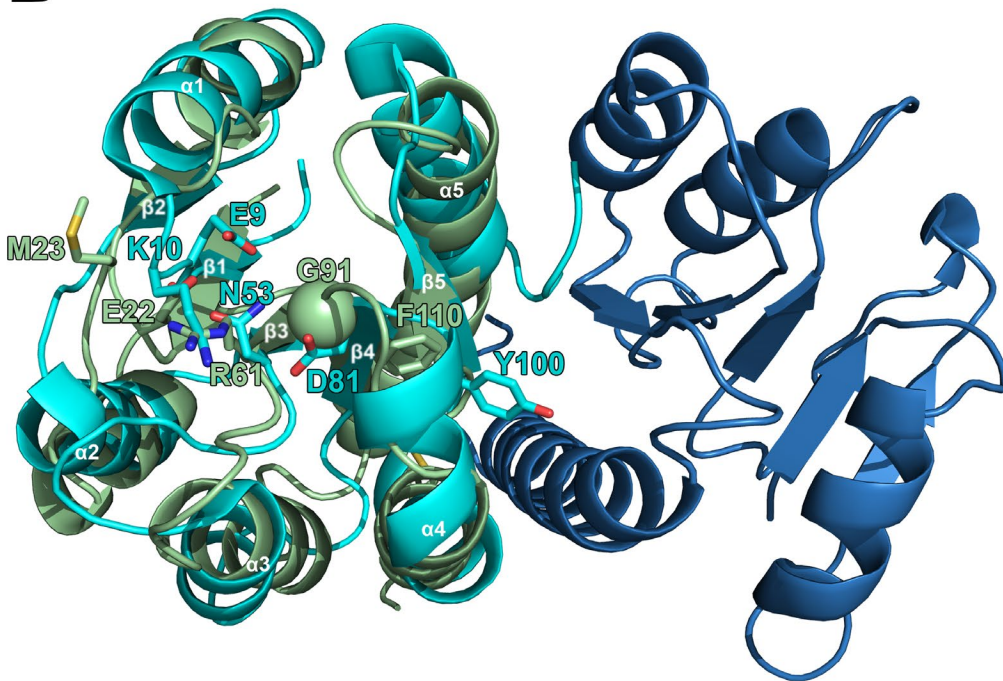
The active site of each DGC domain has one molecule of citrate and a Mg^{2+} ion bound (Figure 3-4). Due to the asymmetric arrangement of the domains—the lack of any inter-domain interface in chain A, and the weak interface between the ALR and DGC domains in chain B—there some differences in the way GTP binds to the two DGC domains. In chain A, there are no contacts between the ALR domain and the citrate. Without GTP addition in the crystallization condition, the structure of purified Q15Z91 alone has c-di-GMP bound, which must have co-purified during the protein expression in cells. In the I site, c-di-GMP forms a dimer through stacking their guanosine rings, interactions between phosphate group with N11 and N21 on the Hoogsteen-edge of the guanine base of the other chain and hydrogen bond between O61 with N7 of the other c-di-GMP (Figure 3-4).

The DGC domains found in other response regulator proteins also have a Rxx(D/E) motif that is associated with the inhibitory site. In Q15Z91, the residues of this Rxx(D/E) motif are R212 and D215. A shown in Figure 3-4, R212 interacts with the O6 and N7 atoms on the Hoogsteen-edge of the guanine base of one c-di-GMP molecule with 2.9 Å and 3.0 Å, respectively. The gunidinium group of R212 also interacts with the phosphate group of the second c-di-GMP within the c-di-GMP dimer at a distance of 2.8 Å. The carboxyl group of D215 forms hydrogen bonds with the N1 and N2 atoms on the Watson-Crick-edge of the guanine base of the second c-di-GMP molecule. There are also many other residues contributing to interacting with c-di-GMP dimer, including T232-N2 of the second c-di-GMP (2.7 Å), R168-posphate group of the c-di-GMP (2.8 Å), E213-posphate group of the c-di-GMP (2.9 Å).
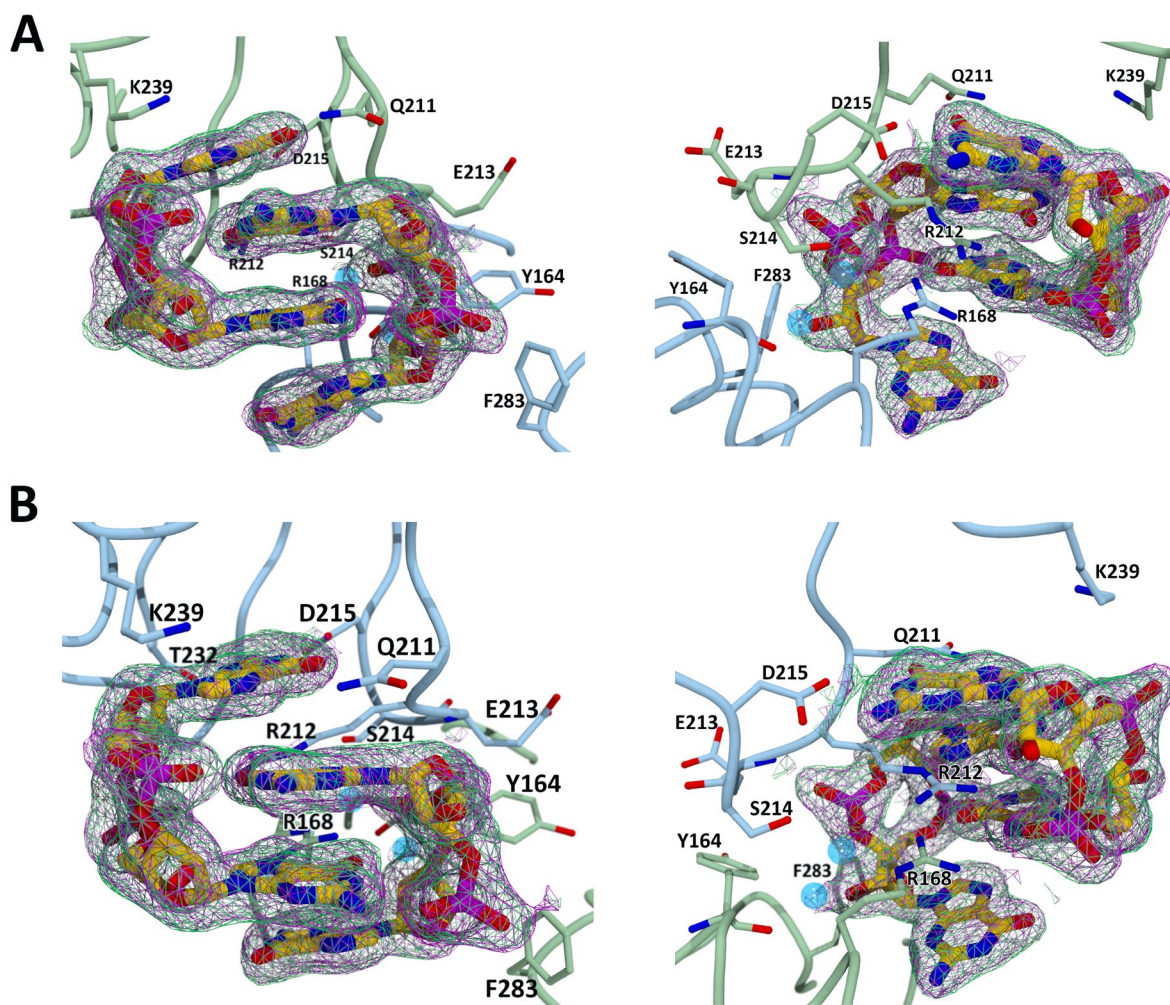
Figure 3-4: Two molecules of c-di-GMP dimerize through stacking of their bases and bind to the inhibitory sites on chain A (A) and chain B (B). The c-di-GMP dimers are shown as sticks with yellow carbon atoms. R212 and D215 are the most conserved residues in the I site. The $2|F_o|-|F_c|$ electron density map, contoured at 1.0σ, is shown as magenta mesh, and the $2|F_o|-|F_c|$ composite omit map, also contoured at 1.0σ, is shown as green mesh.

In addition to the I site with Rxx(D/E) motif, DGC has two active sites (A site) with GGEEF motif. In the structure of Q15Z91, two citrate molecules form the protein storage

buffer binds in the two A site respectively (Figure 3-5). E223, D179, carbonyl group of I180 and the citrate coordinate the Mg^{2+} ion. The basic residues K301 and R305, as well as the amide N atoms of F182, F183 and K184 interact with two carboxyl groups of citrate. In chain B, where there is a small interface between the ALR and DGC domains, Q78 from the ALR domain interacts with H192, which takes on multiple conformations. Due to the small interaction interface, this interdomain construction is probably caused by the whole geometry of this c-di-GMP inhibited dimer.
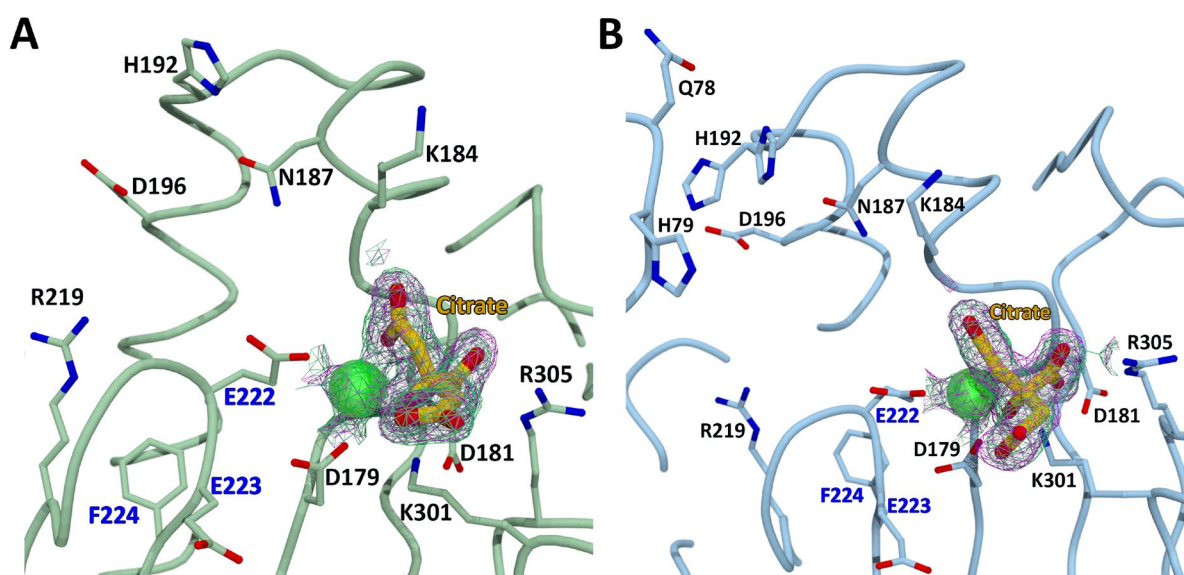
Figure 3-5: Citrate binds in slightly different orientations in the active sites of chain A (A) and chain B (B). Citrate is shown as sticks with yellow carbon atoms. The residues in the A site are also shown as sticks, with carbons colored according to the chain color. The labels of the conserved GGEEF residues are colored in blue. The $2|F_o|-|F_c|$ electron density map, contoured at 1.0σ, is shown as magenta mesh, and the $2|F_o|-|F_c|$ composite omit map, also contoured at 1.0σ, is shown as green mesh.

The structure of Q15Z91 with the substrate GTP was also determined with 1.8 Å data (Figure 3-6). The structure of Q15Z91 with GTP and c-di-GMP is same with the structure of Q15Z91 with citrate and c-di-GMP, except the GTP molecules bound in the A sites. The imidazole of H192 packs on top of the 6-membered ring of the nucleotide. The carboxyl group of D196 interacts with N2 of the guanosine. The guanidine group of R219 interacts with O6 of GTP. K184 forms a hydrogen bonding interaction to the 3' hydroxyl of GTP in chain B, while it interacts with 2'-OH of GTP in chain A. E223, D179, and the phosphate groups of GTP coordinate the Mg^{2+} ion; the metal, therefore, serves to anchor the tri-phosphate moiety of GTP to the enzyme. The basic residues K301 and R305 interact with the terminal phosphate group. In chain B, where there is an interface between the ALR and DGC domains, Q78 from the ALR domain interacts with H192, which takes on multiple conformations and appears to have less of an interaction with the GTP. The relative interactions between D196, K184, and the GTP are different from those observed in chain A: K184 makes a hydrogen bonding interaction with the 3' hydroxyl of GTP rather than the 2' hydroxyl, and D196 loses the interactions with H192 due to the inter-domain interaction between Q78 and H192.

Comparison with the structure of diguanylate cyclase PleD (PDB ID 2V0N[14]), the structure of Q15Z91 with GTP and c-di-GMP is more likely an activated but product-inhibited form of Q15Z91.
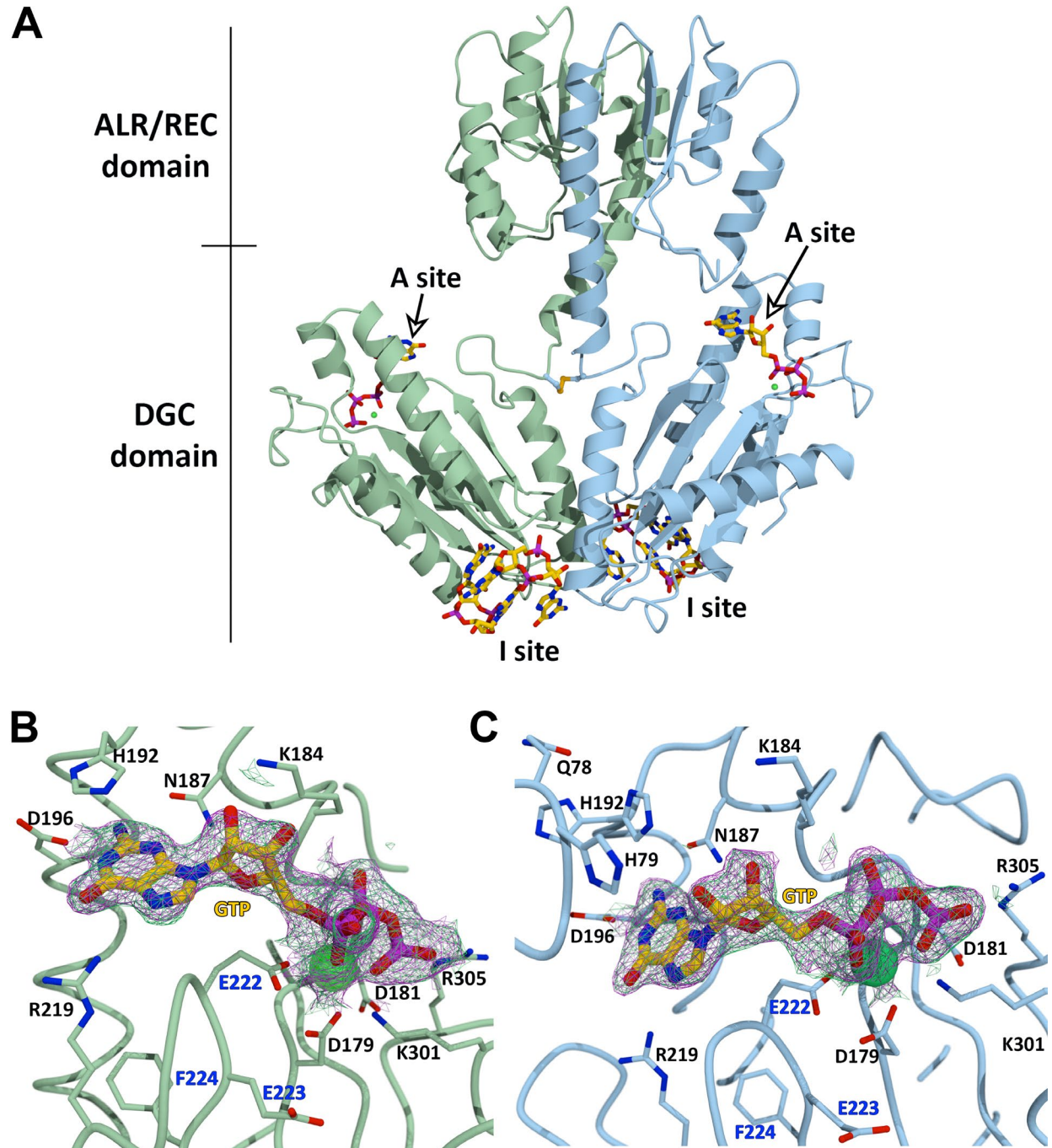
Figure 3-6: The structure of Q15Z91 with c-di-GMP and GTP bound (A). Chain A is colored in light green, while chain B is colored in light blue. Citrate and c-di-GMP are shown as sticks with yellow carbon atoms. GTP binds in slightly different orientations in the active sites of chain A (B) and chain B (C). GTP is shown as sticks with yellow carbon atoms. The residues involved in GTP interactions are also shown as sticks, with carbons colored according to the chain color. The labels of the conserved GGEEF residues are colored in blue. The

$2|F_o|-|F_c|$ electron density map, contoured at 1.0σ, is shown as magenta mesh, and the $2|F_o|-|F_c|$ composite omit map, also contoured at 1.0σ, is shown as green mesh.

It was not possible to obtain crystals of the un-inhibited (c-di-GMP-free) Q15Z91 and the Q15Z91(C142S) variant discussed below. The flexible conformation of the two catalytic domains without the product binding and disulfide bond may be one reason for the failures in crystallization of both c-di-GMP-free Q15Z91 and Q15Z91(C142S).

3.3.4 Probing the roles of conserved residues by site-directed mutagenesis

As mentioned in Chapter 2, the RitR double mutant RitR(L86A/V93A) exhibited activated characteristics (extended conformation and the ability to bind DNA). So the equivalent residues of Q15Z91, I99 and G102, were targeted for mutagenesis. C128 of RitR plays an important role in redox control. To determine whether or not the equivalent residue of Q15Z91, C142, has a similar role, this residue was changed to serine (C142S). Finally, R212 and D215 in the Rxx(D/E) I-site motif were mutated to Asn and Ala to determine the effect of disrupting the I-sites required for product inhibition. As described above, Q15Z91 and its variants only fold correctly in Origami2 *E. coli* cells, which have been engineered to have a less reducing environment in the cytoplasm and thus promote formation of disulfide bonds. After IPTG induction, biofilm formation was observed for in cultures expressing all forms of Q15Z91 we tested (Table 3-4). The fact that all of the Q15Z91 variants promoted biofilm formation indicates that none of them has been inactivated or catastrophically destabilized by the mutations.

Table 3-4: Characteristics of wild-type and mutant forms of Q15Z91.

Proteins	Biofilm formation	c-di-GMP
Q15Z91	+	+
I99A	+	+
G102A	+	+
C142S	+	-
R212N/D215A	+	-

The c-di-GMP produced during expression binds so well to the I-site that it co-purifies with Q15Z91, even persisting through dialysis and desalting steps. To remove the product inhibition, we used the phosphodiesterase SadR from *Pseudomonas aeruginosa* to linearize c-di-GMP into pGpG, which is easily removed by gel filtration. This c-di-GMP-free protein is called Q15Z91$_{free}$. To determine whether or not the mutants had c-di-GMP bound, the absorbance at 260 nm (due to the nucleotide) was measured for Q15Z91 and each of the mutants. The results demonstrate that only the C142S and R212N/D215A variants do not have c-di-GMP bound. This was expected for the double mutant, since both of those residues are involved in key interactions with the c-di-GMP dimer in the I-site. The fact the loss of C142S impacts product binding to the I-site as drastically it does was surprising and likely indicates that reduction of the C142-mediated disulfide bond increases Q15Z91-catalyzed DGC activity by relieving the product inhibition.

3.3.5 The oligomeric state of Q15Z91

As shown in Figure 3-7, Q15Z91 mainly exists as a dimer with c-di-GMP bound (retention time 5.2 min). The peak with a retention time of ~6.2 min is due to c-di-GMP, which is verified by the higher absorption at 260 nm than at 280 nm. After treatment with SadR and gel filtration, Q15Z91$_{free}$ still mainly exists as a dimer; the absence of the peak at ~6.2 min
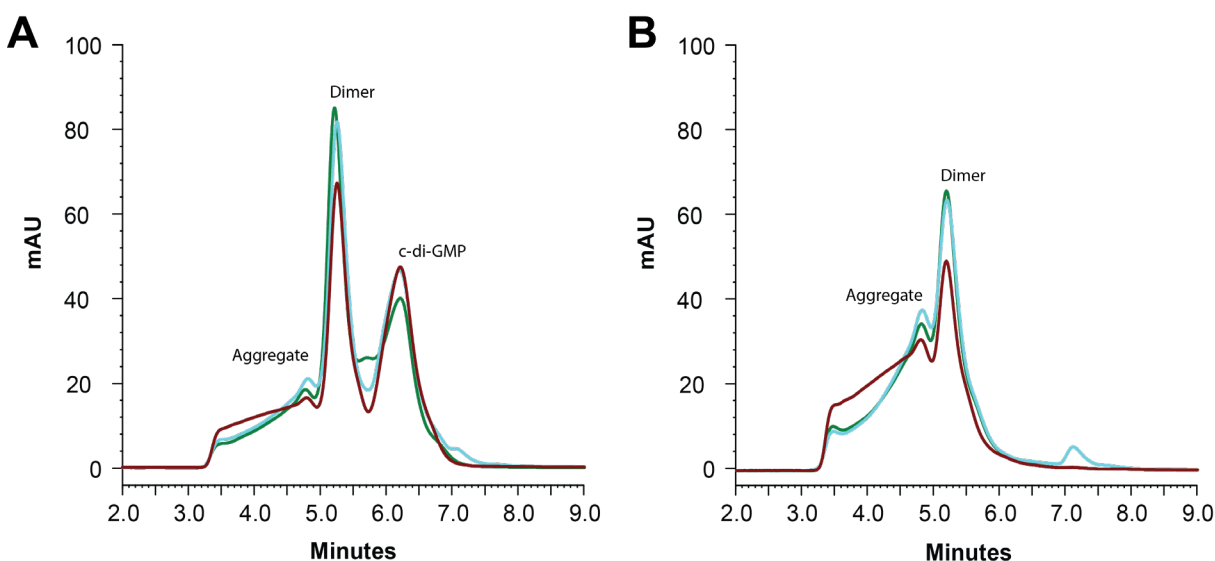
Figure 3-7: The affects of H_2O_2 or DTT treatment on the oligomeric states of Q15Z91 (A) and Q15Z91$_{free}$ (B) were analyzed by SEC. The traces of protein only colored in green, protein with 1 mM H_2O_2 treatment colored in brown red and protein with 1 mM DTT treatment is colored in cyan.

is further evidence that that peak is due to c-di-GMP. The Q15Z91$_{free}$ protein is considerably less stable than Q15Z91 (with c-di-GMP bound), as evidenced by the greater proportion of aggregated protein seen in the chromatogram. The retention time of the Q15Z91 dimer was

longer than expected based on the molecular weight standards. This may be due to shape effects or direct interactions between the protein and column matrix.
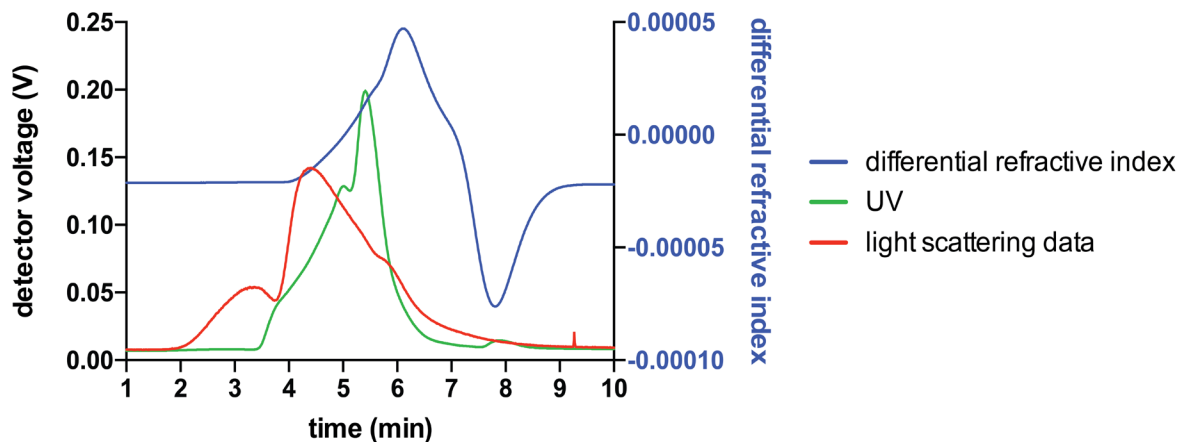
Elution time (min)	2.7-3.1	3.4-4.0	4.2-4.9	5.1-5.7	6.0-6.3
Molar mass (kDa)	101.8e+3	984.3	219.4	79.1	64.7

Figure 3-8: SEC-MALS results for Q15Z91$_{free}$.

In order to get a more accurate mass estimate for the molecule responsible for the peak at 5.2 min, Q15Z91$_{free}$ was analyzed by size exclusion chromatography with multi-angle static light scattering (SEC-MALS). The chromatograms were analyzed using the ASTRA software package (Figure 3-8). The major peak in the UV trace (retention time range 5.1-6.3 min) has a mass of 65-79 kDa, which is close to the mass of the Q15Z91 dimer. The peaks that elute earlier than 5 min were verified to be high molecular weight aggregates. To check the effects of reducing or oxidizing agents on the oligomeric state, DTT and H_2O_2 were added to Q15Z91 and Q15Z91$_{free}$ samples (1 mM each). SEC analysis (Figure 3-7)

shows that oxidizing and reducing C142 have no effect on the oligomeric state of either form of the protein (free or product-inhibited).

The oligomeric state of each of the Q15Z91 point mutants was also checked by size exclusion chromatography (Figure 3-9). The mutants I99A and G102A both existed as dimers with c-di-GMP bound. Like the wild-type protein, both mutants existed as dimers after treatment with SadR to remove the c-di-GMP. Interestingly, in spite of being treated in exactly the same way, SadR was apparently not able to remove all of the c-di-GMP from the I99A mutant. Consistent with the UV/vis data, the other two mutants C142S and R212N/D215A, existed as dimers without c-di-GMP bound. Neither of these mutants is affected by treatment with DTT or H_2O_2.

Figure 3-9: The oligomeric states of Q15Z91(I99A) (A), c-di-GMP free Q15Z91(I99A) (B), Q15Z91(G102A) (C), c-di-GMP free Q15Z91(G102) (D), Q15Z91(C142S) (E), Q15Z91(R212N/D215A) (F) with or without H_2O_2 or DTT treatment were analyzed by SEC. The traces of protein only colored in green, protein with 1 mM H_2O_2 treatment colored in brown red and protein with 1 mM DTT treatment is colored in cyan.
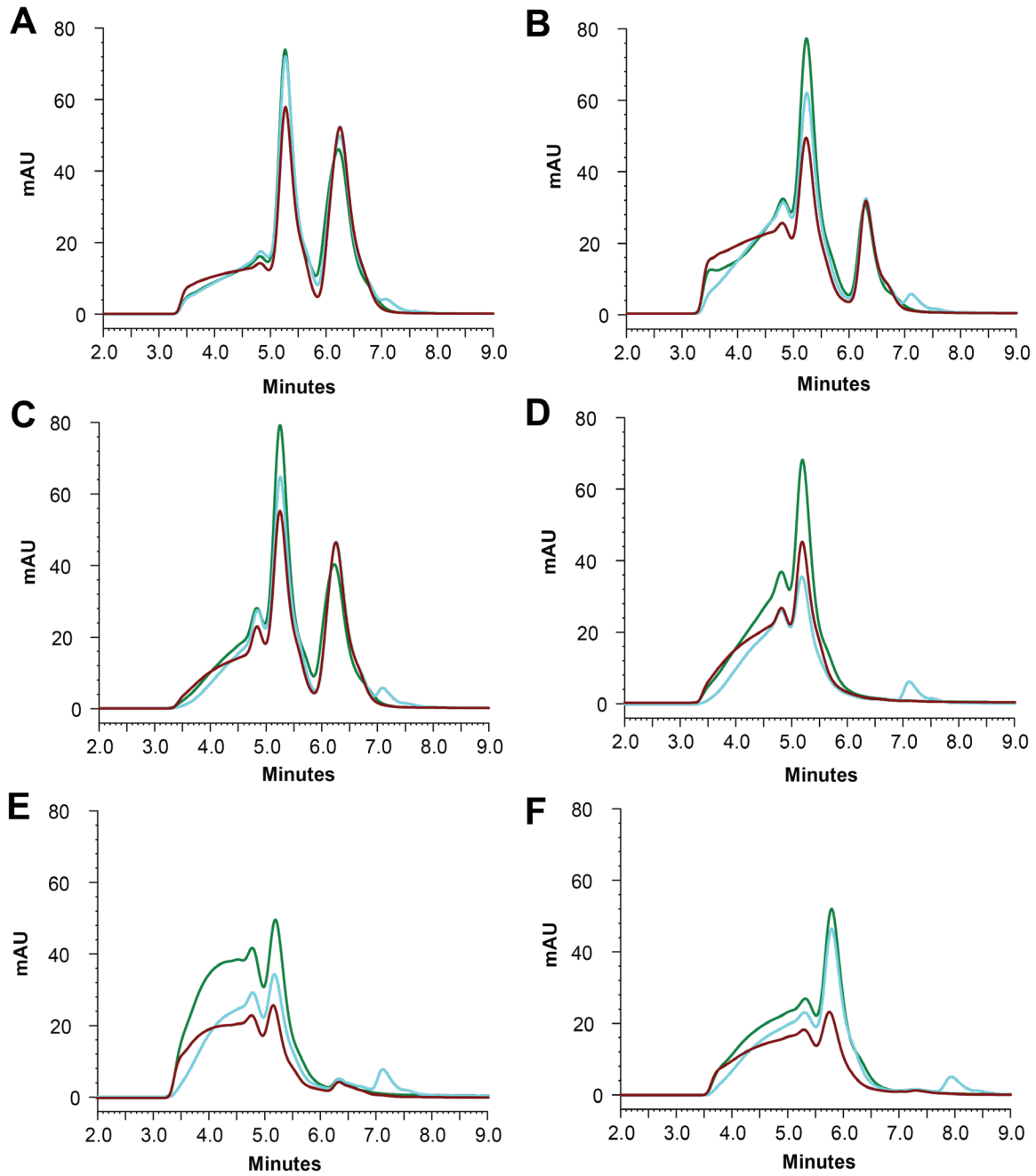
3.3.6 Preliminary quantitation of biofilm phenotype

In order to quantify the observed biofilm formation phenotype of Origami2 cells expressing Q15Z91 or its mutants, the amount of biofilm material deposited on the walls of

culture tubes was measured by the degree of staining with crystal violet as described in the methods section. All proteins were expressed as His$_6$-SUMO fusion proteins; the fact that biofilm formation was observed in response to Q15Z91 expression is evidence that the SUMO tag does not affect the catalytic activity of the protein. After dissolving the crystal violet-stained biofilm in a mixture of ethanol and acetone, the absorbance at 540 nm was measured. The amount of dye, and thus the absorbance, is proportional to the amount of biofilm present. Cultures were induced at an OD$_{600nm}$ of ~0.7, except for the Q15Z91(C142S) culture in LB with kanamycin. After growing for 7 hours at 37 °C, the OD$_{600nm}$ of C142S mutant in kanamycin-containing culture was still about 0.3, so 0.4 mM IPTG was added to induce SUMO-Q15Z91(C142S) expression. It is circumstantial, but this result may suggest that the C142S variant is constitutively active, and therefore, that the reduced form of the protein is the activated form. Not surprisingly, uncontrolled DGC activity is apparently toxic to *E. coli* cells.

In Figure 3-10A, it is clear that all of the other mutants formed biofilm and that the cells are evenly distributed in the medium. The cells carrying the pSUMO-C142S plasmid aggregated at the bottom of the culture tube. The A$_{540}$ of the biofilm extract from the Q15Z91(C142S) culture is the highest one. The double mutant R212N/D215A only formed a small amount of biofilm, which suggests that the I-site is also important for the cyclase activity.
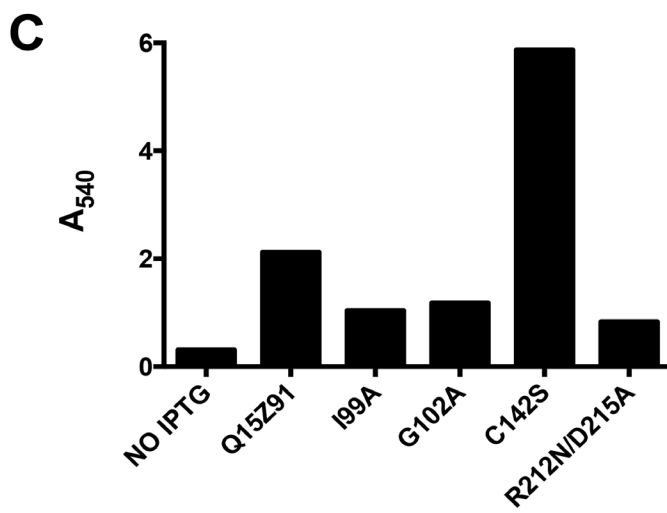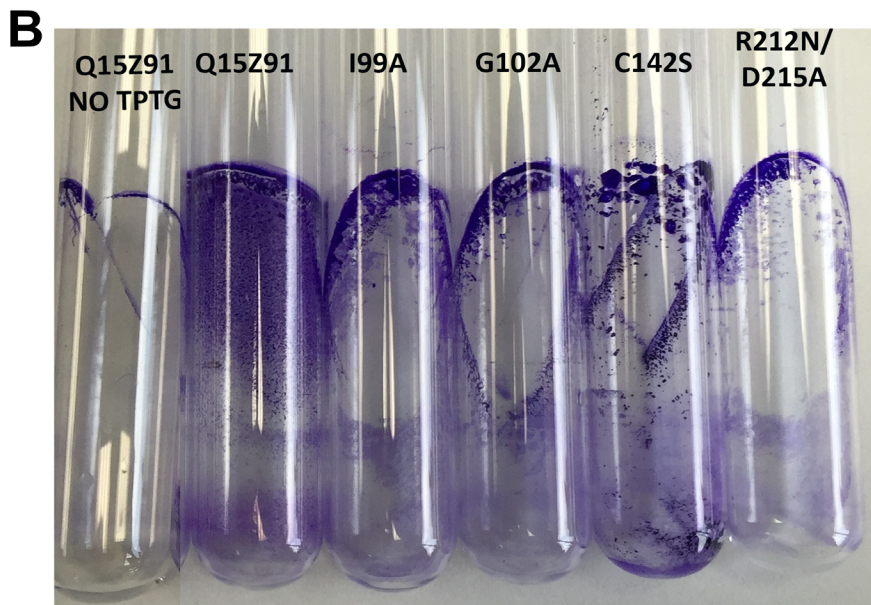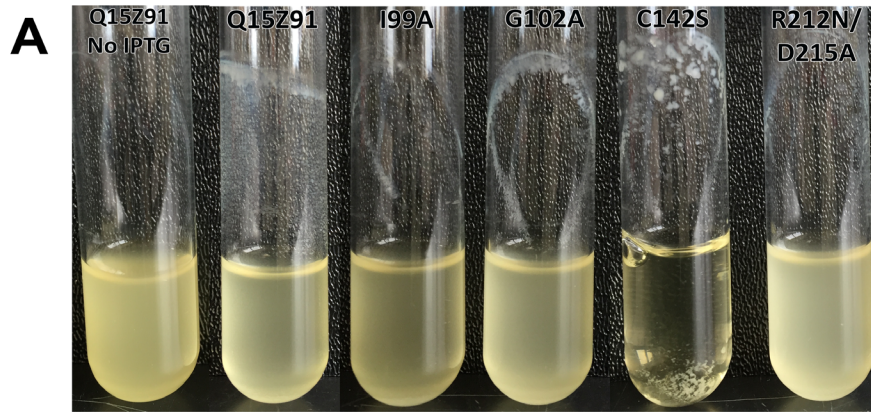
Figure 3-10: Biofilm formation after IPTG induction in Origami2 cells with Q15Z91 related plasmids. (A)

cultures photo (B) After crystal violet stain. (C) Biofilm amounts are quantified by absorbance at 540 nm. The biofilm formed in Q15Z91(C142S) is significantly higher than others.

3.3.7 Preliminary enzymatic activity

Three different assays were attempted in order to characterize the diguanylate cyclase activities of Q15Z91 and Q15Z91$_{free}$, in the presence and absence of DTT and H_2O_2. The first was an end-point assay where the proteins were incubated for 30 min in the presence of GTP and $MgCl_2$. The reaction products were then characterized by reverse-phase HPLC. Extensive method development work failed to identify combinations of stationary and mobile phases whereby GTP could be separated from the product c-di-GMP. In all of the methods tested, the analytes were not retained on the columns.

The second assay used the Biomol Green reagent to quantify the inorganic pyrophosphate released by the DGC. Since Biomol Green can only detect P_i, pyrophosphotase was used to digest the PP_i into two equivalents of P_i. The reactions contained 10 μM Q15Z91 or Q15Z91$_{free}$, 0.2 mM GTP, and 2 mM $MgCl_2$, and were incubated at RT for a total of 180 min. Aliquots were quenched at specific time points and the P_i content was measured to build a time course of the reaction. The preliminary results for Q15Z91 and Q15Z91$_{free}$ with GTP are shown in Figure 3-11. Both Q15Z91 and Q15Z91$_{free}$ exhibit a very low rate of PP_i release. This is not unexpected, since both proteins purify as their product-inhibited forms. Consistent with the SEC results above, the addition of 1 mM DTT results in a much more rapid reaction. The reaction is so rapid, in fact, that it was complete by the time the first aliquot was taken (10 min). The effect of DTT is especially

141

pronounced with the Q15Z91$_{free}$ enzyme, likely because this reaction is not contaminated with c-di-GMP.
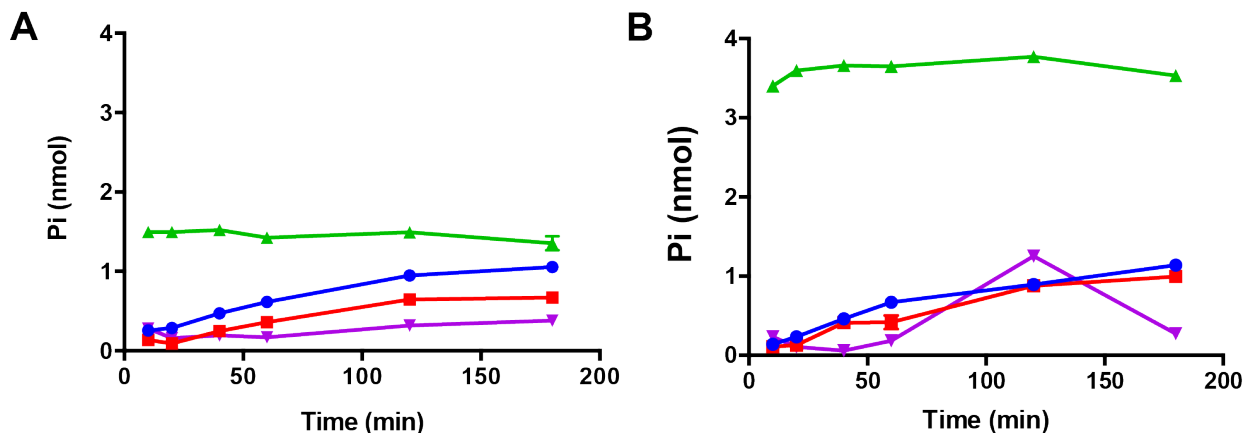
Figure 3-11: Time courses for the reaction of Q15Z91 (A) and Q15Z91$_{free}$ (B) with 0.2 mM GTP in the absence or presence of H_2O_2 or DTT. Blue line represents only. Red line represents 1 mM H_2O_2 treatment. Green line represents 1 mM DTT treatment. Purple line represents the reaction conducted at 4 °C.

Due to the speed of the reaction with the DTT-treated protein, we sought a direct assay that would allow us to monitor earlier time points. The EnzCheck assay kit (Sigma) allows the real-time monitoring of phosphate release by coupling it to purine nucleoside phosphorylase-catalyzed phosphorylation of the synthetic nucleotide MESG, which absorbs light at 360 nm. The first experiments with this assay were inconclusive, since the 10 min pre-incubation with MESG, purine nucleoside phosphorylase appeared to destabilize the enzyme, resulting in extremely small rates.

3.4 Discussion

Q15Z91 is a diguanylate cyclase that is regulated by an ALR domain. Key residues for the phosphorylation-dependent activation of typical response regulator receiver domains, particularly the acidic triad residues, have been lost in Q15Z91. The phosphorylatable aspartate is replaced by arginine, and the other two acidic triad residues are replaced with Met and Glu. These changes exclude the possibility of phosphorylation-dependent activation of Q15Z91. In spite of this, the positions of the switch residues positions and the dimer interface of Q15Z91 are similar to those of the typical, phosphorylation-dependent OmpR/PhoB RRs. Q15Z91, like RitR, the other ALR domain-containing protein discussed here, also has a conserved cysteine residue in the linker between the ALR and effector domains.

Q15Z91 purifies as a product-inhibited form with c-di-GMP bound. The Q15Z91 structure, as well as the SEC results, demonstrates that Q15Z91 exsits as a dimer. After removal of c-di-GMP, it maintains this dimer conformation. Neither H_2O_2 nor DTT have any effect on the oligomeric state of Q15Z91. Preliminary activity tests show that the activities of Q15Z91 and Q15Z91$_{free}$ are dramatically increased in the presence of DTT. To check the role of the conserved cysteine residue, Cys142 was mutated into serine. The variant Q15Z91(C142S) is also dimeric, providing ample evidence that Q15Z91 is not regulated by a switch between dimer and monomer as RitR is. Interestingly, the C142S mutant is no longer able to bind c-di-GMP, or at least only with much reduced affinity compared with the wild-type protein. Cyclic di-GMP, as an important second messenger, regulates biofilm formation, which is an important phenotype to check the cyclase activity in heterogeneous

cell systems. According to the results of our biofilm assays, there is a significantly increment of biofilm formation for Q15Z91(C142S) mutant. In other words, the variant C142S seems to greatly reduce product inhibition by c-di-GMP, which means the disulfide bond formation is likely essential for c-di-GMP binding. In summary, C142 of Q15Z91 is not essential for dimerization, but involved in facilitating c-di-GMP binding and product feedback inhibition.

In RitR, the C128 functions as redox sensor to control the inter-domain interactions for later dimerization and activation. Considering the dimer structure of Q15Z91, Q15Z91 appears to be constitutively active and this activity is modulated by product inhibition. Under oxidizing conditions, where the C142 disulfide bond is formed, product inhibition is potent and the DGC activity of Q15Z91 is low. Under reducing conditions, the C142 disulfide bond is broken, decreasing the affinity for c-di-GMP and largely removing the product inhibition. Taken together, the cysteine located at the C-terminal end of the α5 linker helix is important for the activation mechanism of Q15Z91. This is a departure from the well-studied phosphorylation-dependent diguanylate cyclases PleD[14-16] and WspR[4], where product inhibition is modulated by changes in the oligomeric state (dimer-tetramer equilibrium) that are influenced by additional domains. In short, Q15Z91 replaces the extra domain of PleD and WspR with a redox-active cysteine that directly influences product inhibition.

There are about fifty DGCs in *P. atlantica* T6c. These DGCs tend to be fused to PAS or GAF sensor domains, to additional regulatory domains, or to phosphodiesterase domains. These apparent redundancy but with different activation mechanisms DGCs/PDEs raises a

question that how the small molecule c-di-GMP to regulate the activities of these different

proteins. By studying Q15Z91, we find that c-di-GMP inhibition regulation is dependent on

the C142 disulfide bond formation. So we can assume that there are definitely other

mechanisms combining with product binding to regulate specific DGC/PDE activity upon

different environmental stimulation.

3.5 References

[1] Maule, A. F., Wright, D. P., Weiner, J. J., Han, L., Peterson, F. C., Volkman, B. F., Silvaggi, N. R., and Ulijasz, A. T. (2015) The Aspartate-Less Receiver (ALR) Domains: Distribution, Structure and Function, *PLOS Pathogens 11*, e1004795.

[2] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool, *Journal of molecular biology 215*, 403-410.

[3] States, D. J., and Gish, W. (1994) Combined use of sequence similarity and codon bias for coding region identification, *Journal of computational biology : a journal of computational molecular cell biology 1*, 39-50.

[4] De, N., Pirruccello, M., Krasteva, P. V., Bae, N., Raghavan, R. V., and Sondermann, H. (2008) Phosphorylation-Independent Regulation of the Diguanylate Cyclase WspR, *PLoS biology 6*, e67.

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

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

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

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

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

[10] Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation11Edited by J. Thornton, *Journal of molecular biology 285*, 1735-1747.

[11] Urzhumtseva, L., Afonine Pv Fau - Adams, P. D., Adams Pd Fau - Urzhumtsev, A., and Urzhumtsev, A. Crystallographic model quality at a glance.

[12] Upson, R. H., Haugland, R. P., Malekzadeh, M. N., and Haugland, R. P. (1996) A spectrophotometric method to measure enzymatic activity in reactions that generate inorganic pyrophosphate, *Analytical biochemistry 243*, 41-45.

[13] Webb, M. R. (1992) A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems, *Proceedings of the National Academy of Sciences 89*, 4884-4887.

[14] Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U., and Schirmer, T. (2007) Structure of BeF3−-Modified Response Regulator PleD: Implications for Diguanylate Cyclase Activation, Catalysis, and Feedback Inhibition, *Structure 15*, 915-927.

[15] Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U., and Schirmer, T. (2004) Structural basis of activity and allosteric control of diguanylate cyclase, *Proc Natl Acad Sci U S A 101*, 17084-17089.

[16] Paul, R., Abel, S., Wassmann, P., Beck, A., Heerklotz, H., and Jenal, U. (2007) Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization, *The Journal of biological chemistry 282*, 29170-29177.

Chapter 4

An Unusual PLP-Dependent Enzyme in Enduracididine Biosynthesis

4.1 Introduction

The emergence of antibiotic-resistant pathogens is a serious and growing threat to global health. There is an urgent need for new antibiotics to combat this threat. Unfortunately, the pace of antibiotic development has slowed almost to a standstill (Figure 4-1).
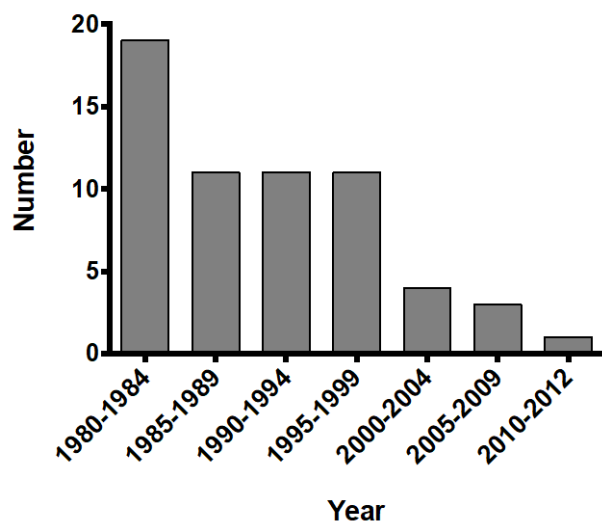
Figure 4-1: The numbers of new antibiotic compounds approved have been declining steadily since 1980. Data from the FDA Center for Drug Evaluation and Research.

In response to the dismal rate of antibiotic development, among other problems, Congress passed the Food and Drug Administration Safety and Innovation Act (FDASIA).

Title VIII of this law, Generating Antibiotic Incentives Now (GAIN), created a fast track for new antibiotic approvals that seems to be paying off. While there have been no new classes of antibiotics, a number of new drugs based on old scaffolds have been introduced since the passage of the FDASIA; in 2014-15 six new antibiotics were approved by the FDA. Happily, it seems that interest in antibiotic development is on the rise. Consequently, any information regarding the biosynthesis or activity of antibiotics or key building blocks of these compounds, is potentially useful in these renewed efforts to discover new antibiotics.

In this context, the non-proteinogenic amino acid L-enduracididine (L-End, **1**, Scheme 4-1) is particularly interesting. This conformationally restricted form of arginine is present in a number of antibiotics, including some with potent activity against antibiotic-resistant strains like methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant enterococci (VRE). Some prominent examples are the cyclic depsipeptide antibiotics enduracidin [3], mannopeptimycin [4] and the recently discovered teixobactin [5].

Enduracidin was first isolated from *Streptomyces fungicidicus* No. B 5477[3]. It is a seventeen amino acid lipodepsipeptide containing one residue of L-End and one of D-End (Scheme 4-2). In spite of having different amino acids and acyl chain, the overall structure enduracidin is similar to the cyclodepsipeptide antibiotic ramoplanin [6]. Both enduracidin and ramoplanin are active against resistant bacteria
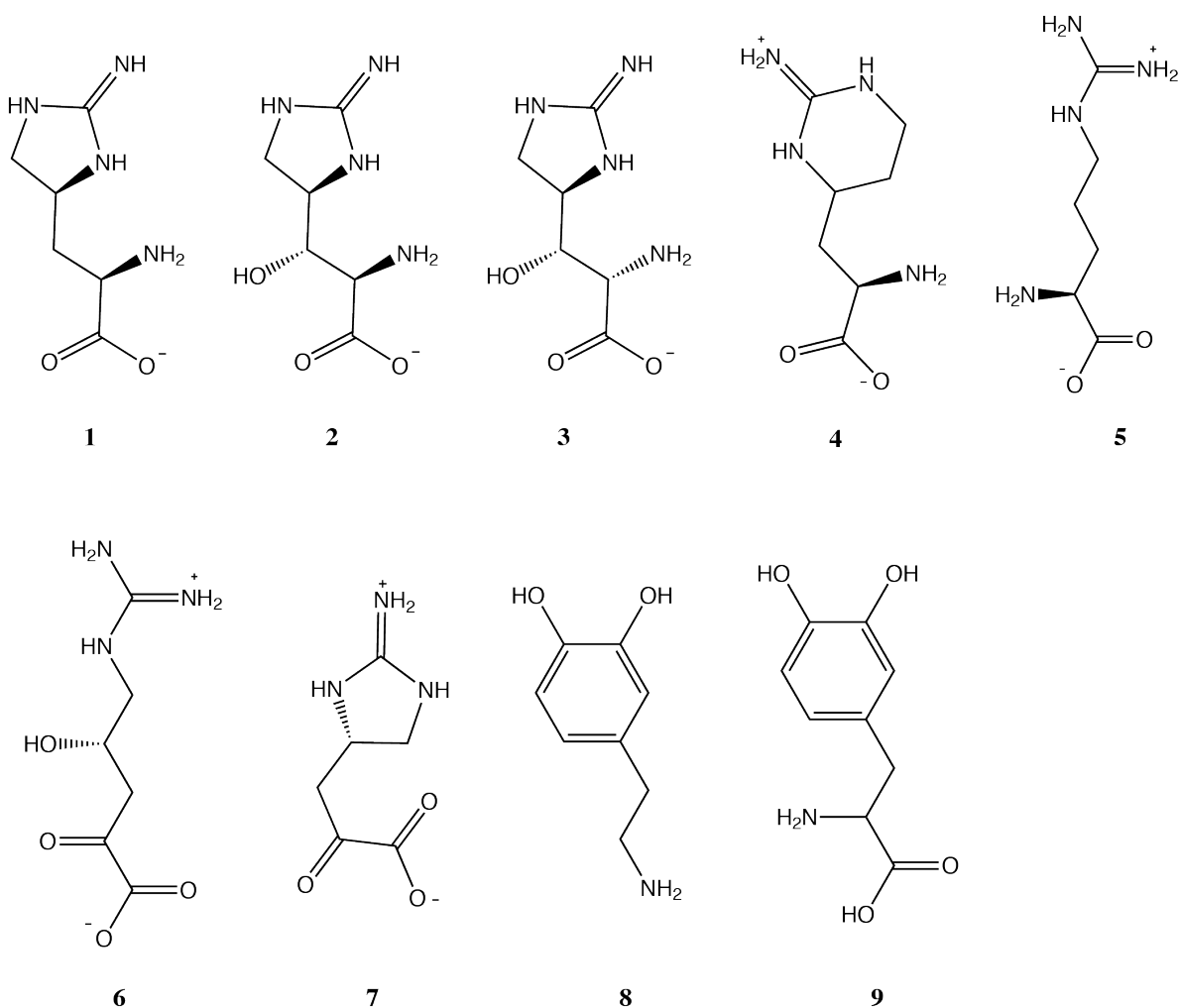
Scheme 4-1: Compounds mentioned in this chapter. **1**, L-enduracididine; **2**, L-β-hydroxyenduracididine; **3**, D-β-hydroxyenduracididine; **4**, capreomycidine; **5**, L-arginine; **6**, 2-oxo-4-hydroxy-5-guanidinovaleric acid (*i.e.* 4-hydroxy-ketoarginine, 4HKA); **7**, 2-ketoenduracididine (2KE); **8**, L-3,4-dihydroxyphenylalanine (L-DOPA); **9**, dopamine.

by binding to the external membrane surface of the lipid II intermediate, inhibiting the transglycosylation reaction of cell wall biosynthesis [7]. Ramoplanin has entered phase III

clinical trials for VRE suppression in gastrointestinal tract infections and prevention of bloodstream infections, but enduracidin is hampered by its low aqueous solubility [8].

Another cyclic End-containing nonribosomal antibiotic, mannopeptimycin, is a glycosylated hexapeptide that also has activity against resistant Gram-positive bacteria, including MRSA and VRE [4, 9]. Mannopeptimycin also inhibits peptidoglycan synthesis by binding to lipid II [9]. Studies of the mechanism of action suggested that mannopeptimycin may first binds to the lipoteichoic acids of the Gram-positive cell wall, and then accumulates to interact with lipid II[9]. The structure of mannopeptimycins α, β, γ, δ, and ε from *Streptomyces hygroscopicus* LL-AC98 were characterized in 2002 [4]. They contain the non-proteinogenic amino acids L-β-hydroxyenduracididine (**2**, L-βhEnd) and D-β-hydroxyenduracididine (**3**, D-βhEnd). Methods for the total synthesis of mannopeptimycins α and β were published recently [10], but these two mannopeptimycins have relatively poor anti-bacterial activities. Much work (a Sci-Finder search returns more than 2000 analogs) has gone into making derivatives in the hope of finding one that has better therapeautic properties than the mannopeptimycins [11-17]. However, almost all of these involve changes to the attached glycans; very few of them alter the peptide core itself.

Enduracidin

Minosaminomycin

Mannopeptimycin

TAN-2177A

N-(1H-indolyl-3-carbonyl)-L-End
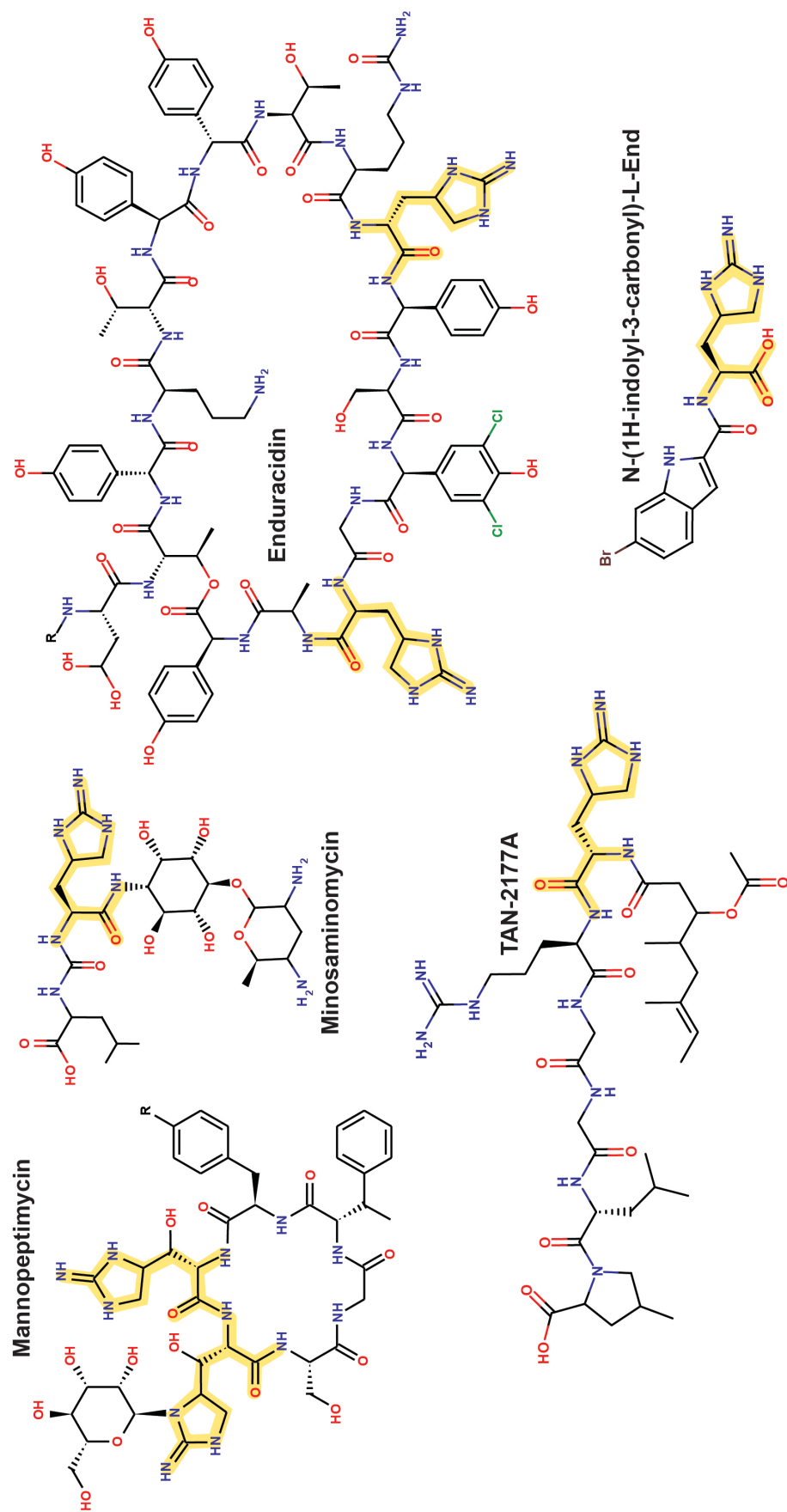
Scheme 4-2: Enduracididine-containing natural products. Enduracididine residues are highlighted in yellow

151

Mannopeptimycin and enduracidin are both antibiotics with potent activity against antibiotic-resistant strains. To improve the physico-chemical properties and/or biological activities for therapeutic applications, they probably require further chemical modifications. However, the total syntheses of enduracidin and mannopeptimycin are limited by the complexity of the structures and lmited commercial availability of some non-proteinogenic amino acid building blocks like enduracididine and its derivatives. Hence, a supply of L-End in quantities suitable to support drug development would greatly facilitate efforts to make derivatives of L-End-containing natural products.

Several synthetic routes to L-End or its β-hydroxy analogs have been published. Shiba et al. began with the methyl ester of L-histidine, which underwent cleavage of the imidazole ring to give a protected amide that was subsequently deprotected, guanidinylated, and treated with HF to give L-End as a diasteriomeric mixture [18]. The whole procedure involves only 4 steps, but it requires HF, which is hazardous, and the yield of the proper diasteriomer is low (22%). Dodd et al. used an aziridine-based method to synthesize 9-phenylfluorenyl-protected L-End with 4 steps, but the yield is still low (33%) [19]. No improvement in the yield was obtained after many attempts. The azide derivatives of β-hydroxy-End were first synthesized by Oberthür et al [20]. It takes 21 steps to synthesize **2** with ~16% overall yield, and 24 steps to synthesize compound **3** with ~10% yield. After optimization, shorter routes were developed (*e.g.* 8 steps from Garner's aldehyde), but with no improvement in the yields (16% for **2** and 8% for **3**) [21]. Recently, the most efficient route to N-mannosyl-D-β-hydroxy-End was published by Cheng et al., which gave 43% yield [22]. Given the difficulties of synthesizing the nitrogen-rich iminoimidazolidine ring of

the L-End, and the presence of two chiral centers, an enzymatic or chemo-enzymatic route to L-End may prove to be the most efficient means of producing large quantities of L-End.

4.2 Biosynthesis of enduracididine

The enduracididine biosynthetic genes were first identified by comparing the enduracidin biosynthetic cluster in *S. fungicidicus* to that of the mannopeptimycins from *S. hygroscopicus* [23, 24]. The enduracidin and mannopeptimycin scaffolds are very different. In fact, aside from the presence of L-End in both products, and the fact that both are non-ribosomally produced depsipeptides, there are no similarities between them. Thus, the presence of 3 pairs of genes in the two biosynthetic clusters with high pairwise sequence identity suggested that these genes are likely responsible for L-End biosynthesis. The products encoded by these gene pairs are MppP/EndP (85%), MppQ/EndQ (66%), and MppR/EndR (73%) (Mpp = mannopeptimycin; End = enduracidin; Figure 4-2). Primary sequence analysis suggested that products P and Q were pyridoxal-5'-phosphate (PLP)-dependent aminotransferases, and that R was likely an acetoacetate decarboxylase.

Experiments by K. Hatano *et al.* [25] used radio-labeled amino acids to investigate the biosynthetic origin of L-End and found that the most radioactivity was incorporated into the L-End fraction of the enduracidin that was produced when cultures were fed with labeled L-Arg. Thus, it was concluded that the biosynthesis of L-End, now presumed to be carried out by Mpp/EndP, Mpp/EndQ, and Mpp/EndR, likely begins from L-Arg [25]. This, in turn, suggests that the iminoimidazolidine ring of End is probably formed by cyclization

153
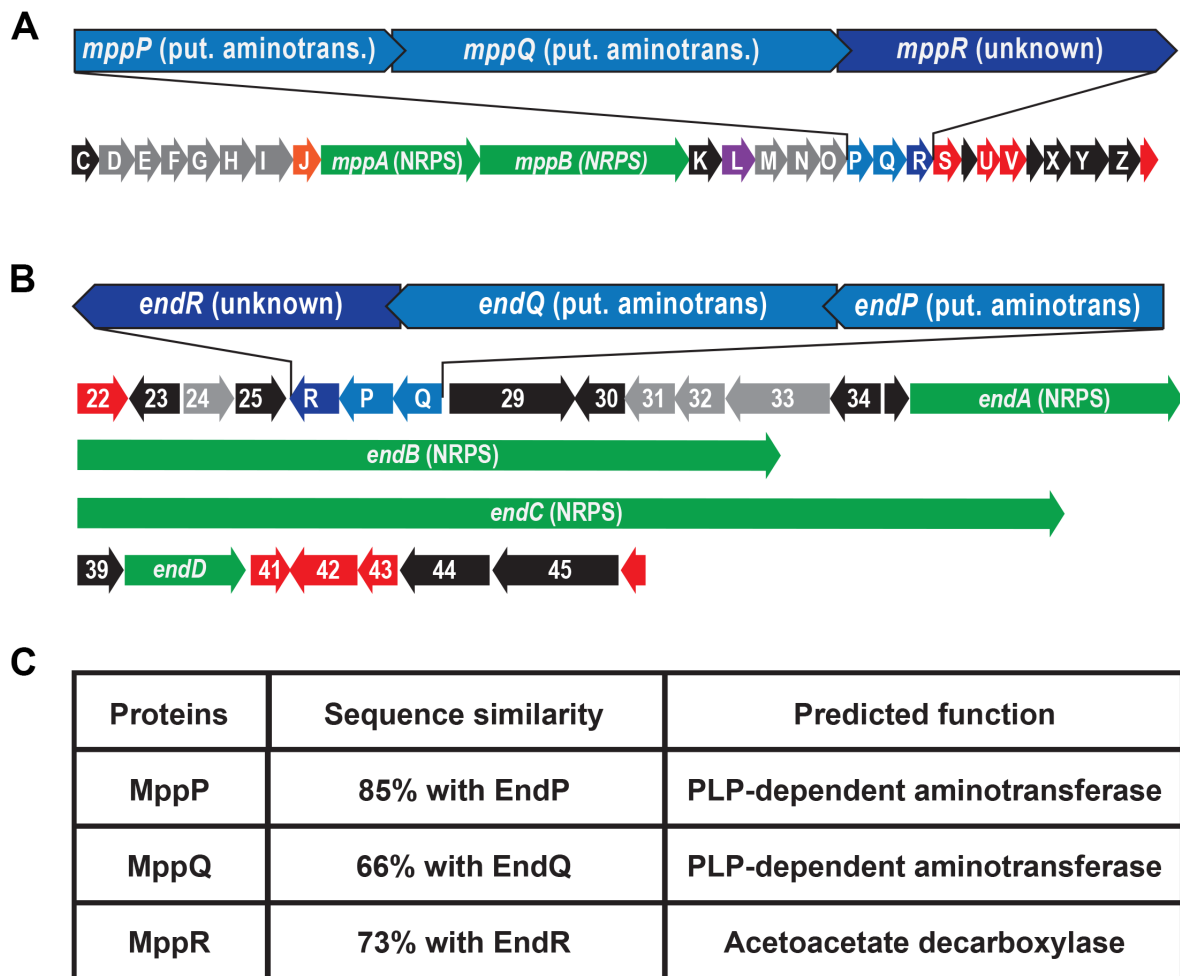
Figure 4-2: Comparison of the mannopeptimycin biosynthetic cluster (A) with that of enduracidin (B) suggested that the three genes *mppP/endP*, *mppQ/endQ* and *mppR/endR* are involved in enduracididine biosynthesis. These genes are enlarged and colored in blue. Comparing the sequences of MppP with EndP, MppQ with EndQ, and MppR with EndR (C) shows a high degree of pairwise sequence identity. On the basis of their sequences, Mpp/EndP and Mpp/EndQ are predicted to be PLP-dependent aminotransferases and Mpp/EndR is predicted to be an acetoacetate decarboxylase.

between the γ-carbon and one of the guanidine N atoms of arginine [25]. There is precedent

for this sort of transformation in the biosynthesis of another arginine-derived non-proteinogenic amino acid, L-capreomycidine (L-Cap, **4**; Scheme 4-3). Capreomycidine is found in the antibiotic viomycin, and its biosynthesis was initially thought to provide a possible chemical logic for L-End biosynthesis. In the viomycin biosynthetic cluster from *S. vinaceus*, L-Cap (compound **4**) is produced from L-Arg (compound **5**) by two enzymes: VioC and VioD. VioC is an Fe(II)/α-ketoglutarate-dependent oxygenase that hydroxylates L-Arg at the β-carbon to give 3S-hydroxy-L-Arg [26]. This product is then cyclized by VioD, a PLP-dependent aminotransferase that catalyzes an intra-molecular transamination to give the
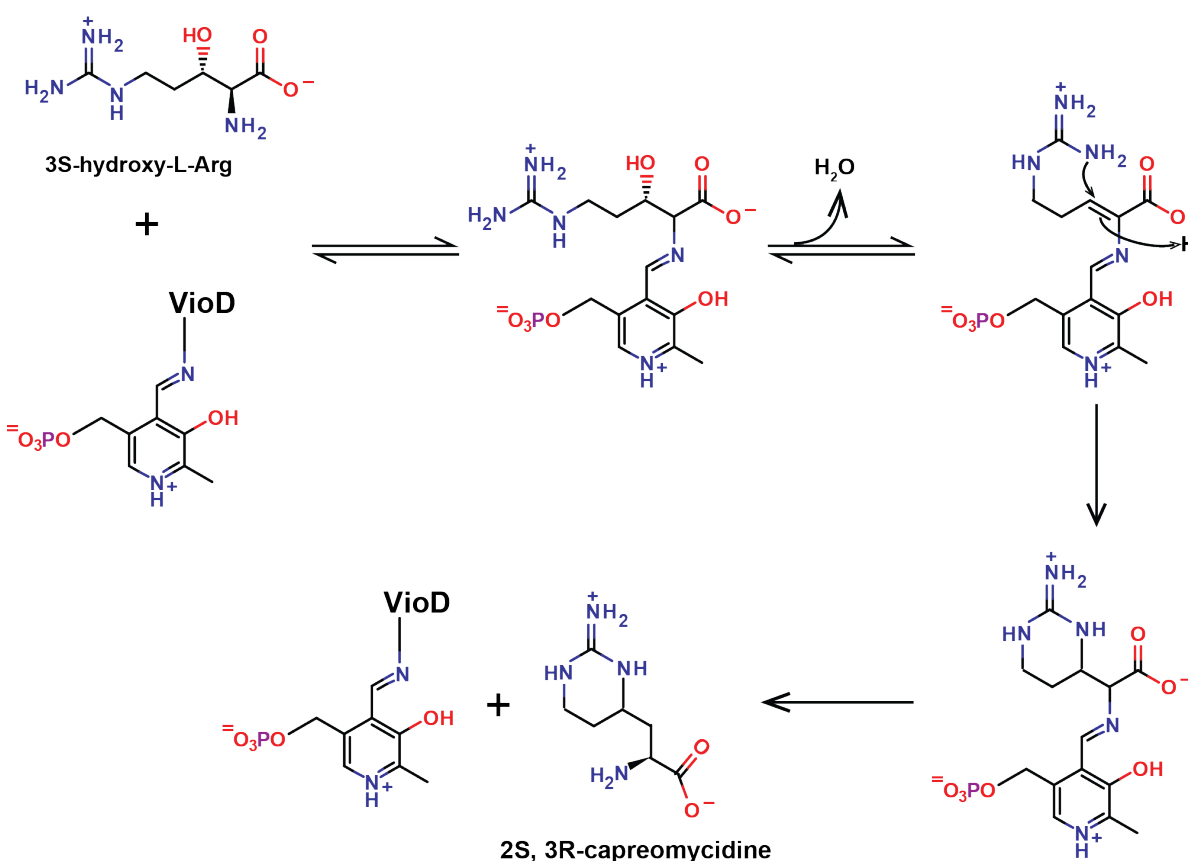
Scheme 4-3: L-capreomycidine biosynthesis mechanism by VioD.

6-membered capreomycidine ring [27, 28].

According to their amino acid sequences, MppP and MppQ were both predicted to be class I/II aminotransferases, while MppR was predicted to be an acetoacetate decarboxylase. Assuming that L-End biosynthesis likely occurs in a similar way to L-Cap biosynthesis, that is by activation of L-Arg by hydroxylation followed by transamination, it was not clear exactly how two aminotransferases and a decarboxylase would accomplish this transformation. Clearly, there had to be a difference in the biosynthetic schemes of L-Cap and L-End.

The first clues came from studies of MppR [29]. In spite of its low (~20%) sequence identity with acetoacetate decarboxylase, MppR was shown to have a nearly identical tertiary structure. Differences in the active site of MppR, such as the loss of key residues for the decarboxylation reaction, suggested that MppR was not a decarboxylase. This was subsequently confirmed experimentally. The real breakthrough came when it was demonstrated that MppR is able to cyclize the ketone derived from γ-hydroxy-L-arginine, to give the ketone form of enduracididine (i.e. 2-keto-End, compound **7**, Scheme 4-1). This product is actually observed bound to the catalytic lysine residue of MppR in crystals soaked with 2-oxo-4-hydroxy-5-guanidinovaleric acid (4HKA, compound **6**). In addition, experiments with MppQ showed that it was a true aminotransferase, capable of transaminating L-Arg and pyruvate or glyolxylate, or the reverse, 2-ketoarginine and alanine or glycine (unpublished results). These observations suggest that MppR generates **7** from the highly oxidized arginine derivative **6**, which is then transaminated by MppQ to give the amino acid, L-End. The L-β-hydroxy-End found in mannopeptimycin is produced

from L-End by the Fe(II)-dependent oxygenase MppO, which hydroxylates the β-carbon of L-End to give L-βhEnd (compound **2**) [30]. D-βhEnd is presumably generated from L-βhEnd by epimerization [31]. All of this was plausible based on the data available at the time, but raised the question of where the highly oxidized arginine derivative **6** comes from.

Extensive work by predecessors on the enduracididine project failed to produce any of the *S. hygroscopicus* MppP in soluble form. Multiple strategies beyond the typical optimization of expression conditions were employed, including for example, truncation of the N- and C-termini, codon optimization, and site-directed mutagenesis. Switching to the *S. fungicidicus* homolog EndP from the enduracidin pathway gave no better results. A final BLAST search preparatory to throwing the MppP project away altogether returned a homolog from *Streptomyces wadayamensis* that was significantly longer than either MppP or EndP. Both of these are listed in GenBank at 293 amino acids, which is considerably shorter than the ~415 residues typical of aminotransferases. The *S. wadayamensis* homolog is 376 amino acids, so it would seem that a non-standard initiation codon fooled the annotation software into thinking the MppP and EndP coding sequences began about 100 codons into the gene. This at least made it clear why the protein would not express—all of the expression constructs tried to date contained at most ~3/4 of the gene! The *S. wadayamensis* enzyme expressed well in *E. coli.*

With soluble protein in hand, we were finally able to begin the work of identifying the catalytic activity of MppP. *S. wadayamensis* is an endophytic strain isolated from mandarin organge trees (*Citrus reticulata*), which produces, among other natural products, mannopeptimycin [32]. Since it was expected from the work with MppR and MppQ that MppP

must catalyze the first step of the pathway, the *S. wadayamensis* MppP (SwMppP) was incubated with L-Arg and monitored by UV-Vis spectroscopy. The results of these first crude experiments, which will be discussed in detail in Chapter 5, indicated that MppP reacted with L-Arg, finally accumulating a colored (pale pink) intermediate when the cuvette had been exhausted of dioxygen. Before delving into the experimental results we ultimately obtained, however, it will be instructive to investigate the chemistry of PLP and some properties of well-characterized PLP-dependent enzymes.

4.3 Pyridoxial 5'-phosphate is a versatile cofactor

PLP is a versatile cofactor catalyzing manifold reactions, including transamination, racemization, C-C bond formation or cleavage, β- and γ-elimination/substitution, and oxidative decarboxylation [33]. Some general reaction mechanisms are shown in Scheme 4-4. Due to its electron sink properties, PLP is inherently reactive and readily catalyzes reactions with amino acids, for example, without the aid of any enzyme. In fact, it has been said that, in general, PLP-dependent enzymes do not so much accelerate the PLP-catalyzed reaction, but rather direct the catalytic power of PLP in a single direction [34]. In other words, PLP-dependent enzymes provide modest rate accelerations and strictly control the reaciton specificity. Toney [35] reviewed and summarized the three major factors that determine the reaction specificity in PLP-dependent enzymes: stereoelectronic control, the electrophilic strength of the Schiff base-pyridine ring π-electron system, and the placement of the substrate's side chain. Certain factors are common to most PLP-dependent enzymes, such as the covalent complex between a catalytic lysine residue on the enzyme and the PLP

158

cofactor; this is the internal aldimine. When substrates bind, they react, normally by attack on the C4' of PLP to displace the enzyme lysine side chain and form the external aldimine.

As shown in Scheme 4-4, the external aldimine is the branch-point for many of the reaction types, such as decaboxylation or Cα-H abstraction. Which bond is going to be broken depends on the orientation relative to the pyridine ring. Thus, the bond aligned parallel to the π-orbitals (perpendicular to the PLP plane) promotes π–σ-orbital overlap, is weakened and is more easily cleaved [35]. Hence, decarboxylases bind their substrates so that the bond to the carboxyl group is located perpendicular to the PLP plane. Conversely, enzymes that begin with Cα-H proton abstraction, such as transaminases and racemases, orient the Cα-H bond perpendicular to the PLP ring.
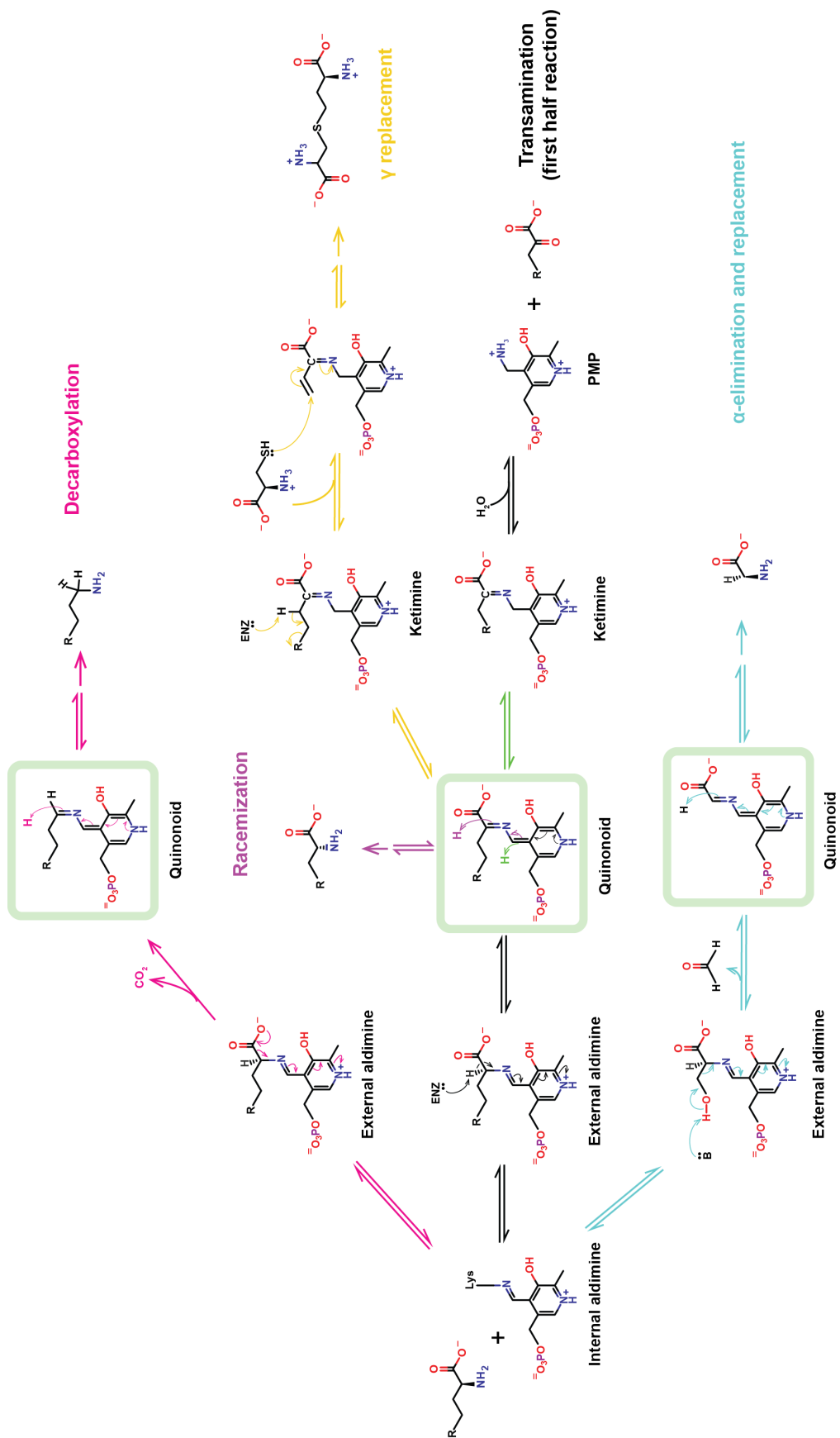
Scheme 4-4: Some examples of reactions involving PLP. Quinonoid intermediates are highlighted by green squares.

160

The protonation state of the pyridine nitrogen atom is another key factor influencing the reaction specificity. In aspartate aminotransferase, for example, this nitrogen is kept in the protonated, pyridinium form by the juxtaposition of the negatively charged carboxyl group of a conserved aspartic acid residue [36]. In enzymes like alanine racemase, where the carbanion at Cα must be stabilized, the pyridine nitrogen is kept deprotonated via interaction with the guanidinium group of a conserved arginine [37].

Percudani and Peracchi have created a B6 database utilizing hidden Markov models to compare the sequences of PLP-dependent enzymes. Their analysis distributes PLP-dependent enzymes into seven separate clusters based on their functions [38]. The representatives of these seven clusters are shown in Figure 4-3. Five of these clusters are essentially identical to the traditional classification system for PLP-dependent enzymes, which is based on 5 distinct topologies or "Fold Types" [39]. The Type I PLP-dependent enzymes are also called the aspartate aminotransferase (ATase) superfamily. Enzymes in this superfamily are typically dimeric (sometimes higher order oligomers) and have three discontinuous domains: the N-terminal extension, the large domain, and the small domain. The active site is located at the interface between the large and small domains, and is composed of residues from two protomers of the oligomer. Fold Type I is the largest and most diverse in terms of function and structure of N terminal domain, including class I, II, III, and V aminotransferases as well as enzymes catalyzing other PLP-dependent reactions like cystathionine γ-lyase, serine hydroxymethyltransferase [39, 40]. Class I ATase mainly contains aspartate aminotransferase, tyrosine ATase, aromatic ATase, alanine ATase and l-aminocyclopropane-1-carboxylate synthase. Class II ATase mainly comprises glycine acetyltransferase, 5-aminolevulinic acid synthase, histidinol-phosphate aminotransferase.

161

Class III mainly contains ornithine ATase, 4-aminobutyrate ATase, ω-amino acid-pyruvate ATase. Class V ATase mainly contains phosphoserine ATase, serine-pyruvate ATase.

The Type II PLP-depedent enzymes are evolutionarily distinct from the Type I enzymes, despite having similar tertiary structures[33, 41]. Fold type II includes the tryptophan synthase family of enzymes that mainly catalyze replacement and elimination reactions at Cβ, while most aspartate ATase family proteins promote reactions on the Cα atom of the substrate[41]. Fold Type II also contains acetyltransferases, ammonia lyases, and some racemases. In contrast to the Type I enzymes, the active sites of Type II enzymes are consist entirely of residues from a single protomer (such as tryptophan synthase[42]). Another distinguishing characteristic of these two enzyme families is the existence of a regulatory domain in the Type II enzymes[43, 44], whereas this is lacking in the Type I proteins.

The other fold types have fewer members than either of those discussed above. Fold Type III, also called the alanine racemase family, contains primarily amino acid racemases and decarboxylases. The structures of these proteins consist of an N-terminal β/α barrel and a C-terminal mostly β sheets domain [39]. Fold Type IV contains the class IV aminotransferases, which take mainly branched-chain amino acids and D-amino acids as substrates. The structures of the Type IV enzymes are very similar to the Type I ATases, but the PLP is bound in the opposite orientation
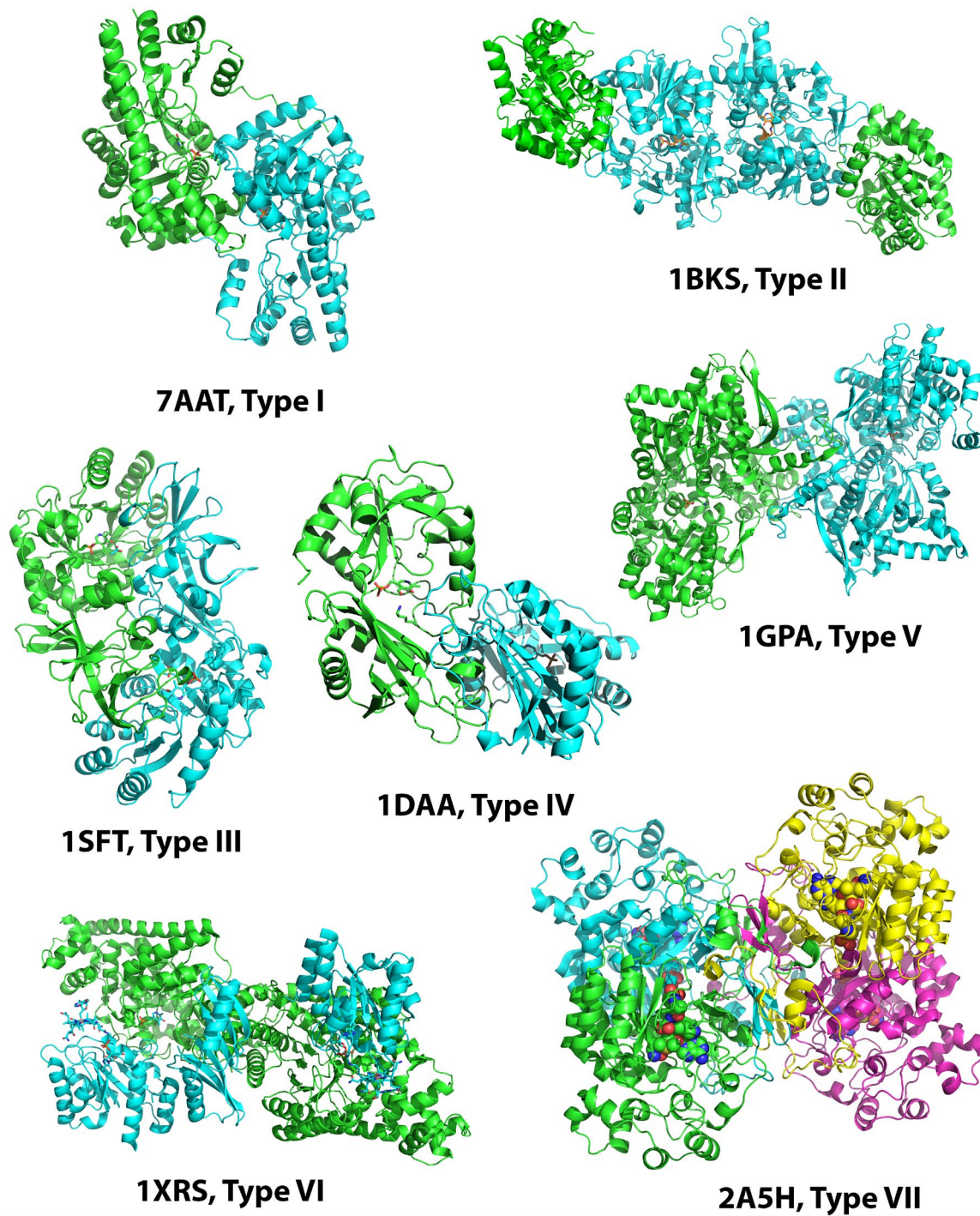
Figure 4-3: Ribbon diagrams of representative structures of the seven Fold Types of PLP-dependent enzymes. Each depicts the functional oligomeric state. The PLP and catalytic lysine, as well as substrates when available, are shown as sticks. PDB ID 7AAT[45] is the structure of mitochondrial aspartate aminotransferase from *Gallus gallus*, which is the representative of fold type I and functions as a dimer (chain

163

A colored in green, chain B colored in cyan). PDB ID 1BKS[46] is the structure of tryptophan synthase from *Salmonella typhimurium*, which functions as an $\alpha_2\beta_2$ complex (α subunit colored in green, β subunit colored in cyan). The α and β subunits catalyze different reactions and indole is the channeled intermediate connecting the two reactions (α and β)[47-49]. The β subunit belongs to PLP Fold Type II. PDB ID 1SFT[37] is the structure of alanine racemase from *Bacillus stearothermophilus*, which is the representative of fold type III and functions as dimer (chain A colored in green, chain B colored in cyan). Each monomer contains two domains: and N-terminal eight-stranded α/β barrel and a C-terminal mostly β-strand domain. PDB ID 1DAA[50] is the structure of D-amino acid aminotransferase from *Baccilus* sp. YM-1, which is the representative of fold type IV and functions as a dimer (chain A colored in green, chain B colored in cyan). Each monomer contains two domains that are very different from the structures of the L-amino acid aminotransferases. PDB ID 1GPA[51] is the structure of glycogen phosphorylase from *Oryctolagus cuniculus*, which belongs to Fold Type V and functions as dimer (chain A colored in green, chain B colored in cyan). PDB ID 1XRS[52] is the structure of lysine 5,6-aminomutase from *Clostridium sticklandii*, which is the representative of fold type VI and functions as an $\alpha_2\beta_2$ tetramer (α subunits colored in green, β subunits colored in cyan). The larger α subunit consists of a PLP binding α/β barrel surrounded with additional α helices and β strands. The smaller β subunit contains two domains: the dimerization domain and the adenosylcobalamin (B_{12})-binding Rossmann domain. PDB ID 2A5H[53] is the structure of lysine-2,3-aminomutase from *Clostridium subterminale*, which belongs to Fold Type VII and functions as a tetramer (chain A colored in green, chain B colored in cyan, chain C colored in magenta, and chain D colored in yellow). This enzyme is S-adenosylmethionine-, $[4Fe-4S]^{2+}$-, and PLP-dependent. S-adenosylmethionine and PLP are shown as sticks; $[4Fe-4S]^{2+}$ is shown in sphere.

compared to the Type I ATases. Fold Type V contains just one enzyme: glycogen phosphorylase (EC 2.4.1.1), which uses the phosphate group of PLP for catalysis[54]. The two new classifications obtained from the hidden markov model analysis by Percudani and Peracchi are designated Type VI and Type VII. Type VI includes the L- and D-lysine 5,6-aminomutases (EC 5.4.3.3 and 5.4.3.4; e.g. PDB ID: 1XRS[55]) and D-ornithine 4,5-

aminomutase (EC 5.4.3.5). Type VII contains the lysine 2,3-aminomutase family enzymes (*e.g.* PDB ID: 2A5H [56]).

4.4 Aminotransferases (ATase)

The PLP-dependent aminotransferases catalyze transamination reactions in which the α amino group of an amino acid is transferred onto the α-carbon of a 2-oxo acid, forming the 2-oxo acid form of the amino acid and the amino acid form of the 2-oxo-acid. The example of the thoroughly studied asparate aminotransferase-catalyzed reaction is shown in Figure 4-4A. Aspartate aminotransferase (AATase) has a large domain, a small domain and an N-terminal extension (Figure 4-4B). It functions as a dimer with the active site residues contributed by both domains and by both chains of the dimer, as discussed above. In L-amino acid aminotransferases, the catalytic lysine residue is located on the *si* face of C4' of the cofactor and the *re* face is accessible by solvent or substrate.
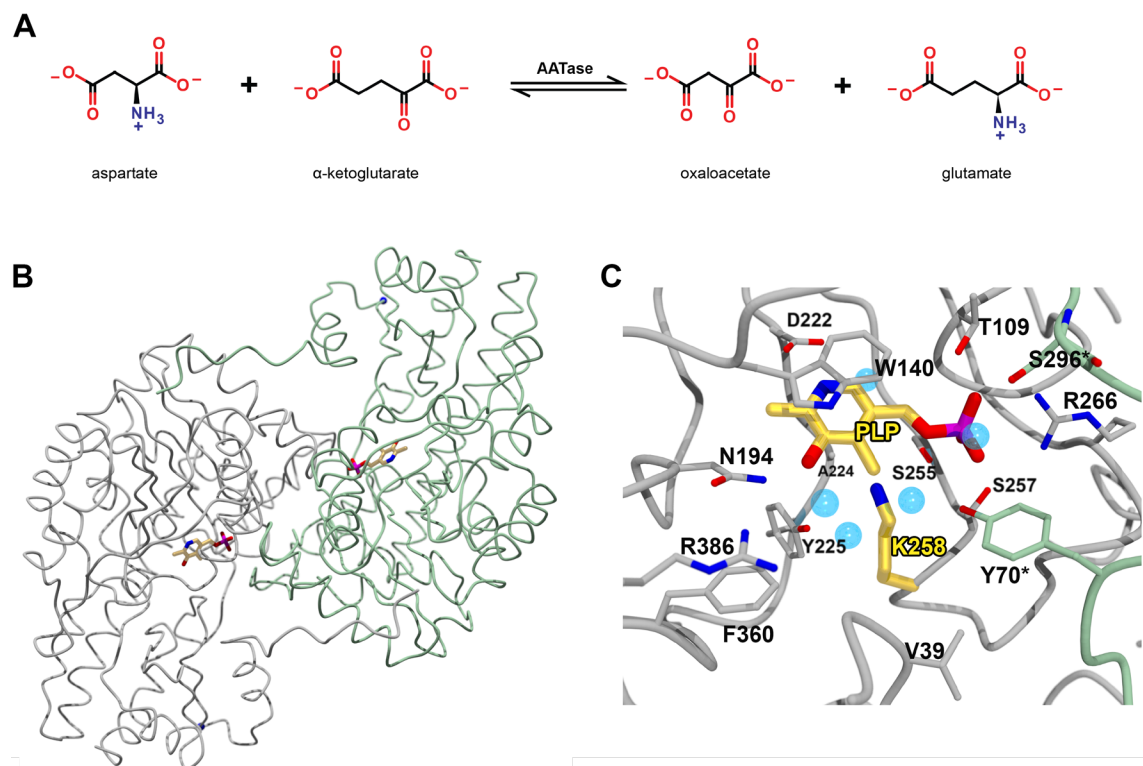
Figure 4-4: The reaction catalyzed by aspartate aminotransferase (A). Ribbon representation of the aspartate aminotransferase dimer (PDB ID: 1ARS[57]). Chain A is colored grey, chain B is colored light green. The active site of AATase shows the binding mode of PLP in this and similar enzymes (C). The catalytic K258 and PLP are shown as sticks with yellow carbon atoms. The asterisk on atom labels denotes that the residue is contributed by the other chain of the dimer.

4.4.1 The AATase active site

As shown in Figure 4-4C, the catalytic K258 of the *E. coli* AATase is covalently bound to the PLP as the internal aldimine. The negatively charged D222 stabilizes the pyridine N atom of PLP as the pyridinium, which allows the ring to accept electrons from the substrate stabilizing the carbanion that forms during catalysis. The C3' hydroxyl group of PLP makes

a hydrogen bonding interaction with N194 and Y225, both highly conserved among the Type I AATases. The phosphate group of the cofactor interacts with R266, T109, S257, Y70* (from the other protomer of the dimer); S296* interacts with the phosphate group through an intervening water molecule. The pyridine ring of PLP is clamped between W140 above and A224 below. W140 is important for retaining PLP in the active site. Mutating W140 to His in the *E. coli* AATase resulted in a 50-fold increase in the dissociation rate constant for PLP[58]. Also, the angled, unstacked imidazole ring of this W140H variant facilitated water access to the *re* face of the substrate Cα, resulting in a faster rate of racemization and decreased aminotransferase activity[58].

4.4.2 Conformational changes associated with catalysis

Because the substrates of AATase are aspartate and α-ketoglutarate, the elimination and replacement reactions cannot happen. In the structure of the *E. coli* AATase with the inhibitor 2-methyl-L-aspartate (PDB ID: 1ART[59]), R386 coordinates the α-carboxylate group of the inhibitor and holds in in the plane of the PLP ring, which discourages decarboxylation. After the deprotonation of Cα, the resulting quinonoid intermediate is capable of undergoing either transamination or racemization. Kochhar and Christen found that human mitochondrial AATase that had been truncated by limited proteolysis with trypsin (residues 27/32-410) had the same racemization activity as the intact protein, but much lower transamination activity[60]. In addition, increasing organic solvent (methanol) proportion in the reaction resulted in decreased racemization but the same transamination activity of AATase[60]. Thus, it appears that AATases have evolved a strategy to discourage

the racemization reaction by limiting the access of water to the catalytic center. This strategy involves closing the active site with N-terminal extension. This excludes water from the active site, preventing water molecules from pronating the *re* side of the quinonoid intermediate.

The small domains of many Type I aminotransferases also shifts towards the large domain upon substrate binding, forming a "closed" conformation[57, 61, 62]. In *E. coli* AATase, this movement of the small domain creates a repulsive interaction between G38 and Y225, which may reduce the pK_a of the aldimine and, thus, enhance catalysis[61]. The lower imine pK_a of the internal aldimine is mainly due to the 90 ° torsion angle of the C4-C4' bond (C3-C4-C4'-Nζ) and the repulsive interaction between G38 and Y225 that destabilizes the unprotonated aldimine in the Michaelis complex [61, 63].

4.4.3 Dual substrate recognition

As mentioned above, the aminotransferase reaction involves two different substrates being transformed, in succession, into two products. In many cases, these two substrates have very different side chains, such as different charges (*e.g.* alanine vs α-ketoglutarate), inverse absosolute configuration (*e.g.* D-p-hydroxyphenylglycine and L-glutamate), and/or different size side chains (*e.g.* aspartate and α-ketoglutarate). Aminotransferases have evolved several ways to solve the problem of having such different substrates binding in the same active site in the same orientation relative to the same PLP cofactor. These

include the "glutamate switch," the "arginine switch," and remodeling of hydrogen bonding networks[33, 64].

The "glutamine switch" works as shown in Figure 4-5. When the substrate, lysine, forms the external aldimine with PLP (PDB ID: 2CJD[65]) in the lysine ε-aminotransferase from *Mycobacterium tuberculosis*, the carboxylate group of E243 neutralizes the guanidinium group of R422. This prevents repulsion due to the juxtaposition of like charges. However, in the structure of the lysine ε-aminotransferase with α-ketoglutarate bound (PDB ID: 2CJH[65]), E243 rotates out of the active site, which allows the interaction between R422 and the α-carboxylate group of α-ketoglutarate. The "arginine switch" works in a similar way, as it is also based on the flexibility of active site residues as demonstrated in the active site of GABA aminotransferase. Comparing the structures of GABA aminotransferase in complexes with either an inactivator CPP-115 with one carboxylate group (similar to GABA) (PDB ID: 4ZSY [66]) or the metabolite of CPP-115 with two carboxylate groups that is similar to glutamate (PDB ID: 4ZSW [66]), R445 that interacts with the one carboxylate group of the CCP-115 metabolite, switches out and interacts with E270 in the CPP-115 bound structure [66, 67].
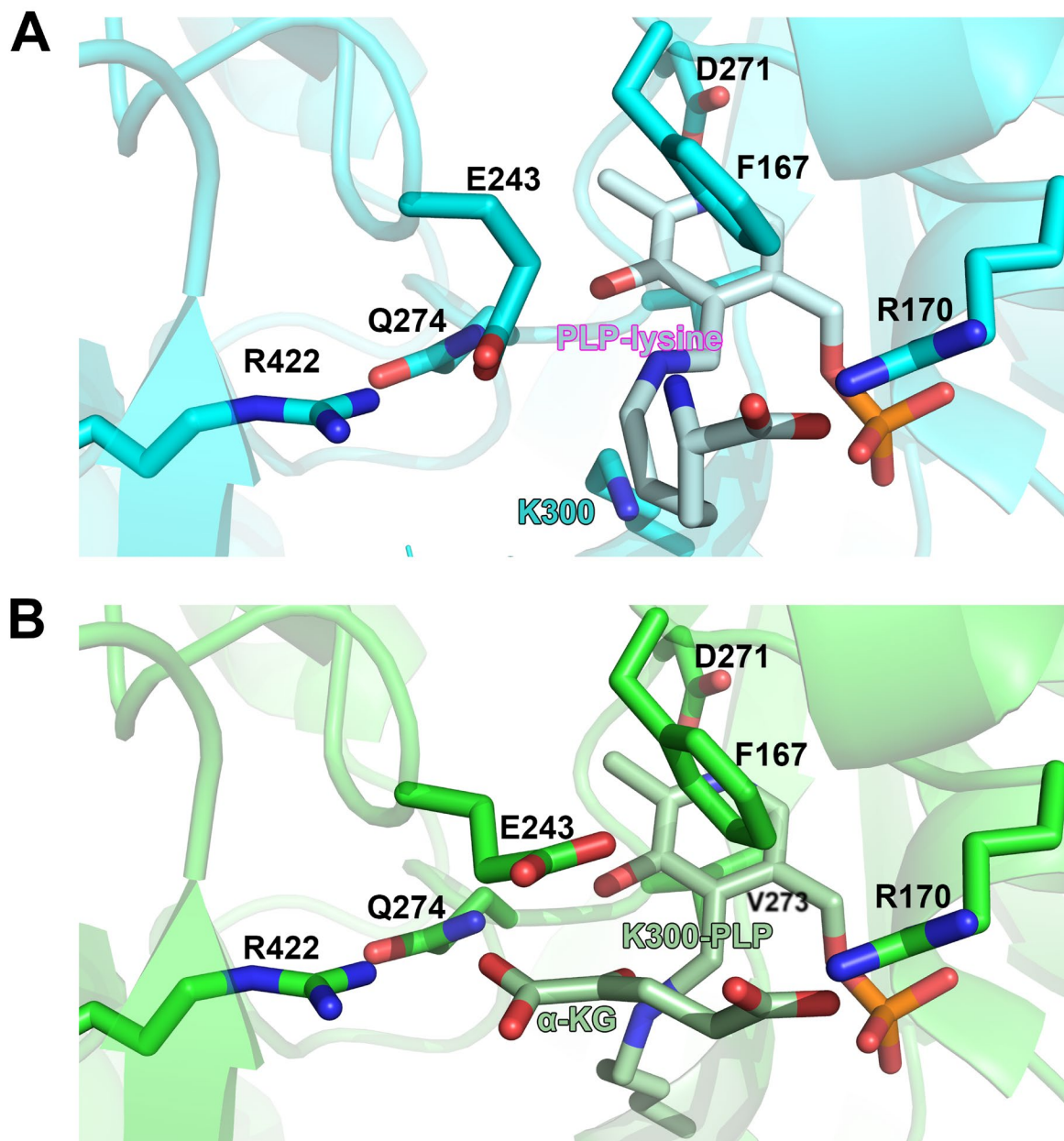
Figure 4-5: Dual substrate recognition in lysine ε-aminotransferase from *Mycobacterium tuberculosis*. When lysine is bound in the active site of lysine ε-aminotransferase as the external aldimine (A), E243 forms a salt bridge with R422. When α-ketoglutarate is bound in the active site (B), E243 rotates away from R422, which allows R422 to interact with the α-carboxylate group of α-ketoglutarate.

4.5 L-3,4-dihydroxyphenylalanine (L-DOPA) decarboxylase

4.5.1 Mechanisms of decarboxylation and oxidative deamination

Some PLP dependent enzymes were found to react with O_2, such as 3,4-dihydroxyphenylalanine (DOPA) decarboxylase/L-aromatic amino acid decarboxylase [68], and so-called "paracatalytic" reactions [69] with O_2 have been described. DOPA decarboxylase catalyzes the conversion of L-DOPA (compound **8**) and L-5-hydroxy tryptophan into two important neurotransmitters, dopamine (compound **9**) and serotonin. The mechanism is shown in Scheme 4-5A. Upon substrate binding, the internal aldimine is converted into external aldimine. Due to the orientation of the substrate carboxyl group perpendicular to the PLP ring, decarboxylation happens. The subsequent quinonoid intermediate is protonated at Cα position, giving the corresponding amine product (**9**). Mutagenesis studies of Y332 demonstrated that this residue is probably the proton donor in the step of quinonoid protonation[70]. Notably, the mutant Y332F has a decarboxylation-dependent oxidative deaminase activity, rather than decarboxylase[70]. In other words, Y332F reacts with L-amino acids and dioxygen to generate the corresponding aromatic aldehydes, CO_2 and ammonia.

DOPA decarboxylase also catalyzes the oxidative deamination of serotonin and dopamine[71, 72]. The proposed oxidative mechanism of **9** is shown in Scheme 4-5B. The internal aldimine is converted into the external aldimine with dopamine, and after deprotonating on Cα, the quinonoid intermediate forms. Based on its close proximity to the substrate Cα and mutagenesis studies, T246 is thought to function as a nucleophile to abstract the α-proton [73]. In the presence of O_2, the quinonoid intermediate reacts with O_2

171

and forms a superoxide anion and semiquinone, followed by a hydroperoxy-PLP intermediate. After several steps, the final aldehyde product, H_2O_2, and ammonia form. In the absence of O_2, the quinonoid intermediate would equilibrate with a ketimine[73].
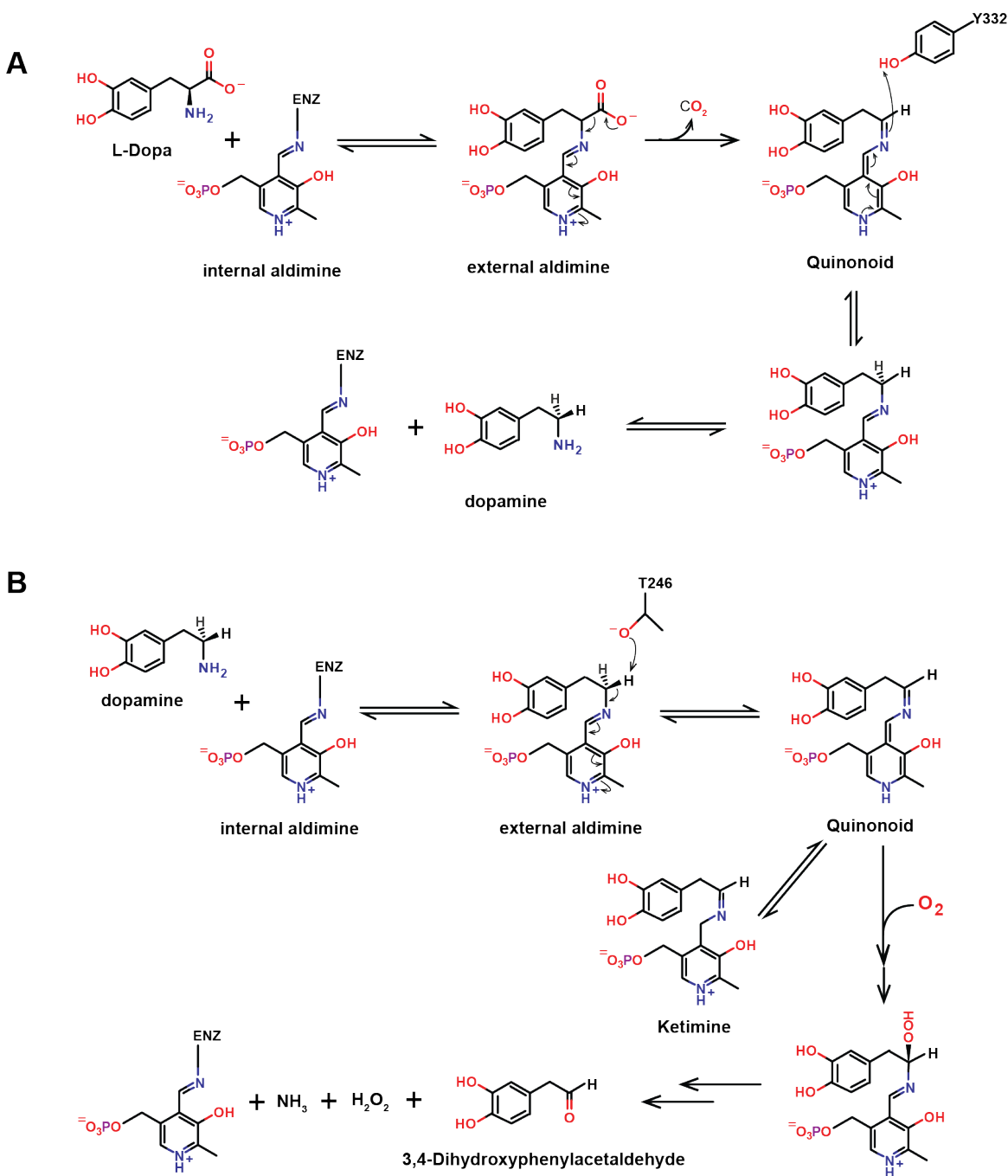
Scheme 4-5: Decarboxylation (A) and oxidation deamination (B) by DOPA decarboxylase.

4.5.2 Structure of human DOPA decarboxylase

DOPA decarboxylase belongs to group II of Fold Type I, which contains the L-amino acid decarboxylases. The group II of L-amino acid decarboxylases include aromatic amino acid decarboxylase, histidine decarboxylase, and glutamate decarboxylase [74]. They share a similar active site conformation and many conserved features. Binding of PLP into the apoenzyme is thought to be the essential regulator of activity [75]. Upon PLP binding, DOPA decarboxylase undergoes a large conformational change into a more compact and less solvent accessible conformation [1].

As shown in Figure 4-7, human DOPA decarboxylase is a homodimer with three domains: an N-terminal domain composed of three α-helices that contact with the other protomer, a large domain with a central seven-stranded β-sheet surrounded by α helices, and a C-terminal domain with a 4-stranded β-sheet covered by α helices on one side. The apoenzyme adopts an open conformation (PDB ID: 3RBL [1]); while after PLP binds in the two active sites, it forms a closed conformation (PDB ID: 1JS6[2]). In the apoenzyme structure, Y274 interacts with H302. After PLP binding, this movement (from open to closed conformation) makes the aromatic stacking interactions among W304, F80 and Y274 (Figure 4-7B). In the active site, the catalytic K303 covalently binds to PLP as the internal aldimine. H192 and A273 are stacked above and below the PLP ring. H192 is also involved in the catalysis of oxidative deamination. In the substrate analog-bound structure (PDB ID: 1JS3[2]), the
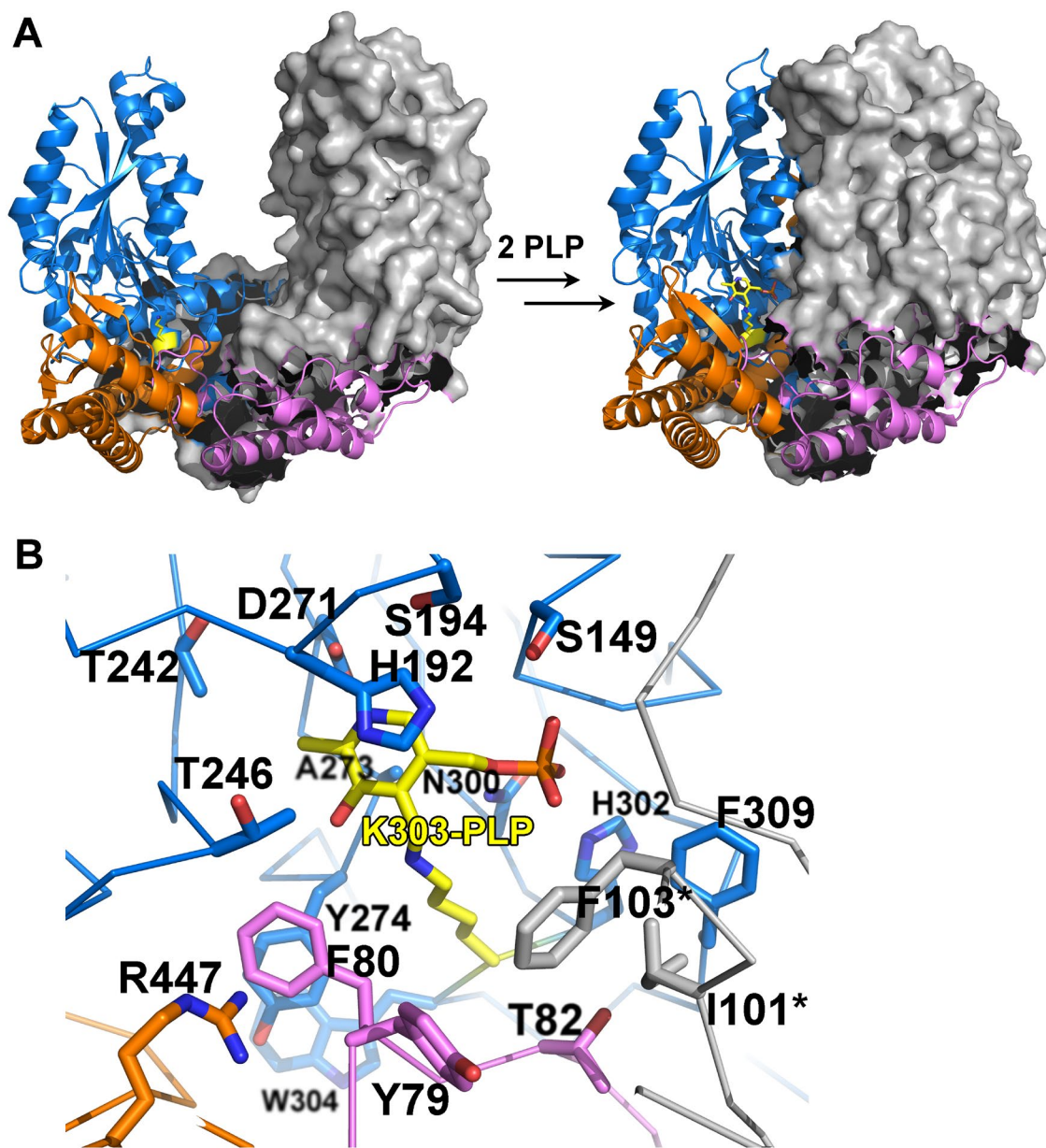
Figure 4-7: (A) The structure of DOPA decarboxylase in an open (PDB ID: 3RBL [1]) and closed (PDB ID: 1JS6[2]) conformation. The domain organization of chain A is highlighted with different colors (N-terminal domain colored in purple, large domain colored in blue, and C-terminal domain colored in orange). Chain B colored in grey. (B) Conserved residues in human DOPA decarboxylase are shown in stick.

iminic nitrogen of H192 is 4.8 Å away from the OH group of T246, which facilitates the deprotonation of T246[76]. As mentioned above, deprotonated T246 would abstract the Cα hydrogen and trigger the formation of the quinonoid intermediate (Scheme 4-5B). D271 makes a salt bridge interaction with the protonated pyridine nitrogen of PLP. Several residues, including S194, S149, H302, F309, are involved in interacting with phosphate group of PLP. I101* and F103* (from the other chain), Y79, F80, T82 are involved in substrate binding.

The sequence of SwMppP suggests that it belongs to the class I/II ATase group of the Fold Type I PLP-dependent enzymes. The sequence identity between SwMppP and a Fold Type I ATase (PDB ID 1ARS[57]) or DOPA decarboxylase (PDB ID 3RBL[1]) is quite low (20% and 17%, respectively). In addition, many of the key active site residues of both the ATases and DOPA decarboxylase are different in SwMppP, suggesting that SwMppP probably has a different function. Two other genes in the L-End biosynthesis gene cluster, MppR and MppQ, are involved in the cyclization of **6** into **7** and the transamination of **7** into L-End (**1**), respectively (Figure 4-2). Hence, we speculate that SwMppP can only be a novel PLP-dependent hydroxylase/deaminase to catalyze the oxidation of L-Arg (**5**) into its γ-hydroxy, 2-oxo derivative **6**. To find out the relationship between the function and structure, we used biophysical methods, as well as enzyme kinetics, to characterize SwMppP. Our findings are detailed in Chapter 5.

4.6 References

[1] Giardina, G., Montioli, R., Gianni, S., Cellini, B., Paiardini, A., Voltattorni, C. B., and Cutruzzola, F. (2011) Open conformation of human DOPA decarboxylase reveals the mechanism of PLP addition to Group II decarboxylases, *Proc Natl Acad Sci U S A 108*, 20514-20519.

[2] Burkhard, P., Dominici, P., Borri-Voltattorni, C., Jansonius, J. N., and Malashkevich, V. N. (2001) Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase, *Nature structural biology 8*, 963-967.

[3] Higashide, E., Hatano, K., Shibata, M., and Nakazawa, K. (1968) Enduracidin, a new antibiotic. I. Streptomyces fungicidicus No. B5477, an enduracidin producing organism, *The Journal of antibiotics 21*, 126-137.

[4] Singh, M. P., Petersen, P. J., Weiss, W. J., Janso, J. E., Luckman, S. W., Lenoy, E. B., Bradford, P. A., Testa, R. T., and Greenstein, M. (2003) Mannopeptimycins, new cyclic glycopeptide antibiotics produced by Streptomyces hygroscopicus LL-AC98: antibacterial and mechanistic activities, *Antimicrobial agents and chemotherapy 47*, 62-69.

[5] Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., Mueller, A., Schäberle, T. F., Hughes, D. E., Epstein, S., Jones, M., Lazarides, L., Steadman, V. A., Cohen, D. R., Felix, C. R., Fetterman, K. A., Millett, W. P., Nitti, A. G., Zullo, A. M., Chen, C., and Lewis, K. (2015) A new antibiotic kills pathogens without detectable resistance, *Nature 517*, 455.

[6] Castiglione, F., Marazzi, A., Meli, M., and Colombo, G. (2005) Structure elucidation and 3D solution conformation of the antibiotic enduracidin determined by NMR spectroscopy and molecular dynamics, *Magnetic Resonance in Chemistry 43*, 603-610.

[7] Fang, X., Tiyanont, K., Zhang, Y., Wanner, J., Boger, D., and Walker, S. (2006) The mechanism of action of ramoplanin and enduracidin, *Molecular bioSystems 2*, 69-76.

[8] Cudic, P., Behenna, D. C., Kranz, J. K., Kruger, R. G., Wand, A. J., Veklich, Y. I., Weisel, J. W., and McCafferty, D. G. Functional Analysis of the Lipoglycodepsipeptide Antibiotic Ramoplanin, *Chemistry & biology 9*, 897-906.

[9] Ruzin, A., Singh, G., Severin, A., Yang, Y., Dushin, R. G., Sutherland, A. G., Minnick, A., Greenstein, M., May, M. K., Shlaes, D. M., and Bradford, P. A. (2004) Mechanism of action of the mannopeptimycins, a novel class of glycopeptide antibiotics active against vancomycin-resistant gram-positive bacteria, *Antimicrob. Agents Chemother. 48*, 728-738.

[10] Wang, B., Liu, Y., Jiao, R., Feng, Y., Li, Q., Chen, C., Liu, L., He, G., and Chen, G. (2016) Total Synthesis of Mannopeptimycins alpha and beta, *Journal of the American Chemical Society 138*, 3926-3932.

[11] Sum, P.-E., How, D., Torres, N., Petersen, P. J., Lenoy, E. B., Weiss, W. J., and Mansour, T. S. (2003) Novel ether derivatives of mannopeptimycin glycopeptide antibiotic, *Bioorganic & Medicinal Chemistry Letters 13*, 1151-1155.

[12] Sum, P.-E., How, D., Torres, N., Petersen, P. J., Ashcroft, J., Graziani, E. I., Koehn, F. E., and Mansour, T. S. (2003) Synthesis and evaluation of ether and halogenated derivatives

of mannopeptimycin glycopeptide antibiotics, *Bioorganic & Medicinal Chemistry Letters 13*, 2805-2808.

[13] He, H., Wang, T.-Z., Dushin, R. G., Feng, X., Shen, B., Ashcroft, J. S., Koehn, F. E., and Carter, G. T. (2004) Structural determination of mannopeptimycin cyclic acetals, *Tetrahedron Letters 45*, 5889-5893.

[14] Dushin, R. G., Wang, T.-Z., Sum, P.-E., He, H., Sutherland, A. G., Ashcroft, J. S., Graziani, E. I., Koehn, F. E., Bradford, P. A., Petersen, P. J., Wheless, K. L., How, D., Torres, N., Lenoy, E. B., Weiss, W. J., Lang, S. A., Projan, S. J., Shlaes, D. M., and Mansour, T. S. (2004) Hydrophobic Acetal and Ketal Derivatives of Mannopeptimycin-α and Desmethylhexahydromannopeptimycin-α: Semisynthetic Glycopeptides with Potent Activity Against Gram-Positive Bacteria, *Journal of Medicinal Chemistry 47*, 3487-3490.

[15] Petersen, P. J., Wang, T. Z., Dushin, R. G., and Bradford, P. A. (2004) Comparative in vitro activities of AC98-6446, a novel semisynthetic glycopeptide derivative of the natural product mannopeptimycin alpha, and other antimicrobial agents against gram-positive clinical isolates, *Antimicrobial agents and chemotherapy 48*, 739-746.

[16] Sum, P.-E., How, D., Torres, N., Newman, H., Petersen, P. J., and Mansour, T. S. (2003) Synthesis and activity of novel benzoxazole derivatives of mannopeptimycin glycopeptide antibiotics, *Bioorganic & Medicinal Chemistry Letters 13*, 2607-2610.

[17] He, H., Shen, B., Petersen, P. J., Weiss, W. J., Yang, H. Y., Wang, T.-Z., Dushin, R. G., Koehn, F. E., and Carter, G. T. (2004) Mannopeptimycin esters and carbonates, potent antibiotic agents against drug-resistant bacteria, *Bioorganic & Medicinal Chemistry Letters 14*, 279-282.

[18] Tsuji, S., Kusumoto, S., and Shiba, T. (1975) SYNTHESIS OF ENDURACIDIDINE, A COMPONENT AMINO ACID OF ANTIBIOTIC ENDURACIDIN, *Chemistry Letters 4*, 1281-1284.

[19] Sanière, L., Leman, L. c., Bourguignon, J.-J., Dauban, P., and Dodd, R. H. (2004) Iminoiodane mediated aziridination of α-allylglycine: access to a novel rigid arginine derivative and to the natural amino acid enduracididine, *Tetrahedron 60*, 5889-5897.

[20] Schwörer, C. J., and Oberthür, M. (2009) Synthesis of Highly Functionalized Amino Acids: An Expedient Access to L- and D-β-Hydroxyenduracididine Derivatives, *European Journal of Organic Chemistry 2009*, 6129-6139.

[21] Fischer, S. N., Schwörer, C. J., and Oberthür, M. (2014) A Short Synthesis of Partially Protected l- and d-β-Hydroxyenduracididines and a Structurally Simplified Dipeptide Analogue, *Synthesis 46*, 2234-2240.

[22] Lin, C.-K., Yun, W.-Y., Lin, L.-T., and Cheng, W.-C. (2016) A concise approach to the synthesis of the unique N-mannosyl d-[small beta]-hydroxyenduracididine moiety in the mannopeptimycin series of natural products, *Organic & Biomolecular Chemistry 14*, 4054-4060.

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

[24] Yin, X., and Zabriskie, T. M. (2006) The enduracidin biosynthetic gene cluster from Streptomyces fungicidicus, *Microbiology (Reading, England) 152*, 2969-2983.

[25] Hatano, K., Nogami, I., Higashide, E., and Kishi, T. (1984) Biosynthesis of Enduracidin: Origin of Enduracididine and Other Amino Acids, *Agricultural and Biological Chemistry 48*, 1503-1508.

[26] Yin, X., and Zabriskie, T. M. (2004) VioC is a non-heme iron, alpha-ketoglutarate-dependent oxygenase that catalyzes the formation of 3S-hydroxy-L-arginine during viomycin biosynthesis, *Chembiochem : a European journal of chemical biology 5*, 1274-1277.

[27] Ju, J., Ozanick, S. G., Shen, B., and Thomas, M. G. (2004) Conversion of (2S)-Arginine to (2S,3R)-Capreomycidine by VioC and VioD from the Viomycin Biosynthetic Pathway of Streptomyces sp. Strain ATCC11861, *Chembiochem : a European journal of chemical biology 5*, 1281-1285.

[28] Yin, X., McPhail, K. L., Kim, K.-j., and Zabriskie, T. M. (2004) Formation of the Nonproteinogenic Amino Acid 2S,3R-Capreomycidine by VioD from the Viomycin Biosynthesis Pathway, *Chembiochem : a European journal of chemical biology 5*, 1278-1281.

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

[30] Haltli, B., Tan, Y., Magarvey, N. A., Wagenaar, M., Yin, X., Greenstein, M., Hucul, J. A., and Zabriskie, T. M. (2005) Investigating beta-hydroxyenduracididine formation in the biosynthesis of the mannopeptimycins, *Chemistry & biology 12*, 1163-1168.

[31] Bodanszky, M., and Perlman, D. (1968) Origin of D-Amino-acids in Microbial Peptides: Rule of [alpha]-Epimerization, *Nature 218*, 291-292.

[32] Araujo, W. L., Marcon, J., Maccheroni, W., Jr., Van Elsas, J. D., Van Vuurde, J. W., and Azevedo, J. L. (2002) Diversity of endophytic bacterial populations and their interaction with Xylella fastidiosa in citrus plants, *Applied and environmental microbiology 68*, 4906-4914.

[33] Eliot, A. C., and Kirsch, J. F. (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations, *Annu. Rev. Biochem 73*, 383-415.

[34] Toney, M. D. (2005) Reaction specificity in pyridoxal phosphate enzymes, *Archives of Biochemistry and Biophysics 433*, 279-287.

[35] Toney, M. D. (2011) Controlling reaction specificity in pyridoxal phosphate enzymes, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1814*, 1407-1418.

[36] McPhalen, C. A., Vincent, M. G., and Jansonius, J. N. (1992) X-ray structure refinement and comparison of three forms of mitochondrial aspartate aminotransferase, *Journal of molecular biology 225*, 495-517.

[37] Shaw, J. P., Petsko Ga Fau - Ringe, D., and Ringe, D. Determination of the structure of alanine racemase from Bacillus stearothermophilus at 1.9-A resolution.

[38] Percudani, R., and Peracchi, A. (2009) The B6 database: a tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families, *BMC Bioinformatics 10*, 273.

[39] Grishin, N. V., Phillips, M. A., and Goldsmith, E. J. (1995) Modeling of the spatial structure of eukaryotic ornithine decarboxylases, *Protein Science 4*, 1291-1304.

[40] Schneider, G., Kack, H., and Lindqvist, Y. (2000) The manifold of vitamin B6 dependent enzymes, *Structure (London, England : 1993) 8*, R1-6.

[41] Mehta, P. K., and Christen, P. (2000) The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes, *Advances in enzymology and related areas of molecular biology 74*, 129-184.

[42] Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., and Davies, D. R. (1988) Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium, *The Journal of biological chemistry 263*, 17857-17871.

[43] Madison, J. T., and Thompson, J. F. (1976) Threonine synthetase from higher plants: stimulation by S-adenosylmethionine and inhibition by cysteine, *Biochemical and biophysical research communications 71*, 684-691.

[44] Wessel, P. M., Graciet, E., Douce, R., and Dumas, R. (2000) Evidence for two distinct effector-binding sites in threonine deaminase by site-directed mutagenesis, kinetic, and binding experiments, *Biochemistry 39*, 15136-15143.

[45] McPhalen, C. A., Vincent Mg Fau - Jansonius, J. N., and Jansonius, J. N. X-ray structure refinement and comparison of three forms of mitochondrial aspartate aminotransferase.

[46] Rhee, S., Parris Kd Fau - Ahmed, S. A., Ahmed Sa Fau - Miles, E. W., Miles Ew Fau - Davies, D. R., and Davies, D. R. Exchange of K+ or Cs+ for Na+ induces local and long-range changes in the three-dimensional structure of the tryptophan synthase alpha2beta2 complex.

[47] Dunn, M. F., Aguilar, V., Brzovic, P., Drewe, W. F., Houben, K. F., Leja, C. A., and Roy, M. (1990) The tryptophan synthase bienzyme complex transfers indole between the .alpha.- and .beta.-sites via a 25-30 .ANG. long tunnel, *Biochemistry 29*, 8598-8607.

[48] Hyde, C. C., and Miles, E. W. (1990) The Tryptophan Synthase Multienzyme Complex: Exploring Structure-Function Relationships with X-Ray Crystallography and Mutagenesis, *Nat Biotech 8*, 27-32.

[49] Dunn, M. F. (2012) Allosteric Regulation of Substrate Channeling and Catalysis in the Tryptophan Synthase Bienzyme Complex, *Archives of biochemistry and biophysics 519*, 154-166.

[50] Sugio, S., Petsko Ga Fau - Manning, J. M., Manning Jm Fau - Soda, K., Soda K Fau - Ringe, D., and Ringe, D. Crystal structure of a D-amino acid aminotransferase: how the protein controls stereoselectivity.

[51] Barford, D., Hu Sh Fau - Johnson, L. N., and Johnson, L. N. Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP.

[52] Berkovitch, F., Behshad E Fau - Tang, K.-H., Tang Kh Fau - Enns, E. A., Enns Ea Fau - Frey, P. A., Frey Pa Fau - Drennan, C. L., and Drennan, C. L. A locking mechanism preventing radical damage in the absence of substrate, as revealed by the x-ray structure of lysine 5,6-aminomutase.

[53] Lepore, B. W., Ruzicka Fj Fau - Frey, P. A., Frey Pa Fau - Ringe, D., and Ringe, D. The x-ray crystal structure of lysine-2,3-aminomutase from Clostridium subterminale.

[54] Fletterick, R. J., and Sprang, S. R. (1982) Glycogen phosphorylase structures and function, *Accounts of Chemical Research 15*, 361-369.

[55] Berkovitch, F., Behshad, E., Tang, K. H., Enns, E. A., Frey, P. A., and Drennan, C. L. (2004) A locking mechanism preventing radical damage in the absence of substrate, as

revealed by the x-ray structure of lysine 5,6-aminomutase, *Proc Natl Acad Sci U S A 101*, 15870-15875.

[56] Lepore, B. W., Ruzicka, F. J., Frey, P. A., and Ringe, D. (2005) The x-ray crystal structure of lysine-2,3-aminomutase from Clostridium subterminale, *Proc Natl Acad Sci U S A 102*, 13819-13824.

[57] Okamoto, A., Higuchi, T., Hirotsu, K., Kuramitsu, S., and Kagamiyama, H. (1994) X-ray crystallographic study of pyridoxal 5'-phosphate-type aspartate aminotransferases from Escherichia coli in open and closed form, *Journal of biochemistry 116*, 95-107.

[58] Vacca, R. A., Christen, P., Malashkevich, V. N., Jansonius, J. N., and Sandmeier, E. (1995) Substitution of Apolar Residues in the Active Site of Aspartate Aminotransferase by Histidine, *European Journal of Biochemistry 227*, 481-487.

[59] Okamoto, A., Higuchi T Fau - Hirotsu, K., Hirotsu K Fau - Kuramitsu, S., Kuramitsu S Fau - Kagamiyama, H., and Kagamiyama, H. X-ray crystallographic study of pyridoxal 5'-phosphate-type aspartate aminotransferases from Escherichia coli in open and closed form.

[60] Kochhar, S., and Christen, P. (1992) Mechanism of racemization of amino acids by aspartate aminotransferase, *European Journal of Biochemistry 203*, 563-569.

[61] Hayashi, H., Mizuguchi, H., Miyahara, I., Nakajima, Y., Hirotsu, K., and Kagamiyama, H. (2003) Conformational change in aspartate aminotransferase on substrate binding induces strain in the catalytic group and enhances catalysis, *The Journal of biological chemistry 278*, 9481-9488.

[62] McPhalen, C. A., Vincent, M. G., Picot, D., Jansonius, J. N., Lesk, A. M., and Chothia, C. (1992) Domain closure in mitochondrial aspartate aminotransferase, *Journal of molecular biology 227*, 197-213.

[63] Islam, M. M., Hayashi, H., Mizuguchi, H., and Kagamiyama, H. (2000) The substrate activation process in the catalytic reaction of Escherichia coli aromatic amino acid aminotransferase, *Biochemistry 39*, 15418-15428.

[64] Steffen-Munsberg, F., Vickers, C., Kohls, H., Land, H., Mallin, H., Nobili, A., Skalden, L., van den Bergh, T., Joosten, H. J., Berglund, P., Hohne, M., and Bornscheuer, U. T. Bioinformatic analysis of a PLP-dependent enzyme superfamily suitable for biocatalytic applications.

[65] Mani Tripathi, S., and Ramachandran, R. Direct evidence for a glutamate switch necessary for substrate recognition: crystal structures of lysine epsilon-aminotransferase (Rv3290c) from Mycobacterium tuberculosis H37Rv.

[66] Lee, H., Le, H. V., Wu, R., Doud, E., Sanishvili, R., Kellie, J. F., Compton, P. D., Pachaiyappan, B., Liu, D., Kelleher, N. L., and Silverman, R. B. (2015) Mechanism of Inactivation of GABA Aminotransferase by (E)- and (Z)-(1S,3S)-3-Amino-4-fluoromethylenyl-1-cyclopentanoic Acid, *ACS Chemical Biology 10*, 2087-2098.

[67] Lee, H., Doud, E. H., Wu, R., Sanishvili, R., Juncosa, J. I., Liu, D., Kelleher, N. L., and Silverman, R. B. (2015) Mechanism of Inactivation of γ-Aminobutyric Acid Aminotransferase by (1S,3S)-3-Amino-4-difluoromethylene-1-cyclopentanoic Acid (CPP-115), *Journal of the American Chemical Society 137*, 2628-2640.

[68] Bertoldi, M., and Borri Voltattorni, C. (2003) Reaction and substrate specificity of recombinant pig kidney Dopa decarboxylase under aerobic and anaerobic conditions, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1647*, 42-47.

[69] Bunik, V. I., Schloss, J. V., Pinto, J. T., Dudareva, N., and Cooper, A. J. L. (2011) A SURVEY OF OXIDATIVE PARACATALYTIC REACTIONS CATALYZED BY ENZYMES THAT GENERATE CARBANIONIC INTERMEDIATES: IMPLICATIONS FOR ROS PRODUCTION, CANCER ETIOLOGY, AND NEURODEGENERATIVE DISEASES, *Advances in enzymology and related areas of molecular biology 77*, 307-360.

[70] Bertoldi, M., Gonsalvi, M., Contestabile, R., and Voltattorni, C. B. (2002) Mutation of tyrosine 332 to phenylalanine converts dopa decarboxylase into a decarboxylation-dependent oxidative deaminase, *The Journal of biological chemistry 277*, 36357-36362.

[71] Bertoldi, M., Moore, P. S., Maras, B., Dominici, P., and Voltattorni, C. B. (1996) Mechanism-based inactivation of dopa decarboxylase by serotonin, *The Journal of biological chemistry 271*, 23954-23959.

[72] Bertoldi, M., Dominici, P., Moore, P. S., Maras, B., and Voltattorni, C. B. (1998) Reaction of Dopa Decarboxylase with α-Methyldopa Leads to an Oxidative Deamination Producing 3,4-Dihydroxyphenylacetone, an Active Site Directed Affinity Label, *Biochemistry 37*, 6552-6561.

[73] Bertoldi, M., Cellini, B., Montioli, R., and Borri Voltattorni, C. (2008) Insights into the mechanism of oxidative deamination catalyzed by DOPA decarboxylase, *Biochem 47*, 7187-7195.

[74] Hayashi, H. (1995) Pyridoxal Enzymes: Mechanistic Diversity and Uniformity, *The Journal of Biochemistry 118*, 463-473.

[75] Rahman, M. K., Nagatsu, T., Sakurai, T., Hori, S., Abe, M., and Matsuda, M. (1982) Effect of pyridoxal phosphate deficiency on aromatic L-amino acid decarboxylase activity with L-DOPA and L-5-hydroxytryptophan as substrates in rats, *Japanese journal of pharmacology 32*, 803-811.

[76] Bertoldi, M., Castellani, S., and Bori Voltattorni, C. (2001) Mutation of residues in the coenzyme binding pocket of Dopa decarboxylase. Effects on catalytic properties, *European journal of biochemistry / FEBS 268*, 2975-2981.

Chapter 5

Streptomyces wadayamensis MppP is a PLP–Dependent L-Arginine Oxidase in the L-Enduracididine Biosynthetic Pathway

5.1 Background

A UV-visible spectrum of SwMppP with L-Arg (Figure 5-1A) has the peak at 415-420 nm associated with the internal aldimine form of Fold Type I PLP-dependent aminotransferases. In buffers with pH < 7.0, like MES pH 6.7, the spectrum of SwMppP was unchanged over a full hour at room temperature. Thus, 10 mM MES pH 6.7 was chosen as the storage buffer for the enzyme. In high pH buffers, like NaPO$_4$ pH 8.4 or BIS-TRIS propane pH 9.0, the enzyme exhibits much higher activity, even though the enzyme loses both its color and the peak at 415 nm ($t_{1/2} \approx 30$ min). So we use high pH buffers for later kinetics study.

Since enduracididine biosynthesis is thought to begin from L-Arg, and the other two enzymes in the pathway, MppR and MppQ, are thought, based on our own results, to be involved in later steps, we tested SwMppP for activity with L-Arg. Upon addition of 1 mM L-Arg to a solution of 30.0 μM SwMppP in 50 mM Bis-TRIS propane, pH 9.0, the peak at 415-420 nm shifted ~10 nm to the right (425-430 nm), indicative of formation of the external aldimine [1,2] (Figure 5-1A). After approximately 70 seconds, two new species absorbing at 510 and 560 nm began to accumulate. The former reached a maximum and plateaued, while the latter only accumulated for a short time and then quickly decayed (Figure 5-1A,

inset). Agitating the cuvette resulted in the disappearance of the 510 nm peak. After a short time, the 510 nm peak would accumulate again. This observation suggested that SwMppP reacts with molecular oxygen as well as L-Arg. To confirm this putative reaction with O_2, we performed the same experiments in a Clark electrode to monitor the consumption of dioxygen directly. As is evident in Figure 5-1B, dioxygen is consumed upon addition of the enzyme, and the rate of O_2 consumption increases with the concentration of L-Arg.

Figure 5-1: SwMppP reacts with L-Arg. Absorbance spectra recorded on a Hewlett-Packard 8453 diode array spectrophotometer for the reaction of 30 μM SwMppP with 2 mM L-Arg and ambient (~250 μM) O_2 (A). Spectra were recorded every 0.5 s, but only the traces every 10 s are shown here for clarity. The purple line represents the enzyme alone in 50 mM BIS-TRIS propane pH 9.0. The blue gradient traces represent the absorbance changes over 300 s (10 s interval data from lighter to darker blue as time goes). The inset shows the absorbance changes at 510 (green), and 560 nm (orange). The consumption of dioxygen was confirmed by monitoring the reaction in a Hansatech Oxygraph Clark electrode (B). Reactions were initiated by adding SwMppP into solutions containing different concentrations of L-Arg (black arrow).

A

B

SwMppP

- 10 uM L-Arg
- 50 uM L-Arg
- 1000 uM L-Arg
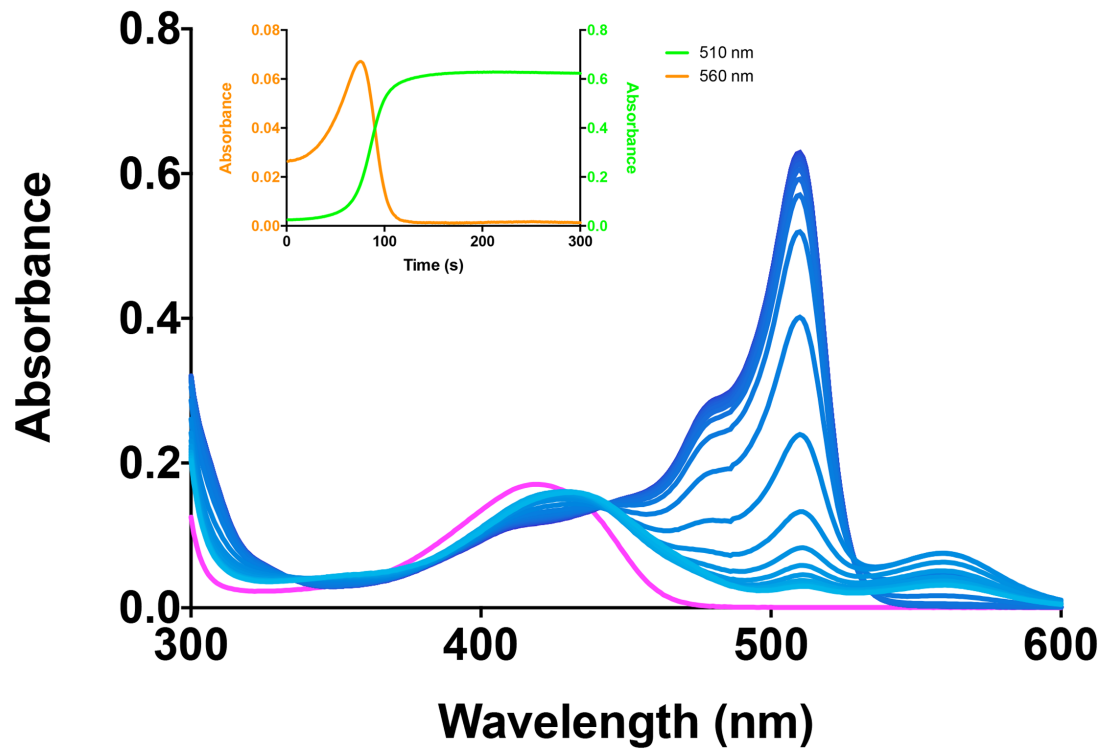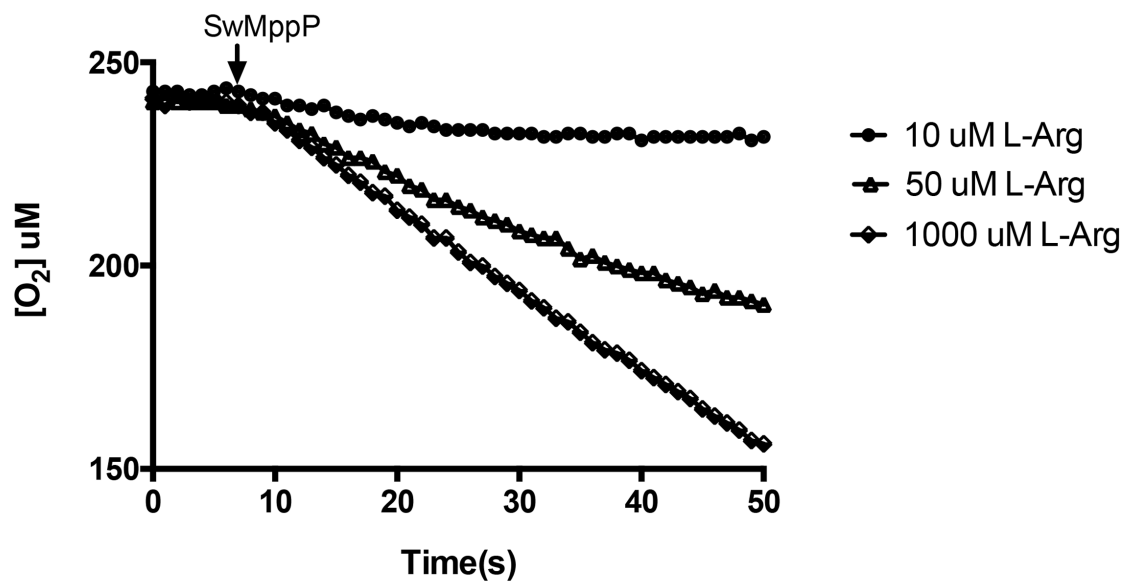
Having established that SwMppP reacts specifically with L-Arg and molecular oxygen, we set about identifying the product(s) of the reaction. Proton NMR analysis of SwMppP reaction mixtures and an L-Arg standard (Figure 5-2), showed that SwMppP converted L-Arg into at least one new product. Preliminary data from ^{1}H COSY, ^{1}H-^{13}C HSQC, and ^{1}H-^{13}C HMBC spectra (Appendix 5-1) showed that the SwMppP-catalyzed reaction between L-Arg and molecular oxygen yields two products: 2-oxo-4(R)-hydroxy-5-guanidinovaleric acid (*i.e.* 4-hydroxy-2-ketoarginine, compound **2** in Figure 5-3) and 2-oxo-5-guanidinovaleric acid (*i.e.* 2-ketoarginine, compound **3** in Figure 5-3).
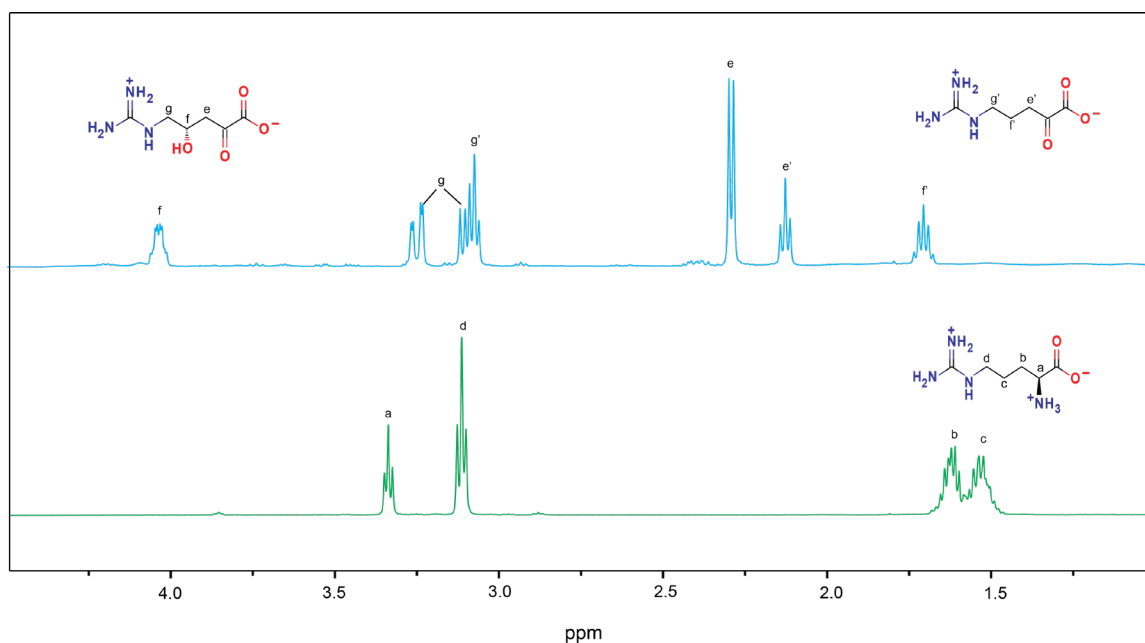
Figure 5-2: Proton NMR results of L-Arg (2.0 mM, bottom) and the reaction of 10 μM SwMppP with 2.0 mM L-Arg (top), both in 20 mM sodium phosphate pH 8.4 buffer.
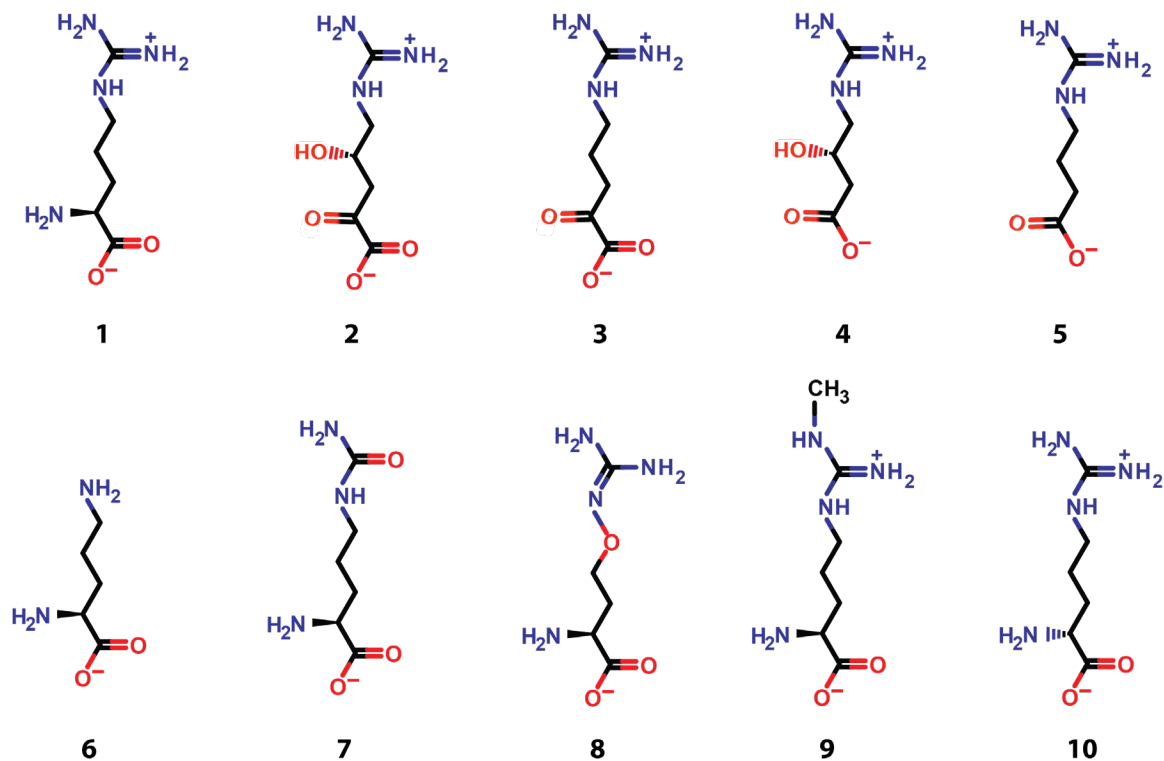
Figure 5-3: Compounds encountered in our studies of SwMppP. **1** is L-arginine. **2** and **3** are the two products of the reaction of SwMppP with L-Arg, 2-oxo-4(R)-hydroxy-5-guanidinovaleric acid (**2**) and 2-oxo-5-guanidinovaleric acid (**3**), respectively. **4** and **5** are the H_2O_2-catalyzed decarboxylation products of **2** and **3** that were observed by mass spectrometry in the absence of catalase, 4-guanidino-(S)-3-hydroxybutyric acid and 4-guanidinobutyric acid, respectively. Compounds **6** -**10** are analogs of L-Arg used to explore the substrate specificity of SwMppP: **6** is L-ornithine; **7** is L-citrulline; **8** is L-canavanine; **9** is N^G-methyl-L-arginine; **10** is D-arginine.

5.2 Methods

5.2.1 Cloning, expression, and purification of *S. wadayamensis* and *Streptomyces globisporus* MppP

The coding sequences of SwMppP (GenBank accession code KDR62041) and *Streptomyces globisporus* MppP (SgMppP; GenBank accession code IH67_RS0105915) were optimized for expression in *E. coli* and synthesized by GenScript Inc. (Piscataway, NJ). The synthetic genes were subcloned into the pE-SUMO$_{kan}$ expression vector (LifeSensors Inc., Malvern, PA) using primers containing BsaI and XbaI restriction sites (Table 5-1). The integrity of the SwMppP coding sequence was confirmed by Sanger sequencing (GeneWiz, Inc). The H29S and F191Y mutants of SwMppP were constructed by the QuickChange mutagenesis method with primers pSUMO-SwPH29A_F/R or pSUMO-SwPF191Y_F/R, respectively, and the pSUMO-SwMppP plasmid as the template. The H29A, F115Y, S190A, D218S, E15A and E15Q mutants of SwMppP were generated with the Q5 site-directed mutagenesis kit (New England Biolabs).

Table 5-1: Primers used for mutagenesis

Names	Primers sequences
pSUMO-SwMppP_F	5′ -GGTCTCAAGGTATGACGACGCAACCGC-3
pSUMO-SwMppP_R	5′ -GCTCTAGATCATTAGCGGGTTTCCAGGAC-3′
pSUMO-SgMppP_F	5′ -GCGGTCTCAAGGTATGAACCGCATTCTGCCGAAAGTC-3′
pSUMO-SgMppP_R	5′ -GCTCTAGATTAGCCTTGACGTTCAGCGGTATC-3′
pSUMO-SwP(H29A)_F	5'-CGCAGACGGCGCCGCCCGTCAGG-3'
pSUMO-SwP(H29A)_R	5'-ATGTTCAGTTCGGAGTTCAGTGC-3'
pSUMO-SwP(H29S)_F	5'-GAACATCGCAGACGGCAGCGCCCGTCAGGCAC-3'
pSUMO-SwP(H29S)_R	5'-GTGCCTGACGGGCGCTGCCGTCTGCGATGTTC-3'
pSUMO-SwP(F115Y)_F	5'-TCATCCGACGTATGATAACATCGCTG-3'
pSUMO-SwP(F115Y)_R	5'-ACCAGTGCCACGCGATCG-3'
pSUMO-SwP(F191Y)_F	5'-CACTGGATACGAGTTATCGTGGTTTCGATG-3'
pSUMO-SwP(F191Y)_R	5'-CATCGAAACCACGATAACTCGTATCCAGTG-3'
pSUMO-SwP(S190A)_F	5'-ACTGGATACGGCTTTTCGTGGTTTCGATGCG-3'
pSUMO-SwP(S190A)_R	5'-GCCAGAACGGTGCCATGT-3'
pSUMO-SwP(D218S)_F	5'-GGTTATTGAATCTACCGGTAAACTGTGGCCGACGC-3'
pSUMO-SwP(D218S)_R	5'-ACCCAGCGACAGCCAGCT-3'
pSUMO-SwP(T12A)_F	5'-CTGAAAGAAAACCTGGCGCAATGGG-3'
pSUMO-SwP(T12A)_R	5'-CCCATTGCGCCAGGTTTTCTTTCAG-3'
pSUMO-SwP(E15A)_F	5'-GACGCAATGGGCATACCTGGCAC-3'
pSUMO-SwP(E15A)_R	5'-AGGTTTTCTTTCAGTTGCGGTTG-3'
pSUMO-SwP(E15Q)_F	5'-GACGCAATGGCAATACCTGGCAC-3'
pSUMO-SwP$_{\Delta1\text{-}22}$_F	5'- TAGGTCTCTAGGTCTGAACATCGCAGACGGC-3'

The His$_6$-SUMO-SwMppP fusion proteins were expressed from *E. coli* BL21 Star (DE3) cells (Invitrogen Inc., Carlsbad, CA) carrying the pE-SUMO-SwMppP plasmid. Cultures were grown in Luria-Bertani medium with 50 μg/mL kanamycin at 37 °C. When the cultures reached an OD$_{600nm}$ of ~1.0, protein expression was induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The temperature was reduced to 25°C, and the cultures were grown overnight while being shaken at 250 rpm. Cells were harvested by centrifugation, resuspended in 5 mL/g of buffer A (25 mM TRIS pH 8.0, 300 mM NaCl, 10 mM imidazole, and 100 μM PLP) supplemented with 0.1 mg/mL DNase I (Worthington Biochemical Corp., Lakewood, NJ). The cells were lysed using a Branson Sonifier S-450 cell disruptor (Branson Ultrasonics Corp., Danbury, CT) for a total of 10 min at 60% amplitude with 30 s pulses separated by 50 s rest periods. The temperature was maintained at or below 4 °C by suspending the steel beaker in an ice bath directly over a spinning stir bar. The lysate was clarified by centrifugation at 39000 x g for 45 min and then applied to a 5 mL HisTrap column (GE Lifesciences, Piscataway, NJ) at a flow rate of 5 mL/min to isolate the His$_6$-SUMO-SwMppP fusion protein. The protein was eluted by a four-step gradient of buffer B (25 mM TRIS (pH 8.0), 300 mM NaCl, and 250 mM imidazole; 5, 15, 50, and 100%). The His$_6$-SUMO-SwMppP fusion protein eluted in the third and fourth steps and was ~90% pure, as judged by Coomassie-stained sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). Peak fractions were pooled and dialyzed overnight against 4 L of 25 mM TRIS pH 8.0, 150 mM NaCl, 100 μM PLP, and ~3 μM SUMO protease (LifeSensors Inc.). The dialysate was passed through the HisTrap column a second time to remove the cleaved His$_6$-SUMO tag as well as the protease. The resulting SwMppP preparation was >95% pure. The protein was desalted using a HiTrap desalting column (GE Lifesciences)

into 10.0 mM MES pH 6.7 and 20 μM PLP and stored at −80°C. All SwMppP mutants were purified by the same protocol as the wild type enzyme. Selenomethionine-labeled SgMppP was expressed from T7 Express Crystal cells (New England Biolabs, Ipswich, MA) carrying the pE-SUMO-SgMppP plasmid. Cells were grown in Selenomethionine Medium Complete (Molecular Dimensions, Newmarket, Suffolk, U.K.) with 50 μg/mL kanamycin. SgMppP was expressed and purified using the same protocols described for SwMppP, except that SgMppP was desalted into three different buffers based on the results of differential scanning fluorimetry experiments with the Slice pH screen (10 mM ADA pH 7.0, 20 μM PLP; 10 mM DL-Malic acid pH 6.0, 20 μM PLP; 10 mM Succinic acid pH 6.0, 20 μM PLP).

5.2.2 UV/Vis spectroscopy of SwMppP and its mutants reacting with L-Arginine and dioxygen

A Hewlett-Packard 8453 diode array spectrophotometer was used for scanning the 700 μL reaction of ~30 μM enzyme (SwMppP or its mutants) with 2 mM L-Arg and ambient (~250 μM) O_2 in 50 mM BisTris propane pH 9.0 buffer. The scan was taken every 0.5 s from 200 nm and 700 nm.

5.2.3 Characterization of MppP reaction products by NMR

Reactions (4.8 mL) were conducted in 50 mL conical tubes with gentle agitation to prevent depletion of dioxygen. All enzymes were stored in 10 mM MES pH 6.7, 20 μM PLP

buffer. To avoid side reactions due to free PLP, the concentration of free PLP was limited by concentrating the proteins to ~600 µM, thus limiting the volume of enzyme (with free PLP) required for the reaction. For wild type SwMppP, ~10 µM SwMppP was added to a solution of 2 mM L-Arg in 20mM sodium phosphate pH 8.4. Due to the reduced catalytic efficiency of the mutants, more enzyme (~65 µM) was used in these reactions. The SwMppP(F115Y) variant was desalted into 20 mM sodium phosphate pH 8.4, 10 µM PLP buffer, since it was not stable in the absence of free PLP. Control reactions contained 2 mM L-Arg in 20 mM sodium phosphate pH 8.4, with the same volume of enzyme storage buffer (*i.e.* no enzyme). All reactions were shaken at 25°C, 200 rpm for 2.5 h and protected from light. At the end of the reaction time, samples were evaporated to dryness using a Centrifan PE Rotary Evaporator (KD Scientific, Holliston, MA) at 40°C for 6 h. The resulting residue was taken up in 600 µL D_2O and centrifuged at 13,000 x g for 10 min to remove any remaining particulates. The supernatants (500 µL) were loaded into NMR tubes for ^1H NMR, ^{13}C NMR, and 2D NMR (^1H-^1H COSY, HSQC, and HMBC) to determine the structures of the reaction products.

5.2.4 Characterization of MppP reaction products by mass spectrometry

Reactions (1.0 mL) contained 2 mM L-Arg in MilliQ H_2O, 10 µM SwMppP in MilliQ H_2O, with or without 53 U of catalase from bovine liver (Sigma-Aldrich, St. Louis, MO). The two negative controls were 2 mM L-Arg in MilliQ H_2O and 2mM L-Arg in water with 53 U catalase. Reactions were run in the dark at 25 °C for 2.5 h with shaking at 200 rpm. All samples were boiled for 10 min and centrifuged at 30,000 x g for 10 min. The supernatants

were diluted 10-fold with MilliQ H_2O and then directly injected into a Shimadzu LCMS-2020.

5.2.5 Ammonium detection by NADH based GDH assay[3]

To detect ammonium formation in the reaction of SwMppP with L-Arg and O_2, I referred to the well-known method summarized in Scheme 5-1. L-glutamate dehydrogenase (GDH) is an NAD(P)H-dependent enzyme that catalyzes the conversion of α-ketoglutarate, NH_4^+ and NADH into glutamate, NAD^+, and H_2O. NADH absorbs UV light at 340 nm, but NAD+ does not. So upon addition of GDH, the absorbance at 340 nm decreases in the presence of NH_4^+.

$$\text{α-ketoglutarate} + NH_4^+ + NADH \xrightarrow{\text{GDH}} \text{Glutamate} + H_2O + NAD^+ + H^+$$
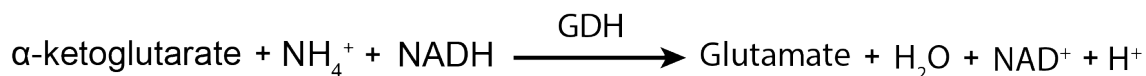
Scheme 5-1: Ammonia assay reaction. The assay detects the absorbance decrease at 340 nm, due to the NADH oxidation.

The 1000 µl reactions were conducted in 50 mM HEPES pH 7.5 buffer and included 20 µM SwMppP, 20 µM or 50 µM L-Arg, 140 µM NADH and 3.5 mM α-ketoglutarate. The reaction components were mixed and incubated for 10 min at room temperature. This was used to blank the spectrophotometer and the absorbance at 340 nm was monitored to ensure a flat baseline. The reaction was initiated by the addition of 10^{-2} units of GDH (Sigma) and the absorbance at 340 nm was recorded for ~15 min, until the absorbance reached a plateau.

5.2.6 Defining the relationship between O_2 usage and H_2O_2 formation

To determine how many O_2 were used to generate H_2O_2, the regeneration of O_2 by catalase was measured in a Hansatech dioxygen electrode. The reaction mixtures contained 25 μM, 50 μM, or 100 μM L-Arg in 50 mM BIS-TRIS propane pH 9.0. When the O_2 concentration was stable, 10 μM SwMppP was added into the solution and O_2 consumption was monitored. When the reaction reached equilibrium (plateau in the dioxygen concentration), 10 μl of 0.5 mg/mL catalase was added to the 1 mL reaction to disproportionate two equivalents of H_2O_2 into one equivalent of O_2 and two equivalents of water.

5.2.7 Reactions with $^{18}O_2$, $H_2^{18}O$

Reactions in the presence of $^{18}O_2$ were conducted in a tonometer, a closed, airtight reaction vessel. The body of the tonometer contained 10 μM SwMppP in 100 mM ammonium bicarbonate pH 9.0 with 50 ng/mL catalase. One short arm of the tonometer contained 20 μL of 500 mM L-Arg, and the other was connected to a 500 mL bottle of $^{18}O_2$ with an air-tight valve. The tonometer was connected to a Schlenk line, which was used to purge the tonometer of oxygen by 45 cycles of vacuum (~5 s) and flushing with argon (45 s). Prior to removing the tonometer from the Schlenk line, the atmosphere was removed with the vacuum pump to enable the labelled dioxygen to fill the tonometer. The tonometer was closed and removed from the Schlenk line. The reaction was started by mixing the

enzyme solution in the tonometer body with the L-Arg in the side arm. Without O_2, the quinonoid I intermediate accumulated, which has a pale pink to orange color. Then the $^{18}O_2$ was allowed to fill the tonometer. The reaction was carried out at room temperature for 2 h. After the reaction, the tonometer was opened and samples were analyzed by ESI-MS.

The 200 µl reaction of SwMppP and L-Arg in $H_2^{18}O$ was conducted in 1.7 mL eppendorf tubes at room temperature overnight. The reaction contained 10 µM SwMppP, 2 mM L-Arg and ~95.4% $H_2^{18}O$. The products of the reaction were identified by ESI-MS.

5.2.8 Crystallization of wild-type and mutant forms of SwMppP, as well as SgMppP

Initial crystallization conditions were identified by screening 15 mg/mL SwMppP and its mutants against the Index HT screen (Hampton Research). After optimization, diffraction-quality SwMppP crystals were obtained by the hanging drop vapor diffusion method from 30% polyethylene glycol monomethyl ether 550 (PEG MME 550), 50 mM $MgCl_2$, and 0.1 M HEPES pH 8.0. Drops contained 2 µL of protein solution at 15 mg/mL and 1 µL of crystallization solution. Crystals formed as plates or rods after 3–4 days and grew to maximum dimensions of ~200 µm × ~50 µm × ~5 µm. Initial crystallization conditions for the homologous enzyme SgMppP (*S. globisporus*) were identified in the same way by screening 15 mg/mL SgMppP against the Index HT screen. Crystals of SeMet-substituted SgMppP in 10 mM ADA pH 7.0, 20 µM PLP were grown from 10% PEG MME 5000, 5% Tacsimate pH 7.0, and 0.1 M HEPES pH 7.5. SgMppP crystallized as small rods (~100 µm × ~10 µm × ~10 µm). Both types of crystals were cryo-protected by sequential soaks in

holding solutions containing 5, 10, or 20% glycerol and flash cooled by plunging into liquid nitrogen.

The structure of SwMppP with D-Arg bound was obtained by transferring crystals into 30 μL drops of the crystallization solution supplemented with 35 mM D-Arg. After being soaked for 5 min at RT, crystals were treated with paratone N and flash-cooled. The structure of SwMppP with L-Arg bound was obtained by transferring crystals into 30 μL drops of the crystallization solution supplemented with 10 mM L-Arg and 20% glycerol. After being soaked for 5 h at RT, crystals were plunged into liquid N2. The structure of SwMppP with **2** bound was obtained by transferring crystals into 30 μL drops of the crystallization solution supplemented with 10 mM 2-keto-Arg (contaminated with ~10% L-Arg) with 20% glycerol. After being soaked for 2 h at RT, crystals were plunged into liquid N2.

SwMppP(H29A) and SwMppP(H29S) crystalized under the same conditions as the wild-type protein. SwMppP(H29A) and SwMppP(H29S) at 12 mg/ml in 10 mM MES pH 6.7, 20 μM PLP were mixed 2:1 with 30% (w/v) PEG MME 550, 0.1 M HEPES pH 7.5, 0.1 M MgCl2. The volume of the resulting drops was 3 μl. Crystals of SwMppP(F115Y) were grown by mixing 1 μl of enzyme at 12 mg/ml in 10 mM MES pH 6.7, 20 μM PLP with 1 μl of 17.5% PEG 3350, 0.04 M citric acid, 0.06 M BIS-TRIS propane, pH 6.4. Crystals of SwMppP(D218S) were grown by mixing 2 μl of enzyme at 12 mg/ml in 10 mM MES pH 6.7, 20 μM PLP with 1 μl of 20% (w/v) PEG 3350, 0.1 M magnesium formate. The well solution contained microseeds derived from poor crystals grown during screening. The structures of Ala and Ser mutants of H29 (SwMppP(H29A/S)) with L-Arg bound were obtained by transferring

crystals into 30 µL drops of the crystallization solution supplemented with 10 mM L-Arg and 20% glycerol. After being soaked for 5 h at RT, crystals were plunged into liquid N_2.

X-ray diffraction data for SeMet SgMppP and native SwMppP were collected at beamline 21-ID-D of the Life Science Collaborative Access Team (LS-CAT) at the Advanced Photon Source (APS). The SwMppP·D-Arg data set was collected at LS-CAT beamline 21-ID-G. All data were processed with HKL2000[4].

The structure of SgMppP was determined by the single wavelength anomalous diffraction (SAD) method using 3.5 Å resolution data collected from a crystal of SeMet-substituted SgMppP at 0.97889 Å, 61.0 eV below the tabulated K-edge wavelength for Se (0.97950 Å). AutoSHARP[5] was used to determine the Se substructure, which contained 14 of the 16 Se atoms in the asymmetric unit, and calculate density-modified electron density maps. An initial model comprising ~90% of the asymmetric unit content was built using BUCCANEER[6].

Chain A of this partially refined model (with B factors set to 20.0 Å^2) was used as the search model for molecular replacement in PHASER[7] to phase the unliganded SwMppP data set. The molecular replacement solution was subjected to iterative cycles of manual model building in COOT[8] and maximum likelihood-based refinement using the PHENIX package (phenix.refine[9]). Ordered solvent molecules were added automatically in phenix.refine and culled manually in COOT. Hydrogen atoms were added to the model using phenix.reduce[10] and were included in the later stages of refinement to improve the stereochemistry of the model. Positions of H atoms were refined using the riding model with a global B factor. Regions of the model for translation-libration-screw (TLS)

refinement were identified using phenix.find_tls_groups (P. V. Afonine, unpublished work), and the TLS parameters were refined in phenix.refine. Once the refinement converged (R = 0.136, and R_{free} = 0.162), the model was validated using the tools implemented in COOT and PHENIX[9, 11]. Sections of the backbone with missing or uninterpretable electron density were not included in the final model. Side chains with poor or missing electron density were modeled in favored rotameric conformations, and the B factors were allowed to refine without additional restraints. This was done to alert end users, who are more likely to display a model colored by B factors than to pay attention to atom occupancy values or download electron density maps, that these residues are not well defined in the electron density.

The final, refined model of SwMppP, stripped of water molecules and H atoms, and with all B factors set to 20.0 $Å^2$, was used to determine the structure of the enzyme with L–Arg bound, **2**, **3** and D-Arg bound by difference Fourier. The model was also used to determine the mutants' structures. Restraints for the L-Arg-PLP, **2**, **3** and D-Arg-PLP external aldimine were generated with phenix.elbow[12]. Data collection and model refinement statistics are listed in Appendix C. Coordinates and structure factors for both SwMppP and SwMppP·D-Arg models have been deposited in the Protein Data Bank (www.rcsb.org) with accession codes 5DJ1 and 5DJ3.

5.2.9 Preparation of 2-keto-arginine (3)

The 2-oxo acids used in this study were prepared by a modification of the procedure first described by Meister[13]. For example, compound **3** was prepared by dissolving 2 g of L-Arg in 50 mL of water and the pH was adjusted to 7.2 with concentrated HCl. Next, 100 mg of *Crotalus atrox* L -amino acid oxidase (LAAO) and 1.5 mg of catalase were dissolved in 10 mL of water and added to the amino acid solution. The volume was adjusted to 100 mL with water, and the reaction mixture was incubated in the dark at room temperature while being vigorously stirred. After 18 h, the 100 mL reaction mixture was concentrated to 7 mL at 50 °C under vacuum. The 2-oxo acid crystallized from the concentrated reaction mixture at 4 °C after approximately 1 week. The crystalline material was harvested, dried, and stored at − 20 °C. The compound was characterized by nuclear magnetic resonance (NMR) and ESI-MS and found to be predominantly the desired product.

5.2.10 Steady State Enzyme Kinetics

All 1.0 mL kinetic assays were conducted in triplicate at 25 °C in 50 mM BIS-TRIS propane, pH 9.0. The initial velocity of SwMppP-catalyzed hydroxylation of L-Arg was measured directly by monitoring the decrease in the dioxygen concentration using a dioxygen electrode (Hansatech Instruments, Norfolk, U.K.). The concentration of L-Arg was varied from 10 to 10,000 μM. The k_{cat} and K_M values were determined from the initial velocity data using the equation $v_0 = V_m[A]/(K_M + [A])$, where [A] is the concentration of L-Arg, v_0 is the initial velocity, V_m is the maximal velocity, and K_M is the Michaelis constant.

5.2.11 The stoichiometry of L-Arg and O$_2$ in the reaction

The stoichiometry of the reaction was determined by measuring the quantity of O$_2$ consumed in reactions with a constant enzyme concentration (10 μM SwMppP or its mutants) and different L-Arg concentrations. Reactions were monitored using the Hansatech dioxygen electrode. The concentration of dioxygen remaining when the reaction reached equilibrium was subtracted from the starting concentration to obtain the concentration of dioxygen consumed. The stoichiometry was computed using the following formula:

$$Eq.\ of\ O_2\ consumed\ per\ eq.of\ Arg\ added = \frac{[O_2]initial - [O_2]remaining}{[L - Arg]}$$

5.2.12 SwMppP substrate specificity

Reactions containing 686 μl of 30 μM SwMppP in 50 mM BIS-TRIS propane pH 9.0 were initiated by adding different L-Arg analogs (compounds **6**, **7**, **8**, **9** and **10**) to a final concentration of 2 mM. The reactions were monitored by recording spectra from 200-700 nm (0.5 s between scans) for 300 s in a Hewlett-Packard 8453 diode array spectrophotometer. The products of these reactions were identified by NMR and ESI-MS as discussed previously.

The inhibition constant of D-Arg was determined using the O$_2$ consumption assay

described above for the steady state kinetics. D-Arg concentrations ranging of 0, 250, 500, 1,000, 2,500, 5,000, and 10,000 μM were added to different L-Arg solutions (5, 10, 25, 50, 100, 250, 500, 1000, 2500 and 5000 μM) in 50 mM BIS-TRIS propane pH 9.0. Reactions were initiated by the addition of SwMppP to a final concentration of 10 μM and initial velocities were recorded. The inhibition constant was determined using the following equation:

$$vo = \frac{\Delta[O_2]}{t} \qquad\qquad \text{Equation 1}$$

Each data set of rate vs L-Arg concentrations was fitted with Michaelis-Menten equation (Equation 2).

$$v = (v_{max} * [\text{L-Arg}])/(K_M + [\text{L-Arg}]) \qquad\qquad \text{Equation 2}$$

Reciprocal plot of resulting v_{max} and L-Arg concentration gives a series of linear plot at different D-Arg concentrations. The slope is $K_M/v_{max}(1+[\text{D-Arg}]/K_i)$, and the intercept is $1/V_{max}$. Reploting the slope vs [D-Arg], and the data were fitted linearly, which gives K_i (-x intercept).

Fluorescence titration experiments were conducted to measure the binding constant of D-Arg. SwMppP (10 μM after mixing) was mixed with different concentrations of D-Arg (40 μM, 80 μM, 160 μM, 320 μM, 640 μM and 1280 μM) and the total fluorescence was measured with excitation at 415 nm. The results are shown in Figure 5-25. By fitting the fluorescence data into the following single exponential equation:

$$y = A1^\wedge \exp(-k_{obs} * x) + C \qquad\qquad \text{Equation 3}$$

In this equation, A1 is the amplitudes associates with the phase rate constants k_{obs}. The C is the constant of fluorescence intensity at the end of the reaction. The dependence of the observed rate constants k_{obs} was fitted into the following hyperbola equation to give the $K_{d,app}$:

$$k_{obs}= k_{max}*[D\text{-}Arg]/(K_{d,app}+[D\text{-}Arg]) \qquad\qquad \text{Equation 4}$$

5.2.13 Testing for communication between the two binding sites of SwMppP(D188A)

5.2.13.1 Isothermal Titration Calorimetry (ITC)

The D188A mutant was used for the determination of binding constants, since this mutant is catalytically dead. About 3 mL of ~300 μM SwMppP(D188A) was prepared for ITC analysis by dialysis against 2 L of 15 mM MES pH 6.7 overnight at 4 °C. The dialysis buffer was used to dilute the protein to 90 μM, and was also used to prepare a solution of 80 mM L-Arg. All solutions were degassed at 4°C for 30 min before loading into the Nano ITC instrument (TA Instruments, New Castle, DE). The protein solution was placed in the reaction chamber, and the 80 mM L-Arg solution in the syringe. Injections of 2.5 μl were made every 300 s for a total of 20 injections. The titration experiments were done in triplicate at 4, 10, 12, 18 and 20 °C in order to find the optimal temperature (maximal signal-to-noise and best baseline).

5.2.13.2 Determination of binding constants by fluorescence titration

Binding constants for the D188A mutant of SwMppP were also determined by fluorescence titration. Reactions were prepared by mixing 1.5 mL of 14 μM SwMppP(D188A) in 100 mM BIS-TRIS propane pH 9.0 with 1.5 mL of solutions with concentrations of L-Arg ranging from 5 to 2560 μM in the same buffer. The reactions were mixed by pipetting and incubated at room temperature for 8 min. Fluorescence data were collected on a XXX fluorimeter (Shimadzu, XXX, Japan). Samples were excited at 415 nm (5 nm bandwidth) and the emission was measured from 436-700 nm (15 nm bandwidth).

The delta fluorescence data at 480 nm on each L-Arg concentration was obtained by subtract the fluorescence intensity with 0 μM L-Arg in 100 mM BIS-TRIS propane pH 9.0 buffer. Then the fraction of sites occupied and the concentration of D188A•L-Arg complex were calculated. The plot of calculated [D188A•L-Arg] complex versus [L-Arg] at 480 nm was fitted to hyperbolic equation:

$$[EL]_{480} = ([EL]_{480,\,max} * [L\text{-}Arg])/(K_{d,app} + [L\text{-}Arg]) \qquad \text{Equation 5}$$

5.2.14 Pre-steady state enzyme kinetics

Pre-steady state kinetics data were collected on a Model SF-xxDX double-mixing stopped flow spectrophotometer. Since the reaction of SwMppP with L-Arg consumes oxygen, the substrate binding and related changes at the PLP center could be monitored by watching the total fluorescence signal when exciting the internal aldimine at 415 nm in samples of anaerobic enzyme and L-Arg. Likewise, the rate of formation of the first quinonoid intermediate (Q1) could be measured by monitoring the absorbance at 510 nm

under anaerobic conditions. The instrument was prepared for anaerobic work by overnight incubation with a solution of 1 mM dextrose, 10 U/mL glucose oxidase, which had been flushed through the entire system. Solutions for anaerobic stopped flow experiments were prepared in a tonometer. Briefly, 7 ml of 20 μM SwMppP in 10 mM MES pH 6.7, 20 μM PLP and 100 μM glucose (with or without 5 ng/mL catalase) were placed in the main body of the tonometer and 25 μl (1 U/μl) glucose oxidase were placed in the side arm. The tonometer was attached to a Schlenk line and dioxygen was removed from the solution using 45 cycles of vacuum (~5 s) and flushing with argon (~30 s). Because the SwMppP protein is light-sensitive and somewhat unstable at room temperature, this procedure was conducted on ice and the tonometer was covered with foil to protect it from light. Different concentrations of L-Arg in 100 mM BIS-TRIS propane pH 9.0 with 200 μM glucose were placed in glass syringes and sparged with argon gas for 5 min to remove dioxygen. Before removing the argon nozzle and mounting the syringe on the stopped flow instrument, 10 μl of glucose oxidase were injected through the luer tip. Single mixing experiments were conducted at 25 °C with 10 μM SwMppP and 10 to 1280 μM L-Arg (all concentrations are after mixing 1:1)

To measure the rates of the O_2-dependent steps, double mixing experiments were performed where anaerobic enzyme was mixed 1:1 with a saturating amount of anaerobic L-Arg. This was then mixed 1:1 with buffer containing various concentrations of dioxygen. As above, the instrument was made anaerobic overnight with the glucose/glucose oxidase treatment described above. The first tonometer contained 40 μM enzyme, 10 ng/mL catalase in 10 mM MES pH 6.7, 20 μM PLP under anaerobic conditions. The second tonometer, made anaerobic as described above, contained 2,560 μM L-Arg in 100 mM BIS-

TRIS propane pH 9.0. The third tonometer contained 100 mM BIS-TRIS propane pH 9.0 with different concentrations of dioxygen (20, 40, 78, 158, 296, 465 μM). Defined dioxygen concentrations were obtained using a Maxtec Maxblend gas blender to mix XX% pure O_2 gas with gaseous N_2. The concentrations of dioxygen produced at different settings of the mixing valve were determined using the Clark-type electrode. Samples of oxygenated buffer were prepared by sparging the 100 mM BIS-TRIS propane pH 9.0 buffer with set concentrations of dioxygen for 5 min. After the first mixing step, the reactions were aged for 20 s prior to the second mixing step in order to convert all of the enzyme to the Q1 intermediate. The second step mixed the anaerobic reaction mixture with buffer containing different concentrations of dissolved O_2. The reactions were monitored at 510 nm and 560 nm for 100 s for each data set.

Data were analyzed by Origin (OriginLab) or Kinetic Studio (TgK Ltd.). For the two phases fluorescence data excited at 415 nm under anaerobic condition, it was fitted into the following equation:

$$y=A1\text{^}\exp(-k_1*x)+A2\text{^}\exp(-k_2*x)+C \qquad \text{Equation 6}$$

In this equation, A1 and A2 are the amplitudes associates with the first and second phase rate constants k_1 and k_2, respectively. The C is the constant of fluorescence intensity at the end of the reaction. The dependence of the observed rate constants k_1 for the first phase (based on the amplitude) was fitted linearly. The dependence of the observed rate constants k_2 for the second phase was fitted to the following hyperbolic equation:

$$k_2= k_{max}*[\text{L-Arg}]/(K_{app}+[\text{L-Arg}]) \qquad \text{Equation 7}$$

In the equation, K_{app} is the apparent equilibrium constant for the conversion of the external aldimine into Quinonoid I intermediate. The rate constant k_{max} is the maximum rate constant.

The Quinonoid I intermediate accumulation was also monitored by recording the absorbance at 510 nm under anaerobic condition. This data has two phases and was fitted into Equation 6. The dependence of the observed rate constants k_1 and k_2 were both fitted into Equation 7.

With different concentrations of O_2, the kinetics of Quinonoid I intermediate decay, Quinonoid II formation and decay were studied. The first 15 s of Quinonoid I decay step corresponding to one turnover was fitted to Equation 6. The dependence of the observed rate constants k_1 for the largest amplitude phase was fitted linearly. The second phase rate constants cannot be fitted meaningfully. The absorbance at 560 nm corresponding to Quinonoid II formation was fitted to the following three exponential equation:

$$y=A1\text{^}exp(-k_1*x)+A2\text{^}exp(-k_2*x)+A3\text{^}exp(-k_3*x)+C \qquad \text{Equation 8}$$

In this equation, A1, A2 and A3 are the amplitudes associates with the first, second and third phase rate constants k_1, k_2 and k_3, respectively. The C is the constant of absorbance intensity at the end of the reaction. The dependence of the observed rate constants k_1 was fitted linearly. The dependence of second phase rate constants was fitted linearly. The dependence of third phase rate constants was fitted to Equation 6.

5.3 Results

5.3.1 Side product H$_2$O$_2$ catalyzed the decarboxylation of two products (3, 4)

The products of SwMppP with L-Arg was also confirmed by mass spectrometry with electrospray ionization (ESI-MS), which detected peaks in the positive mode spectrum at m/z=146 and m/z=162. These masses correspond to the five-carbon products derived from **2** and **3**, 4-guanidino-3(S)-hydroxybutyric acid (**4**) and 4-guanidinobutyric acid (**5**). The result was also verified by ^{13}C-NMR (Figure 5-4). These products are the result of oxidative decarboxylation of **2** and **3**, like catalyzed by H$_2$O$_2$ as has been observed for other α-keto acids [14, 15]. This observation was the first suggestion that H$_2$O$_2$ might be a product of SwMppP-catalyzed oxidation of L-Arg. Therefore, the experiment was repeated, this time with catalase in the reaction mixture. The masses for both products are 174 and 190, corresponding to **3** and **2**. The fully oxidized product **2** is produced at a ratio of ~1.7:1 compared to the partially oxidized product **3**. Thus, once in every two to three turnover cycles, the enzyme loses its substrate to hydrolysis before it can be fully oxidized. This observation likely has important implications for the mechanism of this novel enzyme. Note that **2** is the same compound that MppR was able to cyclize to give the iminoimidazolidine ring of L-End observed in a crystal structure of MppR in the presence of **2**. MppP is therefore the first enzyme in L-End biosynthesis, reacting directly with L-Arg to create the substrate for MppR.
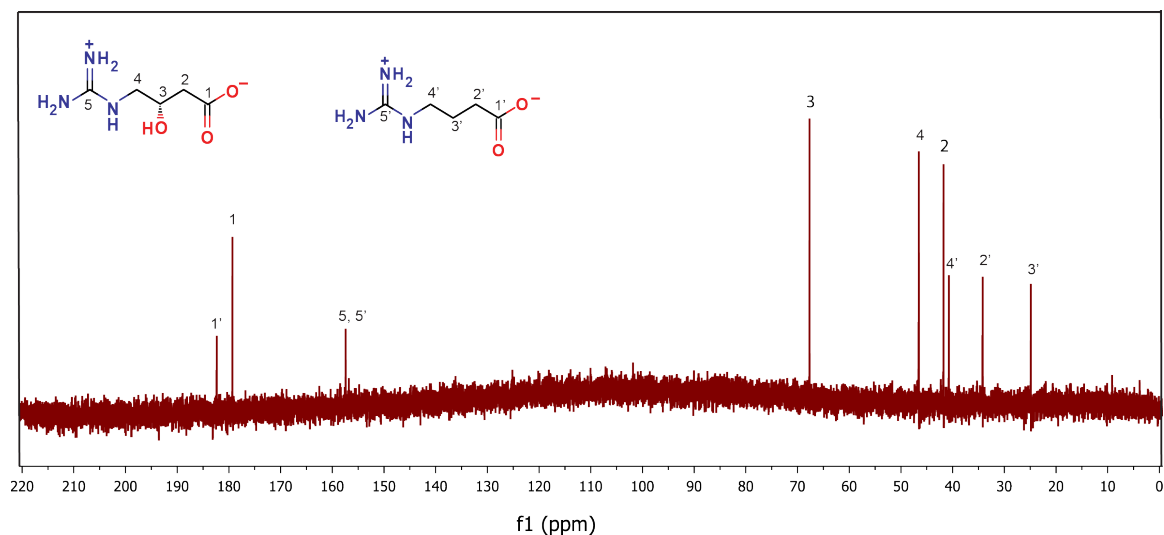
Figure 5-4: ^{13}C-NMR spectrum of the reaction of 10 μM SwMppP with 2.0 mM L-Arg in 20 mM sodium phosphate pH 8.4 buffer.

5.3.2 Ammonia is also a product of SwMppP with L-Arg and O$_2$

The 4-electron oxidation of L-Arg suggested by the results detailed above would require ammonia to also be a product of the reaction, since the PLP cofactor needs to be converted from the pyridoxamine form to the aldehyde (PLP) at the end of the reaction cycle. A glutamate dehydrogenase (GDH)-coupled assay was employed to determine whether or not the reaction of SwMppP with L-Arg produces ammonia. The assay takes advantage of the fact that GDH can be run in reverse, where α-ketoglutarate is reduced by ammonia and NADH to give glutamate, water, NAD$^+$, and a proton. By constructing a standard curve of the rate of loss of absorbance from NADH at 340 nm vs the concentration of ammonia, it is possible to estimate the concentration of ammonia in, for example, SwMppP reaction mixtures. The results of these experiments are shown in Figure 5-5. The correlation

207

coefficient (R^2) of the standard curve is 0.9992, so the response of the GDH initial velocity to the ammonia concentration is linear. In the reaction of 20 µM SwMppP and different concentrations of L-Arg, after the GDH addition, the absorbance at 340 nm decreased and the rate of this decrease is proportional to the centration of L-Arg. In 15 min, approximately 80% of the L-Arg was consumed and an equal amount of ammonia was released.
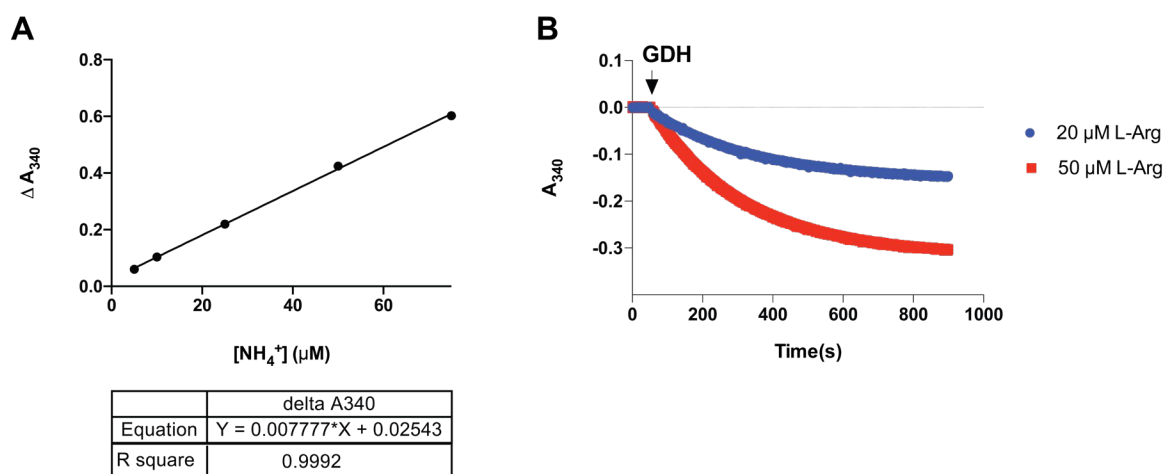
Figure 5-5: (A) The standard curve of NADH based GDH assay. (B) The ammonium formation is proportional to the L-Arg concentration according to the GDH assay.

5.3.3 SwMppP is an oxidase and not an oxygenase

Reactions of SwMppP with 3 concentrations of L-Arg were monitored using an oxygen electrode (Figure 5-6). When the reactions reached equilibrium, catalase was added. The addition of catalase resulted in the regeneration of dioxygen amounting to half of that originally consumed by SwMppP in each reaction. This result indicates that H_2O_2 is another formal product of the SwMppP reaction, and that each molecule of O_2 consumed by

SwMppP is reduced to H_2O_2 during the catalytic cycle.
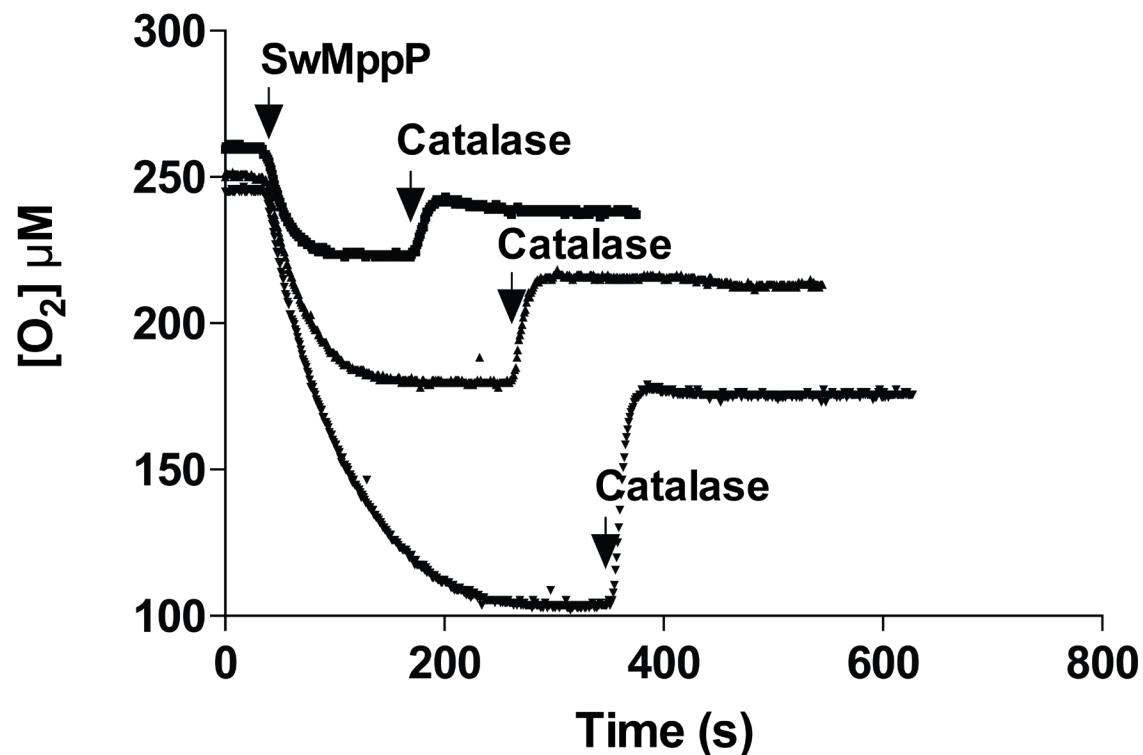
L-Arg (µM)	[O$_2$ consumed] after SwMppP addition (µM)	[O$_2$ generated] after catalase addition (µM)
25	38.07 ± 1.60	19.00 ± 1.00
50	69.33 ± 1.16	35.67 ± 0.58
100	140.33 ± 2.31	71.67 ± 2.08

Figure 5-6: O$_2$ regeneration after adding catalase into equilibrated SwMppP with varied [L-Arg] solutions. The concentration of dioxygen was monitored in a Hansatech dioxygen electrode. The reactions were initiated by add 10 µM SwMppP into solutions containing 25 µM (■, top), 50 µM (▲, middle), and 100 µM (▼, bottom) L-Arg in 50 mM BisTris Propane pH 9.0 buffer. When the reaction was equilibrated, 50 ng/mL catalase were added into the solution. As the table shown, the regenerated O$_2$ is around half of the O$_2$ consumed by SwMppP, which suggests that all O$_2$ was consumed to form H_2O_2.

The observation that each molecule of dioxygen consumed is reduced to hydrogen peroxide logically dictates that none of the oxygen atoms incorporated into the products (**2** or **3**) can come from dioxygen. SwMppP is therefore an oxidase, and not an oxygenase. To confirm that the oxygen atoms incorporated into the products do indeed derive from water and not dioxygen, reactions of SwMppP with L-Arg were conducted with either $^{18}O_2$ or $H_2^{18}O$. Catalase was included in the reactions to prevent the H_2O_2-mediated decarboxylation of the products that would complicate the interpretation of the results. The products of these reactions were analyzed by ESI-MS. This analysis showed that in the presence of $^{18}O_2$, the products were detected at m/z=174 and m/z=190, indicating that neither product contained ^{18}O. When the reactions were run in $H_2^{18}O$, the products were detected at m/z=176 and m/z=194. This is consistent with the incorporation of two atoms of ^{18}O into the fully oxidized **2**, and a single atom of ^{18}O into **3** (Figure 5-7).
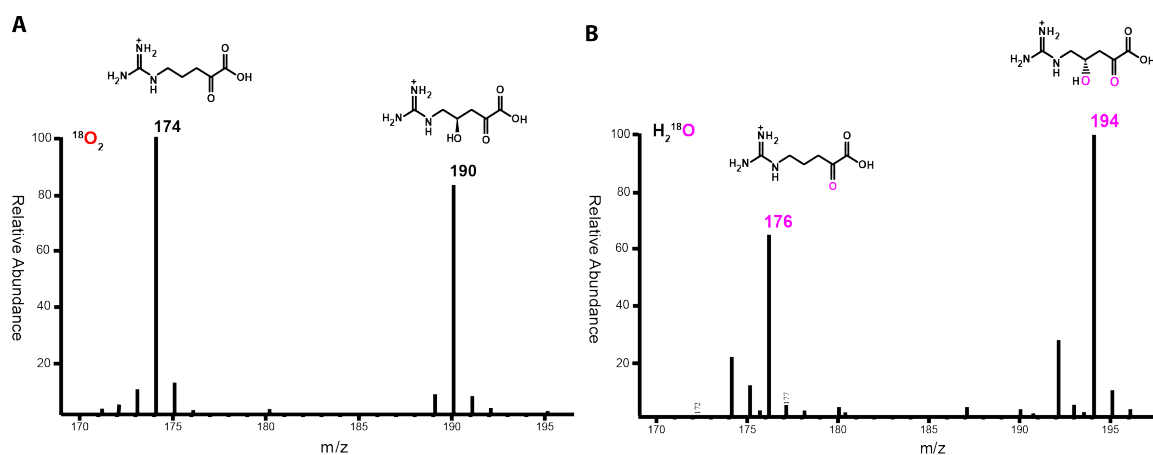
Figure 5-7: Mass spectrometric analysis results. A) Spectrum of SwMppP with L-Arg in the presence of $^{18}O_2$. The mass of both products did not change. B) Spectrum of SwMppP with L-Arg in $H_2^{18}O$. The hydroxy oxygen atom of **2** product originated from $H_2^{18}O$. And the keto oxygen atom of both products also came from the $H_2^{18}O$ by hydrolysis.

210

While other PLP-dependent enzymes are known to react with molecular oxygen, such as DOPA decarboxylase/aromatic L-amino acid decarboxylase [16, 17], and so-called "paracatalytic" reactions with O_2 have been described [18], SwMppP is the first PLP-dependent enzyme to produce a hydroxylated product. Given the sequence similarity between SwMppP and more typical PLP-dependent aminotransferases, it would be particularly interesting to know how the structure of SwMppP differs, and why these differences result in the enzyme acting as an L-Arg γ-hydroxylase rather than a decarboxylase or aminotransferase.

5.3.4 The structures of SwMppP

The experiments described thus far show that SwMppP is an L-Arg oxidase. However, analysis of the primary sequence suggests that SwMppP likely has a very similar structure to the Fold Type I aminotransferases. So what structural features of SwMppP allow it to catalyze a very difficult reaction (oxidation of an unactivated carbon center) using a scaffold normally associated with simple aminotransfer chemistry? To explore the relationship between the structure and function of SwMppP, we set out to determine its structure by X-ray diffraction. SwMppP crystallized by the hanging drop vapor diffusion method from 2 μL of protein solution at 15 mg/mL in 10 mM MES pH 6.7, 20 μM PLP and 1 μL of crystallization solution (30% polyethylene glycol monomethyl ether 550 (PEG MME 550), 50 mM $MgCl_2$, and 0.1 M HEPES (pH 8.0)). These crystals diffracted to 2.1 Å resolution. A number of aminotransferases, including the human kynurenine aminotransferase II (PDB ID 5EUN [19]) had ~30% sequence identity to SwMppP, so we first

sought to obtain initial phases for the structure by molecular replacement. Numerous search models were tested, including those based on different aminotransferase structures, with separated domains (in case the domains had a different orientation in MppP), and ensemble structures built from superpositions of several models. None of these resulted in suitable phase estimates. We then attempted SAD phasing, but were stymied by the unusually low number of methionine residues (3 methionines of 376 AA) in the *S. wadayamensis* enzyme. Rather than attempt to soak in compounds with anomalously scattering atoms, we decided to target a highly similar homolog of MppP that had more methionine residues.

The MppP from *S. globisporus* (SgMppP) was identified through BLAST searches of the SwMppP sequence and is XX% identical to SwMppP. This enzyme crystallized from XXX, but the crystals were not well ordered and only diffracted to ~3.5 Å resolution. However, this was sufficient to build a nearly complete model of the protein that was then used as the search model for successful molecular replacement phasing of the SwMppP data set.

SwMppP crystallized in space group $P2_12_12_1$ with the following unit cell dimensions: a = 85.7 Å, b = 108.3 Å, and c = 195.4 Å. The asymmetric unit contained four molecules that formed two independent homodimers (Figure 5-8). The overall structure of SwMppP, as expected from the sequence analysis, is very similar to typical Type I aminotransferases like aspartate aminotransferase. As in AATase, there are also two domains in SwMppP: a large domain (A33-G272), a small domain (L23 – Q32 and G273 – R376) and an N-terminal extension (M1-E22). The large domain contains a seven-stranded mixed β–sheet flanked by ten α–helices (α-β-α sandwich), while small domain has a mixed α-β topology.
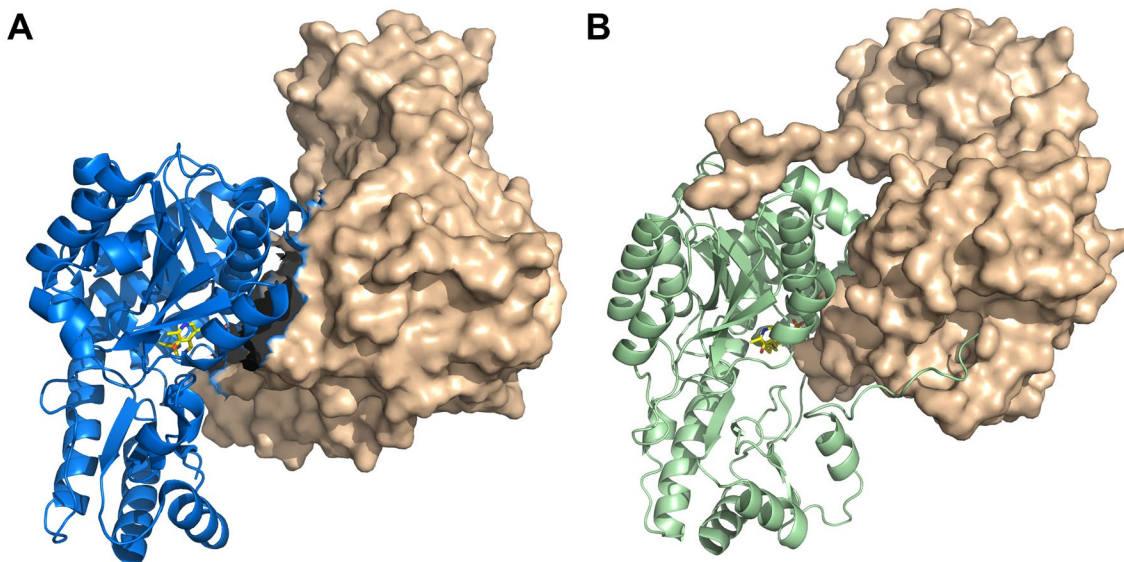
Figure 5-8: (A) SwMppP dimer with one chain shown as a blue ribbon representation and the other shown as a wheat surface representation. The K221-PLP internal aldimine in the active sites is shown in yellow stick. The N terminus (1-22 aa) is disordered in the crystal. (B) *E. coli* aspartate aminotransferase (PDB ID: 1ARS[20]) with one chain shown as a light green ribbon representation and the other shown as a wheat surface representation. The K258-PLP internal aldimine in the active sites is shown in yellow stick. The N terminus (1-30 aa) ordered and interacts with the large domain of the other chain.

Three of the four chains in the asymmetric unit are missing 22 residues from the N-terminus, while the fourth (chain D) is missing only 10 residues. The extra 12 residues in chain D form a loop and α helix that points outward, away from the enzyme and interacts with a symmetry-related molecule (25 Å from the active site). There is nothing in these interactions to suggest that they are physiologically relevant. All four chains are missing one to three residues from the C-terminus.

As alluded to above, and as shown in Figure 5-8, the tertiary structure of SwMppP is

similar to the type I aminotransferases (ATases)[21]. The *E. coli* aspartate ATase (PDB entry 1ARS[20]) superimposes well onto the structure of the SwMppP dimer. The root-mean-square deviation (rmsd) value obtained using secondary structure matching (SSM)[22] of 3.4 Å for 567 of the 704 Cα atoms in SwMppP underscores the similarity of these two enzymes. For reference, different structures of *the same* protein can have rmsd values of more than 0.5 Å. The most obvious difference between the tertiary structures of SwMppP and AATase is that the N-terminal extension is disordered or points outward in SwMppP, which leaves the active sites open to the solvent, whereas this part of the AATase structure closes over the active site, covering the cofactor almost completely.

As in the ATases, the active site is located at the interface between the large and small domains of each protomer. Unlike the ATases, however, there is minimal contribution from the other chain in the dimer. The position of the cofactor is also similar to that seen in all known Fold Type I aminotransferases, in a cleft formed by the central β-sheet and the N-termini of helices α4 and α5 (Figure 5-8). In the absence of bound ligands, PLP is covalently attached to K221 by an aldimine linkage. The C5' phosphate of PLP is held by fewer direct interactions than in other PLP-dependent enzymes. In SwMppP, the only direct interactions between the enzyme and the C5' phosphate are an electrostatic interaction with K229 and a hydrogen-bonding interaction with S91. There may also be a weak hydrogen bond to H29. The remaining interactions constituting the C5' phosphate-binding site are water-mediated hydrogen bonding interactions with D218 as well as Y88* and the main chain amide of L252* from the other protomer of the dimer (Figure 5-9). As in all fold type I aminotransferases, there is a negatively-charged residue, D188, placed near the pyridine N atom of the cofactor to stabilize the positive charge that forms during the catalytic cycle [23-

214

[25]. The C3' hydroxyl of PLP is involved in a hydrogen-bonding interaction with N160, which also makes a hydrogen-bonding interaction with R352, an absolutely conserved residue in the Type I aminotransferases. In the *E. coli* aspartate aminotransferase (EcAAT), for example, the equivalent Arg residue (R386) interacts with the α-carboxylate group of the substrate. Finally, the pyridine ring of PLP is clamped between F115 above and S190 below. Aminotransferases typically have an aromatic side chain packed against the "top" surface of the PLP ring, though the residue below the ring is more typically an Ala.

While there are a number of similarities between the active sites of SwMppP and EcAAT, there are a number of interesting differences that likely play a role in the divergent activity of SwMppP. As mentioned above, the phosphate-binding site is anchored by Lys instead of Arg, and whereas most ATases have 2-3 Ser/Thr residues binding the phosphate group, and SwMppP has only one. One of the positions typically occupied by a Ser/Thr is taken by D218 in SwMppP. The carboxylate group of D218 is within hydrogen bonding distance of S190; this interaction is not observed in most other PLP-dependent enyzmes, most of which have Ala in the position occupied by S190. In many ATases, such as EcAAT, there is a Tyr residue that makes a hydrogen bond to the C3' hydroxyl of PLP and affects the electron distribution in the cofactor [26]. This residue is a Phe in SwMppP. The last unusual feature of the SwMppP active site is H29. This position is not conserved in other PLP-dependent enzymes, but is absolutely conserved among MppP homologs[27]. H29 is adjacent to the catalytic K221 and makes hydrogen bonding interactions with D27 and, perhaps, with the phosphate moiety of the cofactor. In two of the four molecules in the asymmetric unit, H29 takes on at least two conformations.
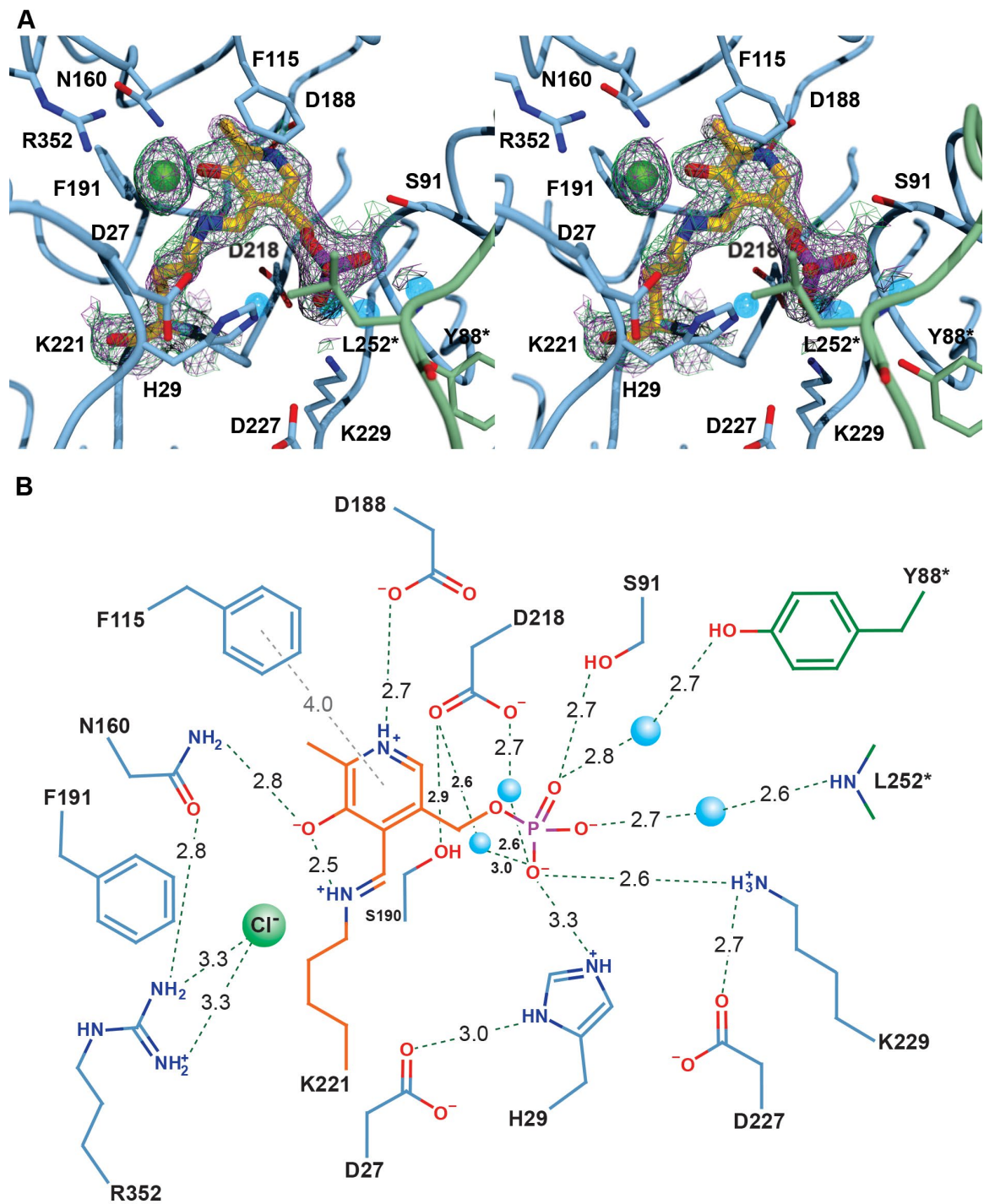
Figure 5-9: The active site of SwMppP. The stereoview of the SwMppP active site (A) shows the PLP covalently bound to enzyme, forming internal aldimine. The 2|F$_o$| - |F$_c$| electron density map contoured at 1.0σ, is shown as magenta mesh, and the 2|F$_o$| - |F$_c$| composite omit map also contoured at 1.0σ, is shown as green mesh. The schematic view of the active site (B) shows potential hydrogen bonding interactions as

green, dashed lines. Water molecules are shown as blue spheres. The chloride ion is shown as a green sphere. Panel A was generated using POVSCRIPT and POV-Ray. Panel B was created using a combination of MarvinSketch (https://www.chemaxon.com) and Adobe Illustrator CS6.

5.3.5 The structure of the SwMppP•L-Arg complex

The structure of SwMppP•L-Arg was obtained by soaking SwMppP crystals in a solution containing 10 mM L-Arg. Unlike the unliganded structure, where the N-terminus is disordered and leaves the active site exposed to the solvent, in the structure with L-Arg bound, the N-terminal extension orders and covers the active site (Figure 5-10). Superimposition of the unliganded SwMppP and SwMppP•L-Arg structures gives a root-mean-square deviation of 0.356 for Cα atoms in the SwMppP dimer, indicating that, except for the N-termini, there are no significant structural alterations as a result of ligand binding. So unlike many ATases, there is no closing movement of the small domain.

Once the substrate binds in the active site, the α-amino group attacks C4' of internal aldimine to form the external aldimine, displacing K221. Compared with the unliganded SwMppP, there are no significant changes in the active site, only a few subtle side chain movements, as shown in Figure 5-11. The Cl$^-$ ion bound to R352 is replaced by the α-carboxylate of L-Arg. T12 and E15 in the N-terminal helix form hydrogen bonds with guanidine group and carboxyl group of L-Arg, respectively. The guanidine group also forms hydrogen bonds with N117, S91, and the carbonyl groups of S248 and D249 from the other protomer. The side chains of H29, D27 and R31 are shifted towards the external aldimine. The ring of F115 rotates away from the PLP ring and is closer to L-Arg.

217
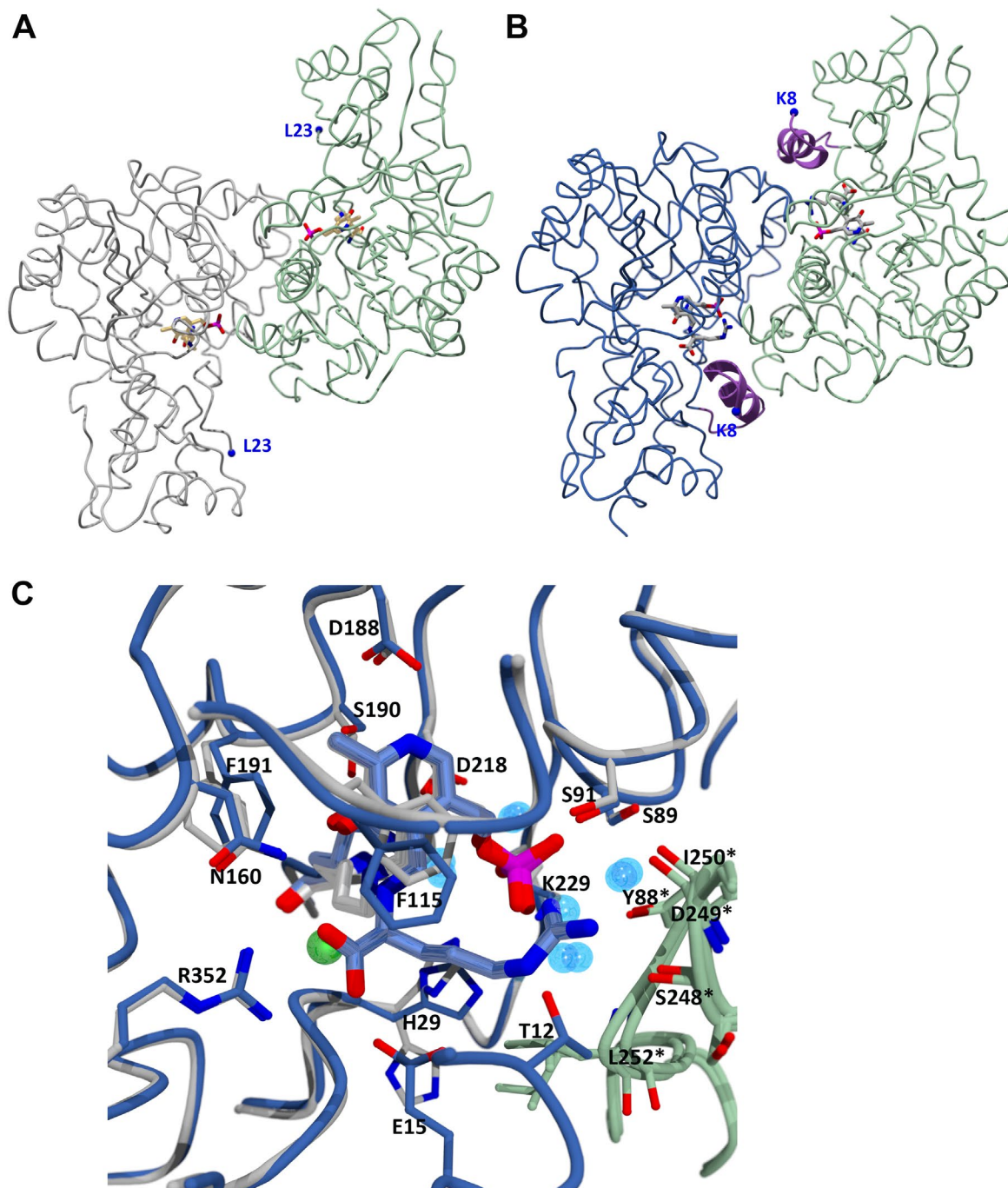
Figure 5-10: Ribbon representation of SwMppP (A) and SwMppP•L-Arg (B) dimers. (A) Chain A is colored in grey, chain B is colored in light green. The N termini of both chains are disordered. The L23 Cα atoms in both chains are shown as blue spheres. (B) Chain A colored dark blue, chain B colored light green and the N-terminal helices colored purple. The PLP cofactors in the active sites are shown in stick. The N-terminal helix

ordered and covers the active site. The K8 Cα atoms in two chains are shown as blue dots. (C) Overlaying the active site of SwMppP and SwMppP•L-Arg. The major change is the E15 and T12 interactions with L-Arg. Also, the rings of F115 and F191Y shift.
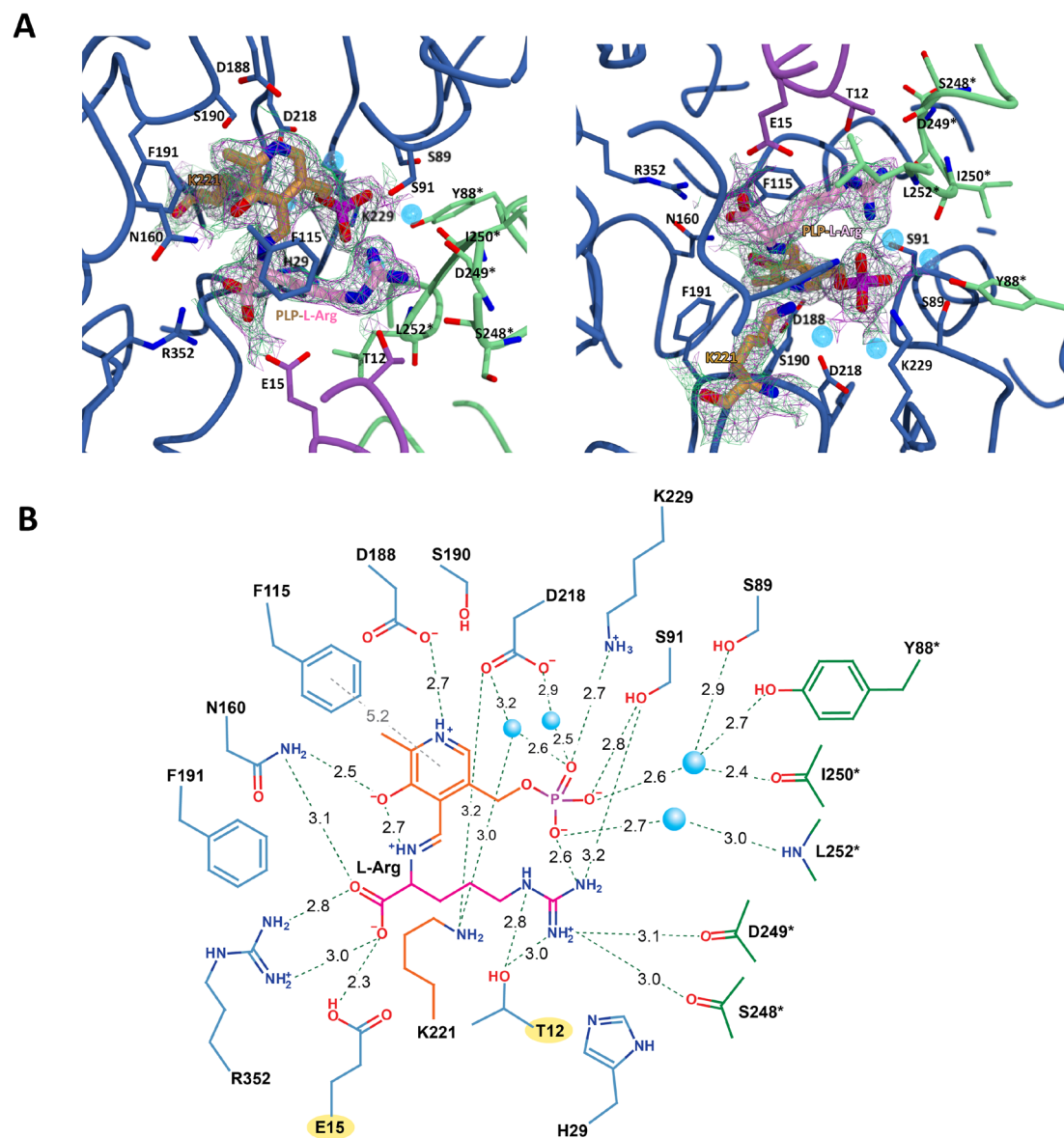
Figure 5-11: The external aldimine structure in active site of SwMppP•L-Arg. (A) SwMppP•L-Arg active site showing the L-Arg covalently bound to the PLP as the external aldimine. The $2|F_o| - |F_c|$ electron density map contoured at 1.0σ, shown as magenta mesh, and the $2|F_o| - |F_c|$ composite omit map also contoured at 1.0σ, is shown as green mesh. (B), Schematic view of the active site showing potential hydrogen bonding interactions

as green, dashed lines. Water molecules are shown as blue spheres. Panel A was generated using POVSCRIPT and POV-Ray. Panel B was created using a combination of MarvinSketch (https://www.chemaxon.com) and Adobe Illustrator CS6.

5.3.6 Structures of SwMppP•product complexes

The structures of the SwMppP•**2** complex (Figure 5-12) and the SwMppP•**3** complex (Figure 5-13) have no big difference on the guanidine is bound from SwMppP•L-Arg structure, but the fact that the products are not external aldimine is a big difference. Also there are several residues side chain movements (overlay results are shown in Appendix F and Appendix G). S190 side chain shifted to hydrogen bond with D218. In the structure of SwMppP•**2**, 4HKA hydroxyl group made hydrogen-bonding interactions with the phosphate group and H29.
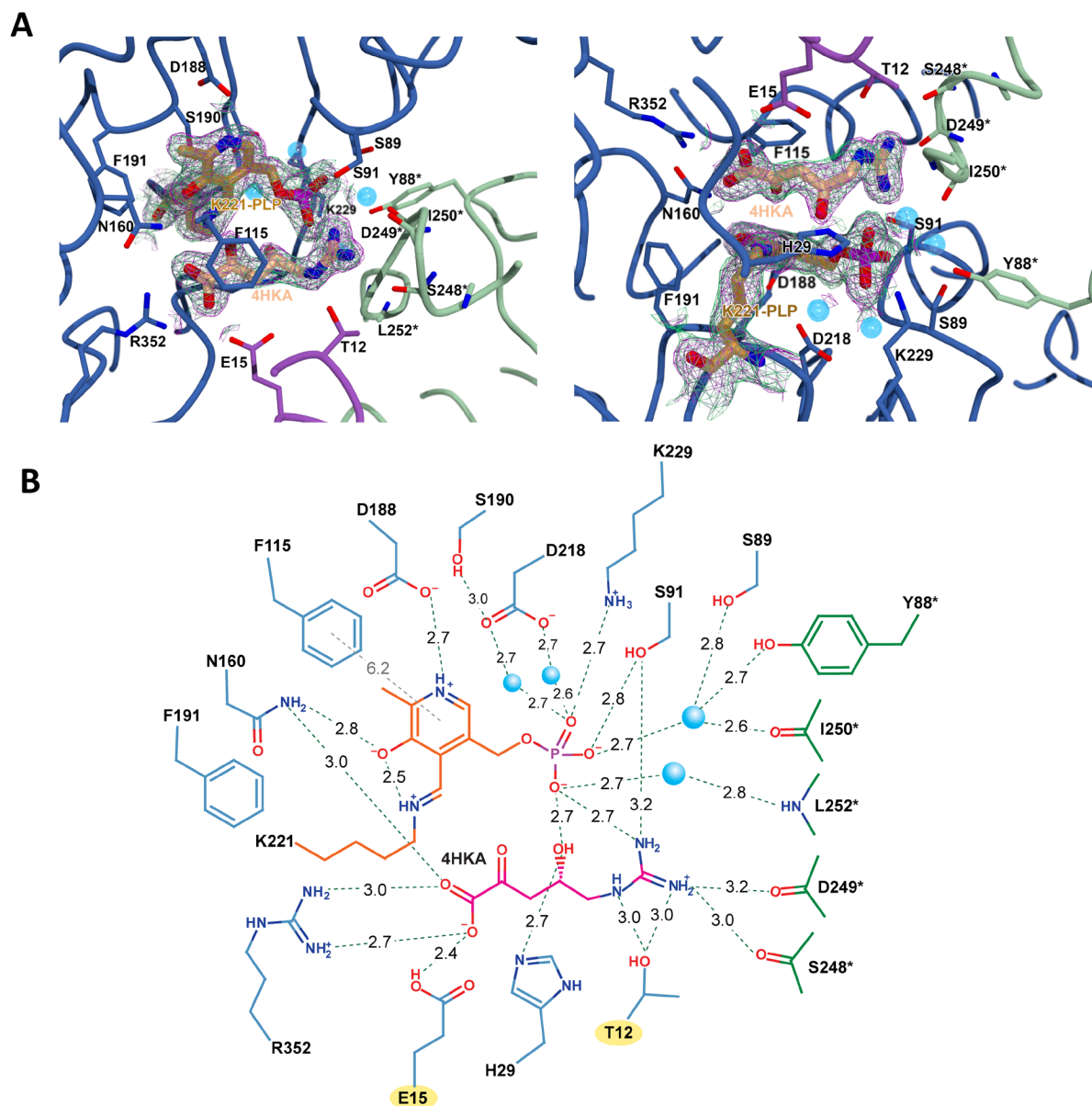
Figure 5-12: The active site of SwMppP•**2**. (A) SwMppP•**2** active site showing the PLP covalently bound to the enzyme as the internal aldimine. E15 and T12 interact with **2** carboxylate group and guanidine group, respectively. His29 bonds to hydroxyl group of **2**. The $2|F_o|\text{-}|F_c|$ electron density map contoured at 1.0σ, shown as magenta mesh, and the $2|F_o|\text{-}|F_c|$ composite omit map also contoured at 1.0σ, is shown as green mesh. (B) Schematic view of the active site showing potential hydrogen bonding interactions as green, dashed lines. Water molecules are shown as blue spheres. Panel A was generated using POVSCRIPT and POV-Ray. Panel B was created using a combination of MarvinSketch (https://www.chemaxon.com) and Adobe

221

Illustrator CS6.
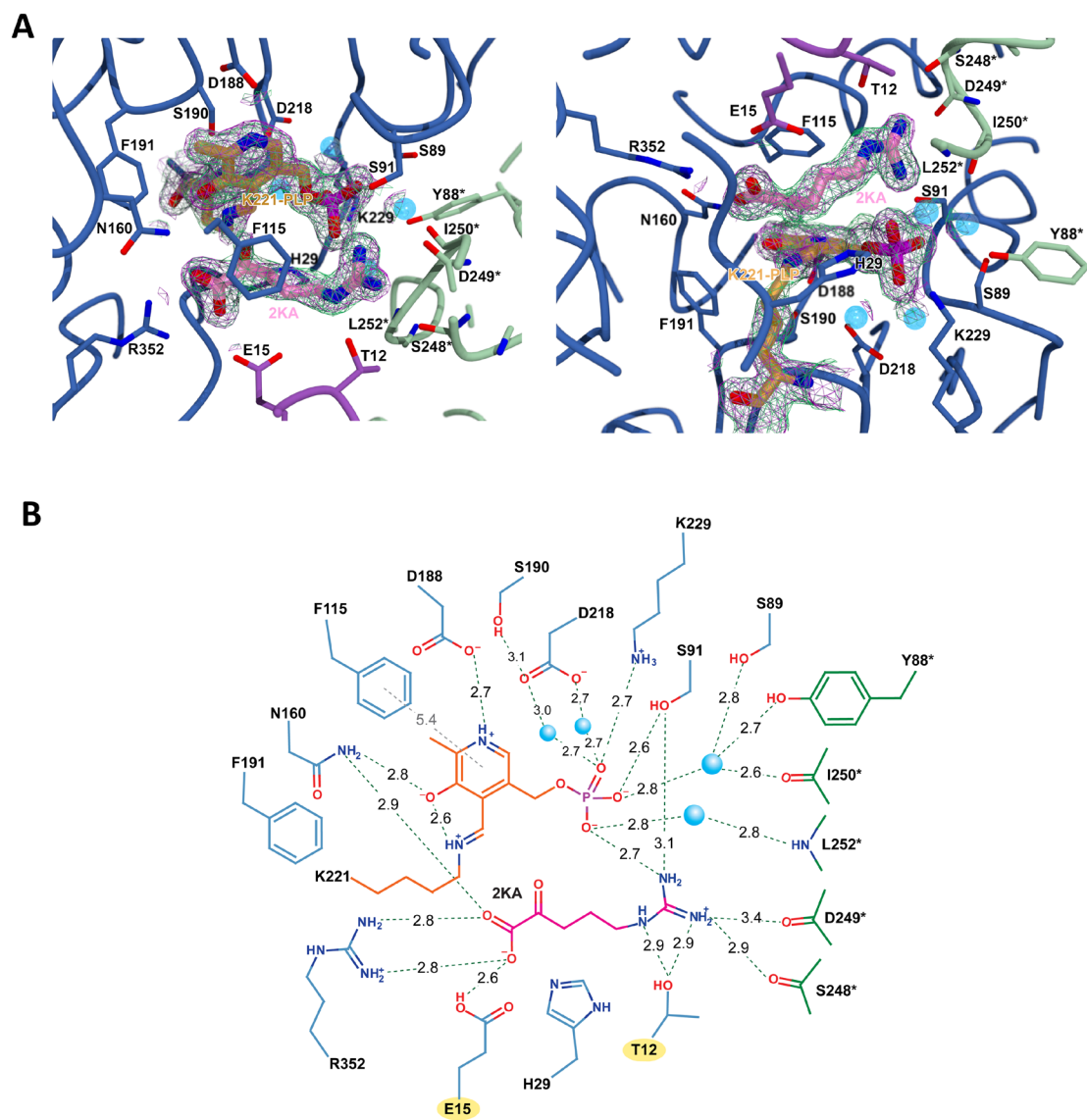
Figure 5-13: The active site of SwMppP•**3**. (A) SwMppP•**3** active site showing the PLP covalently bound to the enzyme as the internal aldimine. E15 and T12 interact with **3** carboxylate group and guanidine group, respectively. The $2|F_o|-|F_c|$ electron density map contoured at 1.0σ, shown as magenta mesh, and the $2|F_o|-|F_c|$ composite omit map also contoured at 1.0σ, is shown as green mesh. (B) Schematic view of the active site showing potential hydrogen bonding interactions as green, dashed lines. Water molecules are shown as blue spheres. Panel A was generated using POVSCRIPT and POV-Ray. Panel B was created using a combination of MarvinSketch (https://www.chemaxon.com) and Adobe Illustrator CS6.

222

5.3.7 Characterization of N-terminus and active site residues

The tertiary structure of SwMppP is very similar with type I aminotransferase, there are several residues differences in the active site, which probably results in the hydroxylase-deaminase activity rather than aminotransferase. Bioinformatic analysis[27] showed that H29, F115, F191, and D218 are conserved in all of the MppP homologs but not in the Type I aminotransferases. To explore the specific roles of these different residues, site-directed mutagenesis was used to make point mutants that were characterized both structurally and kinetically.
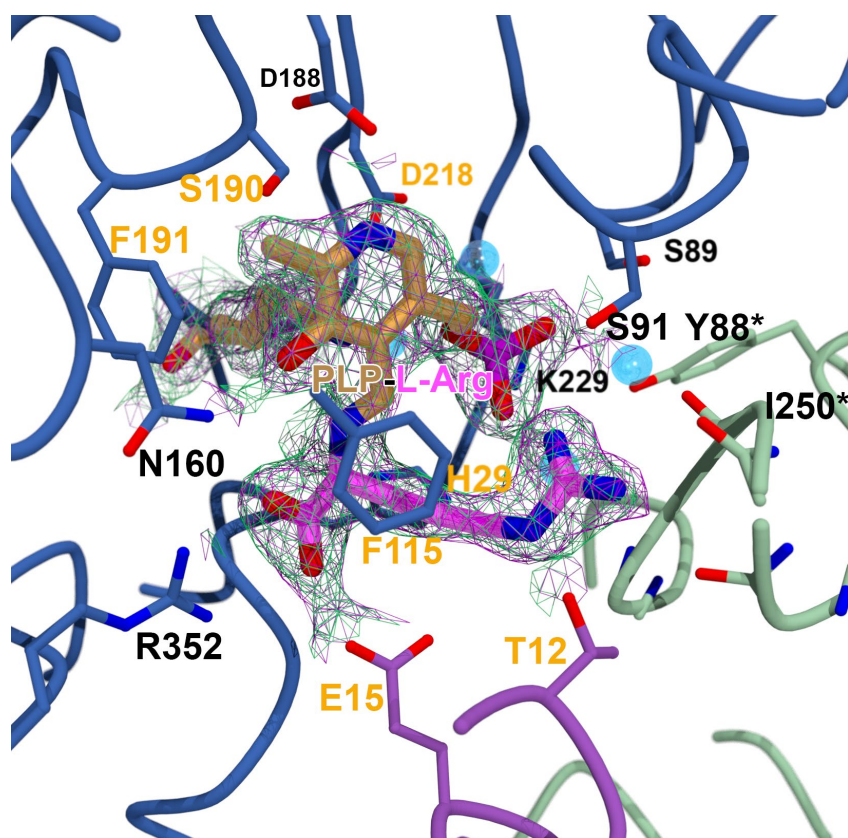
Figure 5-14: Different residues (colored in orange) in the active site of SwMppP•L-Arg. The $2|F_o| - |F_c|$ electron density map contoured at 1.0σ, shown as magenta mesh, and the $2|F_o| - |F_c|$ composite omit map also contoured at 1.0σ, is shown as green mesh.

5.3.8 Characterization of SwMppP and mutants

5.3.8.1 Stability of SwMppP and mutants by differential scanning fluorimetry

The T_m value for each mutant is listed in Table 5-2. Except SwMppP(F191Y), other mutants' stability are same with that of wild-type SwMppP. The T_m value of SwMppP(F191Y) is so low that only flat curve was got.

Table 5-2: T_m values of SwMppP and mutants

Proteins	T_m (°C)
SwMppP	53.3 ± 0.1
SwMppP(E15A)	55.75 ± 0.1
SwMppP(E15Q)	53.3 ± 0.1
SwMppP(H29A)	47.5 ± 0.1
SwMppP(H29S)	46.6 ± 0.1
SwMppP(F115Y)	51.4 ± 0.2
SwMppP(F191Y)	Not applicable
SwMppP(D218S)	50.0 ± 0.1
SwMppP$_{\Delta 1-22}$	50.2 ± 0.1

5.3.8.2 C Determination of the reaction products of the mutants

To determine what reaction these mutants catalyzed, the 4.8 mL reaction contains ~65 μM mutants protein and 2 mM L-Arg in 20 mM sodium phosphate pH 8.4 buffer were shaked at 25°C, 200 rpm overnight. Then the solutions were concentrated for 8 fold and analyzed by proton NMR. The results are summarized in Table 5-3.

Table 5-3: The products of SwMppP and its mutants with L-Arg and O_2

	Products	[2]/[3]
SwMppP	**2** and **3**	1.7:1
SwMppP(T12A)	**3**	0:1
SwMppP(E15A)	**3**	0:1
SwMppP(E15Q)	**2** and **3**	1:2.4
SwMppP(T12A/E15A)	**3**	0:1
SwMppP(H29A)	**3**	0:1
SwMppP(H29S)	**3**	0:1
SwMppP(F115Y)	**3**	0:1
SwMppP(F191Y)	**2** and **3**	1.6:1
SwMppP(D218S)	Trace **3**	-
SwMppP$_{\Delta 1-22}$	Trace **3**	-

SwMppP catalyzes the oxidation of L-Arg and O_2 into **2** and **3** in the ratio of 1.7 : 1. Except SwMppP(F191Y) and SwMppP(E15Q), all mutants lost the hydroxylase activity.

225

They only produce the side product **3**. The products ratio of variant F191Y is 1.6 :1, which is same with that of wild-type SwMppP. However, because the T_m of F191Y is low, the amounts of substrate F191Y catalyzed are much less. Another mutant E15Q still have the hydroxylase activity, although it does make less hydroxylated products (with the [**2**] : [**3**] at the ratio of 1 : 2.4). The substitution of same residue E15 into alanine lost the hydroxylase activity. To figure out why the E15A variant lost the hydroxylase activity, I crystallized the E15A proteins and soaked the crystals with 10 mM L-Arg to get the E15A•L-Arg structure (shown in Figure 5-15). In the structure of E15A•L-Arg structure, L-Arg covalently bonds to PLP, forming external aldimine. The N terminus still covers the active site when substrate binds. And T12 residue interacts with the guanidine group of L-Arg. The main chain of A15 is same with that of E15 in wild type protein. But the shorter side chain of A15 probably creates a void where water can be retained, and which could hydrolyze the quinonoid intermediate prematurely.
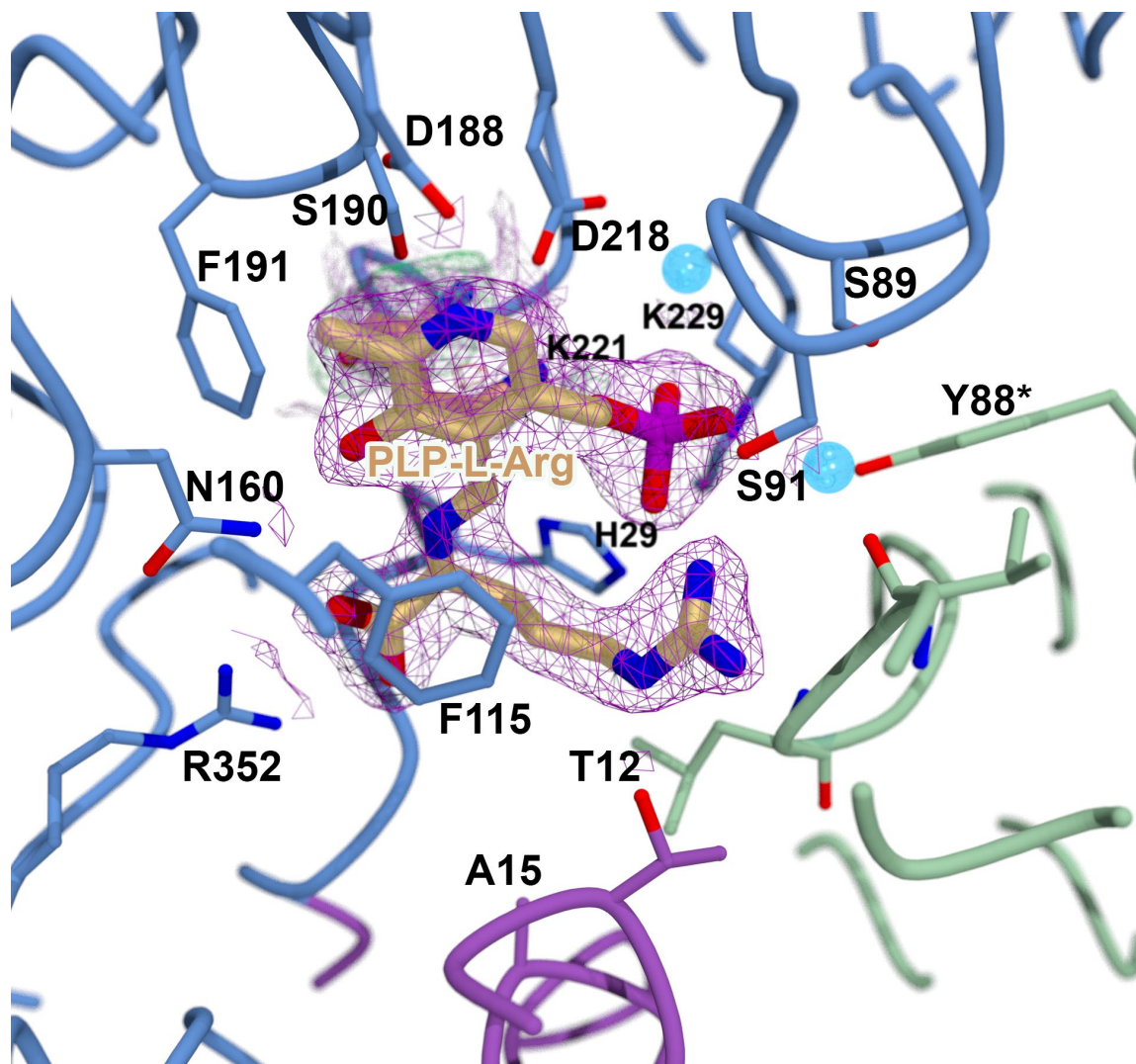
Figure 5-15: The active site of E15A•L-Arg. The N terminus colored in purple. Chain A colored in blue, chain B colored in light green. PLP-L-Arg is shown in golden stick. The $2|F_o| - |F_c|$ electron density map contoured at 1.0σ, shown as magenta mesh, and the $2|F_o| - |F_c|$ composite omit map also contoured at 1.0σ, is shown as green mesh.

5.3.8.3 The steady state parameters monitored by O2 consumption

SwMppP reacts with L-Arg and O_2 forming **2** and **3**. The catalytic parameters (K_M, k_{cat}) of O_2 usage were determined and are listed in Table 5-4.

Table 5-4: Steady state parameters monitored by O_2 consumption

Proteins	$K_{M,L-Arg}$ (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1} s^{-1}$)
SwMppP	26.5 ± 4.2	0.20 ± 0.006	7641.5 ± 1232.8
T12A	105.8 ± 12.1	0.05 ± 0.0016	483.2 ± 57.4
E15A	28.0 ± 4.0	0.24 ± 0.006	8424.4 ± 1212.7
E15Q	19.0 ± 3.5	0.24 ± 0.007	12538.3 ± 2340.2
T12A/E15A	74.5 ± 7.7	0.035 ± 0.002	465.2 ± 54.3
Δ1-22	234.8 ± 33.4	0.072 ± 0.002	306.0 ± 44.6
H29A	25.4 ± 5.6	0.11 ± 0.004	4472.9 ± 1003.2
H29S	16.9 ± 3.6	0.07 ± 0.003	4327.6 ± 931.3
F115Y	66.5 ± 4.8	0.20 ± 0.003	2940.3 ± 217.7
F191Y*	8.9 ± 2.4	0.05 ± 0.003	4983.2 ± 1349.4
D218S	353.2 ± 53.5	0.07 ± 0.003	183.5 ± 28.7

* means inaccurate data due to the instability of protein F191Y.

The K_M of the T12A and T12A/E15A mutants were 3- to 5-fold higher than that of wild type. The turnover numbers (k_{cat}) of the two mutants were also lower than the wild type, which makes the catalytic efficiency (k_{cat}/K_M) ~15-fold less than the wild type. In Figure 5-16, the spectra of both the T12A and T12A/E15A variants exhibit a shift from the external

aldimine peak (435nm) back to the internal aldimine peak (415 nm) as the reaction goes. The truncation mutant, however, showed an approximately 10-fold increase in K_M and a 3-fold decrease in k_{cat}, which reduced the pseudo-second order rate constant by almost 25-fold. The truncation mutant also produced only 2-ketoarginine. According to the spectrum of SwMppP$_{\Delta 1-22}$ with L-Arg and O_2 (Figure 5-16), the external aldimine also shifts into internal aldimine as the reaction goes. The mutants E15A and E15Q have no effect on K_M, k_{cat} or catalytic efficiency. Interestingly, although the steady state kinetics (as measured by dioxygen consumption) is indistinguishable for the E15A and E15Q variants, the E15A variant does not produce **2**, only the abortive product **3** based on the NMR results (Table 5-4).

Alteration of the N-terminus altered the ratios of products produced, but had no, or only a modest effect on the steady state kinetics as monitored by dioxygen consumption. However, as the spectra shown in Figure 5-16 demonstrate, the external aldimine peak (435 nm) gradually shifted to internal aldimine peak (415 nm) as reaction went, demonstrating that the N-terminal helix, especially T12, plays a key role in stabilizing the external aldimine and Quinonoid intermediate.
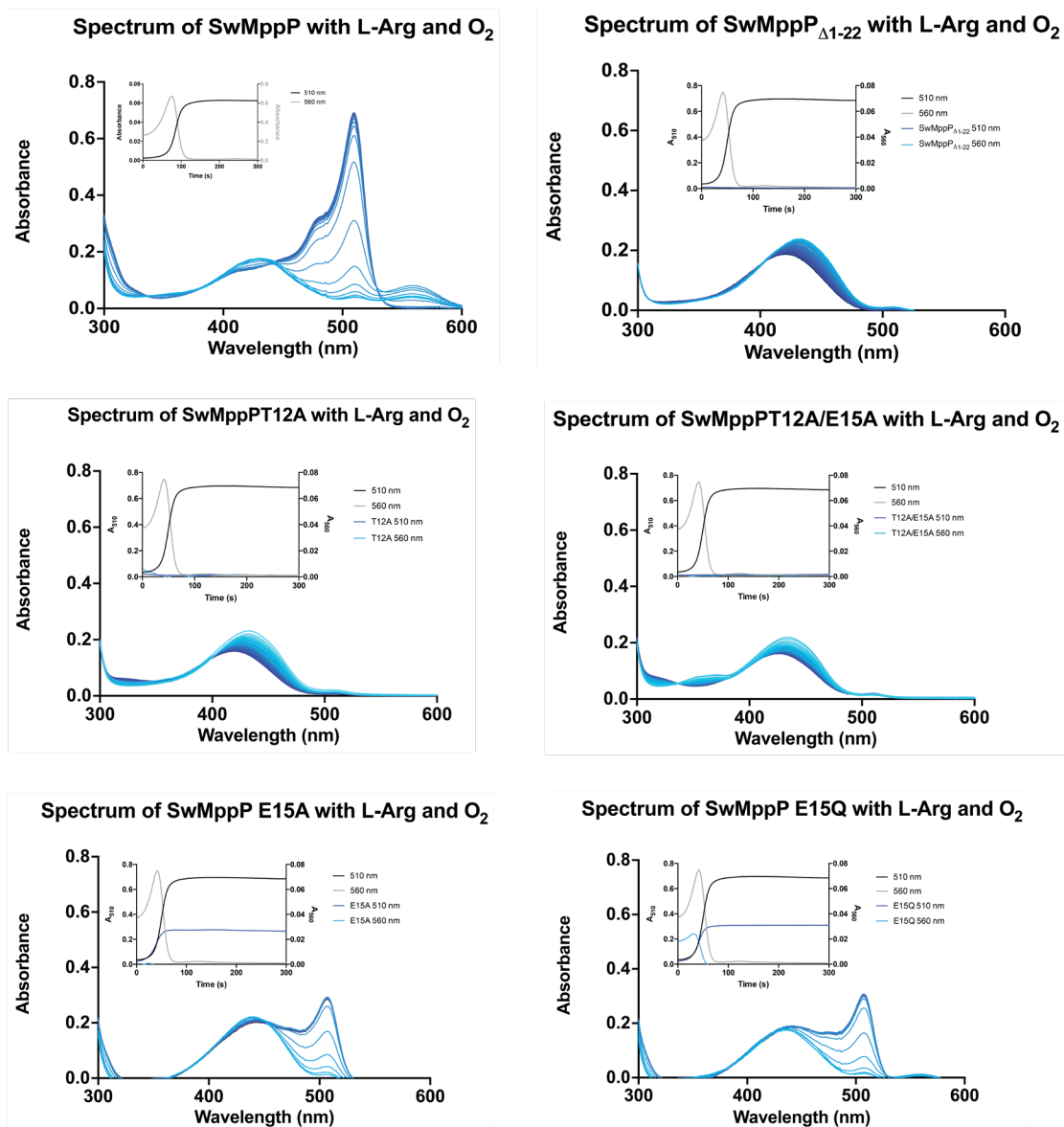
Figure 5-16: The spectrum (300 nm-600 nm) of SwMppP and its N terminus mutants with L-Arg. The insets are the changes at 510 nm and 560 nm for each mutant. There are 3 major peaks (430 nm, 510 nm, 560 nm) for wt SwMppP with L-Arg. The blue lines from lighter to darker represents the data collect from 1.4 s to 300 s. The insets also describe the trends of these mutants reactions in a timer order in comparison with wt SwMppP. For the truncated N-terminal helix mutant SwMppP$_{\Delta 1-22}$, the only peak (430 nm) shifted to 420 nm as the reactions goes. For SwMppP (E15A), there are only 2 peaks (430 nm, 510 nm). But for the mutant SwMppP (E15Q), there are three peaks (430 nm, 510 nm and small 560 nm).

In unliganded protein structure, H29 is close to the catalytic K221 and make hydrogen bonding with D27. In the product **2** bound structure, H29 hydrogen bonds to the hydroxyl group. To study the role of this special histidine residue, it was substituted into aliphatic alanine and polar serine. The steady state parameters on O_2 consumption are very similar to wild type SwMppP (Table 5-4). The structures of H29A and H29S are both obtained. The structures and spectra of both mutants are identical, so only the spectrum of H29A is shown in Figure 5-17. Only the external aldimine and quinonoid I peak formed. No absorbance at 560 nm was observed. In the first 20 s, the Quinonoid I intermediate accumulated are more than wild type SwMppP (the inset in Figure 5-17A). However, the Quinonoid I intermediate accumulation rate and final amounts of H29A/S are less than those of wild type SwMppP. Overlaying the H29A structure with wild type SwMppP, (Figure 5-17B), there is no difference on the active site residues. However, in the H29A mutant active site, there are more water molecules around the PLP, which suggests that H29, together with F115 and F191, may be involved in the water exclusion in the active site. Given that H29 coordinates the hydroxyl group of **2** and the H29A/S mutants lose the hydroxylase activity, H29 may also be involved in positioning a water molecule for hydroxylation.
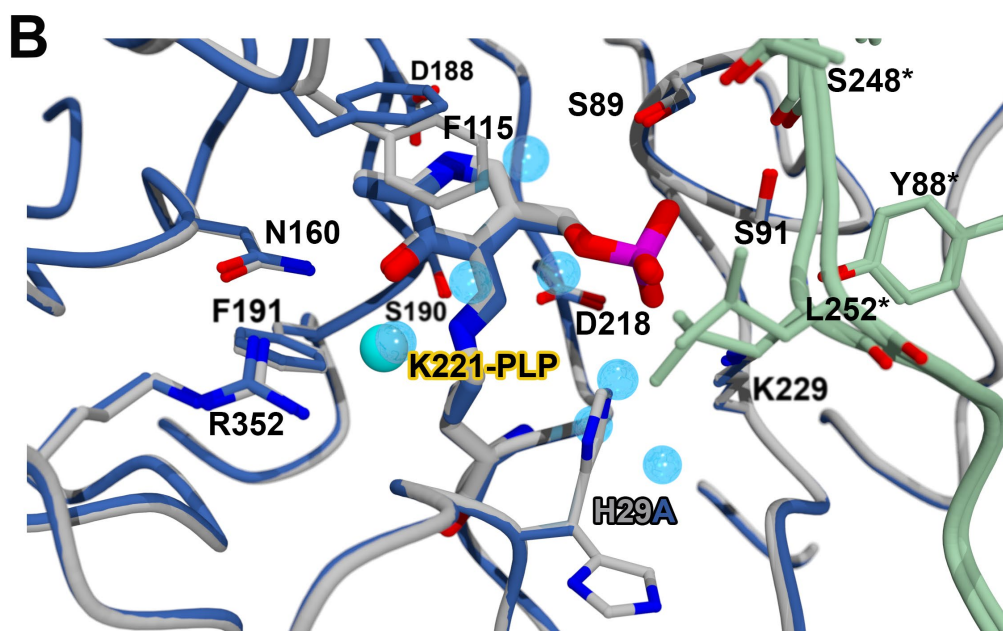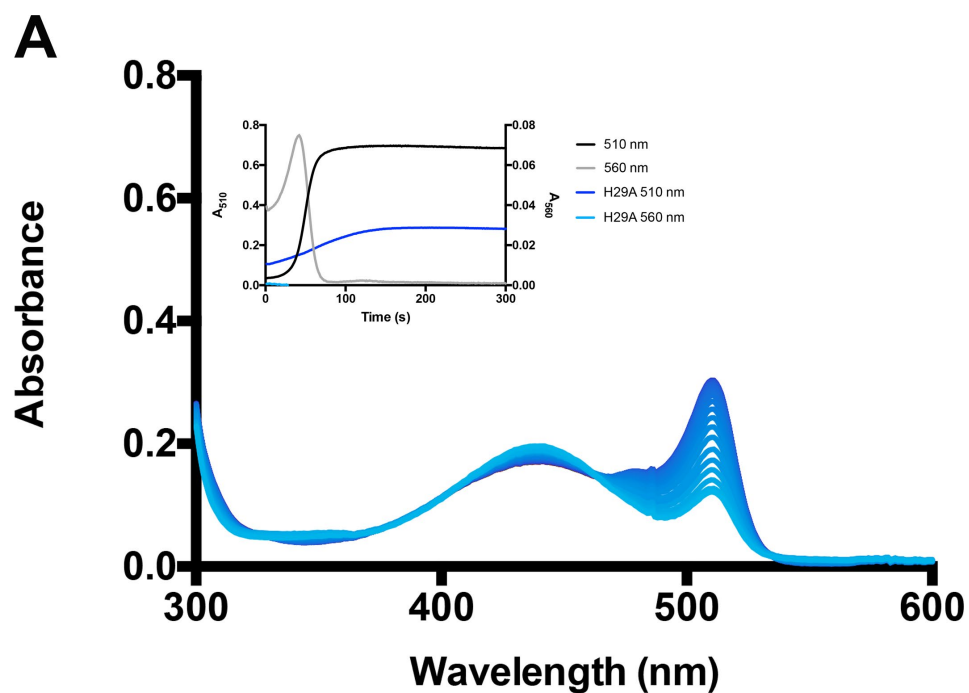
Figure 5-17: (A) The spectrum of H29A with L-Arg. The 700μl solution containing 30 μM H29A proteins with 1.4 mM L-Arg were scanned on diode array spectrophotometer every second over 300 s. The blue color gets darker as time goes. The inset is the 510nm 560 nm change comparison between SwMppP and H29A. (B)

Superposition of the active sites of SwMppP (colored in grey) with H29A (dark blue). Water molecules all from H29A are shown in blue sphere. Cl⁻ ion in SwMppP is shown as cyan sphere. In the active site, the H29A has more water molecules.

The structure of H29A with L-Arg was obtained by soaking 5 mM L-Arg. The asymmetric unit contained four molecules that form two independent homodimers. Three of the four chains have L-Arg bound forming external aldimine with PLP, while there is nothing bound in the active site of chain D and PLP covalently bound to catalytic K221 that forms internal aldimine. As mentioned above, there are more water molecules in the active site of H29A mutant. When the substrate L-Arg bound into the active site, L-Arg covalently bonds to PLP and excludes the water molecules. The water molecules in Figure 5-18A are from H29A internal aldimine structure. The ring of F151 moves close to L-Arg and packed with the Cα of L-Arg. The ring of F191 rotates about 90 degree and packed to the oxygen atom of carboxylate group. Except these two residues, other residues in the active site are same with apo enzyme SwMppP. Overlaying theH29A•L-Arg structure with SwMppP•L-Arg, the active site residues are exactly same (Figure 5-18B). However, the short side chain of A29 cannot exclude the water molecules from the active site, so there are two water molecules close to A29 that are not found in the structure of wild-type SwMppP with L-Arg bound.
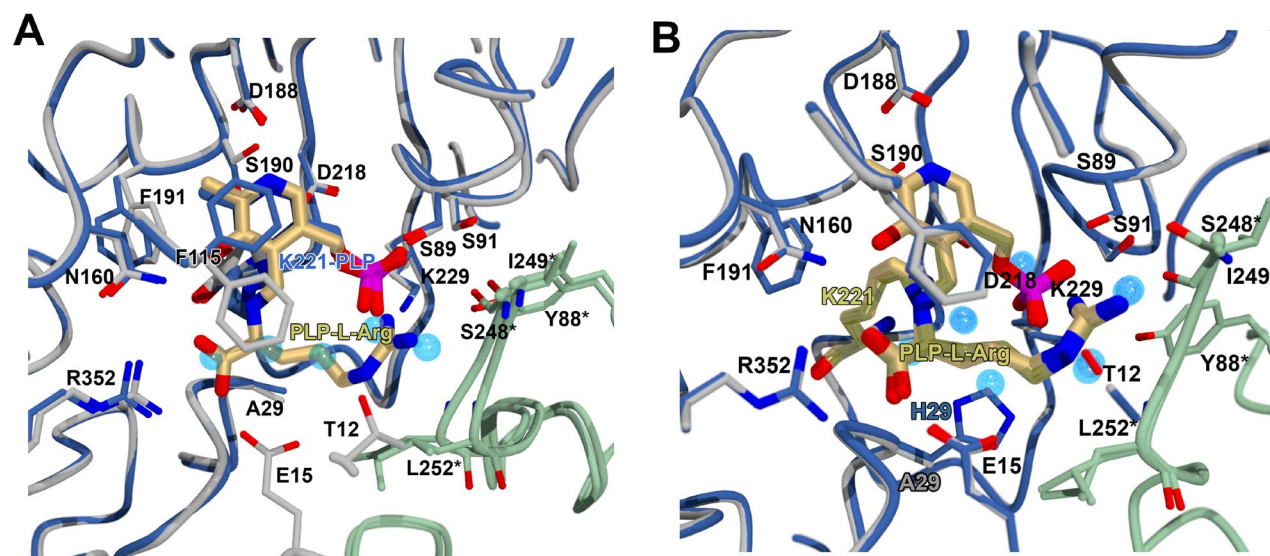
Figure 5-18: (A) Overlaying the active site of H29A (chain A colored in blue) and H29A•L-Arg (chain A colored in grey). Both chain B colored in light green. Upon the substrate bound, N terminus cover the active site. E15 and T12 interact with the carboxylate group and guanidinium group of L-Arg, respectively. The ring of F115 and F191 moves closer to the L-Arg. (B) Overlaying the active site of SwMppP•L-Arg (chain A colored in blue) and H29A•L-Arg (chain A colored in grey) demonstrates that there is no difference between them except two more molecules in the active site due to the short side chain of A29. Both chain B colored in light green.

F115 is another conserved residue in MppP homologs. To study the specific role of F115 play in the mechanism, it was mutated into tyrosine. The K_M of F115Y mutant were 3 folds higher than that of wild type SwMppP. The turnover numbers (k_{cat}) of F115Y is same with that of swMppP, which makes the catalytic specificity (k_{cat}/K_M) were ~3 fold less than the wild type SwMppP protein. As time goes, the external aldimine peak shifted back to internal aldimine peak (Figure 5-19A). The quinonoid I intermediate peak firstly accumulated, then decayed gradually. These features suggest that mutated Y115 probably

Figure 5-19: (A) The spectrum of F115Y with L-Arg. The 700μl solution containing 30 μM F115Y proteins with 1.4 mM L-Arg were scanned on diode array spectrophotometer every second over 300 s. The blue color gets darker as time goes. The inset is the 510nm 5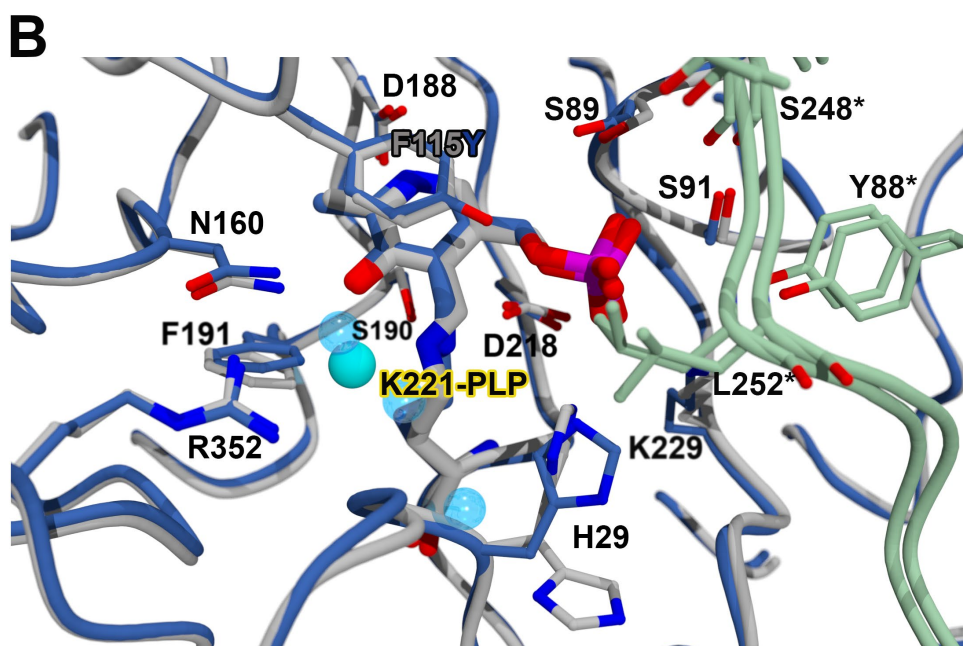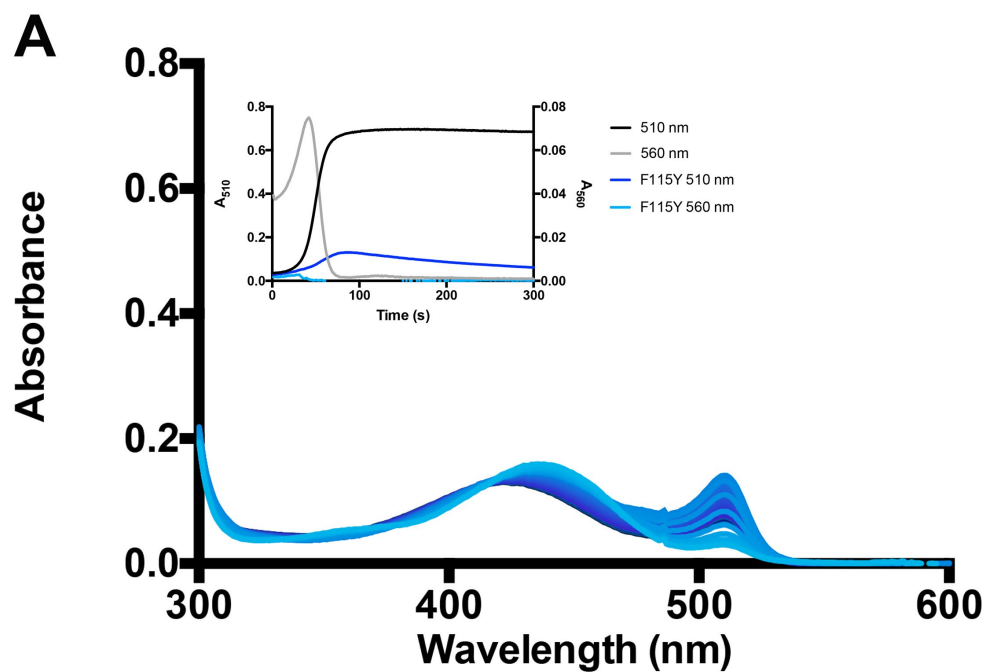60 nm change comparison between SwMppP and F115Y. (B) Superposition of the active sites of SwMppP (colored in grey) with F115Y (dark blue). Water molecules all

from F115Y are shown in blue sphere. Cl⁻ ion in SwMppP is shown as cyan sphere. In the active site, the F115Y has more water molecules.

disrupts the stability of Quinonoid I and external aldimine, thus the equilibrium shifted into internal aldimine. The quinonoid II intermediate did not accumulate. Overlaying the F115Y structure with wild type SwMppP (Figure 5-19B), there are no differences in the active site residues with the exception of H29. In the structure of SwMppP(F115Y), H29 only adopts one conformation. Water molecules align differently between the active site of chain A and chain B. In chain B, one water molecule mediates the interaction between OH of mutated Y115 side chain and NE2 of H29.

S190 and F191 are two adjacent conserved residues in MppP homologs. To study the specific role of S190 and F191, they were mutated into alanine and tyrosine, respectively. The purification of S190A failed because this mutant does not fold correctly and forms inclusion bodies. The mutant F191Y is also not stable due to the low T_m value mentioned above. The turnover numbers (k_{cat}) of F115Y is much slower than that of SwMppP. The K_M of F191Y mutant were 3 folds lower than that of wild type SwMppP. which makes the catalytic specificity (k_{cat}/K_M) were ~2 fold less than the wild type SwMppP protein. These parameters are based on the enzyme concentration, but the accurate concentration of active F191Y enzyme in the solution is unknown due to the instability property. The F191Y mutant is not stable, so I did not get the protein crystallized. As shown in Figure 5-20, there are four small peaks (358 nm, 430 nm, 510 nm, 560 nm) formed in the reaction of F191Y with L-Arg. The Quinonoid I intermediates accumulate much less than that in SwMppP,

while the Quinonoid II intermediates accumulates much more than that in SwMppP (Figure 5-20 inset). The peak with maximum absorbance at ~358 nm probably corresponds to ketimine intermediate that forms in aminotransferase (Figure 4-5). This peak first accumulated and then decreased as time goes. The position of F191 is usually tyrosine in aminotransferase (Figure 4-7C). These facts suggest that F191 is probably involved in exclusion of water close to C5. Substitution of F191 into tyrosine, the water may exist close to C5 and results in the ketimine intermediates formation.
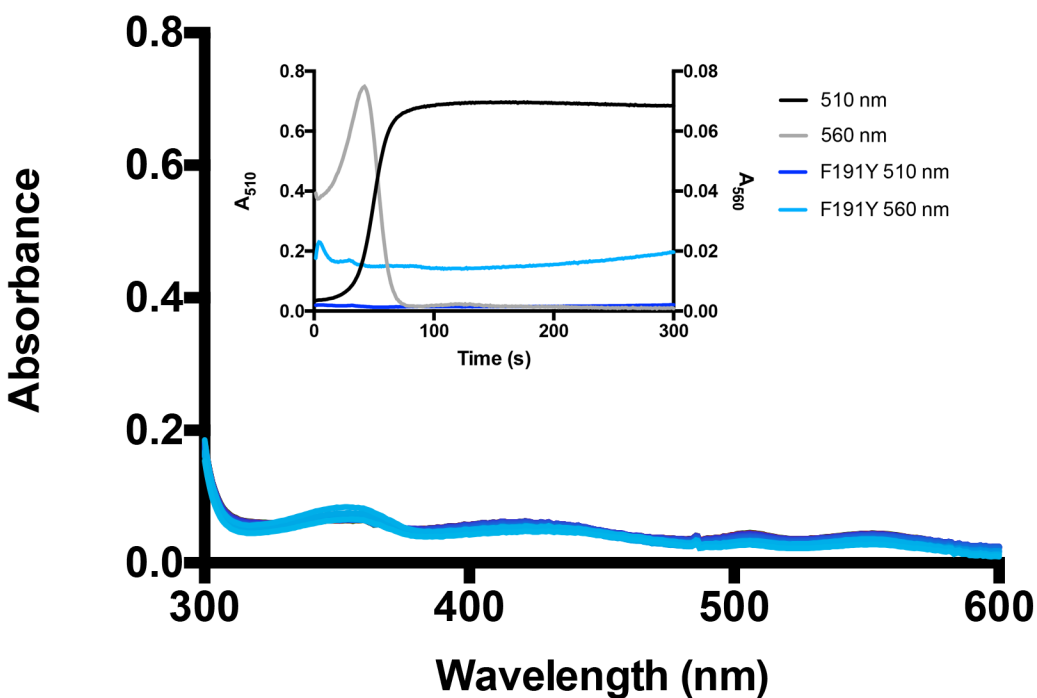
Figure 5-20: The spectrum of F191Y with L-Arg. F191Y enzyme (30 μM) was mixed with 1.4 mM L-Arg and recorded every second in 300s. There are 4 small peaks (358 nm, 430 nm, 510 nm, 560 nm) formed. Based on the thermo-Fluor assay results, F191Y mutant is not stable, which probably resulted in the small peaks.

In the PLP bound internal aldimine structure of SwMppP, D218 hydrogen bonding to
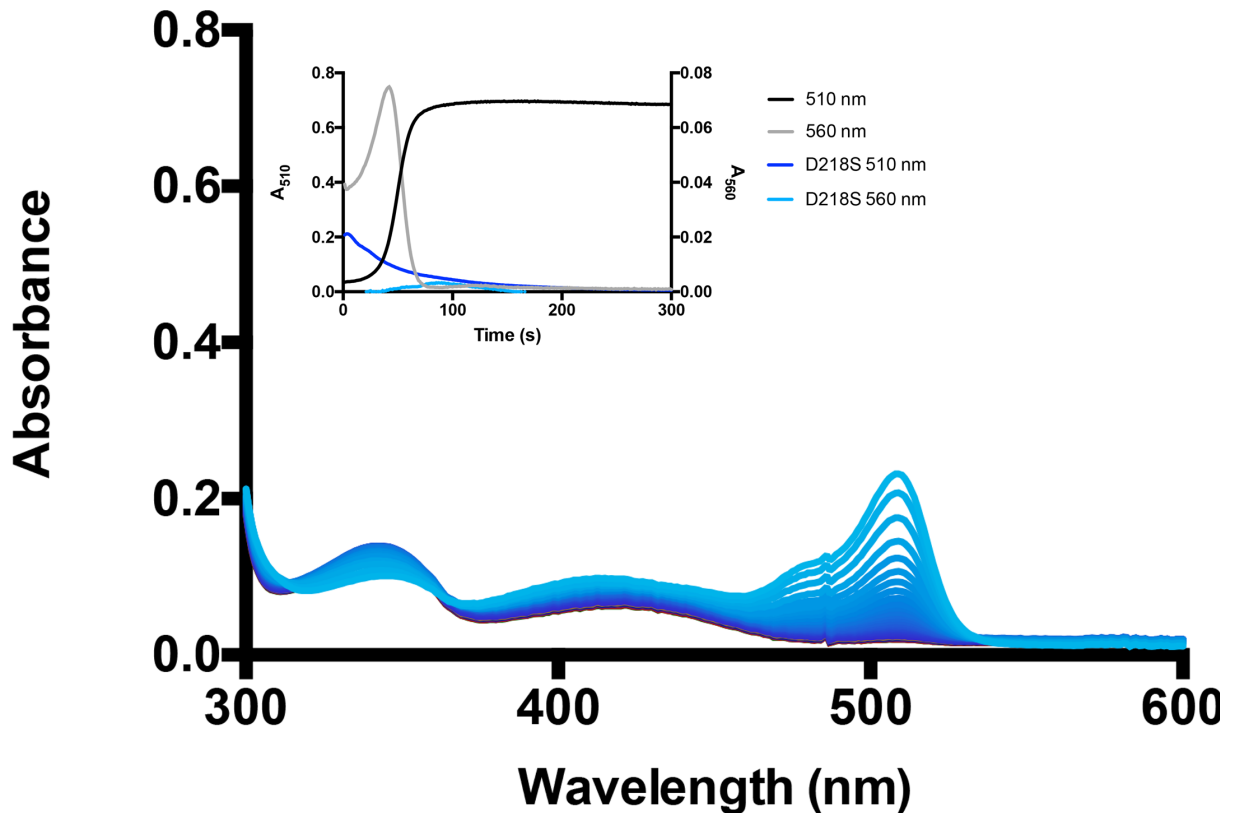
Figure 5-21: The spectrum of D218S with L-Arg. D218S enzyme (30 µM) was mixed with 1.4 mM L-Arg and recorded every second in 300s. There are 3 peaks (344 nm, 430 nm, 510 nm). The inset is the 510 nm and 560 nm comparison with wild type SwMppP.

the OH of S190, C5 of PLP and the main chain amine of T219 and G220. In the substrate bound external aldimine structure, D218 also interacts with the free K221. The SwMppP protein is yellow with PLP bound. However, the color of mutant D218S is less yellow even they are stored in the same buffer. The structure of D218S with poor electron density map showed that there is no PLP in the active site. The crystallization condition needs to be optimized for better diffraction. The spectrum of D218S demonstrates that there are three peaks (maximum absorbance at 340 nm, 430 nm and 510 nm) formed. As shown in the

inset in Figure 5-21, the external aldimine and Quinonoid I peaks decrease as time goes, while the peak with maximum absorbance at 340 nm keeps increase. This peak range from 310 nm to 370 nm may be pyridoxamine phosphate (PMP) or ketimine. Given that no PLP bound in the D218S active site and little products formed by NMR, the peak probably represents PMP (or both by hydrolyzing ketamine into PMP). If PMP cannot be hydrolyzed into PLP and form internal aldimine with K221, the PMP may dissociate out of the active site. So the D218 probably participates in PMP hydrolysis or internal aldimine formation.

5.3.9 Full oxidation of L-Arg requires two equivalents of O_2

The stoichiometry of O_2 consumed and L-Arg added was explored by O_2 electrode (Table 5-5). For wild type SwMppP producing **2** and **3** in the ratio of 1.7:1, the ratio of [O_2] consumed to [L-Arg added] is 1.4:1. The variant E15A, which only produces one product **3**, only needs 1 equivalent of O_2. But for E15D that producing both products, the ratio of [O_2] consumed to [L-Arg added] is >1, suggesting that the **2** branch requires more than one (probably two) equivalent of O_2 and the **3** branch only needs one. The ratios for mutants T12A, T12A/E15A and SwMppP$_{\Delta1-22}$ were lower than 1:1, which was probably due to the lower catalytic efficiency as shown in Table 5-4.

For wild-type SwMppP, E15Q and F191Y, producing both products, the ratio is > 1:1. And for those mutants only forming keto-Arg as the product, the ratio is ≤ 1:1. Because both **3** and **2** are the products of SwMppP with L-Arg and O_2, we assumed that **3** pathway only consumed one equivalent of O_2 and **2** pathway consumed two equivalents of O_2.

239

Table 5-5: The ratio of [O_2 consumption]/[L-Arg added] of SwMppP and its mutants reactions

[L-Arg]/μM	SwMppP	T12A	E15A	E15Q	T12A/E15A	H29A	H29S	F115Y	F191Y	D218S	SwMppP$_{\Delta 1\text{-}22}$
25	1.40:1	0.64:1	1.05:1	1.44:1	0.76:1	0.70:1	0.84:1	0.85:1	1.54:1		0.61:1
50	1.43:1	0.54:1	0.95:1	1.28:1	0.65:1	0.77:1	0.82:1	0.77:1	1.38:1	little activity	0.45:1
100	1.38:1	0.59:1	0.95:1	1.29:1	0.72:1	0.81:1	0.80:1	0.80:1	1.37:1		0.47:1

5.3.10 Substrate specificity

Reacting with L-Arg and O_2 showed the Cγ-hydroxylation activity of SwMppP[27]. To check if the L-Arginine precursors L-ornithine and L-citrulline are the possible substrates, the hydroxylation reaction was determined by spectrometric scan of the intermediates and MS analysis of the products. No intermediate peak formed by spectroscopy (Figure 5-22), even the external aldimine peak. As shown in Figure 6A, when SwMppP reacted with L-ornithine, the broad internal aldimine peak became smaller and a new peak at ~320 nm showed up. When SwMppP reacts with L-citrulline, the broad internal aldimine peak got smaller and red shifted (~425 nm to ~410 nm) as shown in Figure 5-22B. To detect the substrate specificity in more detail, the reactions of SwMppP with L-Arg derivatives L-canavanine, methyl-L-arginine and D-arginine were also determined (Figure 5-22). Upon D-Arg addition, the external aldimine formed and gradually shifted back holoenzyme (internal aldimine). There is no intermediate and product formed by NMR and MS (data not shown). The inhibitory constant (K_i) of D-arginine is 325 μM by monitoring the fluorescence signal at 415 nm on stop-flow. But for the L-canavanine reaction, the intermediate quinonoid I, which has maximum absorbance at 510 nm, accumulated when O_2 was used up in the cuvette. And there is one species formed at ~370 nm and decayed when another species formed which have broad absorbance around 330 nm. After several hours, the quinonoid I intermediate peak disappeared and two broad peaks at ~330 nm and ~425 nm still exist. For the reaction of SwMppP with methyl-L-arginine, quinonoid I peak quickly accumulated and gradually decreased. The NMR results showed that only keto-methyl-L-arginine formed. SwMppP did not catalyze the hydroxylation of methyl-L-arginine. Apparently, SwMppP exhibits very strong substrate specificity for hydroxylation.
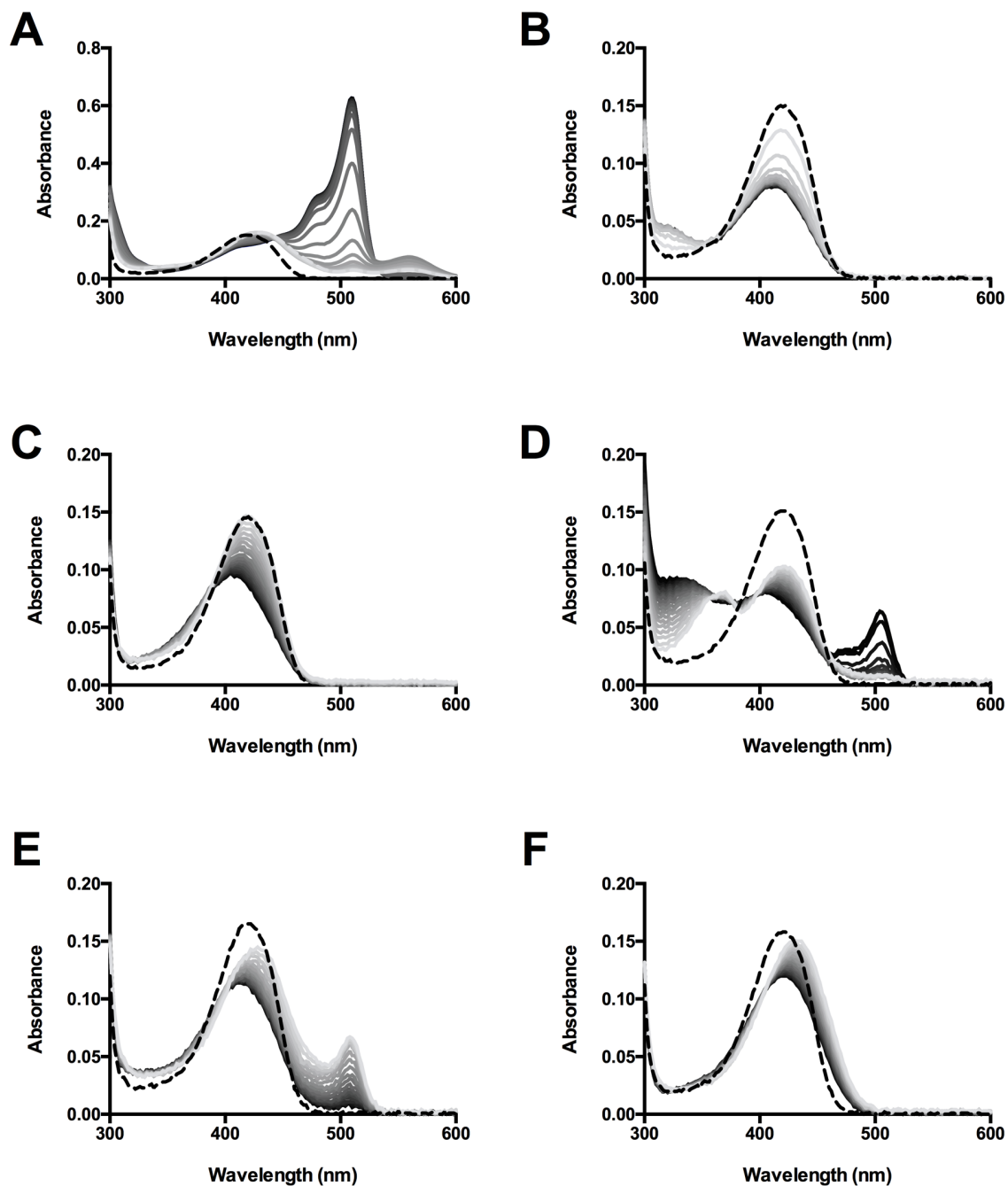
Figure 5-22: The spectrum of SwMppP with L-arginine (A), L-ornithine (B), L-citrulline (C), L-canavanine (D), methyl-L-arginine (E) and D-arginine (F). The spectrum of enzyme only was shown as dash line. After compounds addtion, the spectrum was recorded (from light grey to black).

5.3.11 The structure of SwMppP with D-Arg

As shown in Figure 5-22F, D-Arg is not the substrate of SwMppP, but it can bind to PLP forming external aldimine (peak shiftment from internal aldimine to external aldimine). Based on this, SwMppP crystals were soaked in 35 mM D-Arg to obtain the SwMppP•D-Arg external aldimine structure (Figure 5-23). The active site explains why D-Arg is not a substrate: the α-proton points toward the bulk solvent and away from any suitable general base catalyst (probably K221). Except the aldimine covalent bond, the carboxyl group of D-Arg forms salt bridges with R352 (Figure 5-23B). The guanidine group interacts with D27 and D227 and located in the position of N-terminal helix lid, which disrupts the N terminus covering the active site. Hence, unlike the L-Arg binding, N terminus disorders in the structure of SwMppP•D-Arg. Comparison of the SwMppP•L-Arg (Figure 5-11) and SwMppP•D-Arg (Figure 5-23), there are less interactions from the other chain in SwMppP•D-Arg. Because the side chain of L-Arg well aligned in the active site, the S248* and D249* from the other chain interacts with the guanidinium group of L-Arg. And the S91 also interacts with the guanidinium group. However, the side chain of D-Arg points away from the active site, so it loses these interactions.
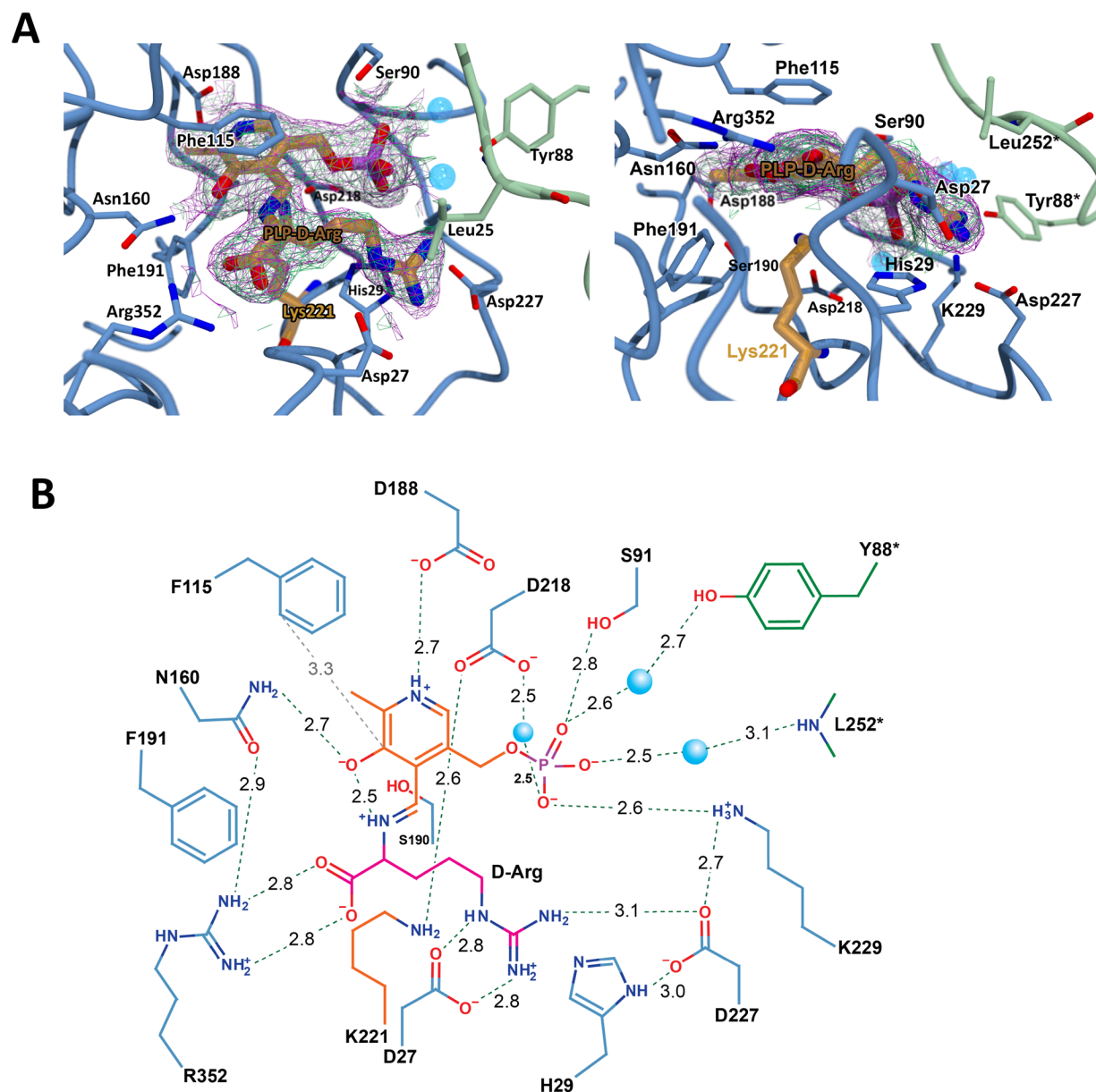
Figure 5-23: The active site of SwMppP•D-Arg. (A) SwMppP•D-Arg active site showing the PLP covalently bound to the enzyme as the internal aldimine. The $2|F_o| - |F_c|$ electron density map contoured at 1.0σ, shown as magenta mesh, and the $2|F_o| - |F_c|$ composite omit map also contoured at 1.0σ, is shown as green mesh. (B) Schematic view of the active site showing potential hydrogen bonding interactions as green, dashed lines. Water molecules are shown as blue spheres. Panel A was generated using POVSCRIPT and POV-Ray. Panel B was created using a combination of MarvinSketch (https://www.chemaxon.com) and Adobe Illustrator CS6.

244

D-Arg, as a competitive inhibitor, forms external aldimine with PLP in SwMppP active site. The dissociation constant of D-Arg (Ki) is determined by both O_2 probe assay and fluorescence change at 415 nm on stop flow. For O_2 probe assay, 10 μM SwMppP firstly incubated with 0 μM, 250 μM, 500 μM, 1 mM, 2.5 mM, 5 mM and 10 mM D-Arg in 50 mM BisTris propane pH 9.0 buffer for 5 min. For each data set (e.g 10 μM SwMppP with 250 μM D-Arg), a series of L-Arg were added and the O_2 consumption rate was determined. The result is shown in Figure 5-24. Each data set of rate vs L-Arg concentrations was fitted into Equation 2. Reciprocal plot of resulting v_{max} and L-Arg concentration gives a series of linear plot at different D-Arg concentrations. The slope is $K_M/v_{max}(1+[D-Arg]/K_i)$, and the intercept is $1/v_{max}$. Linearly fitting the slope vs [D-Arg] data gives K_i of 253.4 μM.

.
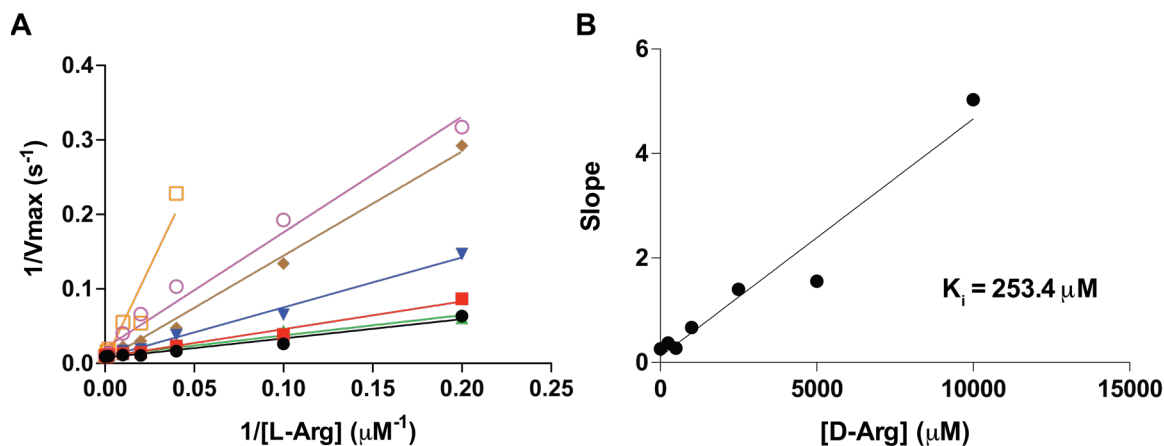
Figure 5-24: The D-Arg dissociation constant (Ki) determination. (A) The reciprocal plot of Vmax and L-Arg concentration gives a series of linear plot at different D-Arg concentrations. ● represents 0 μM D-Arg. ■ represents 250 μM D-Arg. ▲ represents 500 μM D-Arg. ▼ represents 1000 μM D-Arg. ◆ represents 2500 μM D-Arg. ○ represents 5000 μM D-Arg, □ represents 10000 μM D-Arg. (B) The slope from (A) was plotted versus D-Arg concentration and the data is fitted linearly. The absolute x-intercept is the Ki (253.4 μM).

Fluorescence titration experiments were conducted to measure the kinetics of external aldimine formation step with D-Arg. SwMppP (10 µM after mixing) was mixed with different concentrations of D-Arg (40 µM, 80 µM, 160 µM, 320 µM, 640 µM and 1280 µM) and the total fluorescence was measured with excitation at 415 nm. The results are shown in Figure 5-25. The fluorescence data of SwMppP with different concentrations of D-Arg only have one phase and were fitted into Equation 3. The dependence of kobs were fitted into Equation 4 and gave $K_{d,app}$ of 329.3 µM.
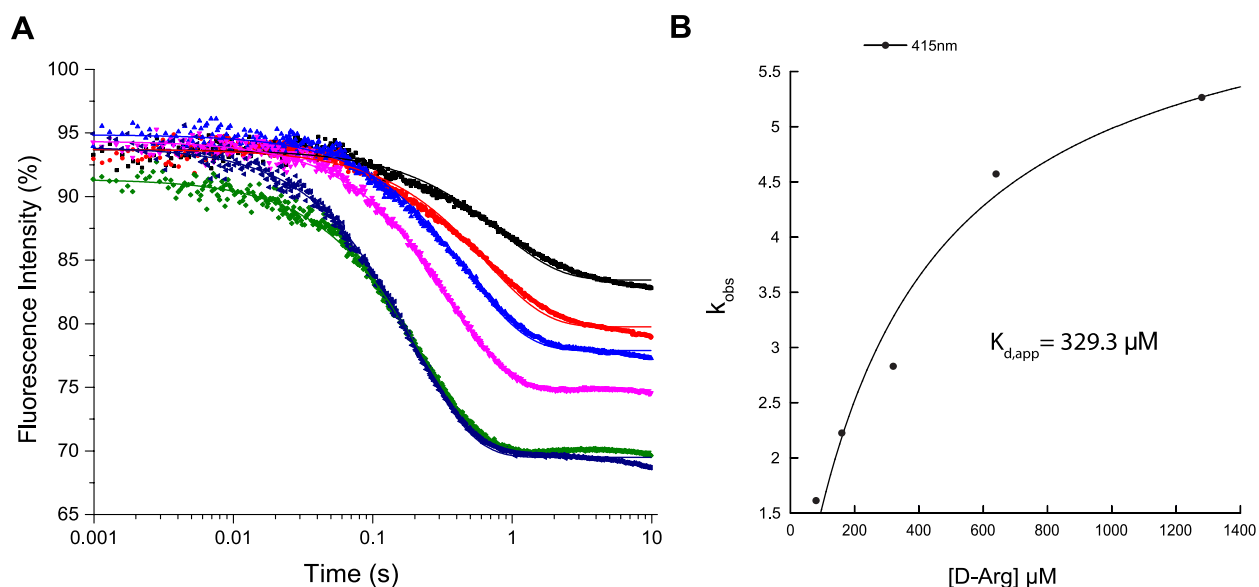
Figure 5-25: Dissociation constant (K_i) determination of D-Arg into SwMppP by fluorescence determination excited at 415 nm. (A) Fluorescence signal fitting by first exponential. . ■ represents 40 µM D-Arg. ● represents 80 µM D-Arg. ▲ represents 160 µM D-Arg. ▼ represents 320 µM D-Arg. ◆ represents 640 µM D-Arg. ◄ represents 1280 µM D-Arg. (B) The plot of k_{obs} from data (A) versus D-Arg concentration gives the $K_{d,app}$ value of 329.3 µM.

5.3.12 Presteady state kinetics of SwMppP

Changes at the PLP center upon binding of L-Arg were monitored using stopped flow spectroscopy under anaerobic conditions, and exciting the internal aldimine at 415 nm. When the enzyme (10 μM after mixing) is mixed with various concentrations of L-Arg in the absence of dioxygen, there is a biphasic decay of fluorescence (Figure 5-26). The total amplitude change of the two phases is independent of the substrate concentration. The data were fitted to Equation 6 using Kinetic Studio (TgK Ltd.). The dependence of the observed rate constants k_1 for the first phase (60% amplitude at 640 μM L-Arg) was fitted linearly (Figure 5-26B). The forward rate constant k_1 and reverse rate constant k_{-1} of internal aldimine decay step are given by the slope and y-intercept, respectively. The external aldimine absorbs UV light at 425-430 nm, so the second decrease phase is probably corresponding to the step of internal aldimine to external aldimine. The dependence of the observed rate constants k_2 for the second phase (40% amplitude at 640 μM L-Arg) was fitted to the Equation 7. According to the hyperbolic fitting, K_{app} for external aldimine intermediate formation equals to 17.7, while the maximum rate constant is 2.3 s^{-1}.
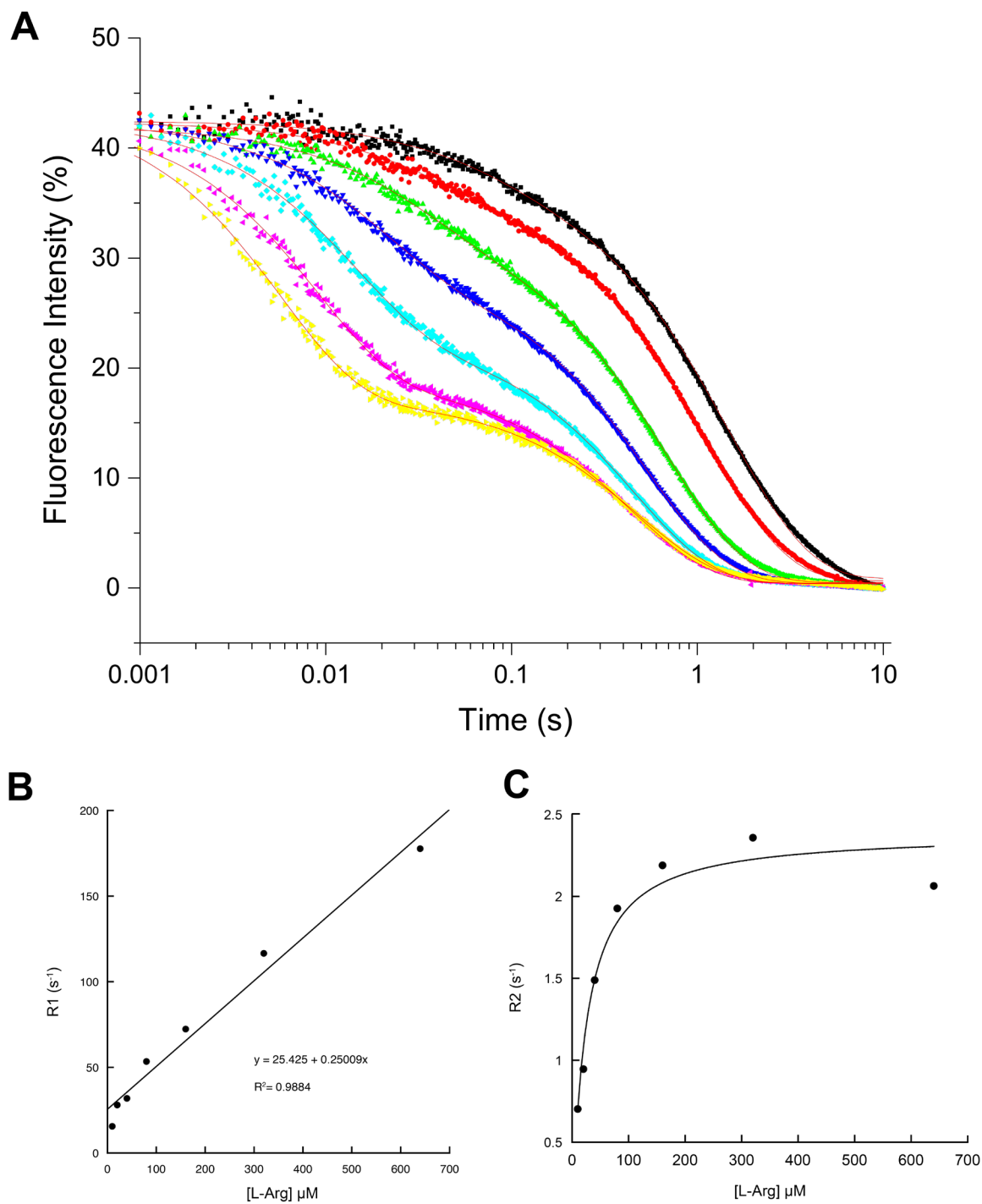
Figure 5-26: (A) Fluorescence intensity change upon L-Arg binding. The data is fitted into two exponential. (B) The plot of rate constant (R1) generated by (A) versus L-Arg concentration. ■ represents 10 μM L-Arg. ● represents 20 μM L-Arg. ▲ represents 40 μM L-Arg. ▼ represents 80 μM L-Arg. ◆ represents 160 μM D-Arg. ◀ represents 320 μM L-Arg. ▶ represents 640 μM L-Arg. The data is fitted linearly and gives k_1 equals 0.25 μM^{-1}

s^{-1} and k$_{-1}$ equals to 25.4 s^{-1}. (C) The rate constant (R2) for the second phase generated by (A) has a hyperbolic dependence on [L-Arg], saturating at 2.3 s^{-1} with an apparent equilibrium constant of 17.7.

The Quinonoid I intermediate accumulation was observed at the maximum absorbance at 510 nm under anaerobic condition. The experiment was set up similar with the method mentioned above, but collect the absorbance rather than fluorescence. The results are shown in Figure 5-27A. The Quinonoid I forms in a first order process with a ~100 ms lag that corresponding to the first phase and saturated when external aldimine totally decay observed in fluorescence data (Figure 5-29). The data are fitted into two exponential equation (Equation 6) using Kinetic studio (TgK Ltd.). The dependence of the observed rate constants k$_1$ for the tiny phase was fitted hyperbolically (Figure 5-27B). So the calculated equilibrium constant for this step is 44.2 with the maximum rate constant is 2.5 s^{-1}. The dependence of the observed rate constants k$_2$ for the largest amplitude phase was fitted to the Equation 7 (Figure 5-27C). In the equation, K$_{app}$ and k$_{max}$ are the apparent equilibrium constant and rate constant for the step of external aldimine into Quinonoid I intermediate, which match that in the second phase in the fluorescence experiment above. According to the hyperbolic fitting, K$_{app}$ in this absorbance data for external aldimine to Quinonoid I intermediate equals to 36.8, while the maximum rate constant is 9.8 s^{-1}.
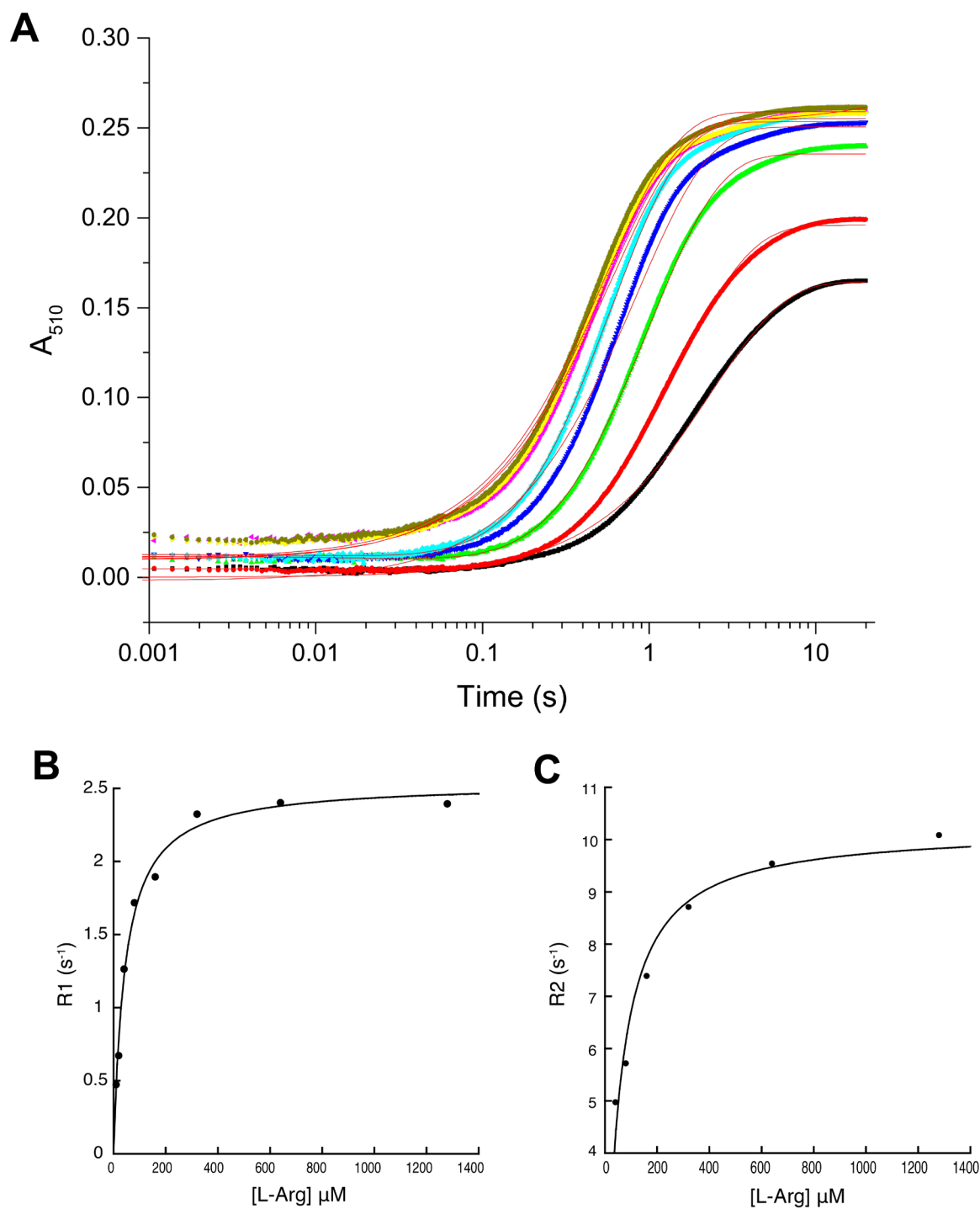
Figure 5-27: (A) Quinonoid I (maximum absorbance at 510 nm) accumulation curves with different L-Arg concentrations under anaerobic condition. ■ represents 10 μM L-Arg. ● represents 20 μM L-Arg. ▲ represents 40 μM L-Arg. ▼ represents 80 μM L-Arg. ◆ represents 160 μM D-Arg. ◄ represents 320 μM L-Arg. ► represents 640 μM L-Arg. ● represents 1280 μM L-Arg. The data is fitted into two exponential. (B) The plot of

250

rate constant (R1) generated by (A) versus L-Arg concentration. The amplitude of this data is negative and the rate constants show a hyperbolic fitting with equilibrium constant of 44.2. (C) The hyperbolic fitting of rate constant (R2) of the largest phase gives the rate constant of 9.8 s^{-1} with apparent equilibrium constant (K_{app}) of 36.8 for external aldimine to Quinonoid I step.

The reaction of Quinonoid I with O_2 and Quinonoid II formation were monitored by double-mixing stop flow spectrophotometry. Prepared anaerobic SwMppP (40 μM) in 10 mM MES pH 6.7, 20 μM PLP buffer was firstly mixing with 1280 μM L-Arg in 100 mM BisTris Propane pH 9.0 buffer and sit for 20 s prior to the second mix of different defined concentrations of dissolved O_2 in 100 mM BisTris propane pH 9.0 buffer. In future, to get a single turnover of Quinonoid I decay, more enzyme and less L-Arg should be used. Otherwise, the extra substrates would react with the recycled enzyme again, which would make the data too complicated to analyze. The dissolved O_2 concentration in the buffer was defined by sparging an inverted syringe containing 100 mM BisTris propane pH 9.0 buffer for 5 min with blended N_2 and O_2 of known partial pressures supplied by a Maxtec Maxblend gas blender. The concentration of dissolved O_2 was determined by firstly sparging the reaction chamber of a Hansatech dioxgen electrode filled with same buffer to define the equilibrium concentration of dissolved O_2. Then the Quinonoid I intermediate observed at 510 nm and quinonoid II intermediate observed at 560 nm was recorded upon the second mixing with varied O_2.
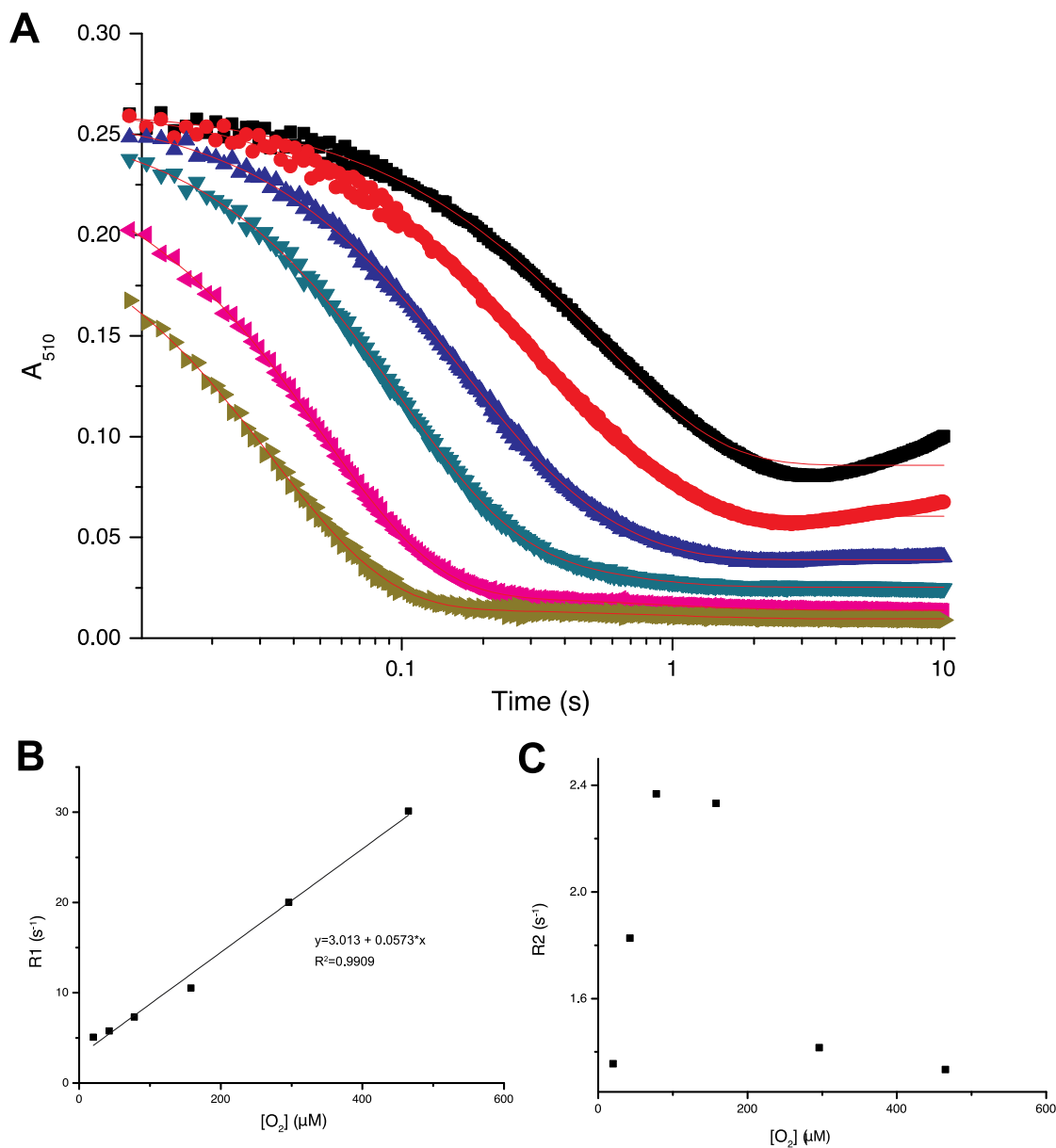
Figure 5-28: (A) Quinonoid I (maximum absorbance at 510 nm) decay curves with different O_2 concentrations by double mix. ■ represents 20 μM O_2. ● represents 40 μM O_2. ▲ represents 78 μM O_2. ▼ represents 158 μM O_2. ◄ represents 296 μM O_2. ► represents 465 μM O_2. The data is fitted into two exponential and the fitting line colored in red. The first mix was obtained by mixing 40 μM SwMppP with 2560 μM L-Arg. The second mix is mixing the first mixture (Quinonoid I) with different concentrations of O_2. (B) The plot of rate constant (k_1) generated by (A) versus varied O_2 concentrations. The data is fitted linearly and gives k equals to 0.0573 μM^{-1} s^{-1}. (C) The amplitude of this tiny phase cannot be fitted meaningfully.

The absorbance at 510 nm was monitored for 100 s and the curve first decay as the reaction goes, then it accumulated again due to the extra L-Arg catalyzed by recycled enzyme. The first 15 s data corresponding to one turnover was fitted to Equation 6. The dependence of the observed rate constants k_1 for the largest amplitude phase (99% at 465 µM O_2) was fitted linearly. The rate constant k of Quinonoid I react with O_2 step is the slope (0.0573 µM^{-1} s^{-1}). The first phase of Quinonoid I decay is the step before the Quinonoid II formation by overlay the figure of Quinonoid I decay and Quinonoid II formation (Figure 5-29). Hence, Quinonoid II only accumulates after Quinonoid I has decayed. The second phase rate constants with tiny amplitude change cannot be fitted meaningfully (Figure 5-28C).
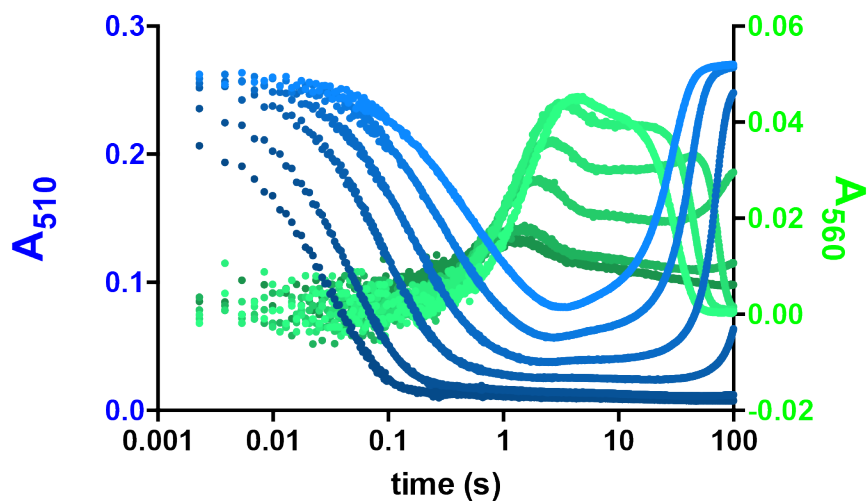
Figure 5-29: The phase comparison of Quinonoid I decay (A_{510}) and Quinonoid II formation (A_{560}). The Quinonoid I decay data colored in blue in a gradient of low O_2 concentration (20 µM, lighter blue) to high O_2 concentration (465 µM, darker blue). The Quinonoid II formation data colored in green in a gradient of low O_2 concentration (20 µM, lighter green) to high O_2 concentration (465 µM, darker green). The reaction of

Quinonoid I with O_2 is mostly complete before QII begins to appear at 560 nm.

The formation and decay with different concentrations of O_2 of Quinonoid II (maximum absorbance at 560 nm) was monitored for 100 s with same setting with the data of Quinonoid I decay at 510 nm. The absorbance at 560 nm first accumulated then decay, followed by a linear decrease, whose slope and time course depend on the O_2 concentration (Figure 5-29). Because the extra L-Arg exists, the absorbance at 560 nm goes up again. The first 15 s data (Figure 5-30A) probably corresponding to one turnover was fitted to Equation 8. The dependence of the observed rate constants k_1 for the increased, largest amplitude phase was fitted linearly (Figure 5-30B). The forward rate constant k_1 and reverse rate constant k_{-1} of Quinonoid II formation step are 0.0036 $\mu M^{-1} s^{-1}$ and 0.9215 s^{-1}, respectively. The dependence of second phase rate constants was fitted linearly (Figure 5-30C). The rate constant k of Quinonoid II with O_2 step is 0.0032 $\mu M^{-1} s^{-1}$. The dependence of third phase rate constants (Figure 5-30D) was fitted to Equation 7. According to the hyperbolic fitting, K_{app} in this absorbance data for VII to VIII intermediate equals to 53.7, while the maximum rate constant is 1.5 s^{-1}.
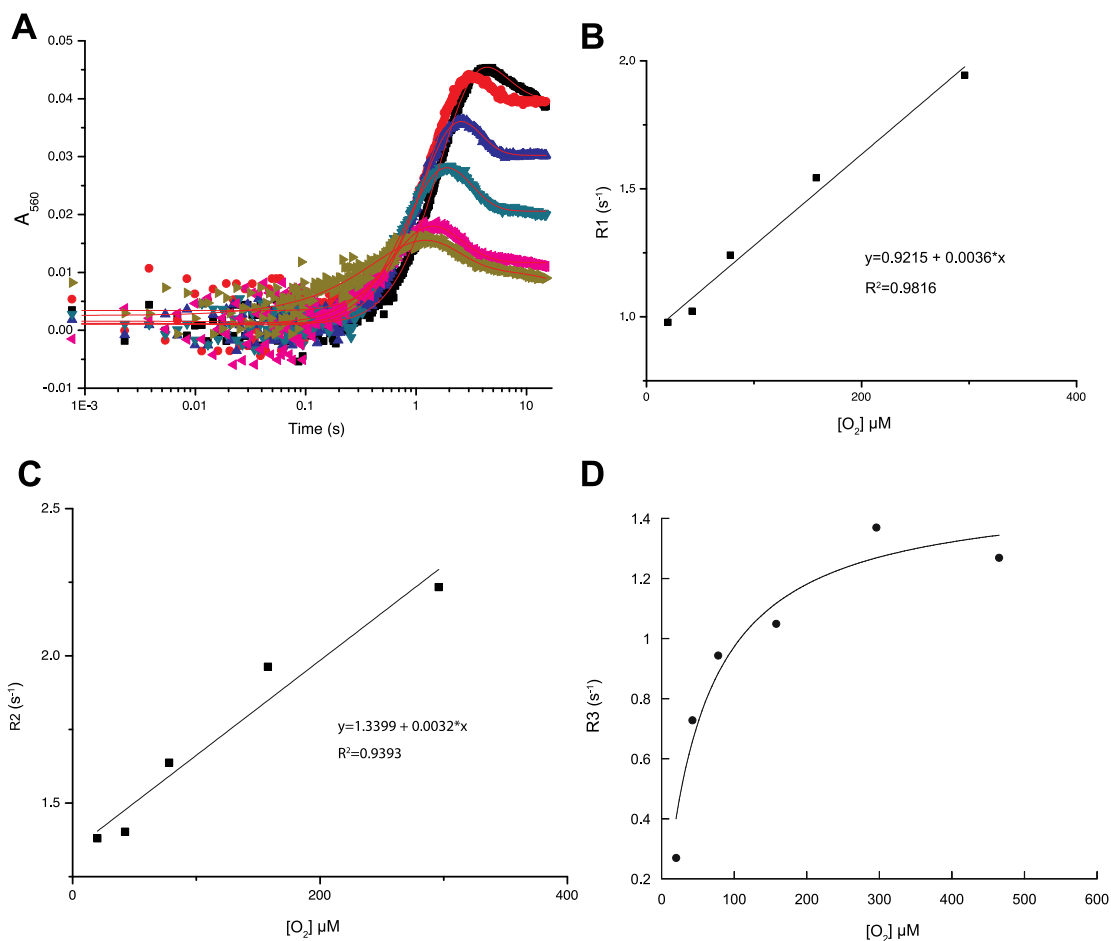
Figure 5-30: (A) Quinonoid II (maximum absorbance at 560 nm) formation curves with different O_2 concentrations by double mix stop flow. ■ represents 20 μM O_2. ● represents 40 μM O_2. ▲ represents 78 μM O_2. ▼ represents 158 μM O_2. ◄ represents 296 μM O_2. ► represents 465 μM O_2. The data is fitted into two exponential and the fitting line colored in red. (B) The plot of rate constant (k_1) generated by (A) versus varied O_2 concentrations. The largest amplitude phase corresponding to the Quinonoid II formation is fitted linearly and gives k_1 equals to 0.0036 μM^{-1} s^{-1} and k_{-1} equals to 0.9215 s^{-1}. (C) The plot of rate constant (k_2) generated by (A) versus varied O_2 concentrations. The data is fitted linearly and gives k equals to 0.0032 μM^{-1} s^{-1}. (D) The plot of rate constant (k_3) generated by (A) versus L-Arg concentration gives the K_{app} of this phase equals to 53.7 with the maximum rate constant is 1.5 s^{-1}.

Table 5-6: Summary of kinetic and equilibrium constant of SwMppP with L-Arg calculated above

Reactions	k^+ ($\mu M^{-1} s^{-1}$)	k^- or k (s^{-1})	$K_{d,app}$ (μM)	K_{app}	k_{max} (s^{-1})
E + L-Arg = E•L-Arg (I) E•L-Arg (I) = EA (II)	0.25 [a]	25.4 [a]	101.6 [a]		
EA (II)=Q1				17.7[b] 36.8 [c]	2.3[b] 9.8 [c]
Q1 + O₂ = IV	0.0573[d]				
IV = V				ND	ND
V = Q2 + H₂O₂	0.0036[e]	0.9215[e]			
Q2 + O₂ = VII	0.0032[f]				
VII = VIII				53.7 [g]	1.5 [g]

[a] represents the results from fluorescence data under anaerobic condition shown in Figure 5-26B.

[b] represents the results from fluorescence data under anaerobic condition shown in Figure 5-26C.

[c] represents the results from absorbance at 510 nm under anaerobic condition shown in Figure 5-27C.

[d] represents the results from absorbance at 510 nm under aerobic condition shown in Figure 5-28B.

[e] represents the results from absorbance at 560 nm under aerobic condition shown in Figure 5-30B.

[f] represents the results from absorbance at 560 nm under aerobic condition shown in Figure 5-30C.

[g] represents the results from fluorescence data shown in Figure 5-30D.

Note: The Roman labels correspond to those in Scheme 5-3 below. E represents wild type SwMppP here. EA represents external aldimine. Q1 represents Quinonoid I. Q2 represents Quinonoid II. ND: not determined.

5.3.13 Proposed mechanism of SwMppP

Based on all structural and kinetics studies shown above, we proposed the mechanism

of SwMppP with L-Arg and O_2 (Scheme 5-2). Upon the L-Arg bound, it attacks the internal
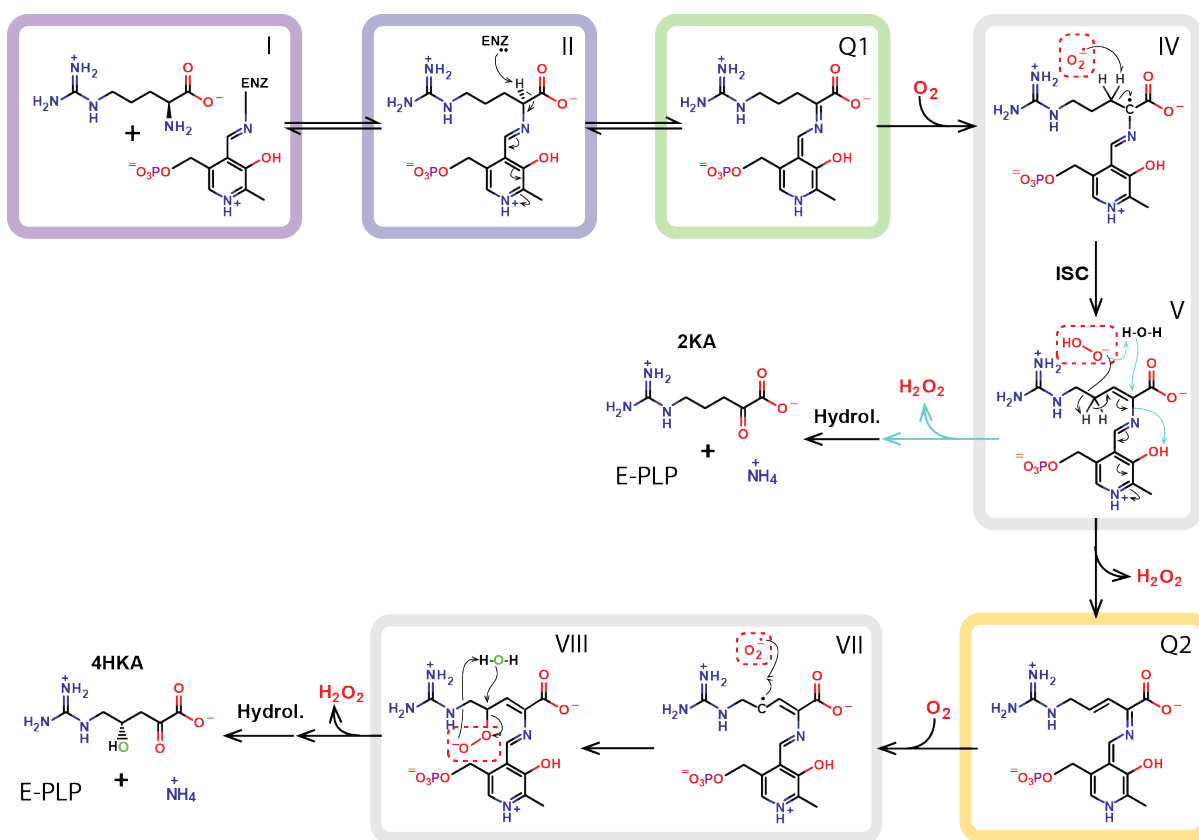
Scheme 5-2: Proposed mechanism of SwMppP.

aldimine, forming external aldimine. The free Lys221 on the enzyme would attack the H at Cα position, resulting in the formation of Quinonoid I (Q1) intermediate, which has maximum absorption at 510 nm. Those un-conserved residues probably contribute to stabilize the e⁻-rich intermediate Q1. In the presence of O_2, the H at β position is attacked by O_2 radical. After rearrangement, it would branch into 2 ways. If the water molecule closes to the hydroperoxide anion (HOO⁻), the anion would attack the H from water and release H_2O_2 and **3**. Final hydrolysis leads to the regeneration of PLP and release of ammonia. In

the **2** pathway, the hydroperoxide anion would attack the H at γ position, which results in the release of H_2O_2 and the formation of Quinonoid II (Q2). Another dioxygen molecule is involved in the oxidation of Q2. After several steps, **2** is formed and the second H_2O_2 molecule is released. The PLP is regenerated by hydrolysis of PMP.

5.3.14 Presteady state kinetics of SwMppP(H29A)

As shown in Figure 5-31, the first phase of SwMppP(H29A) is longer than that in wild type SwMppP. The 510 nm absorbance peak of H29A accumulated faster than that in wild type SwMppP, but accumulated much lower amounts of Q1.
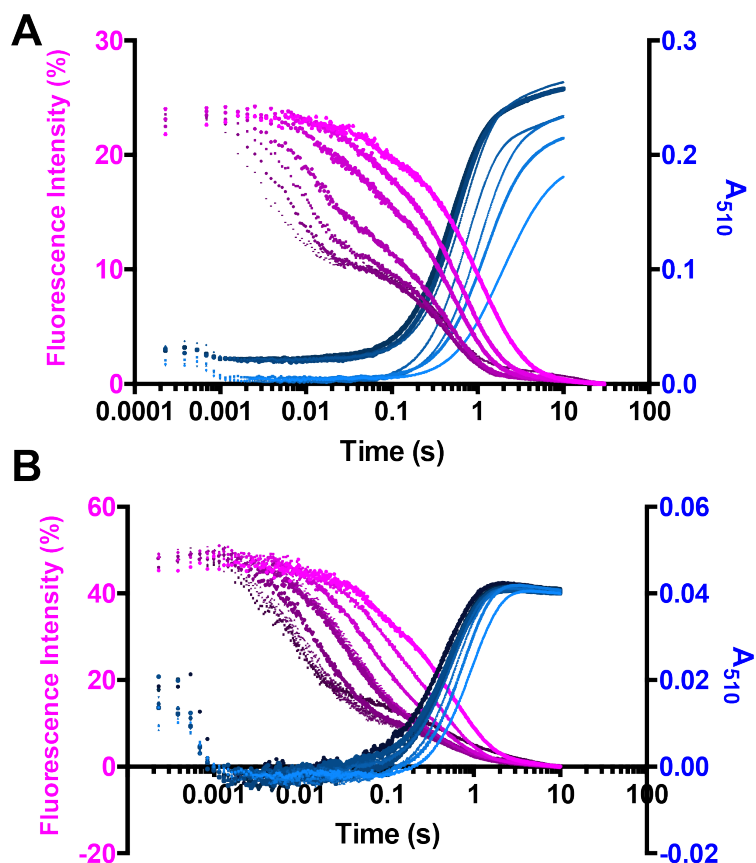
Figure 5-31: The phase comparison of fluorescence intensity data and absorbance at 510 nm of SwMppP with L-Arg (10, 25, 50, 125, 250, 500, 1000 μM)(top) with H29A with L-Arg (25, 50, 100, 150, 200, 400, 600, 800, 1200 μM) (bottom). Both 415 nm data sets colored in magenta in a gradient of low L-Arg concentration (lighter magenta) to high L-Arg concentration (darker magenta). Both 510 nm data sets colored in blue in a gradient of low L-Arg concentration (lighter blue) to high L-Arg concentration (darker blue). The first phase (internal aldimine to external aldimine) of H29A lasts longer than that of wild type SwMppP. However, the quinonoid I accumulated at very close time.

 The fluorescence and the absorbance at 510 nm data for SwMppP (H29A) was fitted to the mechanism depicted in Scheme 5-2 using KinTek Explorer simultaneously to obtain the dissociation constant and equilibrium constant for the four-step reactions under anaerobic condition. The global fitting figure is shown in Figure 5-32. The kinetic and equilibrium constants are summarized in Table 5-7.
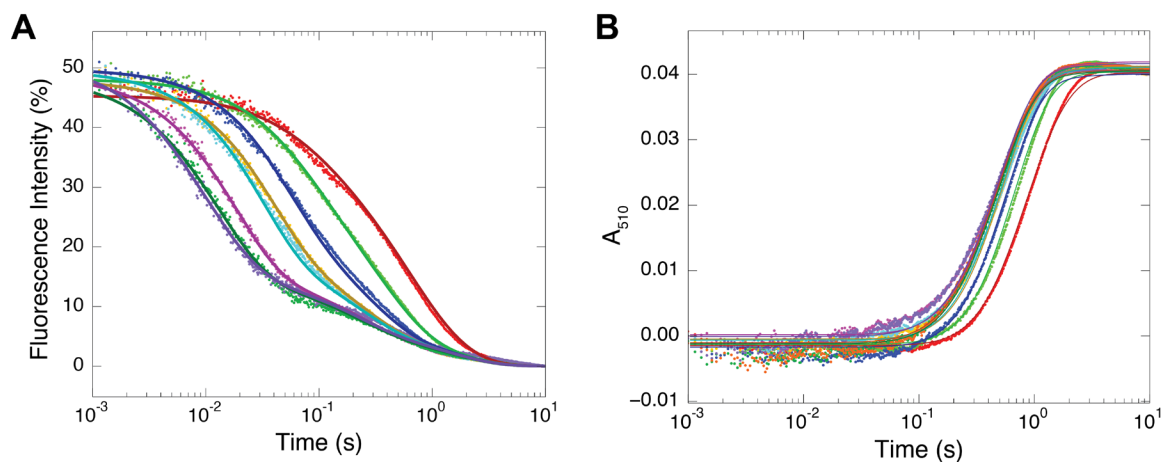
Figure 5-32: Fluorescence excited at 415 nm (A) and absorbance at 510 nm (B) of H29A mutant (10 μM) with varied L-Arg substrate (● 25, ● 50, ● 100, ● 150, ● 200, ● 400, ● 600, ● 800 μM) under anaerobic condition were fitted globally in KinTek Explorer. The data were shown in dots with different colors, while the fitting was shown in lines with same color with the data.

Table 5-7: Summary of kinetic and equilibrium constant of SwMppP(H29A)with L-Arg under anaerobic condition

Reactions	k^+ (μM^{-1} s^{-1})	k^- (s^{-1})	K_d (μM)
E + L-Arg = E•L-Arg	0.252	79.8	316.7
E•L-Arg = EA	121	2.07	
EA=Q1	1.84	0.0682	

Note: E represents SwMppP (H29A) here. EA represents external aldimine, while Q1 represents Quinonoid I

5.3.15 Presteady state kinetics of SwMppP$_{\Delta 1\text{-}22}$

The fluorescence data for SwMppP$_{\Delta 1\text{-}22}$ was shown in Figure 5-33. The amplitude of fluorescence intensity is much lower than wild type SwMppP and SwMppP(H29A). I try to fit the data in Origin/Klaidagraph/KinTek Explorer with the mechanism depicted in Scheme 5-2, but it does not fit well. The global fitting figure is shown in Figure 5-33.
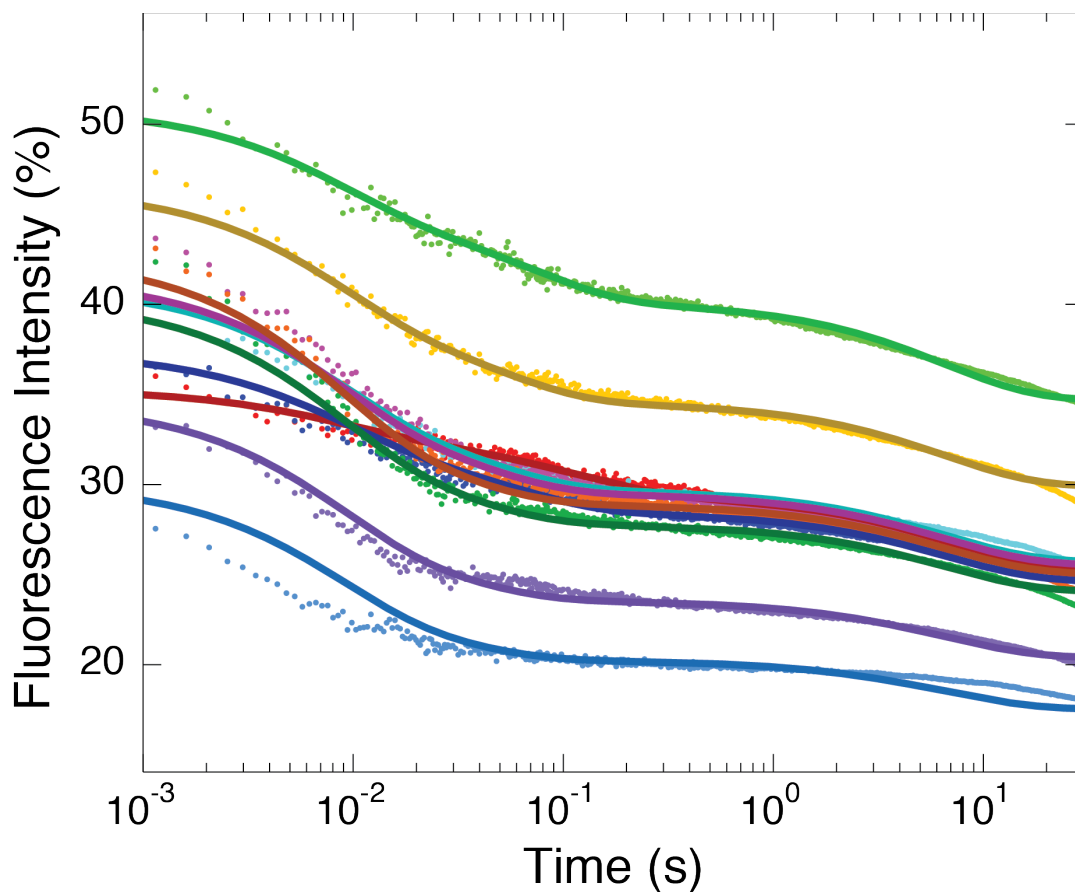
Figure 5-33: Fluorescence excited at 415 nm of SwMppP$_{\Delta1\text{-}22}$ with L-Arg substrate under anaerobic condition. The fluorescence data (dots) of SwMppP$_{\Delta1\text{-}22}$ (10 µM) versus different L-Arg concentrations (● 25, ● 50, ● 75, ● 100, ● 150, ● 200, ● 400, ● 600, ● 800 and ● 1200 µM) was fitted globally (lines with different colors) to the same model mentioned above.

5.3.16 Independent binding between two binding sites of SwMppP

The structures of SwMppP or mutant H29A/S with L-Arg shows two different states in the two binding sites within one dimer. For example, in the two active sites of SwMppP•L-Arg, chain A has the L-Arg bound to PLP while chain B has the product **2** bound. Some structures have **3** bound and internal aldimine state in chain A and chain B, respectively. In

the structure of H29A•L-Arg, chain A has L-Arg bound but chain B has PLP covalently bound to catalytic lysine. So to check if the two binding sites within one dimer are independent or not, ITC and fluorescence experiment that monitoring the internal aldimine decay was determined. SwMppP would form Quinonoid I eventually under anaerobic condition, which make the data harder to analyze. D188 hydrogen bonds to the NH of PLP ring, which is important for Quinonoid I formation. After mutated D188 into alanine, the variant D188A reacts with L-Arg to form external aldimine. So this mutant was used to study the dependence binding of two binding sites.

ITC is a physical technique that works by determining the heat released or absorbed during a binding event. Before the experiment, the system is chemically calibrated. The sample cell containing 300 μl of 90 μM dialyzed, degassed SwMppP(D188A) protein in 15 mM MES pH 6.7 buffer was incremental titrated with 2.5 μl of 800 μM L-Arg in the same buffer each titration. After load both sample cell and the burret with L-Arg, the system would keep equilibrate the system with stirrer on, which usually takes 1.5 h to get acceptable heat change deviation. And the later titration experiment takes about 1 hour. The titration heat rate curve has a shifted baseline due to the instability of D188A protein. After continuous agitating for 2.5 h, the protein D188A precipitated and the baseline shifted upwards. I tried to decrease the temperature to 4°C, the baseline is almost flat but periodic noise peak show up at this low temperature. Then I tried 10°C, 12°C and 18°C, they all have shifted baseline. After many attempts, the ITC results are not satisfactory. So fluorescence experiments with short time and gentle treatment was used to monitor the internal aldimine decay and determine the binding relationship. The reaction containing 1.5 mL of SwMppP (D188A) (14 μM) and equal volume of different concentrations of L-Arg

(0 - 2560 µM) in 100 mM BisTris propane buffer was incubated for 8 min before excitation at 415 nm. As shown in Figure 5-34A, there are two species changing during the experiment. As substrate concentration increases, the peak with maximum absorbance at 480 nm decreased, while the peak with maximum absorbance at 542 nm increased. Hence, the species that emissions maximally at 480 nm corresponds to internal aldimine, and the species that emissions maximally at 542 nm corresponds to external aldimine. According to the data, SwMppP (D188A) is almost saturated with 1280 µM L-Arg. Both 480 nm and 542 nm delta fluorescence data are not fitted into Hill equation, which suggested independence binding between two binding sites. The plot of calculated [D188A•L-Arg] complex versus [L-Arg] at 480 nm was fitted to Equation 4 and the K_d for internal aldimine decay is 92.8 µM (Figure 5-34B).
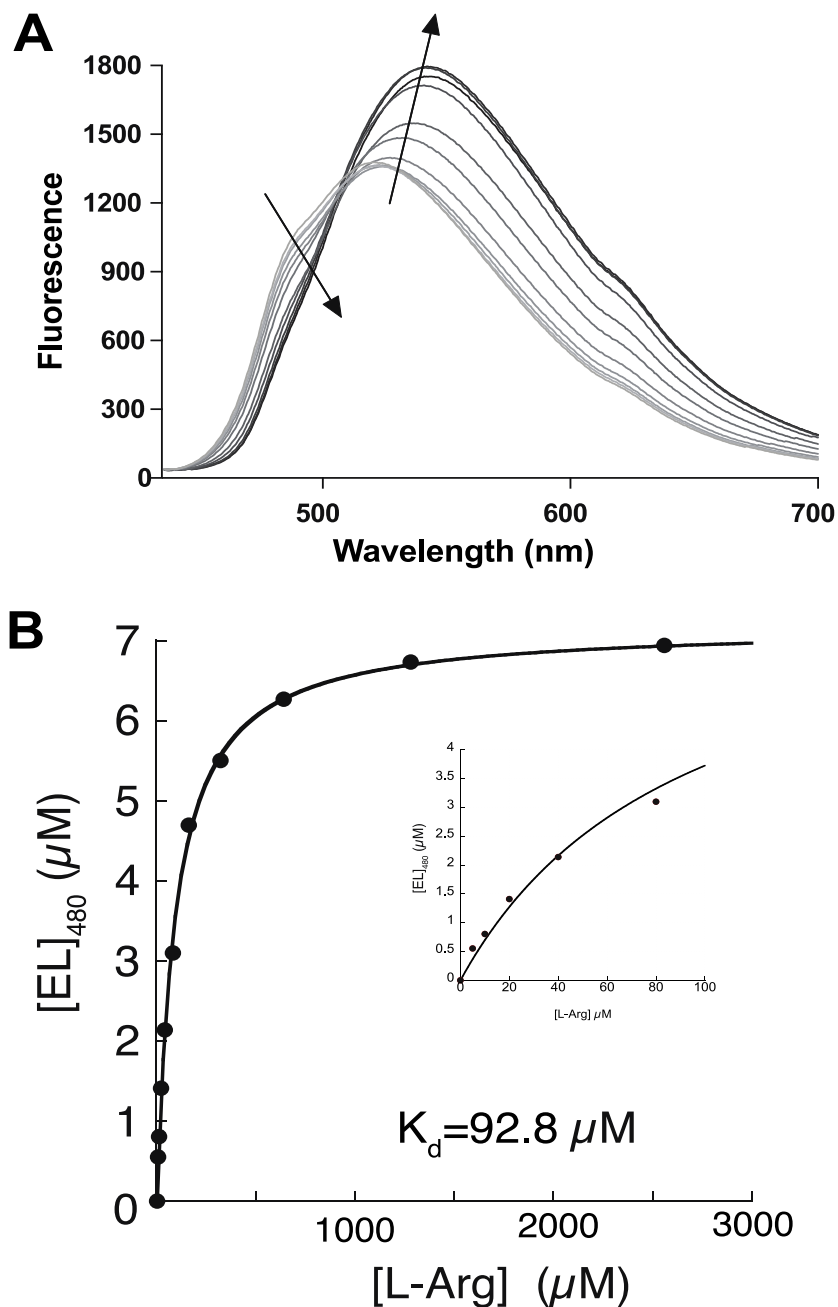
Figure 5-34: (A) Fluorescence change of SwMppP (D188A) with different concetrations of L-Arg (0, 5, 10, 20, 40, 80, 160, 320, 640, 1280 and 2560 μM). There are two major peaks changing as the substrate concentration increases. The peak with maximum absorbance at ~480 nm decreased as [L-Arg] increases, which suggests this species is internal aldimine. The other peak with maximum absorbance at ~542 nm corresponds to external aldimine increases as substrate increases. (B) The fitting data of enzyme and L-Arg

complex versus [L-Arg] at 480 nm was fitted hyperbolically, rather than Hill plot. The inset is the zoomed initial data points and fitting. The data fits hyperbola pretty well, rather than a sigmoid curve.

5.4 Discussion

The nonproteinogenic amino acid L-enduracididine is a critical component of the mannopeptimycins, a family of cyclic glycopeptide antibiotics with potent activities against drug-resistant pathogens like MRSA. Creating derivatives of mannopeptimycin and other L-End-containing natural products is hampered by the limited availability of L-End. We are investigating the L-End biosynthetic pathway in order to develop an efficient enzymatic or chemo-enzymatic route to produce this unusual amino acid. Recently, we found that MppP is a PLP-dependent L-Arg oxidase in L-End biosynthesis in *S. wadayamensis*, not an aminotranferase as the sequence suggested. The quaternary structure of SwMppP shares high similarity with Type I aminotransferases. The major difference is the N-terminus. For type I aminotransferases, N-terminal helix interacts with the small domain of the other chain, and the small domain shifted towards the large domain forming a "closed" conformation when the substrate bound[20, 28, 29]. However, SwMppP mainly exists as homodimer with the active site exposed to the solvent in unliganded state. The fluorescence data of SwMppP (D188A) with L-Arg demonstrated that the two active sites within SwMppP dimer are independence binding. Upon the substrate binding, the N-terminal helix of SwMppP is ordered to cover the active site and interacts with L-Arg. In addition, there is no small domain movement. PLP-dependent cystathionase MalY in *E. coli* (PDB entry: 1D2F[30]) and some aminotransferases (PDB entry 4DQ6 and 3DZZ) have high

structural similarity with SwMppP. Overlaying SwMppP structure with them demonstrates that the small domain of SwMppP is closer to the large domain, demonstrating the small domain and large domain arrangement of unliganded SwMppP is more likely a "closed" conformation.

In *E. coli* aspartate aminotransferase, the small domain inward movements push the repulsion interaction between G38 and Y225, which results in a lower pKa of the aldimine and enhanced catalysis[28]. The lower imine pKa of the internal aldimine is mainly due to the 90° torsion angle of C3-C4-C4'-Nζ bond and the repulsive interactions between G38 and Y225[28, 31]. However, the corresponding torsion angle of SwMppP is ~9°. And SwMppP has a higher catalysis activity at higher pH[27], which indicates a pronated internal aldimine form in SwMppP.

SwMppP reacts with L-Arg and dioxygen to yield two products, **3** and **2** in a ratio of ~1.7:1. The **2** is the products that can be cyclized by MppR (Figure 4-4B). The **3** side pathway may be one approach for self-resistance by limiting the antibiotics level. L-Arg is an important amino acid, which is involved in many pathways and serves as carbon and nitrogen source[32]. MppQ can catalyze **3** into L-Arg, which seems to be a way to regulate L-arginine level in the cells.

The two N-terminal residues, T12 and E15, interact with L-Arg in the active site. To understand what roles these two residues and the N terminus play, we made mutations on the two residues as well as truncation on the N terminus (Δ1-22 aa). The N-terminal variants T12A, E15A, T12A/E15A and SwMppP$_{Δ1-22}$ only produce **3**, which indicates the N-terminal helix is essential for hydroxylation. In addition, T12A, T12A/E15A and SwMppP$_{Δ1-}$

266

[22] have lower catalytic efficiency. As the spectrum shown in Figure 5-16, external aldimine peak (435 nm) gradually shifted to internal aldimine peak (415 nm) as reaction went, demonstrating that N-terminal helix, especially T12, plays a key role on stabilizing external aldimine. The mutant E15A has similar K_M and k_{cat} values with wild type SwMppP but is deficient in the hydroxylase activity. In the structure of E15A•L-Arg, the N terminus also orders and covers the active site. T12 in N terminus still interacts with the guanidine group of L-Arg. But the short side chain of Ala15 creates more room for water retainment, which probably disrupts the hydroxylation pathway. E15Q variant still produced **3** and **2**, although it does make less of the hydroxylated product. Hence, the N terminus interaction is crucial for the stability of external aldimine and later hydroxylation activity.

The reaction of SwMppP with L-Arg and O_2 forms two quinonoid intermediates (Q1 at 510 nm and Q2 at 560 nm). Under anaerobic conditions, the accumulated Quinonoid I is very stable, which suggests this carbanion is stabilized by SwMppP to some extent. When the O_2 is present, Quinonoid I quickly reacts with O_2, which is similar to the PLP-dependent DOPA/aromatic L-amino acid decarboxylases[33]. When the substrate covalently bonds to PLP, the free catalytic K221 forms a hydrogen bonding with D218, which is similar in DOPA/aromatic L-amino acid decarboxylases (PDB ID 1JS3[34]). To study the role of this residue, D218 was substituted into Ser. There is a PMP peak accumulated in the reaction of D218S with L-Arg and D218S mutant only produced little products. In addition, there is no PLP in the low quality structure of D218S. These results suggest that D218 probably participates in the PMP hydrolysis. In the structure of H29A, F115Y and H29A•L-Arg, there are more water molecules retaining in the active site. In the reaction of F191Y with L-Arg, there is a ketimine intermediate formed. The ketimine formation needs a water to provide

H to imine. In conclusion, the water molecule is an important factor to regulate the product pathway. In addition, E15, H29, F115 and F191 are all involved in water exclusion. In SwMppP with L-Arg spectrum, QII firstly accumulate and then quickly decay. But according to the spectrum of F191Y with L-Arg, QII keeps the accumulation, which suggests that F191 is also involved in the reaction of QII with O_2.

According to the spectra of all proteins mentioned above, QII only accumulates in the enzymes producing both products. For those mutants producing **3** only, no QII intermediate peak was observed. These facts suggest that QII formation belongs to the hydroxylation pathway. Pre-steady state data of Quinonoid I decay and Quinonoid II formation shows that Quinonoid II only forms after Quinonoid I decay.

The portion of O_2 consumed to form H_2O_2 was determined by adding catalase after the reaction attained equilibration. Catalase catalyzes 2 moles of H_2O_2 into 1 mole of O_2, and the ratio of O_2 consumed over O_2 regenerated is 2:1, which indicated that each equivalent of O_2 consumed is converted to 1 equivalent of H_2O_2. In addition, the products mass of SwMppP with L-Arg and O_2 conducted in $H_2^{18}O$ demonstrates that the hydroxyl O atom of **2** comes from water. So SwMppP is an oxidase.

The mutant E15A, producing **3** only, has similar catalytic efficiency with wild type. The ratio of [O_2 consumed] over [L-Arg added] for variant SwMppP (E15A) that produces **3** only is 1:1, which suggested that **3** formation only requires one equivalent of O_2. For the other mutants (T12A, T12A/E15A, SwMppP$_{\Delta 1-22}$, H29A/S and F115Y) that producing **3** only, the ratios of [O_2 consumed] over [L-Arg added] are smaller than 1. The ratio is based on the amounts of L-Arg added, not L-Arg consumed, so the mutants with lower catalytic

efficiency would give a smaller ratio value. The ratios of [O_2 consumed] over [L-Arg added] of wild type SwMppP, E15Q and F191Y, producing both **3** and **2**, is 1.4:1, which suggested that **2** formation requires 2 equivalents of O_2.

The dissociation constant for Quinonoid I with O_2 is 52.6 μM by presteady state kinetics, while the dissociation constant for Quinonoid II with O_2 is 418 μM. So Quinonoid II has a weak binding affinity.

The structural and kinetic characterization of the wild-type and several variants of SwMppP allowed us to propose a model where the oxygen atom incorporated in the **2** product is derived from water rather than from dioxygen. And **2** formation requires 2 equivalents of O_2, while the abortive product, **3**, requires only 1 equivalent.

Based on the model, the H29A fluorescence data and absorbance data at 510 nm were globally fitting in Kintek Explorer. The apparent dissociation constant of L-Arg to SwMppP is 101.4 μM, while the K_d of L-Arg to H29A is 316.7 μM.

As described above, SwMppP catalyzes the Cγ-hydroxylation of a non-activated methylene moiety on L-Arg with very high stereospecificity. The stereochemistry of the Cα atom is crucial for the reaction. The α proton of D-arginine is located opposite to the general base catalyst for the Q1 intermediate formation[27], and the external aldimine conformation is not stabilized by the interactions in the active site. Hence, the external aldimine gradually shifted into internal aldimine by the K221 attack. In addition, the position of guanidinium group is essential for the hydroxylation activity. The guanidinium group of L-Arg is hydrogen bonding with PLP phosphate group, T12, S91 and the carbonyl

groups of S248 and D249 from the other chain. The L-Arg homologs, L-citrulline, L-canavanine and methyl-L-arginine, disrupt the coordinations and resulted in the loss of hydroxylation. The side chain of D-Arg pointed away from the PLP cofactor and disrupted the N terminus interaction, which also leads to the inactivity. The length of the substrate is also important for SwMppP activities. Shorter L-Arg analog, L-Orn cannot be deaminized or hydroxylated.

5.5 Reference

[1] Goldberg, J. M., and Kirsch, J. F. (1996) The Reaction Catalyzed by Escherichia coli Aspartate Aminotransferase Has Multiple Partially Rate-Determining Steps, While That Catalyzed by the Y225F Mutant Is Dominated by Ketimine Hydrolysis, *Biochemistry 35*, 5280-5291.

[2] Lu, Z., Nagata, S., McPhie, P., and Miles, E. W. (1993) Lysine 87 in the beta subunit of tryptophan synthase that forms an internal aldimine with pyridoxal phosphate serves critical roles in transimination, catalysis, and product release, *The Journal of biological chemistry 268*, 8727-8734.

[3] Han, L., Schwabacher, A. W., Moran, G. R., and Silvaggi, N. R. (2015) Streptomyces wadayamensis MppP Is a Pyridoxal 5'-Phosphate-Dependent L-Arginine alpha-Deaminase, gamma-Hydroxylase in the Enduracididine Biosynthetic Pathway, *Biochemistry 54*, 7029-7040.

[4] Pacheco, A. A., McGarry, J., Kostera, J., and Corona, A. (2011) Techniques for investigating hydroxylamine disproportionation by hydroxylamine oxidoreductases, *Methods in enzymology 486*, 447-463.

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

[6] Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007) Automated structure solution with autoSHARP, *Methods in molecular biology (Clifton, N.J.) 364*, 215-230.

[7] Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains, *Acta crystallographica. Section D, Biological crystallography 62*, 1002-1011.

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

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

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

[11] Word, J. M., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation11Edited by J. Thornton, *Journal of molecular biology 285*, 1735-1747.

[12] Urzhumtseva, L., Afonine Pv Fau - Adams, P. D., Adams Pd Fau - Urzhumtsev, A., and Urzhumtsev, A. Crystallographic model quality at a glance.

[13] Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L.-W. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard, *Acta Crystallographica Section D 65*, 582-601.

[14] Meister, A. (1952) Enzymatic preparation of alpha-keto acids, *The Journal of biological chemistry 197*, 309-317.

[15] Vlessis, A. A., Bartos, D., and Trunkey, D. (1990) Importance of spontaneous alpha-ketoacid decarboxylation in experiments involving peroxide, *Biochemical and biophysical research communications 170*, 1281-1287.

[16] Li, C., and Lu, C.-D. (2009) Arginine racemization by coupled catabolic and anabolic dehydrogenases, *Proceedings of the National Academy of Sciences 106*, 906-911.

[17] Bertoldi, M., and Borri Voltattorni, C. (2000) Reaction of dopa decarboxylase with L-aromatic amino acids under aerobic and anaerobic conditions, *The Biochemical journal 352 Pt 2*, 533-538.

[18] Bertoldi, M., and Borri Voltattorni, C. (2003) Reaction and substrate specificity of recombinant pig kidney Dopa decarboxylase under aerobic and anaerobic conditions, *Biochim Biophys Acta 1647*, 42-47.

[19] Bunik, V. I., Schloss, J. V., Pinto, J. T., Dudareva, N., and Cooper, A. J. (2011) A survey of oxidative paracatalytic reactions catalyzed by enzymes that generate carbanionic intermediates: implications for ROS production, cancer etiology, and neurodegenerative diseases, *Advances in enzymology and related areas of molecular biology 77*, 307-360.

[20] Nematollahi, A., Sun, G., Harrop, S. J., Hanrahan, J. R., and Church, W. B. (2016) Structure of the PLP-Form of the Human Kynurenine Aminotransferase II in a Novel Spacegroup at 1.83 A Resolution, *International journal of molecular sciences 17*, 446.

[21] Okamoto, A., Higuchi, T., Hirotsu, K., Kuramitsu, S., and Kagamiyama, H. (1994) X-ray crystallographic study of pyridoxal 5'-phosphate-type aspartate aminotransferases from Escherichia coli in open and closed form, *Journal of biochemistry 116*, 95-107.

[22] Eliot, A. C., and Kirsch, J. F. (2004) Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations, *Annual review of biochemistry 73*, 383-415.

[23] Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta crystallographica. Section D, Biological crystallography 60*, 2256-2268.

[24] Kirsch, J. F., Eichele, G., Ford, G. C., Vincent, M. G., Jansonius, J. N., Gehring, H., and Christen, P. (1984) Mechanism of action of aspartate aminotransferase proposed on the basis of its spatial structure, *Journal of molecular biology 174*, 497-525.

[25] Mehta, P. K., Hale, T. I., and Christen, P. (1993) Aminotransferases: demonstration of homology and division into evolutionary subgroups, *European Journal of Biochemistry 214*, 549-561.

[26] Yano, T., Kuramitsu, S., Tanase, S., Morino, Y., and Kagamiyama, H. (1992) Role of Asp222 in the catalytic mechanism of Escherichia coli aspartate aminotransferase: the amino acid residue which enhances the function of the enzyme-bound coenzyme pyridoxal 5'-phosphate, *Biochemistry 31*, 5878-5887.

[27] Yano, T., Mizuno, T., and Kagamiyama, H. (1993) A hydrogen-bonding network modulating enzyme function: asparagine-194 and tyrosine-225 of Escherichia coli aspartate aminotransferase, *Biochemistry 32*, 1810-1815.

[28] Hayashi, H., Mizuguchi, H., Miyahara, I., Nakajima, Y., Hirotsu, K., and Kagamiyama, H. (2003) Conformational change in aspartate aminotransferase on substrate binding induces strain in the catalytic group and enhances catalysis, *The Journal of biological chemistry 278*, 9481-9488.

[29] McPhalen, C. A., Vincent, M. G., Picot, D., Jansonius, J. N., Lesk, A. M., and Chothia, C. (1992) Domain closure in mitochondrial aspartate aminotransferase, *Journal of molecular biology 227*, 197-213.

[30] Clausen, T., Schlegel, A., Peist, R., Schneider, E., Steegborn, C., Chang, Y. S., Haase, A., Bourenkov, G. P., Bartunik, H. D., and Boos, W. (2000) X-ray structure of MalY from Escherichia coli: a pyridoxal 5'-phosphate-dependent enzyme acting as a modulator in mal gene expression, *The EMBO journal 19*, 831-842.

[31] Islam, M. M., Hayashi, H., Mizuguchi, H., and Kagamiyama, H. (2000) The substrate activation process in the catalytic reaction of Escherichia coli aromatic amino acid aminotransferase, *Biochemistry 39*, 15418-15428.

[32] Cunin, R., Glansdorff, N., Piérard, A., and Stalon, V. (1986) Biosynthesis and metabolism of arginine in bacteria, *Microbiological Reviews 50*, 314-352.

[33] Bertoldi, M., Cellini, B., Montioli, R., and Borri Voltattorni, C. (2008) Insights into the mechanism of oxidative deamination catalyzed by DOPA decarboxylase, *Biochemistry 47*, 7187-7195.

[34] Burkhard, P., Dominici, P., Borri-Voltattorni, C., Jansonius, J. N., and Malashkevich, V. N. (2001) Structural insight into Parkinson's disease treatment from drug-inhibited DOPA decarboxylase, *Nature structural biology 8*, 963-967.

Appendices

Appendix A: Crystallographic data collection and refinement statistics.

	RitR(C128S) (5U8K)	RitR$_{ox}$ (5U8M)	RitR(C128D) (5VFA)	rdlv	rd2lv
Space group	C121	P212121	C121	C121	P21212
Cell dimensions					
a, b, c (Å)	141.1, 60.1, 53.3	74.3, 74.8, 102.8	142.8, 59.7, 52.5	93.9, 50.0, 63.0	40.2, 115.3, 125.4
α, β, γ (°)	90.0, 96.0, 90.0	90.0, 90.0, 90.0	90.0, 96.3, 90.0	90.0, 121.4, 90.0	90.0, 90.0, 90.0
Resolution (Å) (last shell)[a]	40.33-1.69 (1.75-1.69)	50.00-2.10 (2.14-2.10)	50.00-1.45 (1.48-1.45)	50.00-1.83 (1.86-1.83)	50.00-2.60 (2.64-2.60)
Wavelength (Å)	0.97872	0.97872	0.97857	0.97857	0.97872
No. of reflections					
Observed	276131(24543)	206898(8842)	396912(14482)	81689(3266)	263987(10059)
Unique	49527(4856)	33913(1667)	76689(3384)	21809(1033)	18645(919)
Completeness (%)[a]	99.6 (97.9)	99.9 (99.6)	99.0 (89.1)	99.1(97.0)	100.0 (100.0)
R$_{merge}$ (%)[a,b]	0.090 (0.340)	0.071 (0.586)	0.046 (0.498)	0.082 (0.436)	0.147 (0.890)
Multiplicity	5.6 (5.1)	6.1 (5.3)	5.2 (4.3)	3.7 (3.2)	14.2 (10.9)
$<I/\sigma(I)>$[a]	16.6 (8.1)	25.2 (2.5)	28.6 (2.9)	13.3 (2.6)	16.9 (3.3)
Model Refinement Statistics					
Reflections in work set	91682	64031	76664	23890	17730

Reflections in test set	3718	3805	2000	1204	1770
R_{cryst} (R_{free})	0.1435 (0.1685)	0.1881 (0.2309)	0.1641 (0.1854)	0.2357 (0.2767)	0.1976 (0.2886)
No. of residues	459	458	443	234	499
No. of solvent atoms	733	185	610	242	49
Number of TLS groups	18	17	16	1	1
Average B-factor (Å^2)^c					
Protein atoms	12.9	46.3	25.9	24.1	43.3
Solvent	27.4	40.7	36.4	30.5	38.0
RMS deviations					
Bond lengths (Å)	0.010	0.011	0.008	0.012	0.015
Bond angles (°)	1.021	1.053	0.976	1.311	1.474
Coordinate error (Å)	0.11	0.26	0.14	0.22	0.40
Ramachandran statistics (favored/allowed/ outliers)	98.5/1.5/0. 0	98.0/2.0/0. 0	97.0/3.0/0. 0	95.7/3.0/1. 3	91.3/4.7/4. 0

Appendix B: Crystallographic data collection and refinement statistics.

	Q15Z91	Q15Z91⊡GTP
Space group	P212121	P212121
Cell dimensions		
a, b, c (Å)	82.1, 116.0, 167.6	82.9, 116.5, 167.3
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	50.00-1.85	50.00-1.95
(last shell)[a]	(1.88-1.85)	(1.98-1.95)
Wavelength (Å)	0.97872	0.97857
No. of reflections		
Observed	797511(33484)	720622(28138)
Unique	135739 (6671)	117445(5618)
Completeness (%)[a]	100.0 (99.8)	98.8(95.6)
R_{merge} (%)[a,b]	0.090 (0.613)	0.103 (0.540)
Multiplicity	5.9 (5.0)	6.1 (5.0)
$<I/\sigma(I)>$[a]	17.9 (2.7)	13.5 (2.7)
Model Refinement Statistics		
Reflections in work set	135607	117218
Reflections in test set	2000	1996
R_{cryst} (R_{free})	0.1662 (0.1833)	0.2616 (0.3161)
No. of residues	1192	1073
No. of solvent atoms	931	427
Number of TLS groups	8	9
Average B-factor (Å^2) [c]		
Protein atoms	38.2	56.8
Ligand[d]	25.4	41.5
Solvent	41.8	45.1
RMS deviations		
Bond lengths (Å)	0.011	0.013
Bond angles (°)	1.259	1.667
Coordinate error (Å)	0.17	0.27
Ramachandran statistics (favored/allowed/outliers)	96.8/2.5/0.7	90.1/6.2/3.8

[a] Values in parentheses apply to the high-resolution shell indicated in the resolution row.

[b] $R = \Sigma(||F_{obs}|-scale*|F_{calc}||) / \Sigma |F_{obs}|$.

[c] Isotropic equivalent B factors, including contribution from TLS refinement.

[d] In the structure of Q15Z91, "ligands" refers to c-di-GMP and citrate. In the structure of Q15Z91⊡GTP, "ligands" refers to c-di-GMP and GTP.

Appendix C: Crystallographic data collection and refinement statistics.

	SwMppP	SwMppP·D-Arg	SwMppP·L-Arg	SwMppP·2KA	SwMppP·4HKA
Space group	P212121	P212121	P212121	P212121	P212121
Cell dimensions					
a, b, c (Å)	85.7, 108.3, 195.4	86.8, 109.1, 194.8	85.8, 108.4, 196.2	85.9, 108.8, 195.5	86.0, 108.4, 196.5
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å) (last shell)[a]	41.45-2.10 (2.14-2.10)	44.53-2.25 (2.29-2.25)	50.00-2.20 (2.24-2.20)	44.59-1.83 (1.87-1.83)	50.00-1.93 (1.96-1.93)
Wavelength (Å)	0.97852 (D)	0.97856 (G)	0.97872	0.97849	0.97856
No. of reflections					
Observed	1532933(60186)	256371 (12494)	408941 (15090)	1055413 (53434)	650654 (25173)
Unique	106188 (5236)	79816 (3984)	85820 (3839)	160466 (7902)	135654 (6682)
Completeness (%)[a]	100.0 (100.0)	90.2 (91.2)	91.6 (84.1)	99.9 (100.0)	98.4 (98.3)
R_{merge} (%)[a,b]	0.106 (0.734)	0.106 (0.726)	0.057 (0.193)	0.074 (0.567)	0.066 (0.488)
Multiplicity	14.4 (11.5)	3.2 (3.1)	4.8 (3.9)	6.6 (6.8)	4.8 (3.8)
$<I/\sigma(I)>$[a]	30.7 (5.2)	10.0 (1.9)	24.7 (6.1)	14.4 (3.3)	21.8 (2.3)
Model Refinement Statistics					
Reflections in work set	100876	76747	85724	160326	135549
Reflections in test set	5256	3015	1999	2000	2000
R_{cryst} (R_{free})	0.148 (0.177)	0.162 (0.197)	0.145 (0.195)	0.151 (0.175)	0.156 (0.178)
No. of residues	1417	1404	1458	1434	1454
No. of solvent atoms	903	693	745	1032	930
Number of TLS groups	29	33	30	31	20
Average *B*-factor (Å²)[c]					
Protein atoms	32.0	34.1	28.8	31.2	32.4
Ligands	30.9	43.9	31.6[d]	37.6[d]	41.9[d]
Solvent	35.6	36.6	33.0	37.3	36.9
RMS deviations					
Bond lengths (Å)	0.013	0.015	0.010	0.010	0.011

Bond angles (°)	1.395	1.565	1.141	1.140	1.155
Coordinate error (Å)	0.17	0.22	0.17	0.19	0.19
Ramachandran statistics (favored/allowed/outliers)	98.3 / 1.7 /0.0	98.4 / 1.6 / 0.0	97.2 / 2.5 / 0.3	97.6/ 2.0 / 0.4	98.2/ 1.8 / 0

[a] Values in parentheses apply to the high-resolution shell indicated in the resolution row.

[b] $R = \Sigma(||F_{obs}|-scale*|F_{calc}||) / \Sigma |F_{obs}|$.

[c] Isotropic equivalent B factors, including contribution from TLS refinement.

[d] In L-Arg complex structure, "ligands" refers to the L-Arg-PLP unit; in the 2KA complex structure, it refers to the 2KA and L-Arg-PLP molecules; and in the 4HKA complex structure, it refers to the 4HKA molecules. In the protein only, "ligands" refer to Cl$^-$ ion.

|	E15A·L-Arg	H29A	H29S	H29A·L-Arg	H29S·L-Arg	F115Y
Space group	P1 21 1	P212121	P212121	P212121	P212121	I121
Cell dimensions						
a, b, c (Å)	64.4, 200.5, 139.9	85.9, 108.3, 195.7	85.8, 108.0, 195.6	85.8, 108.0, 195.7	85.5, 107.7, 195.5	111.62, 63.25, 243.10
α, β, γ (°)	90.0, 103.3, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.29, 90.0
Resolution (Å) (last shell)[a]	50.00-2.20 (2.28-2.24)	50.00-1.90 (1.93-1.90)	50.00-1.95 (1.98-1.95)	50.00-2.30 (2.34-2.30)	50.00-2.10 (2.14-2.10)	60.77-1.92 (1.96-1.92)
Wavelength (Å)						0.97872
No. of reflections						
Observed	977686(170041)	837303(143854)	252515(12536)	494009(79839)	726298 (27943)	433940 (18551)
Unique	36991(8030)	35984 (7049)	131339(6470)	17981(3685)	104882 (5113)	128901 (6294)
Completeness (%)[a]	97.0 (92.2)	99.4 (99.1)	99.8 (99.3)	97.8 (92.3)	99.8(98.9)	99.7 (98.7)
R_{merge} (%)[a,b]	0.110 (0.588)	0.076 (0.555)	0.083 (0.551)	0.078 (0.369)	0.088 (0.460)	==1.162 (3.647)==
Multiplicity	5.7 (4.6)	5.8 (5.1)	7.1 (6.8)	6.2 (4.9)	6.9 (5.5)	3.4 (2.9)
$<I/\sigma(I)>$[a]	14.9 (2.1)	21.7 (2.8)	22.9 (3.1)	20.5 (3.7)	22.6 (3.4)	==1.9 (0.3)==
Model Refinement Statistics						
Reflections in work set	156239	143740	131221	79748	104781	120037
Reflections in test set	2001	2000	1999	1999	2000	1827
R_{cryst} (R_{free})	0.2054(0.2656)	0.1503(0.1818)	0.1508(0.1838)	0.1607(0.2183)	0.1562(0.1989)	0.2620(0.3453)
No. of residues	2839	1421	1421	1456	1455	1441
No. of solvent atoms	1432	1072	943	472	688	468
Number of TLS groups	72	30	25	26	30	18
Average B-factor (Å^2)[c]						
Protein atoms	46.5	29.1	30.4	41.4	37.8	23.1
Ligands	38.0[d]	25.1[d]	28.1[d]	38.8[d]	35.4[d]	-
Solvent	39.0	37.5	36.9	39.2	40.1	23.1
RMS deviations						
Bond lengths (Å)	0.012	0.009	0.010	0.011	0.011	0.014
Bond angles (°)	1.255	1.055	1.108	1.181	1.173	1.411

Coordinate error (Å)	0.29	0.17	0.15	0.22	0.17	0.38
Ramachandran statistics (favored/allowed/outliers)	97.2 / 2.5 / 0.3	98.1 / 1.6 / 0.4	98.2 / 1.5 / 0.4	97.1/ 2.6 / 0.3	97.2/ 2.4 / 0.4	90.6 / 7.4 / 2.0

# Appendix D: Proton NMR spectra for the products of SwMppP and its N-terminal mutants with L-Arg.

(A) Pure L-Arg (2.0 mM, ━); (B) a reaction mixture containing 10.0 μM SwMppP and 2.0 mM L-Arg (━),; (C) a reaction mixture containing 50.0 μM SwMppP$_{Δ1-22}$ and 2.0 mM L-Arg (━); (D) a reaction mixture containing 50.0 μM SwMppPT12A and 2.0 mM L-Arg (━); (E) a reaction mixture containing 50.0 μM SwMppPE15A and 2.0 mM L-Arg (━); (F) a reaction mixture containing 50.0 μM SwMppPE15Q and 2.0 mM L-Arg (━); (G) a reaction mixture containing 50.0 μM SwMppPT12A/E15A and 2.0 mM L-Arg (━). All in 20 mM sodium phosphate buffer (pH 8.4).
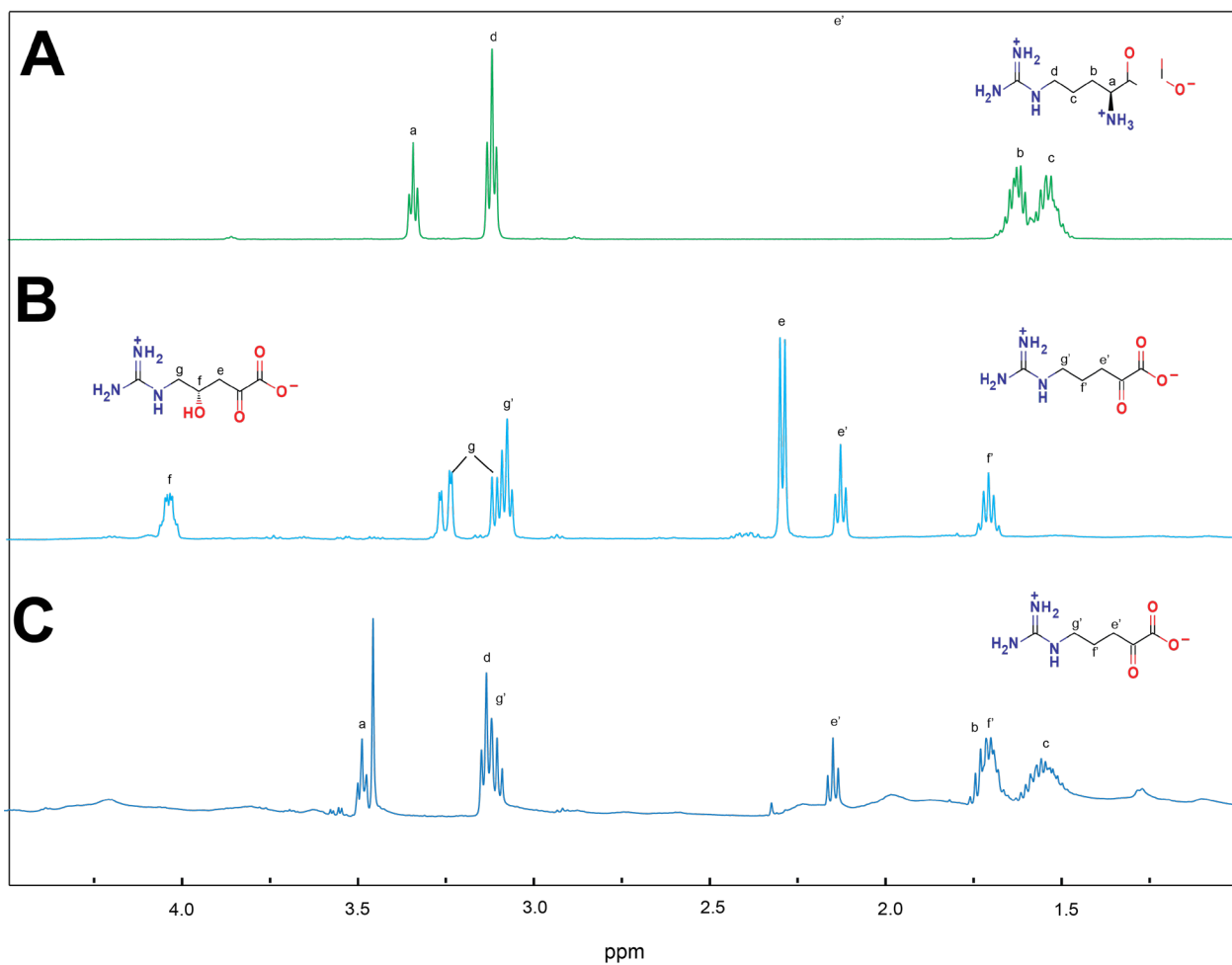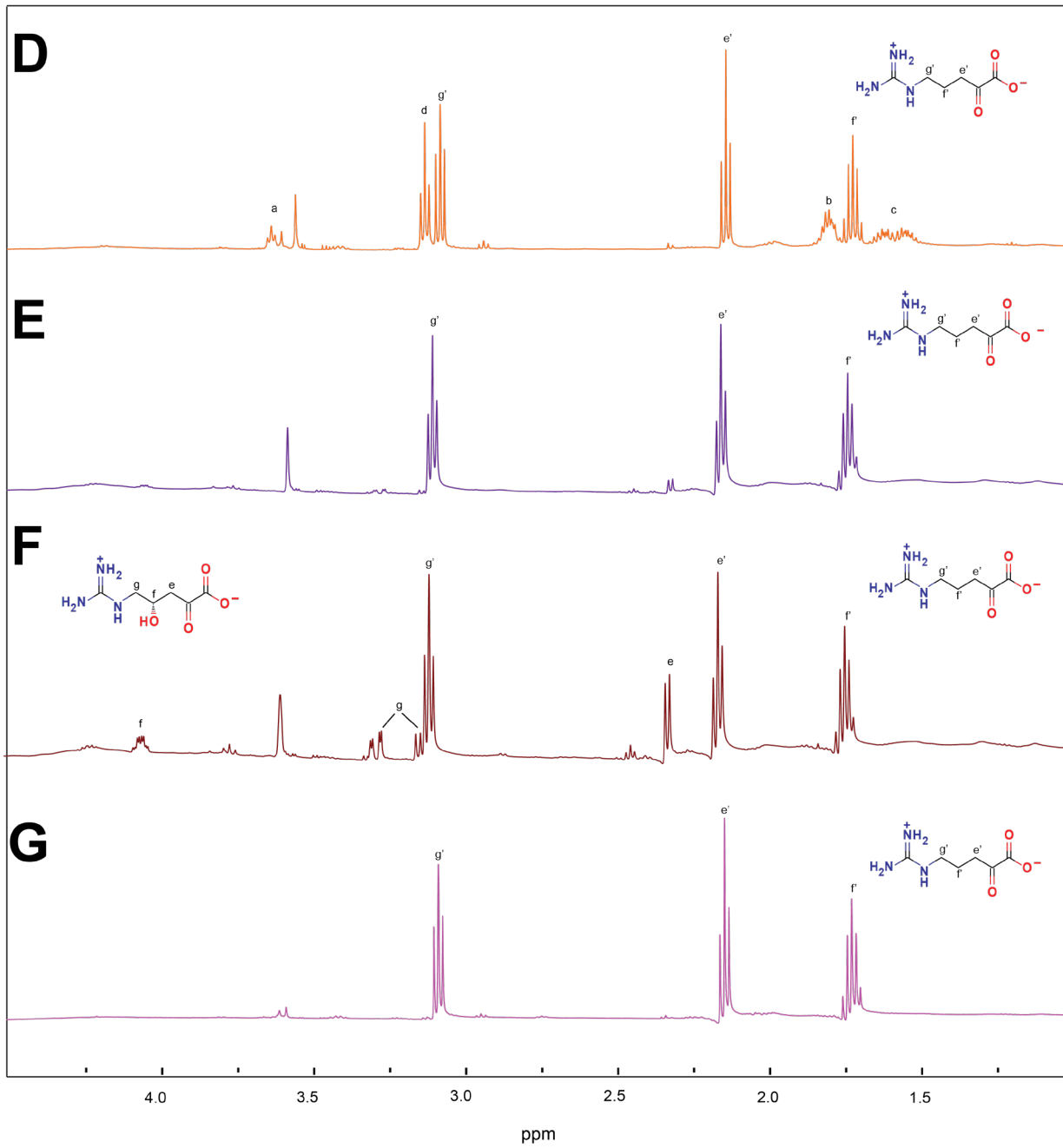
Appendix E: Proton NMR spectra for SwMppP active site mutants with L-Arg.

(A) L-Arg (2.0 mM) with MES (2.0 mM); (B) a reaction mixture containing 10.0 μM SwMppP(H29A) in MES pH 6.7 (2 mM, final concentration) buffer and 2.0 mM L-Arg; (C) a reaction mixture containing 10.0 μM SwMppP(H29S) in MES pH 6.7 (2 mM, final concentration) buffer and 2.0 mM L-Arg; (D) a reaction mixture containing 10.0 μM SwMppP(F115Y) in MES pH 6.7 (2 mM, final concentration) buffer and 2.0 mM L-Arg; (E) a reaction mixture containing 10.0 μM SwMppP(F191Y) in MES pH 6.7 (2 mM, final concentration) buffer and 2.0 mM L-Arg; (F) a reaction mixture containing 10.0 μM SwMppP(D218S) in MES pH 6.7 (2 mM, final concentration) buffer and 2.0 mM L-Arg.
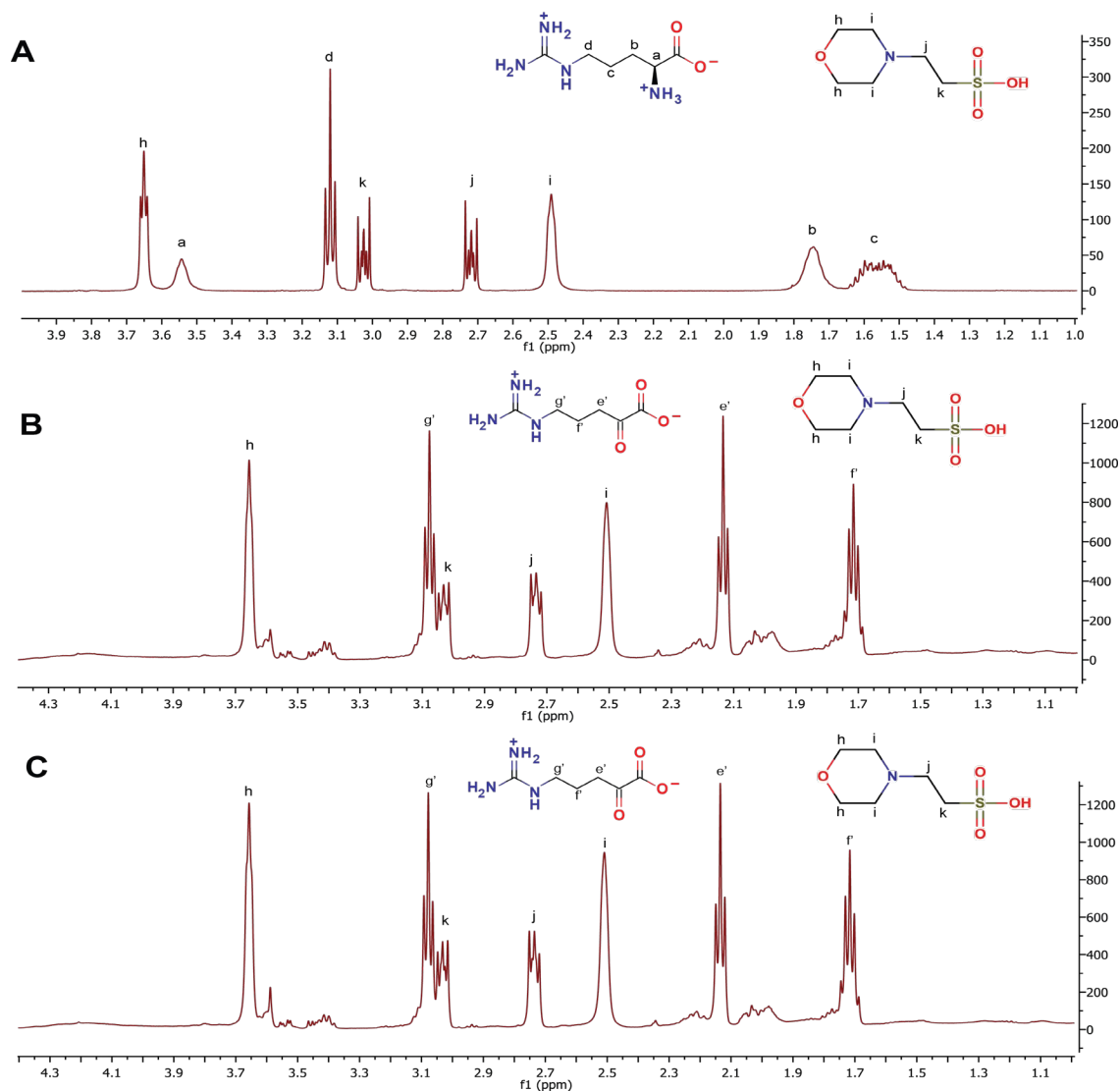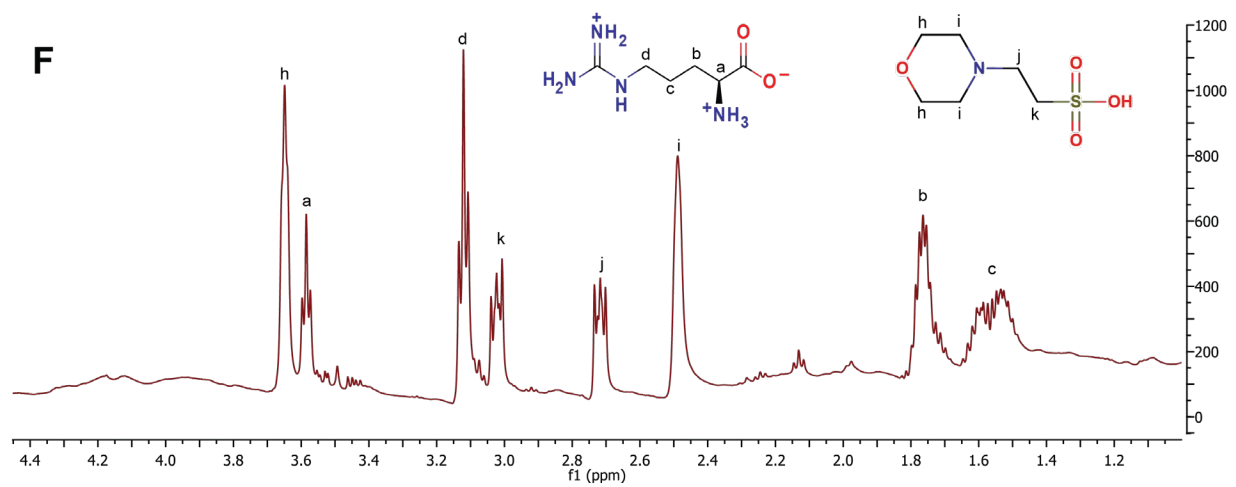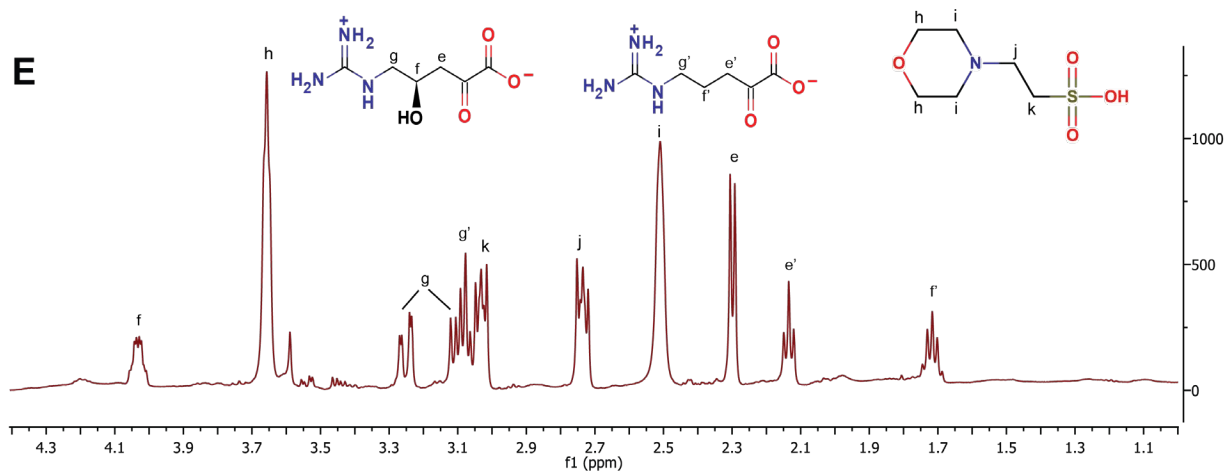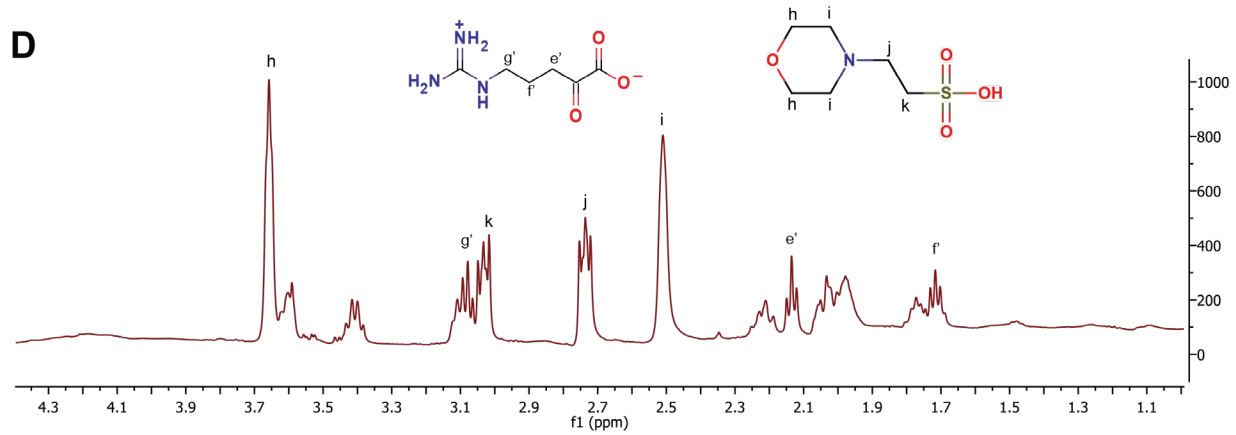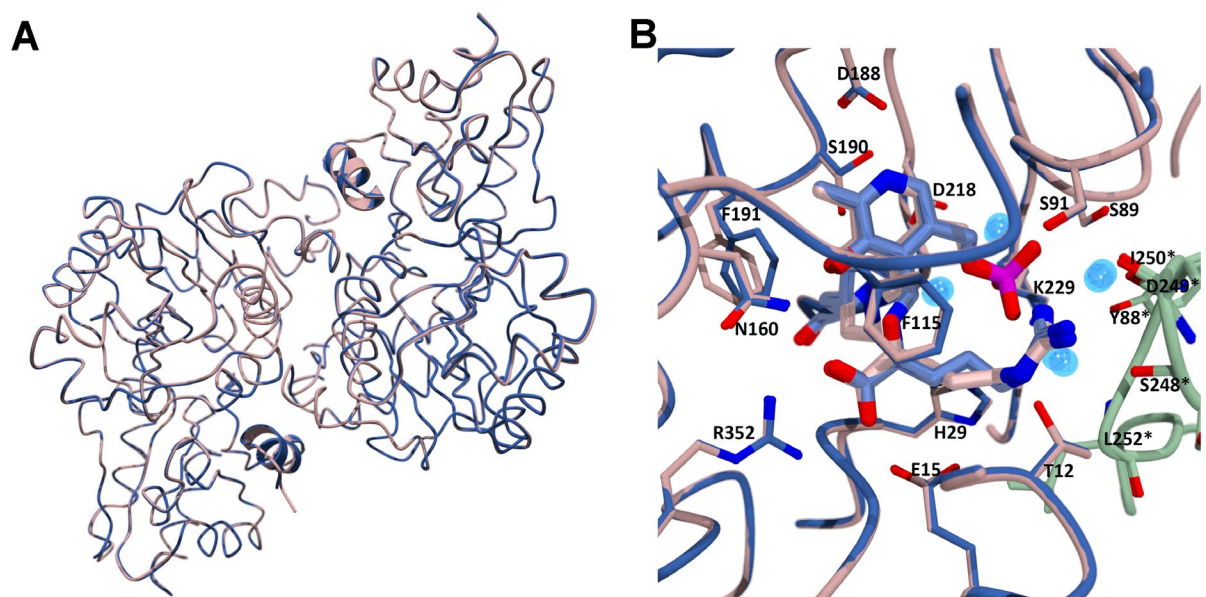
Appendix F: Overlay the structure (A) and active site (B) of SwMppP•L-Arg with SwMppP•2
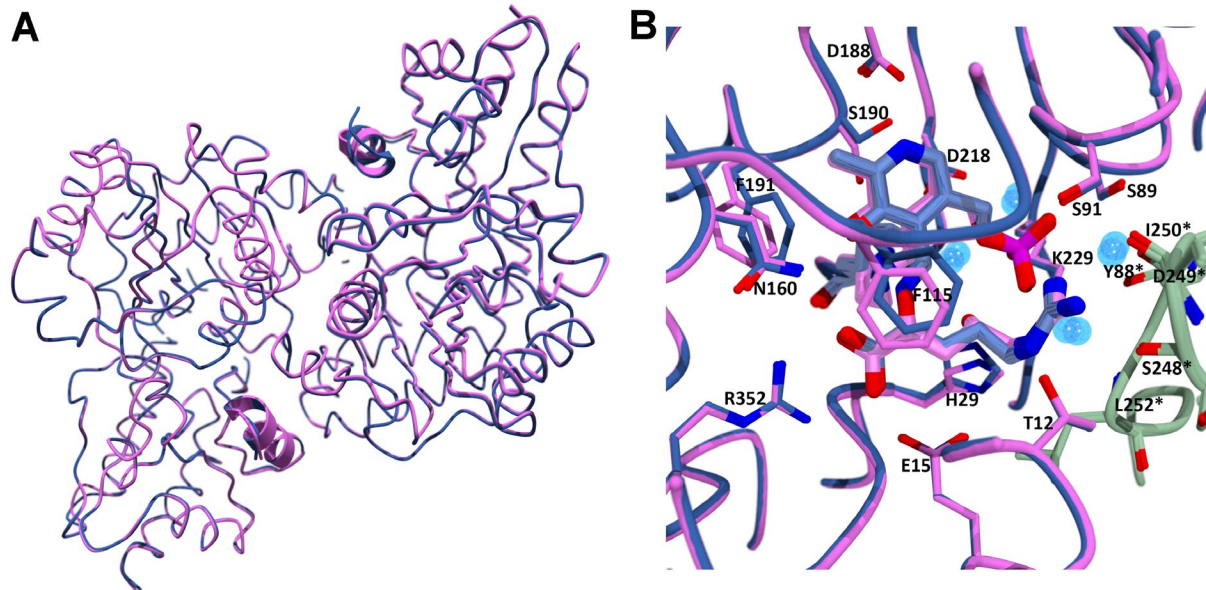
Appendix G: Overlay the structure (A) and active site (B) of SwMppP•L-Arg with SwMppP•3

CURRICULUM VITAE

Lanlan Han

Place of birth: Yancheng, Jiangsu Province, China

Education

Ph.D., University of Wisconsin-Milwaukee, December 2017

Major: Biochemistry

Title of Thesis: Structure-Function Relationships in Bacterial Regulatory Proteins and an Enzyme Involved in Antibiotic Biosynthesis

M.E., Shandong University, June 2012

Major: Bioengineering

Title of Thesis: Screening and Preliminary Separation of Pectinase; Construction of Novel Surface Display Systems of Lactic Acid Bacteria for Oral Vaccines

B. S. in Soochow University, June 2009

Major: Bioscience

Title of Thesis: Screening Bacteria with an Ability of Coaggregation from Town Sewage

Honors and Awards

May 2017   Sosnovksy Award for Excellence in Graduate Research

2016-2017 UWM Graduate Student Excellence Fellowship (GSEF) Award

Sep 2015   The Chair's Prize for my poster titled "A Novel PLP-Dependent L-Arginine Hydroxylase in L-Enduracididine Biosynthesis" on the 35th Midwest Enzyme Chemistry Conference

May 2013   The Student Accessibility Center (SAC) Excellence Award

2008-2009   People's comprehensive scholarship at Soochow University

2007-2008   People's comprehensive scholarship at Soochow University

2006-2007   People's comprehensive scholarship at Soochow University

Publications

1. Zhu Y, Han L, Hefferon KL, Sivaggi NR, Wilson DB, McBride MJ. Periplasmic Cytophaga hutchinsonii endoglucanases are required for use of crystalline cellulose as sole carbon and energy source. *Applied and Environmental Microbiology*. 2016, 82 (15): 4835-4845

2. Han L, Schwabacher AW, Moran GR, Silvaggi NR. Streptomyces wadayamensis MppP is a PLP-Dependent L-Arginine α-deaminase, γ-Hydroxylase in the Enduracididine Biosynthetic Pathway. *Biochemistry*. 2015, 54: 7029–7040

3. Maule AF, Wright DP, Weiner JJ, Han L, Peterson FC, Volkman BF, Silvaggi NS, Ulijasz AT. The aspartate-less receiver (ALR) domains: distribution, structure and function. *PLOS Pathogens*. 2015, 11(4): e1004795

4. Guthrie ML, Sidhu PS, Hill EK, Horan TC, Nandhikonda P, Teske KA, Yuan NY, Sidorko M, Rodali R, Cook JM, Han L, Silvaggi NR, Bikle DD, Moore RG, Singh RK, Arnold LA. Antitumor Activity of 3-Indolylmethanamines 31B and PS121912. *Anticancer Research*. 2015, 35(11): 6001-6007

5. Teske K, Nandhikonda P, Bogart JW, Feleke B, Sidhu P, Yuan N, Preston J, Goy R, Han L, Silvaggi NR, Singh RK, Bikle DD, Cook JM, Arnold LA. Identification of VDR antagonists among nuclear receptor ligands using virtual screening. *Nuclear Receptor Research*. 2014, 1. pii: 101076

6. Sidhu PS, Teske K, Feleke B, Yuan NY, Guthrie ML, Fernstrum GB, Vyas ND, Han L, Preston J, Bogart JW, Silvaggi NR, Cook JM, Singh RK, Bikle DD, Arnold LA. Anticancer activity of VDR-coregulator inhibitor PS121912. *Cancer Chemotherapy Pharmacology*. 2014, 74(4): 787-798

7. Zhang X, Hu S, Du X, Li T, Han L, Kong J. Heterologous expression of carcinoembryonic antigen in Lactococcus lactis via LcsB-mediated surface displaying system for oral vaccine development. *Journal of Microbiology, Immunology, and Infection*. 2016, 49(6): 851-858

8. Hu S, Kong J, Sun Z, Han L, Kong W, Yang P. Heterologous Protein Display on the Cell Surface of Lactic Acid Bacteria Mediated by the S-Layer Protein. *Microbial Cell Factories* 2011, DOI: 10.1186/1475-2859-10-86

Manuscripts in Preparation

1. Han L, Silvaggi NR. Structure and Mechanism of the Novel PLP-dependent L-Arginine Oxidase MppP in Streptomyces wadayamensis. 2017, In preparation for *Structure*.

2. Glanville DG, Han L, Maule AF, Weiner J, Perea-Pallares J, Cheverton A, Lourim D, Morrissey JA, Woodacre A; Clarke TB, Silvaggi NR and Ulijasz AT. A novel family of redox sensors in the streptococci evolved from two-component response regulators. 2017, In preparation for *Molecular Cell*.

3. Han L, Silvaggi NR. The diguanylate cyclase Q15Z91 possesses a non-canonical aspartate-less receiver domain. 2017, In preparation for *Journal of Biological Chemistry*.

Oral Presentations and Posters

Lanlan Han, Nicholas R. Silvaggi. "A Novel PLP-Dependent L-Arginine Hydroxylase in L-Enduracididine Biosynthesis" Poster at the American Society for Biochemistry and Molecular Biology Annual Meeting, Chicago, IL. April 26, 2017

Lanlan Han, Nicholas R. Silvaggi. "RitR, A Response Regulator with an Aspartate-Less Receiver Domain" Presentation to the American Crystallographic Association Annual Meeting, Philadelphia, PA. July 26, 2015

Lanlan Han, Nicholas R. Silvaggi. "A Novel PLP-Dependent L-Arginine Hydroxylase in L-Enduracididine Biosynthesis" Poster at the 35th Midwest Enzyme Chemistry Conference. Illinois Institute of Technology, Chicago, IL, September 12, 2015

Lanlan Han, Nicholas R. Silvaggi. "Structure and Function of the Atypical Response Regulator RitR" Poster at the departmental research symposia (Awards Day), 2015

Lanlan Han, Nicholas R. Silvaggi. "Structure and Function of the Atypical Response Regulator RitR" Poster at the 34th Midwest Enzyme Chemistry Conference. Northwestern University, Evanston IL, September 27, 2014

Lanlan Han, Nicholas R. Silvaggi. "Structure and Function of the Atypical Response Regulator RitR" Poster at the departmental research symposia (Awards Day), 2014

Lanlan Han, Nicholas R. Silvaggi. "Structure and Function of the Atypical Response Regulator RitR" Poster at the departmental research symposia (Awards Day), 2013

Lanlan Han, Nicholas R. Silvaggi. "Structure and Function of the Atypical Response

Regulator RitR" Poster at the departmental research symposia (Awards Day), 2012

Teaching and Research Experience

09/2012-05/2013 Teaching Assistant on CHEM 104: General Chemistry and Qualitative Analysis

06/2013- Now Research Assistant in Dr. Silvaggi Lab