Biochemical Characterization and Structure Determination of a Prolyl 4-Hydroxylase-like protein from *Bacillus anthracis*

By

Submitted to the Department of Chemistry and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Committee Members:

_____
Minae Mure, Ph.D. (chair)

_____
Heather Desaire, Ph.D.

_____
Sue M. Lunte, Ph.D.

_____
Robert C. Dunn, Ph.D.

_____
Emily E. Scott, Ph.D.

Date defended: May 22, 2009

Date approved: May 22, 2009

The Dissertation Committee for Megen A. Culpepper certifies that this is the approved version of the following dissertation.

Biochemical Characterization and Structure Determination of a Prolyl 4-Hydroxylase-like Protein from *Bacillus anthracis*

Committee Members:

_____
Minae Mure, Ph.D. (chair)

_____
Heather Desaire, Ph.D.

_____
Sue M. Lunte, Ph.D.

_____
Robert C. Dunn, Ph.D.

_____
Emily E. Scott, Ph.D.

Date defended: May 22, 2009

Date approved: May 22, 2009

Abstract

Megen A. Culpepper

Department of Chemistry, University of Kansas

Procollagen prolyl-4-hydroxylase (P4H) catalyzes the conversion of peptidyl proline to *trans*-4-hydroxyproline (Hyp). P4H is a member of the α-ketoglutarate-dependent mononuclear non-heme iron oxygenase (αKG/Fe(II)-oxygenase) family of enzymes, which catalyze a wide variety of reactions. The genome of *Bacillus anthracis* (*B. anthracis*), the causative agent of anthrax, contains a gene that is annotated as a *p4h* based on the predicted amino acid sequence. Recently the immunodominant protein of the *B. anthracis* exosporium has been identified as BclA. BclA has collagen-like repeat sequences and has a triple helical structure similar to that of animal collagens. We have detected Hyp from the protein extracts of spores of *B. anthracis* and this is the first report for Hyp detection in bacteria. These results strongly support that *B. anthracis* contains a P4H-like protein, anthrax-P4H.

We have expressed and purified a recombinant form of the putative anthrax-P4H and found that it is a homodimer of 24.3 kDa subunit. The anthrax-P4H exhibits activity and UV-vis spectroscopic properties characteristic of an αKG/Fe(II)-oxygenase, and it can bind a collagen-like substrate, $(Gly\text{-}Pro\text{-}Pro)_{10}$, with an affinity comparable to that of human type (I) collagen-P4H (human-P4H-1). This is the first report of any P4H-like protein from a bacterial source. We propose that anthrax-P4H can serve as a model for human-P4H-1 from its substrate specificity and kinetic parameters comparable to human-P4H-1.

We also report the crystal structure of anthrax-P4H to 1.40 Å. The structure reveals the double stranded β-helix core fold (jellyroll) motif, characteristic of Fe(II)/αKG oxygenases. The oligomerization of the protein is a homodimeric, $\alpha 2$, structure, which makes the anthrax-P4H a new member to the P4H family of enzymes. A putative peptide-binding groove has been proposed based on that identified in the structure of P4H from *Chlamydomonas reinhardtii* (green alga) (Cr-P4H-1). The anthrax-P4H structure provides insight into the dimeric structure of the α-subunits of human-P4H-1, as well as to help understand the mode of substrate recognition that may aide in the development of selective inhibitors.

Acknowledgements

I wish to express my deepest gratitude to my advisor, Dr. Julian Limburg, for his continuous support, encouragement and exchange of ideas to make the work in this dissertation possible. Thank you for believing in me before I believed in myself and providing me with an opportunity to work on such an interesting project over the last five years. I am extremely saddened that he is not here to see his hard work pay off and see me through to completion, but am forever thankful that he was my mentor. He truly had a lasting impression on me both scientifically and personally and will always be in my thoughts.

I would also like to thank my co-advisor, Dr. Minae Mure, for helping me to complete this work even during the hardest of times. Her dedication to me and to the project is inspiring. I would like to thank Dr. Heather Desaire, Dr. Robert C. Dunn, Dr. Sue M. Lunte and Dr. Emily E. Scott for taking the time to serve on my committee, and for their additional assistance in seeing me through to completion. Additional thanks to Dr. Emily E. Scott for being a wonderful collaborator and introducing me to the world of crystallography. Many thanks to Dr. Richard Schowen for insightful discussions and advice on life in science. Your optimism is contagious.

I wish to thank both the former and current members of the Mure and Limburg labs for putting up with me even at my worst. I know I am not always easy to deal with. Many thanks to Dr. Hidehiko Hirawaka for being a patient postdoc and teaching me the world of cloning and protein purifications. You were instrumental in getting me here. Thank you Scott Schrieber for being a hard working technician and always lending a hand even if it meant coming in late or on the weekends. Special thanks to Robyn Moore and Matthew Culpepper for starting and completing this adventure with me. I couldn't have asked for better lab mates. The Triad lives forever!!

A very special thanks to my friends both near and far for continued support and laughs when needed. Mags and Phil, Angelica, Megs, Carl, Wanda-Liz and Nick, Sasha and Alena, Gaywyn and Robyn you are the best and I hope we stay in touch over the years. Thank you Margo for your continued friendship and love. I miss you and hope we are closer in the future. Our trip to Thailand changed both of our lives forever and I am glad I got to experience that with you.

I owe the deepest gratitude to my family. Mom thank you for letting me pursue my dreams even when it meant moving across the country. Thank you for always reminding me that life is a balancing act and what really matters in life. Rose and Mandy thank you for being the best sisters in the world. Rose you know me better than anyone in the world and your unconditional support has always meant so much to me. Mandy you are growing into such a wonderful woman and your enthusiasm and fight for life is truly inspirational. Andy and Mat you are great brother-in-laws and in am lucky to have you in my life. Thank you Grandma Witte for your continuous thoughts and prayers. Patti and Lucy thank you for teaching me about unconditional love, for constant tail-wags and wet kisses and for always being glad to see me and listen to my talks.

Many thanks to my new family for their support and love. Sue and Speck, becoming your daughter-in-law has been a truly wonderful experience and has made my time in Kansas very meaningful. Thank you so much for embracing me into your family with love and laughter. Gloria, Jennifer and Sandy you are all so unique and special in different ways and I love each of you so much.

Lastly I would like to thank Matt for being such a wonderful friend and husband. When I left to come to Kansas my mom said that the man of my dreams was waiting here for me. She was right!!! Going through this experience with you hasn't always been easy, but has shown me how committed you are to us. I have learned more about myself from being with you. Thank you so much for pushing me to fulfill my dreams even if it meant putting yours on hold. I respect you and love you so much more than you know. I can't wait for the new phase in our lives filled with more love and laughter.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Schemes

Chapter 1

Chapter 2

Chapter 3

List of Tables

Lists of Abbreviations

αKG	α-Ketoglutarate
αKG/Fe(II) oxygenase	α-Ketoglutarate Dependent Non-Heme Iron Oxygenase
A. thaliana	*Arabidopsis thaliana*
ACCO	1-Aminocyclopropane-1-Carboxylic Acid Oxidase
ACN	Acetonitrile
AlkB	Alkylation Repair Homolog 1
AtsK	Alkylsulfatase
B. anthracis	*Bacillus anthracis*
BclA	*Bacillus* collagen-like protein of *anthracis*
C. elegans	*Caenorhabditis elegans*
C. reinhardtii	*Chlamydomonas reinhardtii*
CAS	Clavaminate Synthase
CD	Circular Dichroism
CLR	Collagen-Like Region
Cna	*Staphylococcus aureus* adhesion protein
CSM	Collagen-related Structural Motifs
D. melanogaster	*Drosophila melanogaster*
DAOCS	Deacetoxycephalosporin C Synthase
DSBH	Double Stranded β-Helix
DSM	Difco Sporulation Medium
ER	Endoplasmic Reticulum
EXAFS	Extended X-ray Absorption Fine Structure
FI	Fluorescence Intensity
FIH	Factor-Inhibiting HIF
HIFa	Hypoxia Inducible Transcription Factor
HPPD	4-Hydroxyphenylpyruvate Dioxygenase
HRGP	Hydroxyproline Rich Glycoproteins
Hyp	*trans*-4-Hydroxyproline
IPNS	Isopenicillin N Synthase
IPTG	Isopropyl-β-D-Thiogalactopyranoside
MAD	Multi-wavelength Anomalous Dispersion
MLCT	Metal-to-Ligand Charge Transfer
ODDD	Oxygen-Dependent Degradation Domain
OPA	*o*-Phthaldialdehyde
OPD	*o*-Phenylenediamine
ORF	Open Reading Frame
P3H	Proline 3-Hydroxylase
P4H	Prolyl-4-Hydroxylase

PAHX	Phytanoyl-CoA 2-Hydroxylase
PBCV-1	Paramecium bursaria Chlorella virus-1
PBD	Peptide Binding Domain
PDI	Protein Disulfide Isomerase
P-PEK	Partial Bacterial Proteome Extraction Kit
Pro	Proline
SAM	Stanford Automated Mounting
SeMet	Selenomethionine
SSRL	Stanford Synchrotron Radiation Laboratory
TauD	Taurine Dioxygense
TfDa	2,4/αKG Dioxygenase
	The Institute for Genomic Research-Pathogen
TIGR-PFGRC	Functional Genomics Resource Center
TPR	Tetratricopeptide-Repeat
VdW	Van der Waals
VHL	Von-Hippel Lindau
WT	Wild-Type
XAS	X-ray Absorption Spectroscopy

Chapter 1: Introduction

1.1 Collagen Biosynthesis

Collagens are a family of extracellular matrix proteins that function to provide mechanical strength and structural integrity in the body. Collagen comes from the Greek words kolla (glue) and gennan (to produce).[1] Collagens are the most abundant proteins in the human body making up ~ 30 % of total proteins.[2] It is responsible for the formation of hair, fingernails, teeth, eye lenses, bones, and joints in the body, as well as additional roles in cell adhesion, chemotaxis, and migration.[2]

There are at least 27 types of collagen in vertebrates and they are identified by roman numerals according to the order of their discovery.[2] From the 27 diverse types, the form of collagen present in a particular area is determined depending on its role in the body. Collagens most often assemble into supramolecular assemblies, such as fibrils and networks. These can be further divided into eight subfamilies depending on their structural assembly.[2] They include: (1) fibril-forming collagens (types I, II, III, V, XI, XXIV, XXVII), (2) fibril-associated collagens with interrupted triple helices and related collagens (types IX, XII, XIV, XVI, XIX, XX, XXI, XXII, XXVI, XXVIII), (3) collagens forming hexagonal networks (types VIII and X), (4) type IV collagens forming chicken-wire structures that are the major components of basement membranes, (5) type VI collagen, forming beaded filaments, (6)

type VII collagen, forming anchoring fibrils in basement membranes, (7) collagens with transmembrane domains (types XIII, XVII, XXIII, XXV) and 8) collagen types XV and XVIII.[2]

All collagens are made up of three polypeptide chains known as α chains.[2] Some collagens contain identical α chains for all three of the chains and some collagens, most commonly fibrils, contain different types of collagens among the three α chains.[2] The α chains contain a repeat Gly-Xaa-Yaa sequence in which Xaa is often Proline (Pro) and Yaa is often 4-hydroxyproline (Hyp). Gly occupies every third position and the Pro and Hyp pyrrolidine rings have steric constraints all which favor the helical conformation.[3] Hyp is essential to the stability of the triple helix, which will be discussed further in the chapter. The α chains each wind up into a left-handed helix and then wrap up together to form a right handed coiled coil triple helix. This forms an extremely stable rope-like structure, which gives collagen its characteristic strength.

The biosynthesis of collagen is characterized by the post-translational modification of the polypeptide α chains. Collagen processing can be divided into two stages, intracellular, and extracellular. The intracellular modifications include removal of propeptide sequences, hydroxylation of Pro or lysine residues and results in the formation of the triple helices (Figure 1.1).[4] The

extracellular modifications convert the single triple helix into stable, cross-linked fibril collagens (Figure 1.1).[4]

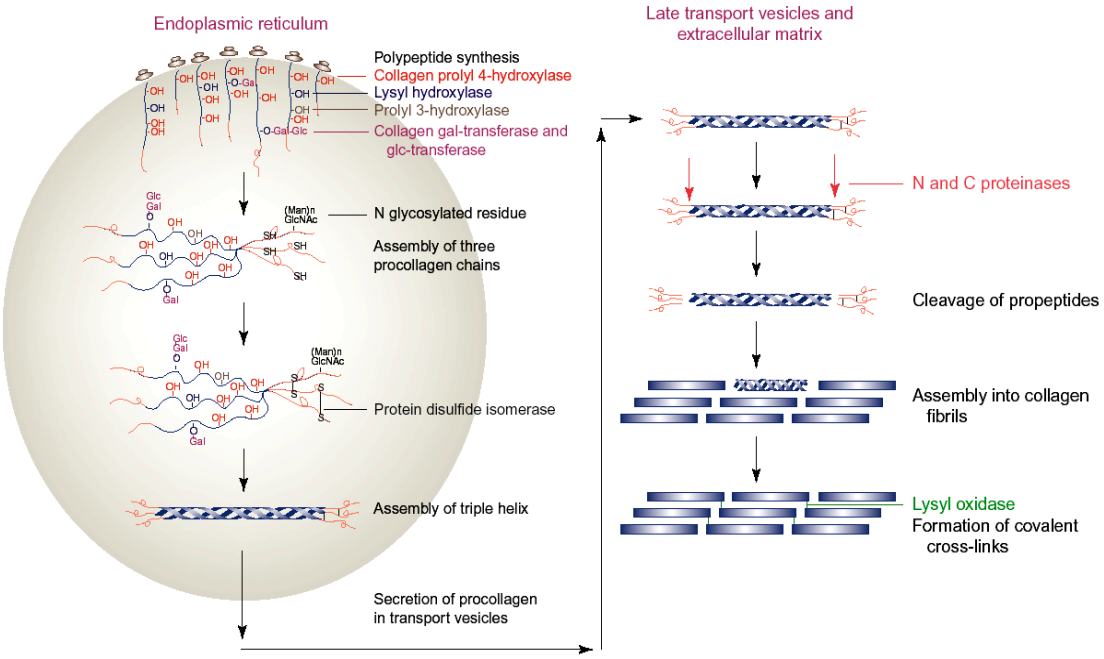


Figure 1.1 The steps involving the biosynthesis of collagen. The intracellular process begins with the hydroxylation of proline or lysine residues. Glycosylation occurs at hydroxylysine and asparagine residues. The three α chains assemble and form intramolecular and intermolecular disulfide bonds. The triple helix assembles in a zipper-like fashion and the procollagen is secreted from the ER to the extracellular matrix. The N- and C- propeptides are cleaved and the collagen self-assembles to form fibrils. The formation of covalent cross-links completes the biosynthesis of collagen. This figure was reprinted from Trends in Genetics, 20, J. Myllyharju, & K.I.Kivirikko, Collagens, modifying enzymes and their mutations in humans, flies, and worms, 33-43, 2004, with permission from Elsevier.

The intracellular events are catalyzed by five specific enzymes; collagen prolyl 4-hydroxylase (P4H), prolyl 3-hydroxylase, lysyl hydroxylase, collagen galactosyltranserase and collagen glucosyltransferase, which are all

located in the lumen of the Endoplasmic Reticulum (ER).[5] The role of 3-hydroxyproline that is formed from prolyl-3-hydroxyase is currently not well defined, but minimal characterization from isolated chick embryo has been reported.[6,7] The product of lysyl hydroxylase, hydroxylysine has two functions in collagen biosynthesis as it provides a site for glycosylation and is essential to intermolecular crosslinking, catalyzed by lysyl oxidase.[7]

As stated previously, the formation of Hyp is essential to the stability of the collagen triple helix. Type I collagen contains 1000 amino acid residues in which ~ 100 are Hyp.[8] Elastin does not form a triple helix but does contain Hyp at 10 - 25 for every 1000 residues.[8] Hyp exists in vertebrates almost exclusively in the Yaa position of the characteristic collagen sequence. Nonhydroxylated collagens can form triple helices at low temperatures but have a Tm of 24 °C.[9,10] Full hydroxylation is therefore required for stable collagen *in vivo* at 37 °C.

The crystal structure of $(GlyProHyp)_4$-AlaProHyp-$(GlyProHyp)_5$ was the first collagen-like peptide structure to be determined.[11] The structure confirmed supercoiled α chains and interchain hydrogen bonding.[11] It also identified that, either the pyrrolidine ring of Pro and Hyp or the C=O for Gly were exposed to the solvent, creating an extensive hydration network. The H_2O molecules establish bridges for hydrogen bonding and the Hyp residues act as anchoring points for these bridges (Figure 1.2).[11] Subsequent work

questioned this rationale with the synthesis of $(GlyProFrp)_{10}$, where Frp is 4-fluoroproline. This peptide showed a characteristic triple helix in the Circular Dichroism (CD) spectrum and a dramatic increase in T_m from 69 °C to 91 °C.[12] The stability of this peptide was explained through inductive effects due to the electronegativity of fluorine and its inability to form hydrogen bonds. The stabilizing effect of Hyp is most likely a combination of both the hydrogen-bonding network and inductive electron effects.

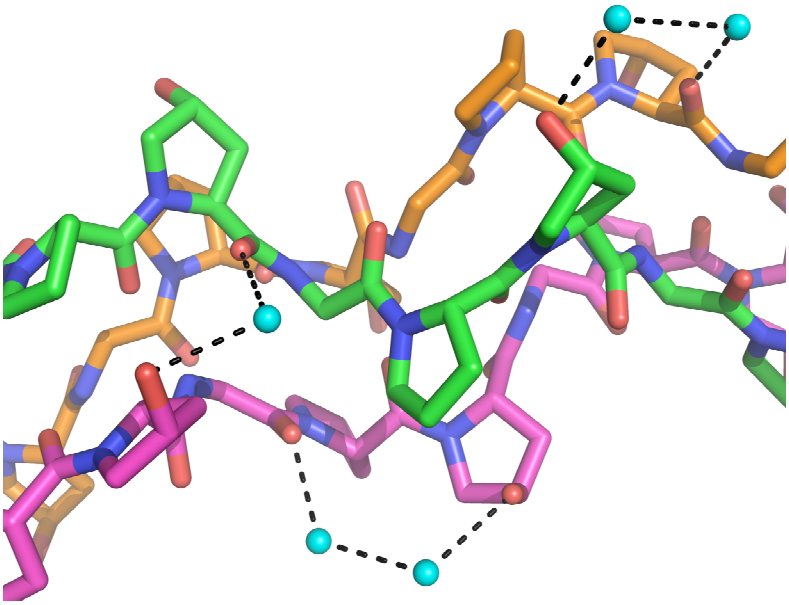


Figure 1.2 Representation of peptide-solvent hydrogen bonding (black lines) network in collagen. H_2O molecules (cyan) stabilize collagen triple helix with both interchain and intrachain hydrogen bonding through H_2O bridges. Figure reproduced from PDB:1CAG with Pymol.

1.2 Collagen Related Diseases

A thousand mutations have been identified in six out of the 42 collagen genes currently known.[2] These include COL1A1, COL1A2, COL2A1, COL3A1, COL4A5, and COL7A1. Most of the mutations are single nucleotide polymorphisms, which convert Gly to an alternate amino acid.[2] This incorporates excess bulk in the triple helix and results in unstable collagen. There are additional mutations such as those that cause the Xaa or Yaa position to change or become a stop codon, but these mutations do not cause as severe a phenotype as the Gly mutations.[2] A vast majority of these mutations have been identified from heritable diseases, as outlined in Table 1.1.[13]

Other diseases, such as fibrosis, occur not from mutations, but from collagen overproduction.[14] The up-regulation of collagen is a normal healing response after injury, during wound healing, tissue remodeling, and repair.[15] However, the overproduction and progressive accumulation of collagen after repeated injury or long-term inflammation can lead to fibrosis.[15] Fibrosis results from chronic inflammation due to chemical, microbial or physical tissue injury. The fibrosis pathway is widespread in the body and exists in many forms including pulmonary, renal, hepatic, myocardial, dermal, and ocular.[16]

Collagen's central involvement in fibrosis makes it a potential drug target for inhibition.[14,17] Specific attention has been paid to the proteins

involved in the post-translational modifications of collagen. Collagen-P4H is of particular interest as it is the rate-limiting step in the biosynthesis of collagen and the formation of Hyp is essential to its stability.[14,17]

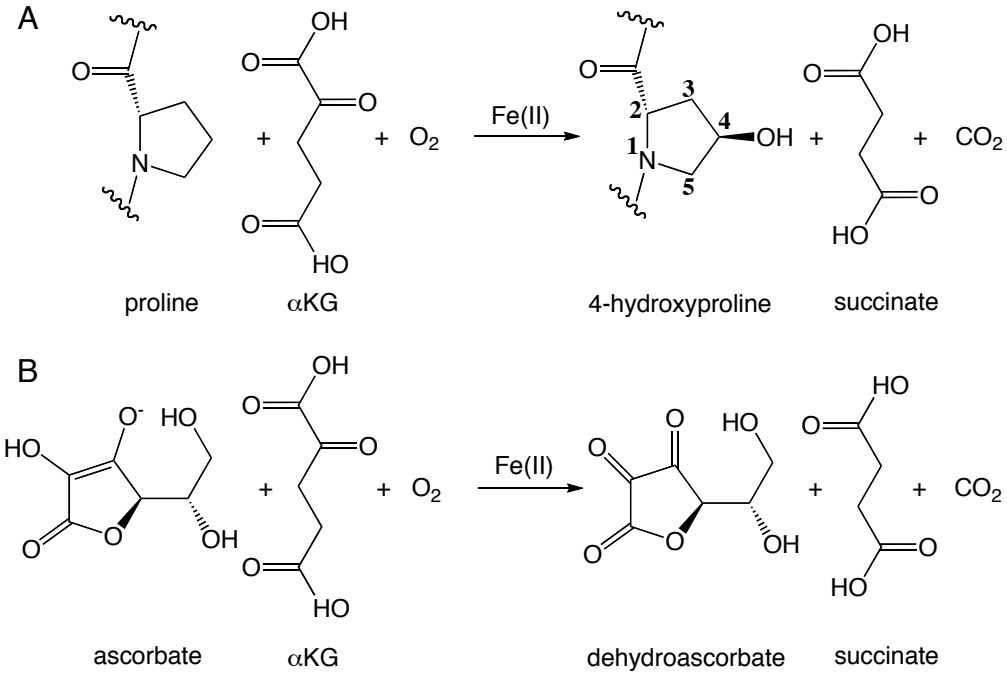
Disease	Gene or enzyme
Osteogenesis imperfecta	COL1A1; COL1A2
Ehlers-Danlos syndrome type VIIA	COL1A1
Ehlers-Danlos syndrome type VIIB	COL1A2
Marfan syndrome	COL1A2
Osteoporosis	COL1A1; COL1A2
Achondrogenesis	COL2A1
Hypochondrogenesis	COL2A1
Spondyloepiphyseal dysplasia	COL2A1
Stickler syndrome	COL2A1
Osteoarthrosis	COL2A1; COL9A1
Ehlers-Danlos syndrome type IV	COL3A1
Aortic aneurysms	COL3A1
Alport syndrome	COL4A5
Epidermolysis bullosa, dystrophic forms	COL7A1
Spondylometaphyseal dysplasia	COL10A1
Ehlers-Danlos syndrome type VI	Lysyl hydroxylase
Ehlers-Danlos syndrome type VIIC	ProcollagenN-proteinase
Ehlers-Danlos syndrome type IX	Lysyl oxidase
Menkes syndrome	Lysyl oxidase

Table 1.1 Mutations identified in collagen that results in heritable disease. Table reproduced from Annals of Medicine, 25, K. I. Kivirikko, Collagens and their abnormalities in a wide spectrum of diseases, 113-126, 1993, with permission from Informa healthcare.

1.3 α-Ketoglutarate Dependent Non-Heme Iron Oxygenases

P4H was identified in seminal work by Udenfriend[18] and Prockop.[19] Udenfriend showed that fetal rat skin exhibited P4H activity that was dependent on Fe(II), α-ketoglutarate (αKG), and O_2 (Scheme 1.1), hereby

identifying it as the first member of a family of enzymes known as the αKG-dependent mononuclear non-heme oxygenases (αKG/Fe(II) oxygenases).[18,20] P4H catalyzes the general reaction shown in Scheme 1.1A, where the active site Fe(II) catalyzes the hydroxylation of peptidyl proline coupled to the decarboxylation of αKG to yield succinate.[18,19] In addition to O_2 and αKG, P4Hs require a pool of ascorbate to be fully active.[21] In the absence of substrate, P4H can still decarboxylate αKG at a slow rate, but the enzyme becomes inactivated.[22] Ascorbate can protect P4H against inactivation by serving as a surrogate substrate to yield succinate and dehydroascorbate



Scheme 1.1 Procollagen-P4H catalyzes the hydroxylation of peptidyl proline to 4-hydroxylproline (Hyp) coupled to decarboxylation of αKG to yield succinate (A). In the absence of peptidyl proline, procollagen-P4H catalyzes the uncoupling reaction to oxidize ascorbate to dehydroascorbate (B).

(Scheme 1.1B), and this reaction is termed uncoupling.[23]

Members of the αKG/Fe(II) oxygenases are diverse in function, catalyzing a wide range of reactions from fatty acid metabolism to DNA repair and antibiotic biosynthesis (Figure 1.3).[24] Though their roles in nature are extremely diverse, they have some common properties that classify them as a superfamily. All of the enzymes contain a H_1-X-D/E-X_n-H_2 facial triad binding

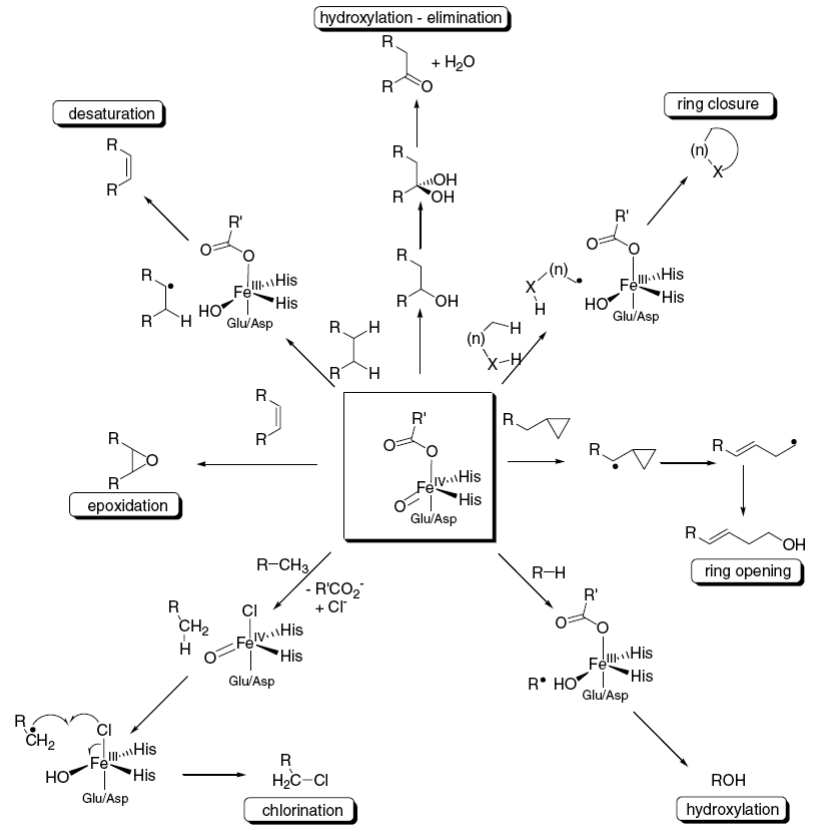


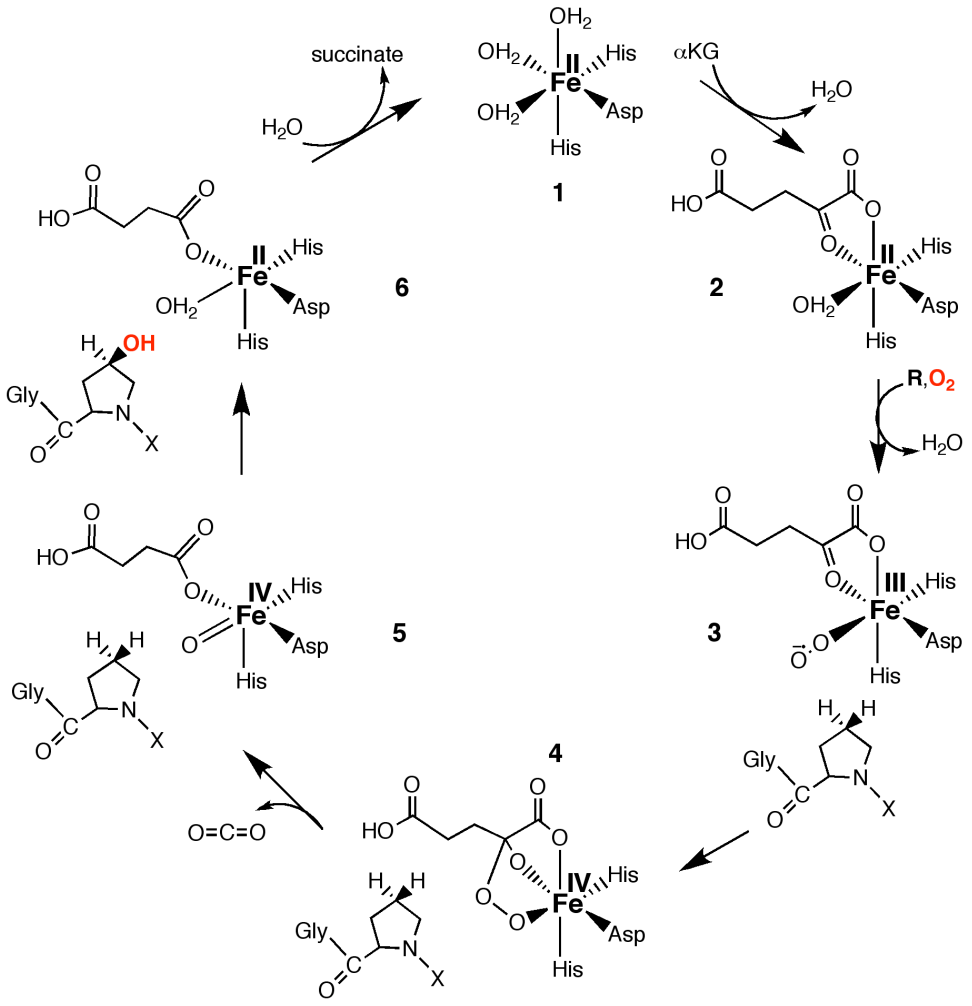
Figure 1.3 Diverse functions catalyzed by αKG/Fe(II) dependent oxygenases. Figure from J. Inorg. Biochem., 100, I.J. Clifton, M.A. McDonough, D. Ehrismann, N. J. Kershaw, N. Granatino, & C. J. Schofield, Structural studies on 2-oxoglutarate oxygenases and related double-stranded β-helix fold proteins, 644-669, 2006, with permission from Elsevier.

site for Fe(II) in their active sites.[20] Pioneering work by Hanauske-Able and Günzler proposed a mechanism of O_2 activation and substrate hydroxylation by P4H, now termed the HAG mechanism (Scheme 1.2).[25] Subsequent characterization of other αKG/Fe(II) oxygenases determined that other members of this family of enzymes undergo the same mechanism.

In the mechanism αKG first binds Fe(II) in a bidentate fashion and displaces two H_2O molecules coordinated to the Fe(II) (2). The substrate then binds and activates O_2 to produce a proposed superoxo-Fe(III)-O_2 intermediate (3). A peroxo-Fe(II)-O_2 (4) is postulated to be formed by direct nucleophilic attack on the uncoordinated O atom at the C2 position of αKG. Decarboxylation of (4) and heterolytic cleavage of the O-O bond results in a reactive Fe(IV)=O intermediate (5). The highly reactive Fe(IV)=O undergoes an H-atom abstraction to hydroxylate the substrate (6). The loss of succinate and binding of two H_2O molecules reduces the Fe(II) back to its resting state.

Gaining a clearer understanding of the mechanistic intermediates can aid in the design of specific inhibitors for fibrosis. Many of the intermediates have been identified experimentally through UV-vis spectroscopy, Resonance raman or X-ray crystallography.[20,26] The initial resting state (1) has been established using X-ray absorption spectroscopy (XAS) for both 2,4-D/αKG dioxygenase (TfdA)[27] and deacetoxycephalosporin C synthase (DAOCS)[28] and using CD for clavaminate synthase (CAS).[29] Intermediate (2) exhibits a

characteristic metal-to-ligand charge transfer (MLCT) absorption transition in the UV-vis visible region at ~ 500 nm.[30] This spectrum is seen in many αKG/Fe(II) oxygenases including taurine dioxygenase (TauD), which shows a lilac colored chromophore with a λ_{max} = 530 nm.[31] The species has also been identified in several crystal structures of the αKG/Fe(II) oxygenase family.[24,30]



Scheme 1.2 HAG mechanism for the hydroxylation of substrate by members of the αKG/Fe(II) oxygenase superfamily of enzymes, including P4H.

The binding of substrates does cause a shift in the UV-vis spectra of some enzymes. In TauD, the binding of taurine causes a blue shift to λ_{max} = 520 nm.[31] The species involved in the reaction of O_2 with the substrate/Fe(II) (intermediates 3 and 4) are very reactive and poorly characterized. The Fe(IV)=O species was first identified in TauD by Mössbauer,[32] resonance Raman[33] and extended X-ray absorption fine structure (EXAFS).[34] This species was identified as an S = 2 ground state and undergoes H-atom abstraction followed by a rebound step to hydroxylate the substrate.

The members of this family of enzymes are also characterized by a double-stranded β-helix core (DSBH) or jellyroll core fold motif in their crystal structures.[24] The fold consists of eight β-sheets that form a β-sandwich composed of two-four stranded anti-parallel β-strands. The cupin and Jmjc family of proteins also adopt this conformation leading to the nomenclature of DSBH.[24] The DSBH consists of a minor β-sheet (usually containing shorter strands) and a major β-sheet (often has extended β-sheets other than in the core fold). Additional secondary structures further stabilize the DSBH domain. As of 2006, over 50 structures of the αKG/Fe(II) oxygenases have been determined and all are seen to adopt this core fold.[30]

1.4 Vertebrate P4H

1.4.1 Collagen-P4H

In humans, P4H exists as an $\alpha_2\beta_2$ tetramer. The hydroxylation activity resides in the α subunit and the β subunit is a protein disulfide isomerase (PDI).[35,36] The function of PDI as the β subunit is to keep the α subunit properly folded and solubilized in the ER. There are three identified isoenzymes of human-P4H, all of which associate with the same PDI β-subunit (Figure 1.4).[37,38] Evidence exists that the tetramers do not form mixed $\alpha(1)\alpha(2)\beta_2$ tetramers, but instead distinct $\alpha(1)_2\beta_2$, $\alpha(2)_2\beta_2$, and $\alpha(3)_2\beta_2$ tetramers.[36,37] The sequence identity among human-P4H-1 and -2 is 65 % and the identity among -1 and -3 and -2 and -3 is 35 % and 37 % respectively.[14,36] The highest degree of similarity occurs at the C-terminal

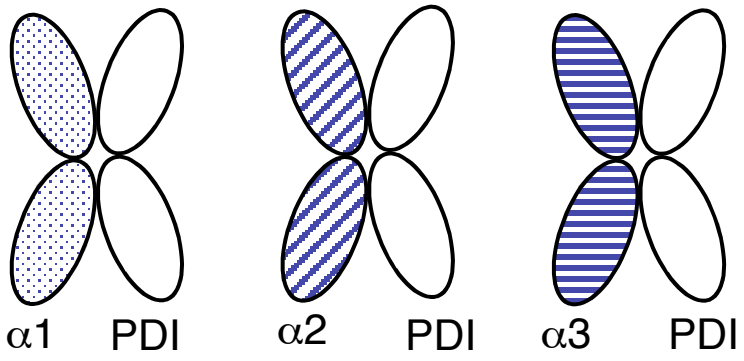


Figure 1.4 A schematic representation of the three forms of human-P4H. The variation in the forms is present in the α subunit. The β PDI subunit is the same for all three forms.

end, which is involved in catalysis and includes the residues involved in the facial triad Fe(II) binding.[39] In this region, α(1) and α(2) have 80 % identity while α(3) has ~ 56 % identity with the other two subunits over the last 120 residues.[14] A peptide-binding domain (PBD) of ~ 100 residues is located at the N-terminal region of each of the isoenzymes and is responsible for the substrate recognition of human-P4H.[14]

The substrate specificity of the three forms differs slightly from each other. All three isoforms bind polypeptide $(GlyProPro)_{10}$ with in one order of magnitude (Table 1.2).[37,38,40] Poly(L-proline) has long been an effective competitive inhibitor with respect to $(GlyProPro)_{10}$ in human-P4H-1.[41] Therefore, it was surprising that human-P4H-2 had K_i values for poly(L-proline) which were 200 and 1000 fold greater for M_r 7,000 and M_r 44,000 poly(L-proline) respectively than for type-1 (Table 1.2).[37] Human-P4H-3 shows intermediate IC_{50} values compared with human-P4H-1 and -2 when tested with M_r 5,000 poly(L-proline).[38] The difference in binding of both $(GlyProPro)_{10}$ and poly(L-proline) is most likely is due to differences in the

Substrate or Inhibitor	Constant	Human-P4H-1[a] (μM)	Human-P4H-2[b] (μM)	Human-P4H-3[c] (μM)
(GPP)10	K_m	15	95	27
Poly(L-Proline)				
Mr 5,000	IC_{50}	6	300	30
Mr 7,000	K_i	0.5	95	N/A
Mr 44,000	K_i	0.02	20	N/A

[a]Ref 37 and 40, [b]Ref 37, and [c]Ref 38

Table 1.2 Comparison of kinetic parameters of the isoforms of human-P4H.

PBD of the three forms. These differences in the substrate specificity present the potential for the development of selective inhibitors to target fibrosis.

Human-P4H-1 is the most common type of C-P4H in cells and tissues.[37,42] It is found highly expressed in less differentiated cell lines including malignant hepatocytes.[42] Human-P4H-2 has been found mostly in chondrocytes, osteoblasts, endothelial cells, and cells in epithelial structures.[42] It represents 70 % of P4H activity in cultured mouse chondrocytes and 80 % activity in cultured mouse cartilage, providing evidence that human-P4H-2 may have a role in the development of cartilage, cartilagenous bone, and capillaries.[37] Human-P4H-3 was found expressed in various tissues, but at much lower levels.[37] High expression levels were seen in the fibrous cap of atherosclerotic human carotid arteries.[43]

The PDI/β subunit of P4H is a multidomain protein that catalyzes the formation, reduction, and isomerization of disulfide bonds during protein folding in the ER.[44] In addition to P4H, PDI acts as the β subunit for the microsomal triglyceride transferase dimer.[45] PDI contains four domains, a, b, b' and a', followed by a C-terminal extension in which the ER retention signal KDEL is located (Figure 1.5).[8,44] The a and a' domains are responsible for catalytic activity, and the b and b' domains are not redox active. The first crystal structure of full-length PDI from yeast has recently been solved to 2.4 Å resolutuion.[46] The structure revealed the four domains adopt a thioredoxin

fold and arrange in a twisted "U" conformation.[46] The active sites of a and a' face each other and are located at the ends of the U, with the b and b' making up the base of the "U". A continuous hydrophobic surface was seen for the structure, which correlates the structure and function of the protein, as PDI's substrates are misfolded proteins with exposed hydrophobic residues.[46]

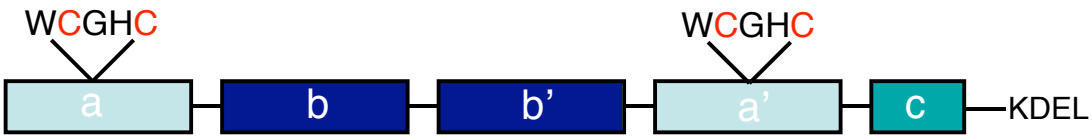


Figure 1.5 Diagram of PDI domains. PDI is the β subunit of human-P4H. The catalytic activity resides in the a and a' domains. Each active site contains two cysteine residues in the sequence WCGHCK. The b and b' domain are not catalytically active.

The concentration of PDI in the ER in the mM range and is found at greater levels than the α subunit of P4H.[44] The roles of PDI and its relationship to the α subunit of P4H have been explored through mutagenesis experiments. Domains a' and b' are significant for minimal tetramer assembly, but assembly greatly increases when all four domains are present.[47] The primary substrate binding site for PDI has been identified as b', but b' is not essential for P4H activity or assembly.[48] The PDI active site cysteine residues are also not necessary for P4H activity or tetramer assembly.[49]

1.4.2 Inhibitors of Collagen Human-P4H

There are several approaches to the development of inhibitors to target P4H. The difficulty in designing inhibitors is specificity to human-P4H-1 as well as delivery. As stated previously, targeting inhibition via the PBD of human-P4H is one strategy to address inhibitor selectivity, as the PBD represents the major difference in the three isoenzymes of human-P4H.[7] Poly(L-proline) is a good example of targeting the selective properties of the PBD because it has a lower *K*i value for human-P4H-1 versus the other two forms. Peptide inhibitors substituting Pro with unnatural amino acids such as 5-oxaproline have also been shown to be effective suicide inhibitors of P4H.[50] The most potent of these inhibitors, benzyloxycarbonyl-Phe-Oxaproline-Gly-benzylester inactivated human-P4H-1 by 50 % in 1 hour at 0.8 μM.[50] Divalent cations are also competitive inhibitors of Fe(II), with Zn(II) being the most effective with a K_i of 1 μM.[7]

Many of the current P4H inhibitors are designed to mimic the binding of αKG (Figure 1.6 A) to Fe(II). N-oxalylglycine is structurally very similar to αKG as it replaces the methylene at the C3 of αKG with an NH group (Figure 1.6 B). These inhibitors have functional groups that interact directly with the binding coordination and active site of αKG. This is powerful inhibitor with a K_i of 0.5-8 μM.[51,52] Pyridine-2',4'-dicarboxyate and pyrindine-2',5'-dicarboxylate (Figure 1.6 C and D) are the most potent competitive inhibitors with respect to

αKG with K_i values of 2 μM and 0.8 μM respectively.[53,54] 3,4-dihydroxybenzoate differs from the other inhibitors in that it is competitive against both αKG and ascorbate and is effective with a K_i of 5 μM (Figure 1.6 E).[55]

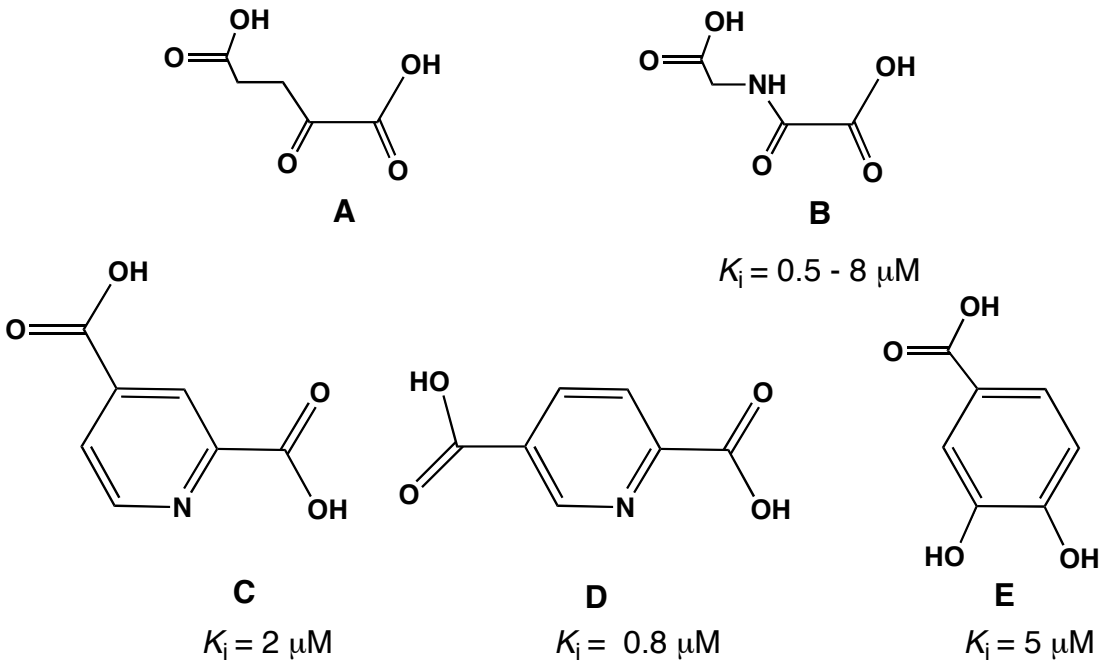


Figure 1.6 Structures of αKG and several competitive inhibitors of P4H. A = αKG, B = N-oxalylglycine, C = Pyridine-2',4'-dicarboxyate D = pyrindine-2',5'-dicarboxylate, and E = 3,4-dihydroxybenzoate.

The difficulty in using these inhibitors clinically is bioavailability. Many of these inhibitors cannot permeate the cell membrane and therefore are not effective as potential drug targets. Further design is necessary to develop a potent C-P4H inhibitor that is both selective and able to be delivered to the cell for therapeutic approaches. We hope to accomplish inhibitor selectivity through the development of a chimeric inhibitor that incorporates an Fe(II)

chelator such as phenanthroline and a peptide tag which will be specific to the human-P4H-1 PBD. The inhibitor delivery issue will be addressed with the addition of polar groups to both the Fe(II) chelator and peptide tag of the chimeric inhibitor to increase biodelivery.

1.4.3 Hypoxia Inducible Factor-P4H

An alternate P4H was identified to regulate oxygen response called the hypoxia-inducible transcription factor (HIFα).[56,57] The lack of oxygen, or hypoxia, causes stabilization of a DNA specific HIF transcription factor, which in turn causes the transcription of hypoxia-responsive genes. HIF is expressed as a $\alpha\beta$ heterodimer in which the HIFα subunit stability is O_2 dependent and HIFβ subunit is not.

Under normal O_2 conditions the HIFα subunit is constitutively expressed but is rapidly degraded by the hydroxylation of one or two conserved proline residues in the Leu-X-X-Leu-Ala-Pro sequence of its oxygen-dependent degradation domain (ODDD) as shown in Figure 1.7.[14] The formation of the Hyp residue causes HIFα to bind to the von-Hippel Lindau (VHL) E3 ubiquitin complex to undergo degradation. The enzyme responsible for the hydroxylation of Pro was identified as a cytoplasmic P4H (HIF-P4H).[57,58]

In hypoxic environments the above pathway is suppressed. HIF-P4H activity is inhibited and hydroxylation of the ODDD sequence is ceased. The

absence of prolyl hydroxylation causes the HIFα to dimerize with HIFβ. The HIFαβ complex is translocated into the nucleus where it binds to one of the 100 hypoxia-induced genes.[14,36] These genes have involvement in angiogenesis, oxidative stress resistance and tumor progression to name a few. HIF-P4H is a member of the αKG/Fe(II) oxygenases, but its function is distinct from C-P4H.

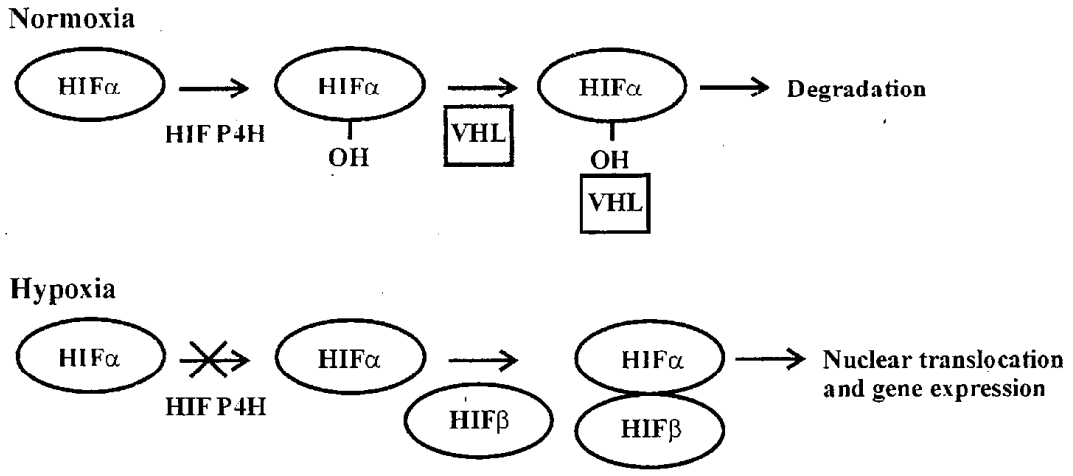


Figure 1.7 Regulatory pathway of hypoxia inducible transcription factor HIFα by αKG/Fe(II) oxygenase HIF-P4H. Under normoxic conditions HIFα binds to von-Hippel Lindau (VHL) complex and undergoes degradation. Hypoxia causes the prolyl hydroxylation to be ceased and the HIFα to bind to the HIFβ. The HIFαβ complex is then translocated to the nucleus and is involved in gene expression. Figure used from Matrix Biology, J. Myllyharju, Prolyl 4-hydroxylases: the key enzymes of collagen biosynthesis, 15-24, 2003, with permission from Elsevier.

1.5 Invertebrate P4H

1.5.1 *Caenorhabditis elegans* P4H

The nematode *Caenorhabditis elegans* (*C. elegans*) contains two genes, *phy-1* and *phy-2*, encoding for collagen-P4Hs that are involved in the synthesis of cuticle collagen.[59,60] The assembly of recombinant P4H from *C. elegans* (PHY-1 and PHY-2) differs from previously described human collagen-P4H. The PHY-1 expressed in Sf9 insect cells was found to form $\alpha\beta$ dimers with either human-PDI or *C.elegans*-PDI (Figure 1.8).[59] The sequence identity between human-P4H-1 and PHY-1 is 45 %, with the identity increasing to 65 % at the C-terminus.[59] PHY-1 mutants were found to play a role in the post-embryonic body morphology.[60]

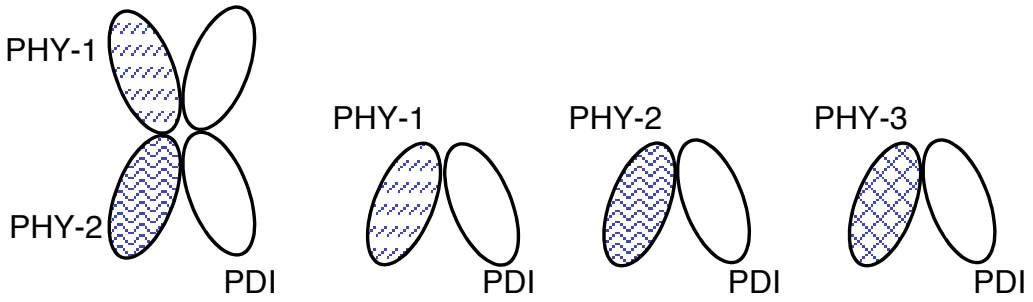


Figure 1.8 Depiction of differing forms of P4H from *C. elegans*. *C. elegans* readily forms PHY-1/PHY-2/PDI_2 tetramers, but also forms PHY-1/PDI and PHY-2/PDI dimers *in vivo*. PHY-3 forms a distinct dimer from the other two P4Hs.

Coexpression of both PHY-1 and PHY-2 with PDI-2 from *C. elegans* were found to form distinct PHY-1/PHY-2/$(PDI\text{-}2)_2$ mixed tetramers most prevalently, with less active PHY-1/PDI-2 and PHY/PDI-2 dimers also formed

(Figure 1.8).[61] The discovery of these heterotetramers was significant, as vertebrate P4Hs do not form mixed tetramers with their α subunits. The assembly of the mixed tetramers was seen both *in vivo* and *in vitro* and was found to be dependent on the *phy-2* gene.[61] Mutations of this gene did not form the PHY-1/PHY-2/$(PDI\text{-}2)_2$ mixed tetramer. The loss of this gene up-regulates the otherwise low expression of the PHY-1/PDI-2 heterodimer.[61] The PHY-1, PHY-2, and PDI-2 are coexpressed in cuticle collagen synthesizing hydrodermal cells during the larval stages 1 - 4 and adult nematodes when the collagen synthesis is at its highest.[61,62]

When expressed with human PDI, recombinant PHY-1 shows catalytic activity almost equal to that of human-P4H-1. However, PHY-1 in the presence of human PDI is not inhibited by poly(L-proline) at concentrations up to 30 μM.[59] PHY-1/PHY-2/$(PDI\text{-}2)_2$ mixed tetramer was shown to be the dominant P4H *in vivo,* and showed similar activity to the human-P4H tetramers when tested with using radiolabeled protocollagen substrate as described previously (10280 ± 1540 dpm/10 μL vs. 17390 ± 4500 dpm/10μL respectively),[63] and had similar K_i values for the αKG analogs, pyridine-2'5'-dicarboxylate (1 μM vs. 0.8 μM) and pyridine-2'4'-dicarboxylate (1 μM vs. 2 μM) for *C. elegans* and human respectivly.[62] Poly(L-proline) was not tested for inhibition with this form.

A third *phy-3* gene has been cloned and expressed as a 295 amino acid residue protein.[64] This differs from the more than 510 residues of the PHY-1 and PHY-2 proteins.[59,60] PHY-3 forms a distinct heterodimer with PDI-1 of *C.elegans* and does not associate with PHY-1 or PHY-2.[64] *phy-3* deletion nematodes had normal morphology at the adult and larval stages but showed a decrease in Hyp levels of collagen in the early embryo stages most likely associated with egg shell development.[64]

1.5.2 *Drosophila melanogaster* P4H

The genome of *Drosophila melanogaster* (*D. melanogaster*) encodes 19 α-subunit related *p4h* genes.[65] Eight of these genes contained significant homology to the C-terminal end of the α subunit of human-P4H-1 and were cloned, including a previously characterized form.[65,66] The eight open reading frames contained conserved residues involved in catalysis. The number of genes in *Drosophila* was unexpected compared with human, as there are usually three to four human homologues for each fly gene. It was speculated that this may be due to the lower number of collagen types found in the *D. melanogaster* genome. *D. melanogaster* has only two collagen genes, which encode for proteins similar to collagen IV in humans. This is considerably lower than the 38 known types of collagen found in humans.[67] The higher number of P4H proteins in *Drosophila* may have evolved to compensate for the fly's low collagen content.[65]

A recombinant P4H from *D. melanogaster* has been expressed in Sf9 insect cells and was found to assemble into a $\alpha_2\beta_2$ tetramer similar to human-P4H.[66] The tetramer can assemble with either the *Drosophila* or human PDI subunit. The α subunit was found to be expressed in larvae, but not in embryos or adults.[66] The K_m values for $(GlyProPro)_{10}$ (260 ± 100 μM) were 12 times higher than for human-P4H-1 (21 ± 5 μM) and 3 times higher than human-P4H-2 (88 ± 25 μM).[66] The K_i values for poly(L-proline), M_r 7000, (18 ± 8 μM) also resemble human-P4H-2 (95 ± 31 μM) more than human-P4H-1 (0.6 ± 0.1 μM), providing evidence that the P4H from *Drosophila* mimics the less common human-P4H-2.[66]

1.6 Monomeric P4Hs

Though collagens are not present in plants they do contain Hyp. The plant cell wall is a self-assembling, interactive network of structural polysaccharides and Hyp rich glycoproteins (HRGP).[68] HRGPs contain Hyp that is O-glycosylated, which is seen only within the plant kingdom. There are no human homologs to these proteins.

The genome of *Arabidopsis thaliana* (*A. thaliana*) contains six open reading frames coding for P4H-like polypeptides.[69] Two forms of P4H from *A. thaliana*, At-P4H-1 and At-P4H-2, have been cloned and characterized.[69,70] Both proteins were found to form simpler monomeric structures, unlike the forms discussed previously. At-P4H-1 and At-P4H-2 both effectively

hydroxylated poly(L-proline) with K_m values of 2 and 30 μM respectively.[69,70] At-P4H-1 binds $(GlyProPro)_{10}$ with a K_m of 60 μM, where as At-P4H-2 has a K_m of 2.8 mM.[69,70] At-P4H-2 was found to act on substrates with three or more Pro residues, like the sequences found in the extensin and arabinogalactan-like peptides of plants.[70]

The cell wall in green algae *Chlamydomonas reinhardtii* (*C. reinhardtii*) is also made up of 25 to 30 different HRGPs. It is built of mostly extensin-like HRGPs. Pro rich peptides have been identified in the HRGPs of the cells wall and Hyp has been identified.[71] The genome of *C. reinhardtii* has 10 *p4h* genes similar to the α subunit genes of other P4H systems. Since Cr-P4H-1 contained the highest identity to the α subunit of human-P4H-1, it was expressed in both Sf9 insect cells and *E. coli* where it purified as a soluble monomer.[72] The *p4h* gene was found to be essential for cell wall assembly, as knockout forms of the gene lead to the inability of the algae to regenerate their walls. Cr-P4H-1 was found to hydroxylate Pro-Pro-Ser-Pro-X and $(Ser-Pro)_{19}$ motifs found in HRGPs of the cell wall. It does not bind human collagen-like peptides tightly and has a K_m for $(GlyProPro)_{10}$ of > 1.5 mM.[72]

Though virus and bacteria are believed not to contain Hyp, both virus and bacteria genomes contain proline-rich, collagen-like sequences.[73-75] Bacterial proline-4-hydroxylases have been identified from *Streptomyces griseoviridus*[76] and *Dactylosporangium sp.*[77], but are distinct from

procollagen-P4Hs as they cannot take peptidyl proline as substrates and do not share significant sequence identity with procollagen-P4Hs.[77] A viral P4H from the eukaryotic algal virus *Paramecium bursaria Chlorella virus-1* (PBCV-P4H) has been cloned and characterized as a soluble monomer in *E. coli*.[73] The function and native substrate of P4H in the virus is not known. Activity assays showed that PBCV-P4H binds $(GlyProPro)_{10}$ substrate poorly with a K_m of 2.9 mM. Poly(L-proline) also bound weakly to the PBCV-P4H. Several synthetic substrates representing proline-rich peptides from the viral genome were screened and $(Pro\text{-}Ala\text{-}Pro\text{-}Lys)_{10}$ had low K_m values of 20 μM and highest activity.[73] This P4H is distinct from the other forms currently known.

1.7 Bacterial Collagens

Because bacteria are believed to not contain Hyp and their collagens must be stabilized with other interactions. Bacteria and phage contain collagen-like sequences in their genomes and trimeric structures have been identified that are similar to animal collagens.[78] The bacterial enzyme pullulanase from *Klebsiella pneumoniae* was the first bacterial protein shown to assist in the generation of a triple helical structure through collagenase digestion.[79] Trimeric proteins have also been identified on proteins located in the tail-fiber and head of bacteriophage.[74]

With recent bacterial genome projects, collagen-like proteins are becoming readily identified and studies are looking into the stabilization of

their collagens. Collagen-related structural motifs (CSMs) have been identified and are characterized by Gly at the primary position and a high Pro content.[80] Bacteria with these motifs most commonly contain Thr at the Yaa position while the Xaa position can be Pro, Ser, or Ala.[80] Collagen-like proteins have been identified in group A *Streptococcus* and *Bacillus anthracis* genomes, and are found to form stable triple helical structures similar to animal collagens.[75,81,82] Glycosylation and electrostatic interactions are believed to stabilize bacterial triple helices, but are still being investigated.[81,83,84]

Hydroxylation of Pro may occur in bacterial systems but has not been identified to date. Model peptides {Gly-Pro-Thr(βGal)}$_{10}$ and {Gly-Hyp-Thr(βGal)}$_{10}$ showed that even when glycosylated, substituting Pro with Hyp further stabilizes the triple helix.[85] Animal collagens have been shown to preferentially hydroxylate Pro at the Yaa position. However, invertebrate systems do show that hydroxylation can occur at the Xaa position as may be occurring in bacteria. Interstitial cuticle collagens from hydrothermal vent worms glycosylate Thr at the Yaa position and amino acid analysis has shown Hyp to be present at the Xaa position.[86] Earthworm cuticle collagens and collagens involved in cell wall assembly from *C. reinhardtii* also preferentially hydroxylate collagen at the Xaa position.[72,87]

1.8 Goals of Current Research

The current available systems for recombinant human-P4H-1 expression give low yield and low purity enzyme, which is not sufficient for spectroscopic studies or structure determination.[88,89] Both spectroscopic and structural knowledge can aid in new strategies for the design of potential therapeutic agents that target fibrosis.

Table 1.3 is a compilation of the information discussed in this introduction on the various forms of P4H. Current model systems for human-P4H-1 include the monomeric PBCV-P4H and Cr-P4H-1. All P4Hs require Fe(II), O_2, and αKG for activity, and mechanistic intermediates have been detected in the reaction of PBCV1-P4H, including the Fe(IV)=O intermediate.[73] A crystal structure of a Cr-P4H-1 has also been recently determined to 1.85 Å.[90] This is the first crystal structure of a collagen-P4H, and identifies the mode of Fe(II) binding to the His-X-Asp/Glu-X_n-His facial triad. However, neither of these forms of P4H bind collagen-like substrate $(GlyProPro)_{10}$ with the same affinity as human-P4H-1. Also both forms take poly(L-proline) as a substrate, which is a known inhibitor for human-P4H-1. Lastly their monomeric structures differ from the $\alpha_2\beta_2$ tetramer of human-P4H. The monomeric structure and differing substrate specificity of PBCV-P4H and Cr-P4H-1 make these systems poor representative models for studying human-P4H-1.

We propose to characterize a collagen-P4H from *Bacillus anthracis* (anthrax-P4H) as a structural and spectroscopic model for human-P4H-1. We will show that anthrax-P4H exists in a homodimer, α_2, form and binds collagen-like substrate with the same affinity as human-P4H-1. The specific aims for this work were:

1. Determine if non-pathogenic *Bacillus anthracis* (ΔSterne) produces Hyp.

2. Expression, purification, and biochemical characterization of recombinant anthrax-P4H. Determine the suitability of anthrax-P4H as a model enzyme for human-P4H-1 by defining its substrate specificity and kinetic parameters in comparison to human-P4H-1.

3. Crystallization and structure determination of the recombinant anthrax-P4H.

In addition:

4. Alternate P4H systems, including two viral forms and four chimeric P4Hs, were cloned and expressed in order to gain further insight of the P4H enzyme.

P4H	# genes	Oligomeric state	Substrate *in vivo*	Substrate *in vitro*	$(GPP)_{10}$ K_m (μM)	Poly(L-proline) *Mr* 40,000 K_m (μM)	Poly(L-proline) *Mr* 40,000 K_i (μM)
Human							
Collagen [a]	3	$\alpha_2\beta_2$	Collagen	$(GPP)_{10}$	18	-	0.02
HIF [b]	3	$\alpha_2\beta_2$	ODDD peptide	$(L\text{-}X\text{-}X\text{-}L\text{-}A\text{-}P)_n$	-	-	-
C.elegans [c]		$\alpha\beta$			-	-	-
	3	$\alpha(1)\alpha(2)\beta_2$	Collagen	$(GPP)_{10}$	20	-	-
D.melanogaster [d]	19	$\alpha_2\beta_2$	Collagen	poly(L-proline)	260	-	2.9
A.thaliana [e]	6	α	PRGP	poly(L-proline)	60	0.2	-
C.reinhardtii [f]	10	α	PRGP	poly(L-proline)	>1500	25	-
PBCV-1 [g]	1	α	?	$(P\text{-}A\text{-}P\text{-}K)_{10}$	2900	100	-

[a] Ref 37 , [b] Ref 39, [c] Ref 62, [d] Ref 65, [e] Ref 69, [f] Ref 72, [g] Ref 73

Table 1.3. Comparison of the differing P4H properties. The K_m and/or K_i data is for the type 1 isoform of each P4H.

Chapter 2: Determination of Hyp in *B. anthracis* Cells

2.1 Introduction

Bacillus anthracis (*B. anthracis*) is an endospore forming, gram-positive bacterium that is the etiological agent of anthrax.[91] Although uncommon in developed countries, anthrax still remains a threat in underdeveloped areas of the world. It is in endemic levels in areas of Central America, South America, Africa, and Asia. Anthrax is primarily found in herbivores, which ingest or inhale soil-containing spores while grazing (Figure 2.1). Most human infections result from contact with contaminated animals or animal products.[92] Three types of human infection include cutaneous,

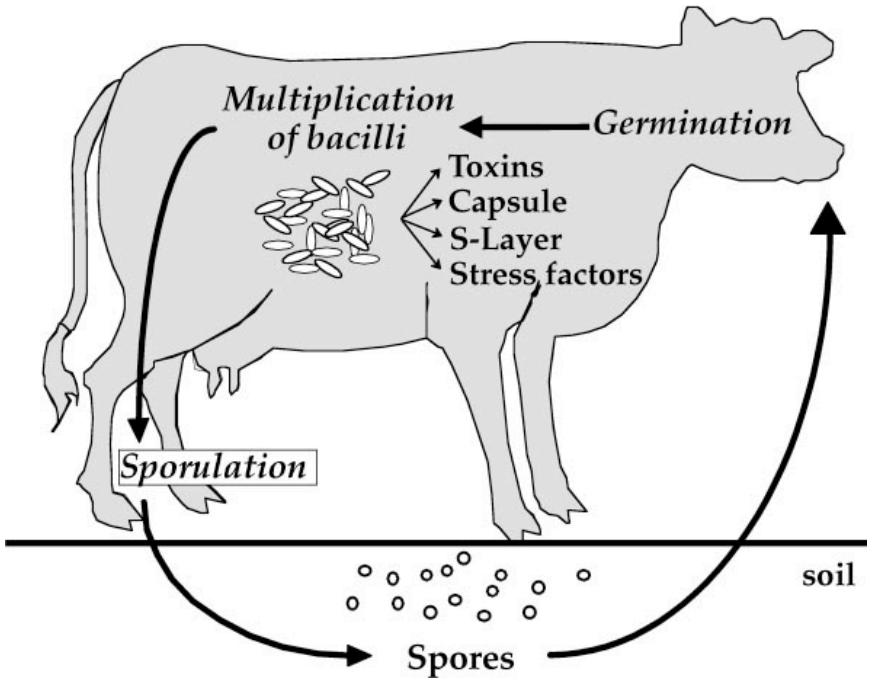


Figure 2.1. Life cycle of *Bacillus anthracis.* Figure used from Annu. Rev. Microbiol., M. Mock & A. Fouet, Anthrax, 2001, 55, 647-671, with permission.

gastrointestinal, and inhalation, which can all progress to fatal systemic anthrax if left untreated.[91]

The prominence of anthrax has gained much attention since the 2001 terrorist attacks. Anthrax spores make an ideal bioterrorism weapon as they can be generated *in vitro* in high quantities and can be aerosolized for exposure to an expansive area.[93] Making the potential to infect large populations fairly easy to do. The hardiness and dormancy of the spore coat also allows for survival of the spore under even the most brutal of environments, making it an ideal bioterrorism weapon.

The anthrax bacterium alternates between two morphologies: the vegetative cell and the dormant endospore. The vegetative rod-shaped cells cannot survive outside of a host system, as they require an environment rich in amino acids, nucleosides, and glucose from the blood and tissues of the animal or human host.[93] When the host's nutrients are exhausted and the bacilli encounter oxygen, the cells begin to undergo sporulation.[93] The spore protects the bacterium from chemical and environmental stressors and allows for survival until the environment is again suitable for stable growth.[94] The spore is protected by a multi-layered coat containing many proteins, and a loose-fitting balloon like layer known as the exosporium (Figure 2.2).[95,96] The exosporium is made of a basal layer and a hair-like nap outer layer, and is chemically complex as it is made up of proteins, polysaccharides, and lipids.[97]

The exosporium is the first part of the bacteria that comes in to contact with the host and so contains many of the antigens to which the antibodies in vaccines react. [91]

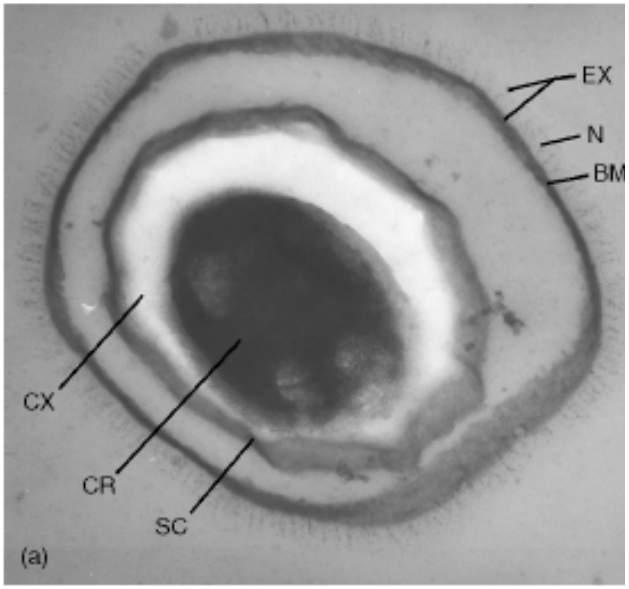


Figure 2.2. Structure of *B. anthracis* spore. The outermost layer is called the exosporium (EX), and is composed of the basal layer (BM) and hairlike nap (N). The inner spore is made up of the spore coat (SC), cortex (CX), and core (CR). Figure used with permission from Levine, S., Tang, Y.W., Pei, Z., Recent advances in the rapid detection of Bacillus anthracis, Rev. Med. Microbiol., 16, 4, page 127.

The immunodominant protein of the exosporium was identified to be a glycoprotein, BclA (*Bacillus* collagen-like protein of *anthracis*), that exists in the exosporium filaments.[98] This protein was found to possess an internal collagen-like region of Gly-Xaa-Yaa repeats of varying length, many which contain Gly-Pro-Thr triplet repeat sequences. These repeats are similar to the Gly-Pro-Hyp repeats found in human collagen. The Thr residues are sites of extensive glycosylation, which add stability to the "collagen-like" triple helix

structure in the absence of Hyp.[83,84,96] The number of Gly-Pro-Thr repeats in the collagen-like region (CLR) of BclA is highly strain specific. BclA from the Sterne strain contains 76 Gly-Pro-Thr repeats, making the protein hydrophobic and hard to solubilize in buffers.[99] Similar Gly-Xaa-Thr repeats have been seen in one of the cuticle collagens of deep-sea vent worms where the Thr residues are glycosylated, and some of the residues in the Xaa position have been identified as Hyp.[86] Much like *Bacillus anthracis*, these worms also survive in harsh conditions with dramatic fluctuation of temperature, low pH, high sulfide levels, and low oxygen levels.[100] Studies on synthetic peptides by Bann and Bachinger showed $\{Gly\text{-}Pro\text{-}Thr(\beta Gal)\}_{10}$ repeats obtained higher thermal stability with the substitution of Pro to Hyp.[85] Earthworm epithelial collagens have also been shown to preferentially hydroxylate Gly-Pro-Ala versus Gly-Ala-Pro.[87] A second collagen-like glycoprotein from *B. anthracis* (BclB) has been identified.[101] Though it does contain collagen-like repeats, these repeats do not contain Pro in the Xaa position as BclA. The repeat in BclB is GITGVTGAT, with sites for glycosylation likely present at the Thr residues as BclA.[102]

A crystal structure of the collagen-binding domain (151-318) of the *Staphylococcus aureus* adhesin protein (Cna) has been determined to 2 Å (PDB: 1AMX), where the binding interface of the domain is built along a groove on a concave β-sheet and has significant structural and chemical

complimentarily to the triple helix of collagen.[103] The subsequent structural study of Cna apo-protein (31-344, N1 and N2 subdomains) and Cna complexed with a synthetic collagen-like peptide revealed that the peptide penetrates through a spherical hole formed by the two subdomains and the N1-N2 linker.[104] Though there is no sequence or structural similarity between Cna and procollagen-P4H, the structure and work on Cna establishes prokaryotic binding of collagen.

Bacterial proline-4-hydroxylases have been identified from *Streptomyces griseoviridus*[76] and *Dactylosporangium sp.*,[77] but are distinct from procollagen-P4Hs as they cannot take peptidyl proline as substrates and do not share significant sequence identity with procollagen-P4Hs.[77] Peptidyl proline has not yet been identified in prokaryotes.[80,81] Therefore, we set out to determine if *B. anthracis* ΔSterne produces Hyp, and if its' production was associated with sporulation.

2.2 Materials and Methods

2.2.1 Growth of *B. anthracis* ΔSterne Cells and Spores and Fractionation of Protein Extracts

B. anthracis strain ΔSterne (spore concentrate in silica) was a kind gift of Dr. George Stewart (University of Missouri, Columbia). ΔSterne is a strain derived from Sterne, but lacks the plasmids necessary for toxicity, pXO1 and pXO2. Initial growing conditions were followed based on the gram-positive bacteria preparation from the Partial Bacterial Proteome Extraction Kit (P-

PEK) (Calbiochem). Spores in silica were streaked on a LB agar plate, and incubated at 37 °C overnight. A single colony was inoculated in 5 mL LB media overnight with gentle agitation. One mL of the overnight culture was diluted to 200 mL fresh LB and incubated at 37 °C, 225 rpm, until the OD_{600} reached 0.8 - 1.0. Growth was stopped in mid-log phase after ~ 4 hrs by placing culture on an ice bath for 15 minutes. Cultures were centrifuged at 20,000 g for 15 minutes at 4 °C. The supernatant was decanted and the pellet resuspended in a total of 2 mL cold Tris-buffered saline (TBS). Centrifugation was repeated, the supernatant decanted, and pellets frozen at - 80 °C.

Due to low detection of Pro and Hyp with the previous method (Calbiochem), we used a protocol that monitors the cell growth and sporulation with phase contrast microscopy. This method is modified from the original spore growth method by Nicholson and Setlow.[105] Twenty-five mL Difco Sporulation Medium (DSM) was inoculated with the same ΔSterne spores absorbed on silica as previously described. The culture was incubated at 37 °C with agitation at 150 rpm until an OD_{600} of 0.5 (mid-log phase) was reached. The starter cultures were diluted 10-fold into 250 mL DSM, and the degree of sporulation was monitored by phase contrast microscopy (Microscopy and Analytical Imaging Laboratory at the University of Kansas) every 24 hours. Vegetative cells were harvested at 24 hours, prior to

sporulation (> 99% vegetative). Complete sporulation was observed after 36 hours. At both time points, cells were harvested by centrifugation at 8,000 g at 4 °C, and the pellets were washed with Milli Q water and centrifuged. Spores were subjected to additional purification to remove residual vegetative debris using Renografin (Bracco Diagnostics) purification according to published methods.[105] Renografin is used in a density gradient to separate the spores from the cells. *B. anthracis* cells and spores were stored at 4 °C until ready for fractionation.

The P-PEK kit from Calbiochem was used to separate the complex protein mixture from the cells and spores. The kit uses reagent mixtures with increasing solubility strength prior to separation by HPLC. Both *B. anthracis* cells and spores were subjected to fractionation resulting in six samples each. Samples were stored at - 20 °C until ready for hydrolysis.

2.2.2 Derivatization and Separation of Fractionated Protein Extracts from *B. anthracis* Cells and Spores

Sample derivatization was based on published methods with minor alterations.[106] Hyp and Pro standards, *o*-phthaldialdehyde (OPA), and reagents for the mobile phases and sample hydrolysis were purchased from Sigma Aldrich, and were of HPLC-grade or equivalent. A mixture of 17 amino acid standards (H-type, 2.5 μmol/ml each in 0.1N HCl; except for L-cystine, which is supplied at 1.25 *μ*moles/ml) and dabsyl-Cl were purchased from Thermo Scientific-Pierce Protein Research Science.

Fractionated cells and spore samples were subjected to acid hydrolysis overnight. 500 μL of each sample was added to 500 μL 12 N HCl in a hydrolysis tube (ChemGlass) and sealed under vacuum. Samples were heated at 110 °C for at least 12 hours to ensure complete hydrolysis. Samples were removed from the hydrolysis tubes, neutralized with NaOH, and freeze dried.

OPA (15 mg) was weighed and dissolved in 500 μL of acetone and 7.5 μL of β-mercaptoethanol. This OPA stock solution (100 μL) was diluted into 6.4 mL of 50 mM carbonate-bicarbonate buffer (pH 9.0). In addition dabsyl-Cl (6.5 mg) was weighed and dissolved in 5.00 mL of acetonitrile (ACN). The mixture was heated to 70 °C in order to completely dissolve dabsyl-Cl in the solution. The diluted OPA solution (500 μL) was added to each ΔSterne *B. anthracis* sample or standard and mixed well. Samples were incubated for 5 minutes at room temperature. The dabsyl-Cl solution (500 μL) was added to each of the samples, mixed well, and incubated at 70 °C for 20 minutes with the lids closed. After the incubation, 1 mL of ACN was added to each tube. The samples were shaken gently and 1 mL of each sample was added to a 5 mL sample loop onto the HPLC.

The amino acid separation was performed with a HPLC system from Shimatzu Scientific Instruments consisting of a SCL-10A system controller, DGU-20A3 degasser, LC-20AT pump, CTO-20A column oven, and SPD-

20AV UV-vis detector. A Phenomenex Ultracarb 5μ ODS C18 column (250 x 4.60 mm) was used for separation with a Phenomenex guard column. The gradient program and mobile phases used for separation of Hyp and Pro amino acids are outlined in Table 2.1 and in previously published methods.[106] Fractions of the Hyp standards and samples were collected and subjected to mass spectrometry for confirmation.

Time (min)	Flow (ml/min)	Solvent A (%)	B(%)	Curve
Initial	1.0	55	45	Isocratic
3.0	1.0	55	45	1
8.0	1.0	40	60	1
12.0	1.0	33	67	1
13.0	1.0	0	100	1
15.5	1.0	55	45	1

Table 2.1. HPLC gradient method for separation of Hyp and Pro from complex protein samples. Mobile phase A made up of 40 mM $Na_2H_2PO_4$ pH 3.0 and mobile phase B made up of 4:1 ACN:H2O + 4 % DMF (v/v). Samples were run at 50 °C at 1 mL/min and detected at 436 nm with a UV-vis detector.

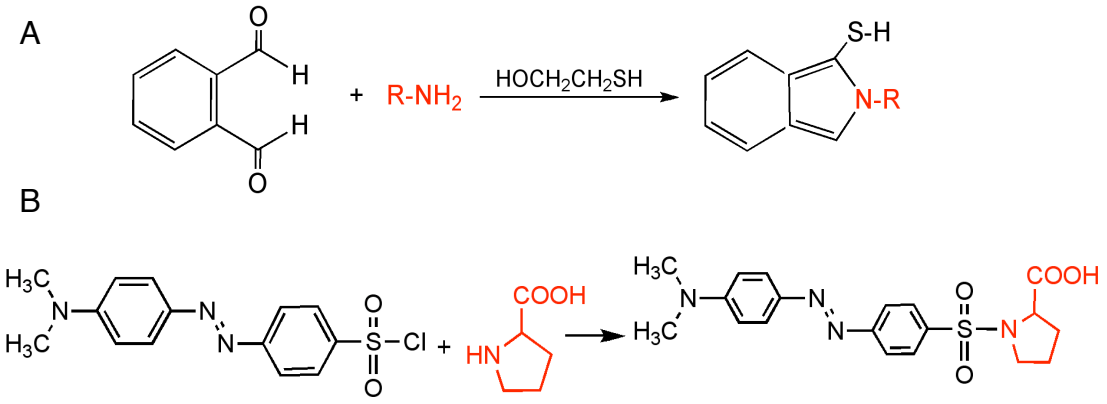
2.3 Results

2.3.1 Growth of *B. anthracis* ΔSterne Cells and Fractionation of the Protein Extracts

We received ΔSterne *B. anthracis* from Dr. Stewart (U of Missouri-Columbia), and grew vegetative cells in LB-agar. Initial cell growth progression was monitored only through OD_{600}, not microscopy. Vegetative cells were grown for 4 hours, and hydrolysis was performed in sealed

ampules under vacuum. These growth conditions were not optimized for maximum growth of cells or to ensure growth of spores and cells separately.

We then subjected the *B. anthracis* cells to a double derivatization with OPA and dabsyl-Cl to allow for detection with UV-vis.[106] The OPA reacts only with primary amines and therefore suppresses the signal from these in the chromatograph. The dabsyl-Cl can then react with the secondary amines present in Hyp and Pro, resulting in a less complex chromatograph. Scheme 2.1 shows the reactions of OPA and Dabsyl-Cl with amino acids for pre-column derivatization.



Scheme 2.1 Reaction schemes for the double derivatization of amino acids. A) Fluorogenic amine-derivatization reaction of *o*-phthaldialdehyde (OPA). B) Detection of secondary amines, Pro and Hyp, by dabsyl-Cl.

Initial experiments were performed to validate the method for selectively determining Hyp and Pro from a complex amino acid mixture.[106] Figure 2.3 shows that Hyp and Pro (2.5 μg/mol) are detected from of a complex mixture of 18 amino acids. Fractions containing the Hyp peak were analyzed by mass spectrometry and confirmed as Hyp with an m/z of 419.1.

Fractionation was performed per protocol from the P-PEK kit (Calbiochem) using buffers of increasing solubility strength (Figure 2.4).

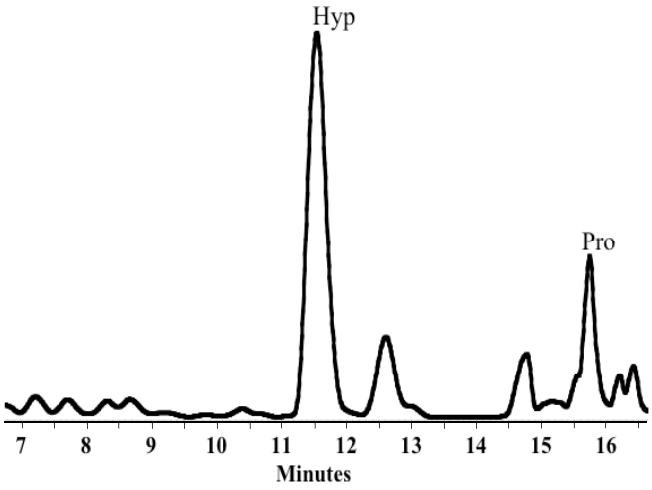


Figure 2.3. HPLC chromatogram of amino acid standards, detecting Hyp and Pro.

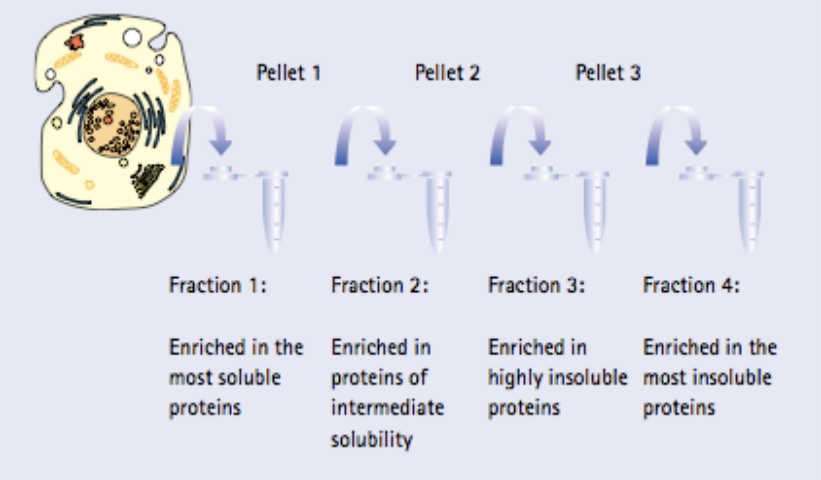


Figure 2.4. Flow chart of partial proteome extraction, P-PEK. Proteins are sequentially extracted with four buffers with increasing solubilization strength. Figure used from Calbiochem Cat. No. 539780 catalog.

Figure 2.5 is an overlay of the HPLC chromatographs from each of the fractions isolated. We were successfully able to detect small levels of Hyp in fraction 2 containing Tris, Urea, Detergent, and DTT (Calbiochem), which corresponds to the same retention time as the Hyp standard. The retention times of these runs differ slightly from more recent data because we were using a different HPLC instrument in the KU-Analytical Lab at the beginning stages of the project.

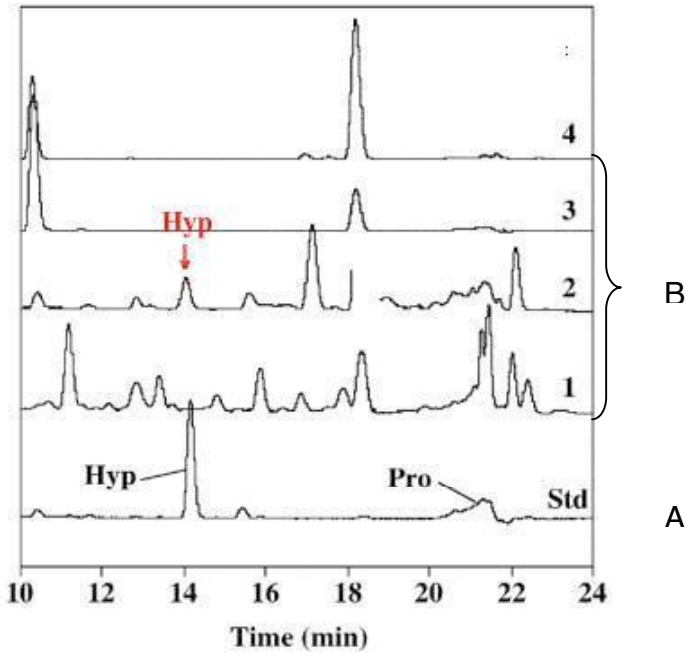


Figure 2.5. HPLC analysis of amino acid standards and fractionated protein extracts from *B. anthracis*. A) Chromatogram of 18 amino acid mixture detecting Hyp and Pro. B) Chromatograms of four samples of varying solubility obtained after fractionation of protein extracts from *B. anthracis* vegetative cells.

The initial method of cell growth resulted in low levels of Hyp that were difficult to detect consistently from a sample consisting of a combination of cells and spores. We therefore set out to grow cells and spores separately and determine if Hyp levels are associated with sporulation. We modified a sporulation protocol from Nicholas and Setlow in which we used a Renografin (Bracco Diagnostics) density gradient to purify the spores from the cells.[105] We monitored cell growth and sporulation with phase contrast microscopy (KU-Microscopy and Analytical Imaging Laboratory), and were able to isolate both cells and spores (Figure 2.6).

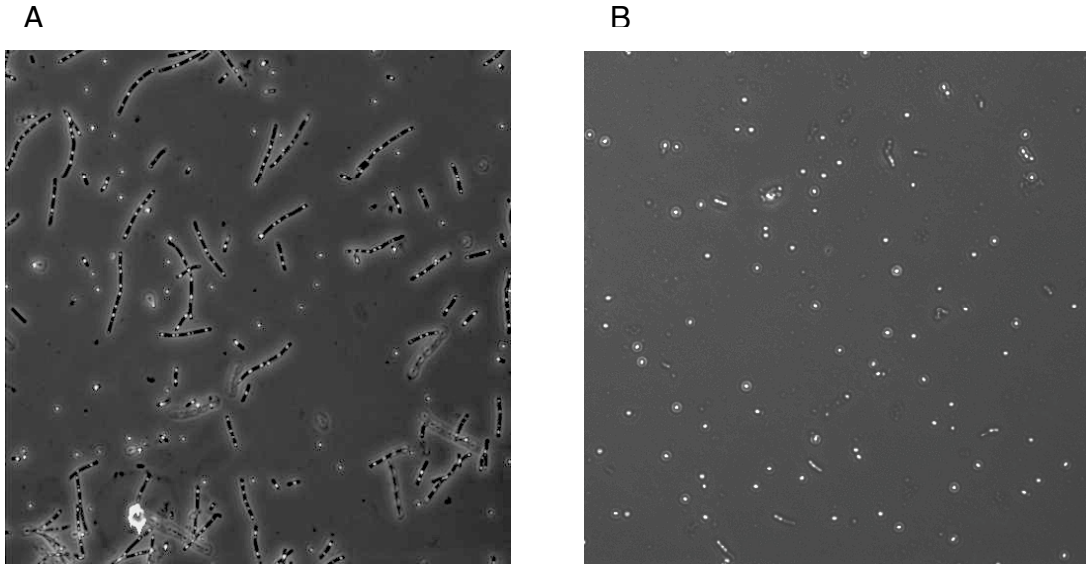


Figure 2.6. Phase contrast microscopy of *B. anthracis* ΔSterne growth. A) *B. anthracis* vegetative cells after 24 hours. B) *B. anthracis* spores after 36 hours.

After the isolation and spore purification we repeated the P-PEK serial extraction, acid hydrolysis, double derivatization, and HPLC separation of the cells and spores separately. We were consistently able to detect Hyp in the

Wash 2 fraction in both the cells and spores. To confirm that the peak was Hyp we spiked the sample with 20 μM Hyp standard and monitored the HPLC chromatogram for co-elution. Figure 2.7 shows the sample run with and without the presence of the Hyp standard. The two peaks co-elute at 11.2 minutes, providing evidence that Wash 2 contains Hyp. Initial attempts to run mass spectrometry on the Hyp fraction from the cells has been unsuccessful due to low concentration. We have combined the Hyp peaks from several runs, but have still not detected anything above background. High-resolution mass spectrometry (FT-MS) is now being investigated.

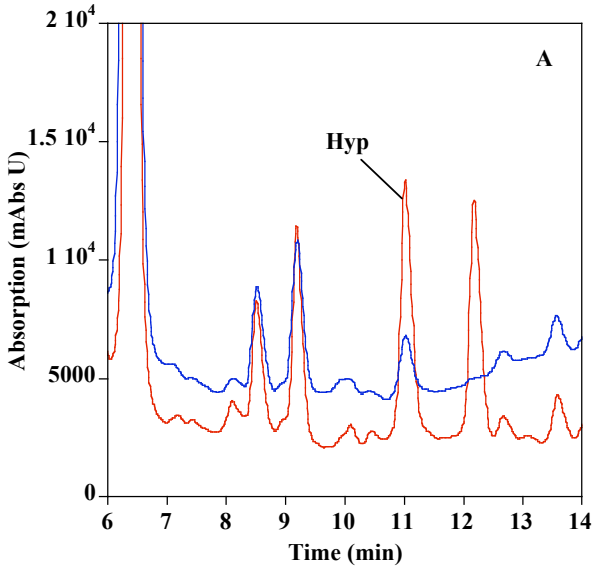


Figure 2.7. HPLC analysis of Hyp in the protein extracts from *B. anthracis* (ΔSterne). HPLC traces of protein extracts of unsporulated cells with (red line) and without (blue line) standard addition of authentic dabsyl-Hyp (elutes at 11.2 min).

We next set out to determine if Hyp production levels were affected from vegetative cells to sporulation. Figure 2.8 shows a comparison of Hyp present in cells (blue line) versus spores (red line). The presence of Hyp in the spore samples is greatly (> 10-fold) increased in the spore containing samples, suggesting that Hyp formation is associated with sporulation. High-resolution mass spectrometry analysis is also being investigated for definite confirmation that these peaks are Hyp.

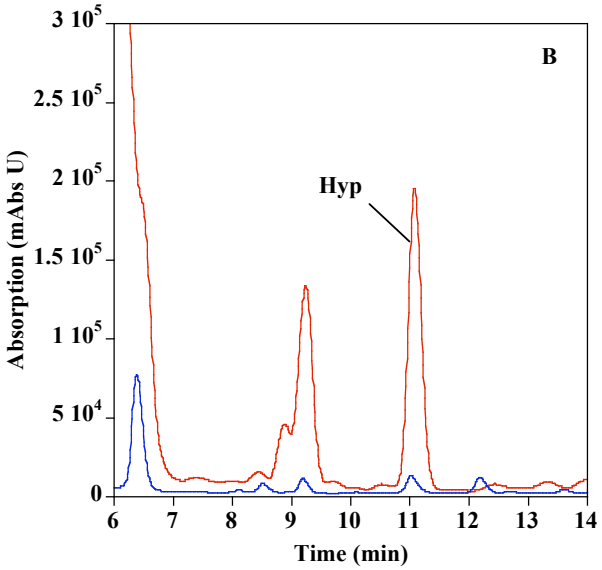


Figure 2.8. HPLC analysis of Hyp comparison of the amount of Hyp in protein extracts from sporulated (red line) and unsporulated cells (blue line).

2.4 Discussion

2.4.1 Absence of Hyp in Prokaryotic Systems

As discussed previously, Hyp is believed essential in providing thermal stability and structural support to the collagen triple helix. In animal collagens Hyp is most commonly located in the Yaa position of Gly-Xaa-Yaa repeat

sequences, with Xaa commonly Pro. Collagenous domains vary in length and can either contain uninterrupted Gly-Xaa-Yaa repeats or contain imperfection in the sequence which cause flexibility in the otherwise rigid triple helix structure.[107]

Though bacteria and viruses were generally believed not to synthesize Hyp, collagen-related structural motifs (CSM) have been previously identified in these organisms.[80] Gly at the primary position and a high Pro content are markers in identifying these proteins.[81] In bacteria, the Yaa position is occupied typically with a Thr residue, and in these Thr-rich CSMs the Xaa is usually Pro, Ala or Ser.[80] The Thr residue can form direct hydrogen bonding with the backbone carbonyl as seen in animal collagens [80]. The closest homologue to the Thr rich CSM is the cuticle collagen found in hypothermal vent worms.[80] Interestingly, these vent worms contain a P4H, which hydroxylates Pro at the Xaa position rather than the conventional Yaa position.[86]

Collagen-like sequences have been identified in several bacteria genomes including group A *Streptococcus* proteins Scl1 and Scl2, which are expressed as cell surface proteins on the gram positive bacteria.[74,75,79] These proteins were the first to be structurally characterized and showed direct evidence that the collagen-like sequences formed triple-helical structures.[75] Collagen-like proteins have also been identified and characterized in the

genome of *Bacillus anthracis*.[96,98,101] None of the proteins are thought to be stabilized by Hyp as with animal collagens. Recent research has been performed to determine stabilization of collagen triple helix in the absence of Hyp.[81,83,84,99]

2.4.2 Stabilization of Triple Helix in the Absence of Hyp

Though the identification of Hyp in prokaryotes has not been previously shown, research on collagen stabilization has been performed to understand how prokaryotes bind and stabilize collagen. Work with the recombinant Scl2 protein from *Streptococcus pyogenes* shows that the collagens are stabilized by electrostatic interactions caused by ion pairing, inter-chain hydrogen bonds and hydration network associated with packing of polar groups.[81] Though Scl2 has only ~ 12 % Pro residues in the Xaa position, versus ~ 20 % found in animal collagens, there are 30 % charged residues compared with ~ 15 % seen in fibrillar collagens.[81] These charged particles lend to the stability of the triple helix due to the hydration network.

Structural studies have also been performed on *Bacillus* collagen-like protein from *anthracis* (BclA). This protein contains three domains: N-terminal domain (NTD) made up of 38 amino acids, a variable amino acid length collagen-like domain located in the center of the sequence (CLR), and a 134 amino acid carboxy-terminal domain (CTD).[96,98] The CLR varies from 17 - 91 amino acids and is responsible for the variation in length of the hair-

like naps on the exosporium of 140 - 608 Å.[108] rBclA in *E. coli* was found to form trimers which are sensitive to collagenase and have characteristic CD spectra similar to animal collagen.[99] The T_m of the CLR was determined to be 37 °C and the entire rBclA trimer has a T_m of ~ 90 °C, which is greater than seen in animal collagens.[99] The high T_m for the triple helix structure is due to the strong interactions of the CTD. It is believed that these strong interactions present in the CTD aid in the spontaneous refolding of the CLR after denaturation.[99]

Within the CLR region of BclA, the majority of Gly-Xaa-Yaa repeats are Gly-Pro-Thr repeats. There is a high Pro content of ~ 25 % in BclA and the Thr residues have been shown to be sites for O-glycosylation.[83,98] Two O-linked oligosaccharides were identified on BclA, a 715 Da tetrasaccharide and a 324 Da disaccharide. The 715 Da sugar is a novel sugar not seen in other members of the *Bacillus* family and was given the name anthrose.[83] This sugar was found solely linked to the Thr residues present in the Gly-Pro-Thr repeat sequences within the CLR. The 324 Da sugar was found to attach outside of the CLR. The function of the glycosylation of BclA is still to be determined, but may be involved in the stability of the triple helix.[83]

Glycosylation involvement in triple helix stability has been studied with model peptides.[84] Proteins with O-glycosylation have been shown to have higher resistance to enzyme degradation. Peptides Ac-(Gly-Pro-Thr)$_{10}$-NH_2

and Ac-{Gly-Pro-Thr(Galβ)}$_{10}$-NH_2 were synthesized and characterized with CD and NMR. Only Ac-{Gly-Pro-Thr(Galβ)}$_{10}$-NH_2 was found to form a triple helix similar to collagen in the absence of Hyp.[84] One possible mechanism for the triple helix formation in the glycosylated peptide is that the galactose restricts the conformational space available in the polypeptide backbone similar to what is seen with Pro and Hyp in the pyrrolidine ring.[84] The increase in stability seen in the glycosylated peptide may also be due to the increase in hydrogen bonding of the sugar moiety to the polypeptide backbone creating a more ordered structure.[84]

Animal collagens are known to require Hyp at the Yaa position for stable, correctly folded collagen. However, bacterial systems have shown that in the absence of Hyp at the Yaa position they can still form stable collagens with characteristics similar to animal collagens. Though rare, evidence of Hyp at the Xaa position does occur in nature, most commonly in invertebrates.[87,109] We believe hydroxylation is occurring in bacterial systems at the Xaa position to further stabilize the triple helix.

2.4.3 Hyp at the Xaa Position in Invertebrates

Vent worms exist in harsh environmental conditions with minimal oxygen and fluctuating temperatures from 20 - 70 °C. The tube worm *Riftia pachyptila* contains both a cuticle collagen localized on the worm surface and an interstitial collagen located in the body wall underneath the epidermis.[86]

The interstitial collagen has a Tm of 29 °C and hydroxylation exclusively at the Yaa position.[86] The cuticle collagen shows a higher T_m of 37 °C, lower Hyp content and higher Thr content (200/1000) with 80 % being glycosylated.[86] Similar to BclA, Gly-Xaa-Thr repeats have been seen in the cuticle collagens of deep-sea vent worms with some of the residues in the Xaa position identified as Hyp.[86,110] Earthworm P4H from *Lumbricus terrestris* has also been shown to hydroxylate at the Xaa in its cuticle collagens.[87] Tripeptides Gly-Hyp-Ala [111] and Gly-Hyp-Ser [112] were both isolated from collagenase digests from earthworms in quantities of 10 % and 15 % Hyp. Earthworm P4H has shown to hydroxylate at the Xaa position if the Yaa position is a primary amino acid. However, if the Yaa position is occupied with a Pro residue it will preferentially hydroxylate the third position.[87] Unhydroxylated natural collagen was used as an *in vitro* substrate as (Gly-Pro-Pro)$_n$ repeats are rare or nonexistent in earthworm cuticle collagen.[87,113]

Further studies by Bann and Bachinger investigated the position of Hyp at the Xaa position with and without glycosylation of Thr residue at the Yaa position.[85] The calculation study has shown that {Gly-Pro-Thr(βGal)}$_{10}$ has a T_m ~ 20 °C lower than {Gly-Hyp-Thr(βGal)}$_{10}$ and Hyp in the Xaa position actually increases triple helix stability when Yaa is occupied by an amino acid other than Pro.[85] This is most likely due to the stereoelectronic inductive effect in which the OH group draws the electrons from the imide

bond and favors the *trans* isomer.[114] All of the peptide bonds in the triple helix are *trans*, and an increase in *trans* isomer increases the overall stability of the triple helix.[85]

2.5 Conclusions and Future Directions

Our results show that there is a significant amount of Hyp associated with sporulation in *B. anthracis.* This strongly suggests that a P4H-like protein able to catalyze the formation of Hyp exists in *B. anthracis*. Studies show that P4H can hydroxylate in the Xaa position in lower order organisms and that Hyp in the Xaa position adds to the stability of these triple helices when Yaa is a primary amino acid.

BclA does form strong triple helices on the hair-like naps of the exosporium, which suggests that this is the natural substrate for anthrax-P4H. The genome of *B. anthracis* does contain several predicted amino acid sequences that contain proline-rich repeats that are annotated as collagen-like proteins which may serve as substrates for anthrax-P4H (http://www.tigr.org/CMR).[115]

We have not been able to identify the protein associated with Hyp to date, but will undertake a proteomics approach to identify binding partners. In collaboration with Dr. George Stewart (University of Missouri), a *p4h* gene knockout of *B. anthracis* has been developed. We believe the levels of Hyp will be negligible with the knockout as the gene responsible for Hyp is

eliminated, providing conclusive evidence that P4H is responsible for the Hyp formation.

Chapter 3: Expression, Purification, and Biochemical Characterization of a Collagen Prolyl-4-Hydroxylase Like Enzyme from *Bacillus anthracis*

3.1 Introduction

Collagens and collagen-like proteins are essential to provide structural integrity and tensile strength for muscles and tissues in animals.[2] The triple helix structure of collagen is stabilized by Hyp at the Yaa position of (Gly-Xaa-Yaa)$_n$ repeats, where Xaa is often Pro.[2] Procollagen-P4H is the enzyme responsible for the post-translational oxidation of Pro. In humans, P4H is a 240 kDa $\alpha_2\beta_2$ tetramer, where the β-subunit is a protein disulfide isomerase (PDI) that serves to keep the α-subunit correctly folded and to retain the tetramer in the endoplasmic reticulum.[35,116] The α-subunit contains the active site for Pro hydroxylation as well as a separate peptide-binding domain (PDB: ITJC).[117] Procollagen-P4Hs are also found in a variety of invertebrates that contain collagen such as nematodes, sponges, mollusks, and worms, all that contain collagen.[36,107,118,119] Earthworm cuticle collagen has been found to have hydroxylation predominantly occurring at the Xaa position rather than the Yaa position.[87,111] Deep sea vent worms have also been shown to contain collagen that has Hyp at the Xaa position, when the Yaa position is Thr.[86,110] Recombinant P4Hs from algae have also been shown to preferentially hydroxylate Pro at the Xaa position.[72]

Several expression systems for a recombinant form of human-P4H-1 have been reported. A baculovirus expression system using an Sf9 insect

cell line and a yeast expression system are reported but the yields of the isolated protein is low (~ 5 mg/L of culture) and no details for the purification procedures are reported.[120,121] In 2004, an *E. coli* expression of human-P4H-1 tetramer was reported.[89] The expression vector contained mutations in the thioredoxin reductase and glutathione reductase genes in order to increase disulfide bond formation in *E. coli*.[89] Site-directed mutagenesis studies have shown that Cys276 and Cys293, and Cys486 and Cys511 (human-P4H numbering) form disulfide bonds that are essential to the formation of an active tetramer.[122,123] The full-length form of human-P4H was expressed in Origami BL(DE3) *E. coli* cells and the isolated proteins showed kinetic properties similar to those obtained for human-P4H expressed in the baculovirus expression system.[89,121] The *E. coli* expression system yields only 1.5 mg/L of protein per liter of culture, which is not a high enough concentration suitable for spectroscopic studies. An alternate *E. coli* expression for human-P4H was reported by Myllyharju and coworkers, in which they obtained 25 mg/L of protein per liter of culture using a fermenter.[88] This human-P4H was expressed in a double mutant of *E.coli* lacking both thioredoxin reductase and glutathione reductase. However, we have not been able to reproduce the protein yield to date.

A recombinant expression system for human-P4H in *E. coli* has also been developed by Dr. Hirakawa of the Limburg Lab (unpublished results).

The vector contains promoters which express both the α and β subunits separately. This expression system yields 5 mg of protein per liter of culture.

A. thaliana,[69,70] *C. reinhardtii,*[72] and PBCV-1[73] all contain a monomeric P4H which have about 30 % sequence identity to the C-terminal catalytic domain of the α subunit of human procollagen-P4H, but have distinct substrate specificities from the human enzyme. These P4Hs are currently used as models to study human-P4H for spectroscopic and structural studies. Plants and green algae contain HRGPs in the extracellular matrix.[69,72] Two forms of P4H from *A. thaliana*, At-P4H-1, and At-P4H-2, have been cloned and characterized. Both of the proteins were expressed in Sf9 insect cells and expressed as the soluble form only at low yields, as most of the protein was expressed as inclusion bodies.[69,70] Neither of the proteins were purified to homogeneity. The crude mixture was used to examine the substrate specificities of both proteins and found to differ significantly from that of human-P4H. Ten P4Hs have been identified in the *C. reinhardtii* genome, the recombinant forms of one of the 10 P4Hs have been expressed both in Sf9 insect and *E.coli* cells, purified, and biochemically characterized.[72] However, the recombinant proteins hydroxylate HRGP motifs and collagen-like peptides, which differ from $(Gly\text{-}Pro\text{-}Pro)_n$ peptide used in assays for human-P4H.

A P4H from PBCV-1 produces soluble, monomeric protein that hydroxylates collagen-like repeats located in the viral genome.[73] The natural substrate and current biological significance of P4H in PBCV-1 is not known. Therefore, hypothetical proteins containing collagen-like sequences from the genome were synthesized and used as potential substrates in the *in vitro* activity assay. The enzyme binds $(Gly\text{-}Pro\text{-}Pro)_{10}$ with a K_m of 2.9 mM, which is 150-fold higher than human-P4H.[40,73] Spectroscopic characterization of PBCV-PH identified two chemically competent intermediates using stopped-flow UV-vis, and Mössbauer spectroscopy.[124] The first intermediate was assigned as an Fe(IV)=O species.[124] The high ^{2}H-KIE also provided evidence that the Fe(IV) activates the substrate through H-atom abstraction.[124] The second intermediate was identified as a high spin Fe(II) and assigned as a Fe(II)·product complex with comparison of the data with TauD.[32,124,125]

In this chapter, we describe the characterization of a P4H-like protein from *B. anthracis* (anthrax-P4H). We have found that anthrax-P4H is an α_2 homodimer αKG/Fe(II)-oxygenase that can catalyze the oxidation of ascorbate and glutathione. Kinetic studies show that a $(Gly\text{-}Pro\text{-}Pro)_{10}$ peptide, a common *in vitro* substrate for mammalian P4Hs, increases the rate of O_2 consumption in the presence of glutathione and exhibits Michaelis-Menten kinetic behavior. This is the first report of a P4H-like protein from a bacterial source.

3.2 Materials and Methods:

3.2.1 Cloning, Expression, and Purification of N-terminally His-tagged Anthrax-P4H.

The 651 bp ORF encoding the putative *p4h* gene (BA 4459) was provided to us in Gateway (Invitrogen) entry vector pDONR21 by the TIGR-PFGRC (Maryland). We initially subcloned the gene into Gateway plasmid pDEST-17 (Invitrogen), which contains an N-terminal His-tag for protein expression in *E. coli* using the LRII clonase reaction. A 150 ng aliquot of the gene was mixed with 300 ng pDEST-17 vector and 2 μL LR clonase II enzyme. This mixture was incubated at 15 °C overnight to ensure completion of the reaction. The reaction was stopped with the addition of 1 μL proteinase K, vortexed briefly and incubated at 37 ºC for 10 minutes according to standard reaction protocols.

After the completion of the LR clonase II reaction, the recombination reaction was transformed into competent DH5α *E. coli* cells (Invitrogen). Transformed cells were analyzed by colony PCR, and the plasmids were isolated from the overnight culture LB-ampicillin (LB-amp) of the positive colonies. The pDEST-17 expression plasmid was transformed into BL21-AI (Invitrogen) competent *E. coli* cells. Starter cultures (100 mL) were made in LB-media containing 100 μg/mL ampicillin. Ten mL of the starter cultures were used to inoculate 1 L of LB-media containing 100 μg/mL ampicillin. Cells were grown at 37 ºC until the OD_{600} reached 0.6 at which point they

were induced with 20% L-arabinose to a final concentration of 0.2%. After induction, cells were incubated at 37 ºC for 3 hours and then harvested by centrifugation at 8,000 rpm at 4 ºC for 20 minutes. Cells were frozen at - 80 ºC until ready for use.

Cells were resuspended in 40 μM imidazole, 0.5 M NaCl and 20 μM KPi (pH 7.4) containing protease inhibitors (1 mM PMSF, 1 mM leupeptin, 1 mM antipain, 1 mM pepstatin), lysozyme, and DNAse and RNAse. Cells were lysed by ultrasonic disruption and centrifuged at 20,000 rpm at 4 ˚C for 1 hour in order to separate the soluble from insoluble portions of the cells. Once separated, both the soluble and insoluble fractions were analyzed by SDS-PAGE to determine if anthrax-P4H was expressed as a soluble protein. The soluble portion was concentrated down to 10 mL using 10,000 MWCO Amicon concentrators.

Recombinant His-anthrax-P4H was purified using a HisTrap HP (GE Biosciences) Ni^{+2} Sepharose column. The protein was loaded onto the HisTrap HP column equilibrated with binding buffer (20 mM sodium phosphate, 0.5 M NaCl, 40 mM imidazole, pH 7.4). Unbound proteins were removed by washing with 5 column volumes of the binding buffer. Increasing the concentration of imidazole to 500 mM over 5 column volumes eluted His-anthrax-P4H. The fractions were analyzed with SDS-PAGE and those containing pure protein were pooled, concentrated using 10,000 MWCO

Amicon concentrators and desalted using a HiTrap Desalting (GE Biosciences) column to remove the imidazole. The fractions containing impurities were loaded on a Superdex 16/60 Size Exclusion column fitted to an FPLC (Akta), and eluted with 50 mM KPi, 150 mM KCl, pH 7.0. Additional purifications were performed as above with the addition of 2 % Triton-X 100 in an attempt to increase solubility. Fractions containing His-anthrax-P4H (> 99% purity) were combined, made 20% v/v with glycerol, concentrated to > 30 mg/mL and frozen at -80 ˚C. Protein concentration was determined with the BCA assay (Pierce) using BSA as the protein standard, and had an average yield of 5 mg of protein per liter of culture.

3.2.2 Cloning, Expression, and Purification of Native Anthrax-P4H.

Precipitation was seen with the N-terminally His-tagged anthrax-P4H protein. Therefore we decided to prepare an untagged version of the putative anthrax-P4H. The *p4h* gene was cloned using traditional methods. PCR primers (forward 5'-CATGCCATGGCAAACAACAATCAAATAGG-3' and reverse 5'-GGAATTCCATATGGCCAACTTTGTACAAGAAAGC-3') were designed to clone the gene between the NcoI and NdeI sites of pET15b (Novagen) in order to express the recombinant protein without the N-terminal His-tag. The NcoI and NdeI restriction recognition sites are underlined. The PCR product was initially ligated into pGEMT-Easy (Promega), and then transformed into competent DH5α *E. coli* cells to select for expression clones.

Transformation colonies were analyzed with colony PCR. Purified plasmid was sequenced in the forward and reverse direction (DNA sequencing facility at University of California, Berkeley) and then subcloned into pET15b.

The pET-15b expression vector encoding for the expression of the putative anthrax-P4H was transformed into *E. coli* strain BL21(DE3) (Novagen). A single colony from an LB-amp agar plate was used to inoculate a 90 mL LB-amp starter culture, which was incubated at 37 ˚C overnight. Aliquots of 15 mL of the overnight culture were used to inoculate 6 x 1.5 L of LB-amp media. The cultures were incubated at 37 ˚C to an OD_{600} of 0.6, and protein expression was induced by addition of 200 μM isopropyl-β-D-thiogalactopyranoside (IPTG). The cultures were incubated at 37 ˚C for 3 hours after induction, at which point the cells were harvested by centrifugation.

Cells were suspended in lysis buffer: 50 mM Tris (pH 7.4), 1 mM EDTA, 5 mM β-mercaptoethanol, containing protease inhibitors (1 mM PMSF, 1 mM leupeptin, 1 mM antipain, 1 mM pepstatin), lysozyme, and DNase and RNase. The cells were lysed using a sonifier (Branson) and then centrifuged at 24,000 g for 1 hour at 4 °C. The crude cell lysate was loaded onto a 100 mL DEAE-sepharose column (Tosoh Biosciences). The column was washed with three volumes of lysis buffer and then the protein was eluted with a four column volume gradient rising from 0 - 250 mM KCl. Fractions containing

anthrax-P4H were identified by SDS-PAGE, pooled, concentrated (Amicon) and then loaded onto a HiLoad Superdex 200 column (Amersham) connected to an Akta FPLC. Anthrax-P4H was eluted with 50 mM Tris buffer (pH 7.4) containing 150 mM KCl and 5 mM β-mercaptoethanol, and the fractions containing highly pure anthrax-P4H (> 99 % purity) were combined, concentrated to > 20 mg/mL, and stored frozen at 4 ˚C. Due to the presence of β-mercaptoethanol as a reducing agent in our buffers, protein concentration was determined with Bradford assay (Pierce) using BSA as the protein standard. Standard yield for this protein prep is 10 mg protein per liter of culture.

3.2.3 Quaternary Structure Determination of Native Anthrax-P4H

3.2.3.1 Gel Filtration

The molecular weight (*Mr*) of anthrax-P4H was estimated by gel-filtration chromatography using a HiLoad Superdex 200 column (GE Biosciences) connected to an Akta FPLC. Low and High molecular weight standards (GE Biosciences) were used to calibrate the column and the approximate *Mr* of anthrax-P4H was determined from the retention time.

3.2.3.2 MALDI-TOF MS

The protein samples were submitted in the presence and absence of β-mercaptoethanol for MALDI-TOF mass spectrometry (KU Proteomics Facility). β-Mercaptoethanol was removed from the purified anthrax-P4H by

dialysis overnight against 50 mM Tris buffer, pH 7.4 at 4 °C. The solution of anthrax-P4H (100 μM) was diluted to 2 μM with Milli Q water. The protein sample was mixed with the matrix (sinapinic acid, 20 mg/mL in 1 M ammonium citrate, adjusted to pH 7 with ammonium hydroxide) at a 1:1 ratio, and spotted on a stainless steel MALDI sample plate. Analysis was performed using a MALDI TOF mass spectrometer (PerSeptive Biosystems, Voyager-DE STR). The instrument was operated in positive linear mode with the following parameters: accelerating voltage 25 kV, grid voltage 90%, guide wire 0.1% and extraction delay time 500 nsec. Acquisition mass range was 5,000-80,000. External mass calibration was performed using m/z 12361.1 and 24722.2 ions of cytochrome c immediately prior to the analysis of anthrax-P4H samples.

3.2.3.3 Monomer-Dimer Equilibrium Constant Determination

The purified anthrax-P4H in 150 mM KCl, 5 mM β-mercaptoethanol in 50 mM Tris buffer at pH 7.4 was placed in Ultracel YM30 membrane, MWCO of 30,000 (Millipore) and centrifuged at 10,000 g for 15 min at room temperature. The experiments were performed at three different concentrations of anthrax-P4H with the corresponding buffer blank. The sample remaining on the membrane and in the flow-through were diluted with the corresponding buffer blank to 100 μL and the absorption at 280 nm was measured. The concentration of each fraction was determined using the

extinction coefficient for anthrax-P4H approximated, ε = 45,100 $M^{-1}cm^{-1}$ where we assume the ε value of the dimer is twice that of the monomeric form of anthrax-P4H. The apparent dissociation constant (K_d^{app}) was determined by equation 1 from samples at three different concentrations.

$$K_d^{app} = [\text{dimer}]/[\text{monomer}]^2 \quad (1)$$

3.2.4 Preparation of C53S and C53A Mutants of Anthrax-P4H

Site-directed point mutations were incorporated into the expression vector of the wild-type (WT) as a template using the QuikChange method (Stratagene). For C53S, two complementary primers, forward primer: 5'-GGAAATGTATTAAGTGATGAAGAAGAGTGATGAATTAATTGAA-3' and reverse primer 5'-TTCAATTAATTCATCACTCTCTTCATCACTTAATACATTTCC-3' were used. For C53A, forward primer: 5'-GTATTAAGTGATGAAGAGGCTGATGAATTAATTGAATTG-3', and reverse primer 5'-CAATTCAATTAATTCATCAGCCTCTTCATCACTTAATAC-3'. The resulting plasmids were fully sequenced and the point mutations were confirmed. The mutant proteins were expressed and purified as described for the WT.

3.2.5 Anthrax-P4H Activity Assays

The activity of anthrax-P4H was assayed using a fluorometric detection of αKG consumption,[126] a coupled assay employing L-glutamic dehydrogenase (Sigma Aldrich) to detect αKG or by an O_2 sensitive Clark-type electrode (Hansatech, UK).

3.2.5.1 Fluorescence Assay

Schofield has reported an assay for monitoring αKG/Fe(II) oxygenase activity based on the formation of a fluorescent product from the reaction of αKG with *o*-phenylenediamine (OPD).[126] This assay was used in the initial screening of anthrax-P4H. Limits of the assay were determined with an αKG standard curve. Two-hundred μL of αKG standards (0-2 mM) were incubated at 37 °C for 15 minutes after which time the reaction was quenched with the addition of 400 μL 0.5 M HCl. Derivatization was initiated with the addition of 200 μL 10 mg/mL OPD in 0.5 M HCl and heated for 10 min at 95 °C. The sample was centrifuged at 10, 000 g for 5 min, and the supernatant (600 μL) was neutralized with 300 μL of 1 M NaOH and fluorescence was measured on a Quantamaster-3 (QM-3) Scanning Luminescence Spectrofluorometer with FeliX32 software (Photon Technology International, Birmingham, NJ) with the excitation filter at 340 nm and the emission filter at 420 nm. Samples were run in triplicate.

Anthrax-P4H activity was measured within the detection limit of the assay as determined above. Ten μM anthrax-P4H was pre-incubated with 2 μM Fe_2SO_4 (prepared as 500 μM stock in 20 mM HCl and diluted with water) at 37 °C for 5 min to ensure complete Fe(II) incorporation into the enzyme. The reaction was initiated with the addition of 20 μl of enzyme/iron mixture to 180 μL of the substrate/αKG mixture pre-equilibrated at 37 °C for 5 min

containing 750 μM DTT, 750 μM αKG, substrate, and 50 mM Tris buffer, pH 7.4. The absorbance change was monitored at various time points, the reaction was quenched, and the product was derivatized and neutralized as described above. Data was also collected as described above.

3.2.5.2 UV-Vis Spectroscopy Coupled Assay

The fluorescence assay as described above did not yield reproducible, consistent results in our hands therefore we developed a new, convenient and highly reproducible assay for enzyme-based detection of αKG that requires neither radioactivity nor derivatization. The coupled assay employs L-glutamic dehydrogenase (Sigma Aldrich), which catalyzes the reductive amination of αKG in the presence of 40 mM NH^{+4} and 80 μM NADH. The reaction can be monitored by the loss of absorbance at 340 nm corresponding to the oxidation of NADH. A standard curve was prepared to correlate the consumption of αKG (5 - 50 μM) to the consumption of NADH. Aliquots of αKG were added to a solution containing NH_4Cl (100 mM), NADH (80 μM), 0.3 mg L-glutamic dehydrogenase, $Fe(NH_4)_2(SO_4)_2$ (10 μM), sodium ascorbate (500 μM), and EDTA (100 μM) in 50 mM KPi (pH 7.0), with a total volume of 800 μL. The Fe(II), ascorbate, and EDTA were added as they would be carried over from the anthrax-P4H activity assays (see below). Reactions were incubated at 30 ˚C in a cuvette while monitoring the absorbance decrease at 340 nm. The reactions were typically complete within

30 seconds.

To assay for the consumption of αKG by anthrax-P4H, anthrax-P4H (20 μM) was added to 50 mM Tris buffer (pH 7.4) containing sodium ascorbate (100 μM), and αKG (750 μM). The reaction was initiated by the addition of 40 μM $Fe(NH_4)_2(SO_4)_2$ from a 20 mM stock solution in 0.2 N HCl. The reaction mixture was incubated for 10 minutes at 37 ˚C and the reaction was quenched by the addition of 1 mM EDTA. The amount of residual αKG was monitored by adding the entire reaction mixture to 750 μL of 50 mM Tris buffer (pH 7.4) containing NH_4Cl (100 mM), NADH (80 μM), and 0.3 mg L-glutamic dehydrogenase as described above.

3.2.5.3 O_2 Electrode Assay

The O_2 consumption by P4H was monitored by PC operated Oxygraph oxygen electrode control unit where the S1 Clark-type electrode disc was mounted in DW1/AD liquid-phase oxygen electrode chambers (Hansatech Instruments Ltd, UK). Temperature control of the sample and S1 Clark-type polarographic oxygen electrode disc are achieved by connecting the water jacket on the electrode chamber to a thermostated circulating water bath at 37 °C. A 500 μL of 50 mM Tris buffer (pH 7.4) containing variable concentrations of αKG, sodium ascorbate, glutathione, and $(Gly\text{-}Pro\text{-}Pro)_{10}$, were equilibrated with air. $Fe(NH_4)_4(SO_4)_2$ was then added and the background O_2-consumption was measured for one minute. An aliquot of

anthrax-P4H was added to a final concentration of 5 μM and the initial rate was determined between 30 seconds to 1 minute of reaction time.

3.2.6 Anaerobic UV-vis Spectroscopy Monitoring Anthrax-P4H Cofactor Binding

UV-vis spectra were recorded on an Agilent 8453 spectrophotometer at 30 °C. A 100 μL aliquot of 1 mM anthrax-P4H in 50 mM Tris buffer (pH 7.4) containing 2 mM sodium dithionite was placed in a septum-rubber sealed 50 mL cuvette and was degassed by gently purging with O_2-free Ar for 30 min. The blank was taken with this anthrax-P4H solution. An equivalent molar of Fe(II) was added by syringe from a degassed 100 mM stock solution in 20 mM HCl and the spectrum of the Fe(II)/anthrax-P4H binary complex was recorded immediately. An equivalent molar of αKG was then added from a degassed 50 mM stock solution in 50 mM Tris buffer (pH 7.4) and the spectrum of the αKG/Fe(II)/anthrax-P4H ternary complex was recorded. After 10 minutes, the cuvette was open to air and additional spectra were taken immediately and 10 minutes later.

The effect of binding of $(Gly\text{-}Pro\text{-}Pro)_{10}$ peptide to the UV-vis absorption spectrum of αKG/Fe(II)/anthrax-P4H ternary complex was also studied under anaerobic conditions. A 25 mM imidazole buffer (pH 7.0) containing 50 μL of 1 mM anthrax-P4H, 50 μL of 2 mM $(Gly\text{-}Pro\text{-}Pro)_{10}$ peptide and 2 mM sodium dithionite was prepared anaerobically as described above, yielding a 1:2 molar ratio of enzyme to substrate. Equal molar

equivalents of Fe(II) and αKG to the enzyme were added under anaerobic conditions as described previously to a final concentration of 0.5 mM. The spectra were recorded repeatedly over a 15 min period. Binding studies were also performed with the inhibitors N-oxaylglycine (Alexis Biochemicals), DMHP (Calbiochem), and ciclopirox olamine (Sigma Aldrich).

3.2.7 UV-vis Spectroscopic Titration of Anthrax-P4H with αKG

An 1 μL of degassed stock solution of 200 mM sodium dithionite in 25 mM imidazole buffer (pH 7.0) and 100 mM $Fe(NH_4)_4(SO_4)_2$ in 2 N HCl was added to the degassed anthrax-P4H solution prepared as above and the spectrophotometer was blanked with this solution. 1 μL of degassed 20 mM αKG stock solution was then added to the enzyme solution by syringe and the spectrum was recorded immediately. Additional aliquots of the buffered αKG stock were added to the sample and spectra were recorded until no further spectral change was observed. The titration data were fitted to Equation 2 using KaleidaGraph, where ΔA_{510}^{max} is the maximum absorbance change at 510 nm, $[E_T]$ is the total enzyme concentration equal to 1.0 mM, $[L_T]$ is the total ligand (αKG) concentration and K_d is the apparent ligand affinity. The error bars represent the standard deviations from three independent experiments. Equation 2 is fit to a single line to determine the binding affinity of αKG to anthrax-P4H.

$$\Delta A_{510} = \Delta A_{510}^{max}\left\{ ([E_T]+[L_T]+K_d) - \sqrt{([E_T]+[L_T]+K_d)^2 - 4[E_T][L_T]} \right\} / 2[E_T] \quad (2)$$

3.2.8 Determination of Fe(II) Binding Stoichiometry

The binding stoichiometry of Fe(II) to anthrax-P4H was assessed by the Fe(II) concentration dependency of the initial rate of the uncoupling reaction. A solution of anthrax-P4H and molar equivalents of Fe(II) were pre-incubated at 30 °C for 10 minutes. In order to minimize the oxidative chemistry of Fe(II), apo-anthrax-P4H was reconstituted with Fe(II) in small quantities, and used within 30 min. The mixture was transferred to the O_2 electrode chamber containing αKG and sodium ascorbate, and initial rates were recorded. The assay mixture contained 10 μM anthrax-P4H, 750 μM αKG, 50 μM sodium ascorbate, and equivalents of $Fe(NH_4)_4(SO_4)_2$ ranging from 0.2 - 3 mM. Initial rates were determined by subtracting background rates due to oxygen consumption in the presence of Fe(II) in the absence of the enzyme. The initial rates were plotted against the concentrations of Fe(II).

3.2.9 pH-Dependency of the Initial Rate of the Uncoupling Reaction Catalyzed by Anthrax-P4H

The assay solution contained 5 μM anthrax-P4H, 750 μM αKG, 50 μM sodium ascorbate, and 25 μM $Fe(NH_4)_4(SO_4)_2$ in 50 mM buffers at varying pHs. NaOAc was used for pH 5.0, MES was used for pH 6.0, HEPES was

used for pH 7.0, Tris was used for pHs 8.0 and 9.0 and CAPS was used for pH 10.0. The reactions were run in triplicate at each pH. Analysis of the pH dependence of the initial rate was done by non-linear least squares fitting of the data to Equation 3 using KaleidaGraph.

$$\text{rate} = k_{\max} \Big/ \left(1 + 10^{\text{p}K_{\text{a}}^{1} - \text{pH}} + 10^{\text{pH} - \text{p}K_{\text{a}}^{2}}\right) \tag{3}$$

3.2.10 Initial Screening of Potential Substrates

Initial activity assays were run using poly(L-proline) (1 mg/mL) as the substrate. Poly(L-proline) acts as an inhibitor in human-P4H, however, plant and viral P4Hs can use it as a substrate. We obtained inconsistent results using poly(L-proline) in our assays. We therefore looked into the genome of *Bacillus anthracis* for potential substrates.

3.2.10.1 Expression and Purification BclA

BclA gene BA1222 from *Bacillus anthracis* cloned into pQE-30 (Qiagen) expression vector containing a N-terminal His-tag was a kind gift from Dr. George Stewart at the University of Missouri-Columbia.

Colonies were transformed into M15 (Qiagen) *E. coli* competent cells and grown in LB-media containing 100 μg/mL ampicillin and 25 μg/mL kanamycin. Cells were grown at 37 °C to an OD_{600} of 0.6 and induced with IPTG to a final concentration of 1 mM. Cells were maintained at 37 °C for 1.5 hours, harvested, and frozen at -80 °C. Though native BclA is known to be

heavily glycosylated, no glycosylation occurred on the protein due to synthesis in *E. coli*. BclA was purified with His-Tag chromatography as previously described. Additional impurities were removed by incubation at 70 °C for 8 min, followed by centrifugation at 10,000 *g* for 10 min at 4°C as previously reported.[99] Solubility of BclA was improved with the addition of 0.1 % Triton X-100 to all purification buffers.

3.2.10.2 Synthetic $(GPT)_5$ and Collagen-Like Sequences from *B. anthracis* Genome

The $(Gly\text{-}Pro\text{-}Thr)_5$ peptide was synthesized at the Biopolymer/Genomics Core Facility at the University of Maryland. Other peptides were synthesized at the Biochemical Research Service Laboratory at the University of Kansas.

peptide 1: H_2N-GPEGPPGPTGKK-COOH

peptide 2: H_2N-GIGGPPGPPGPTGKK-COOH

peptide 3: H_2N-GPQGVQGPAGATGKK-COOH

All were all tested as potential native substrates for recombinant anthrax-P4H using both the UV-Vis coupled assay and O_2 electrode assay.

3.2.10.3 $(GPP)_{10}$ Peptide from *E. coli* Expression and Purification

Expression of polypeptide $(GPP)_{10}$ as a recombinant Thioredoxin fusion protein in *E. coli* was developed in by Dr. Hirakawa in the Limburg Lab (unpublished results). Briefly, the plasmid was transformed into TOP10 (Invitrogen) *E. coli* cells. The fusion protein was grown in TB media

containing 100 μg/mL ampicillin at 37 °C until OD_{600} reaches 0.4. Protein production was induced with L-arabinose to a final concentration of 0.1 %, and incubated overnight at 30 °C.

The fusion protein was purified using a HisTrap HP (GE Biosciences) Ni^{2+} Sepharose column, followed by a Superdex 16/60 sizing column. The thioredoxin fusion protein was cleaved off of the polypeptide with the enzyme Enterokinase (Invitrogen) and incubated overnight at 37 °C. The cleaved thioredoxin and $(GPP)_{10}$ were separated with a Superdex 16/60 sizing column and fractions containing $(GPP)_{10}$ were dialyzed into water. The $(GPP)_{10}$ mass was confirmed with MALDI-TOF MS analysis.

3.3 Results

3.3.1 Sequence Analysis

A BLAST search using the amino acid sequence of the α subunit of type(I) human-P4H against bacterial genomes retrieved a similar sequence from *B. anthracis* that was annotated as a *p4h*. The α subunit of human type(I)-P4H (human-P4H-1) contains a separate collagen binding domain and an active site domain at the C-terminus.[127] Anthrax-P4H lacks the separate collagen-binding domain similar to plant, algae, and viral P4Hs.

Figure 3.1 shows the amino acid sequence alignment for the putative anthrax-P4H with the C-terminal end of the α subunit of human-P4H-1, *A. thaliana*-P4Hs (At-P4H-1, At-P4H-2), PBCV1-P4H, and *C. reinhardtii*-P4H (Cr-P4H-1). Anthrax-P4H shares 30 % sequence identity with PBCV1-P4H and the C-terminal domain of the α subunit of human-P4H-1, and ~ 40 % identity with the catalytic domains of At-P4H-1, At-P4H-2, and Cr-P4H-1.

The residues comprising the facial triad that bind iron are conserved in all of the sequences (His127, Asp129, and His193 in anthrax-P4H). The cationic residues that are essential for binding of αKG in human-P4H-1 (Lys493 and His501, human numbering)[40] are also conserved in anthrax-P4H (Lys203 and Arg211) as well as the other P4Hs. These residues are the minimal set that can be used to annotate a protein sequence as an αKG/Fe(II)-oxygenase. The superfamily generally shares little sequence

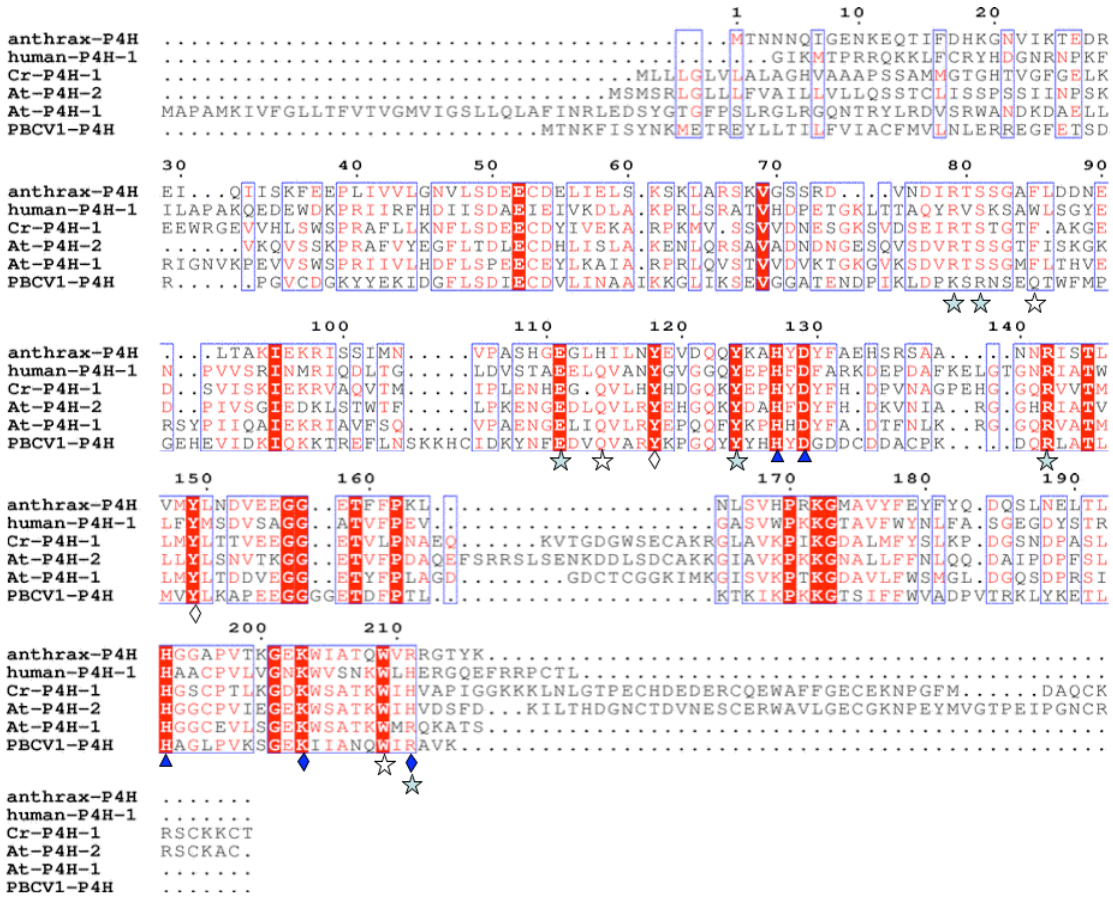


Figure 3.1. Amino acid sequence alignment of P4Hs. Human-P4H-1: C-terminal catalytic domain of the a subunit of type I human-P4H, residues 243 - 512; Cr-P4H-1: P4H from *C. reinhardtii*, At-P4H-1 and At-P4H-2: P4Hs from *A. thaliana*; PBCV: P4H from *Paramecium bursaria Chlorella* virus-1. ▲Iron binding, ♦αKG binding identified in human-P4H-1, ◊ αKG binding identified in Cr-P4H-1, ☆ collagen-like peptide binding identified in Cr-P4H-1, ☆ collagen-like peptide binding identified in Cr-P4H-1. The alignment was generated using the Mutlialin program (http://bioinfo.genotoul.fr/multalin/multalin.html).

homology beyond the Fe(II) and αKG-binding sites. A recent study of the active site structure of Cr-P4H-1 (PDB: 2JIG) suggests that two Tyr residues (Tyr134 and Tyr168, Cr-P4H-1 numbering) are also involved in αKG binding, where the mutation of those residues to Phe increased the K_m for αKG 6- and

3.5-fold respectively, but had no affect on the K_m for poly(L-proline) substrate.[90] Those Tyr residues are conserved in anthrax-P4H (Tyr118 and Tyr149) as well as the rest of the P4Hs shown in Figure 3.1.

The structure of Cr-P4H-1 revealed an extended shallow groove lined by two flexible loops. Point mutation of residues in the groove (Arg93, Ser95, Glu127, Tyr140, Arg161, and His245 in Cr-P4H-1 numbering) to Ala resulted in complete loss of activity. Therefore it is proposed to serve in collagen-like peptide recognition/binding.[90] The corresponding residues are strictly conserved in anthrax-P4H (Arg79, Ser81, Glu111, Tyr124, Arg142, and Arg211, anthrax-P4H numbering), human-P4H-1, At-P4H-1, and At-P4H-2. In the PBCV1-P4H, the first two residues are Lys and Arg but the rest are conserved. In addition, mutation of Trp99, Gln130, and Trp243 in the groove resulted in 1.8- to 2.5-fold increase in K_m for poly(L-proline) and 2- to 12-fold decreases in k_{cat} for the algal-P4H, suggesting that those residues also contribute to peptide substrate binding. The corresponding residues in anthrax-P4H are Phe85, His114, and Trp209. Taken together, these results support that the anthrax enzyme is a procollagen-P4H.

3.3.2 Cloning, Expression, and Purification of Putative *p4h* Gene from *B. anthracis*

The *p4h* gene (Ames) was provided to us in a pDONR221 Gateway entry vector (Invitrogen) by The Institute for Genomic Research-Pathogen Functional Genomics Resource Center (TIGR-PFGRC) (Maryland). We subcloned the gene into pDEST17 (Invitrogen) containing an N-terminal His-tag by performing a LR recombination reaction with the enzyme LR Clonase II (Invitrogen). Unfortunately the purified N-terminally His-tagged protein was found to be not stable.

We subcloned the gene into the pET15b expression vector in order to prepare an untagged recombinant protein (Figure 3.2). We detected the recombinant protein in the cell extract of BL21(DE3) *E. coli* on a SDS-PAGE gel with an apparent *Mr* of ~ 28 kDa compared to a predicted *Mr* of 24.5 kDa. The protein was purified to > 99% purity based off SDS-PAGE gel by anion exchange and gel filtration chromatography (Figure 3.3). The yield of purified recombinant anthrax-P4H was 10 mg/L of culture.

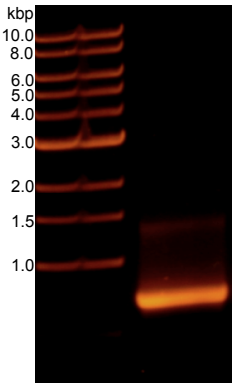


Figure 3.2. PCR amplification of *p4h* gene resulting in 651 bp gene.

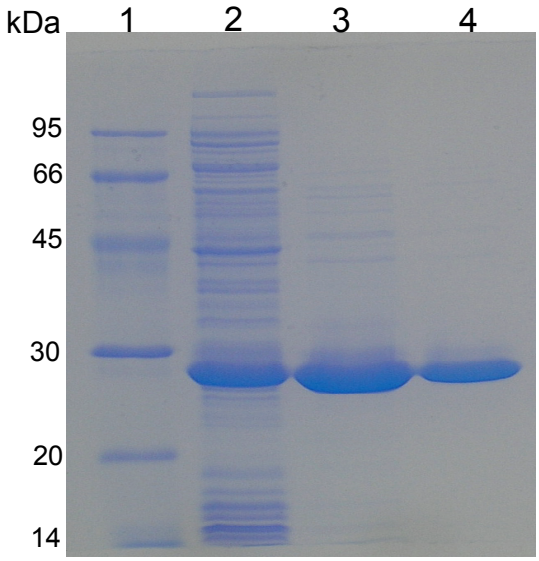


Figure 3.3. A SDS-PAGE gel showing the purification profile of recombinant anthrax-P4H. Lane 1) molecular weight marker; Lane 2) crude lysate; Lane 3) DEAE column fraction; Lane 4) size-exclusion column fraction. The recombinant anthrax-P4H was detected as a single band with an apparent molecular weight of 27 kDa.

3.3.3 Quaternary Structure Determination of Native Anthrax-P4H

The *Mr* of recombinant anthrax-P4H was estimated from the retention time on the gel filtration column to be ~ 57 kDa (Figure 3.4), suggesting the protein is a homodimer of 28 kDa subunits.

MALDI-TOF mass spectrometry was run in the presence and absence of β-mercaptoethanol at room temperature to determine the quaternary structure of the recombinant anthrax-P4H (Figure 3.5). The MALDI-TOF analysis of the recombinant protein gave a dominant signal at m/z 24329.75 that is very close to the predicted mass of the monomer of 24605.53. A signal corresponding to the dimer is present at 48693.92. These results clearly

A

Protein Standard	MW (kDa)	Retention Volume (mL)
Ferritin	440	53.4
Catalase	232	62.9
Aldolase	158	68.9
Albumin	67	74.1
Ovalbumin	43	80.4
Chymotrypsinge A	25	89.4
Ribonuclease A	13.7	95.2

B

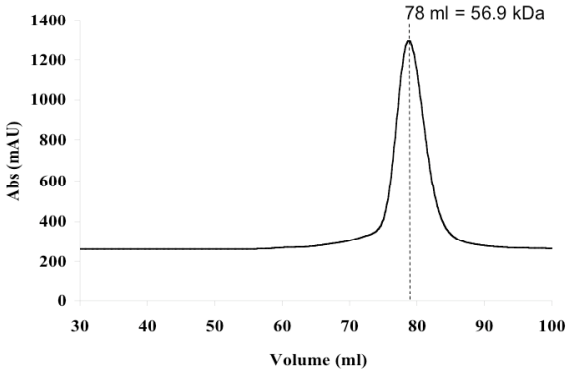


Figure 3.4. Gel filtration analysis of the recombinant anthrax-P4H by Superdex 16/60 S-200. A) Protein standards of known molecular weight and retention volume. B) Gel filtration trace indicating anthrax-P4H retention volume of 78 ml, which corresponds to a molecular weight of 57 kDa.

indicate that the recombinant enzyme is in dimer-monomer equilibrium where the monomer is predominant under the experimental conditions tested. The presence or absence of β-mercaptoethanol does not affect the MS results, indicating that the dimer interactions are not covalent via a disulfide linkage.

We examined the solution properties of the recombinant anthrax-P4H with dialysis run at room temperature and found that anthrax-P4H exists predominantly as the dimer form in the monomer-dimer equilibrium in the assay condition used in this study (Figure 3.6). The equilibrium constant was

determined to be K_d^{app} = 2.76 ± 1.6 μM from the three concentrations evaluated.

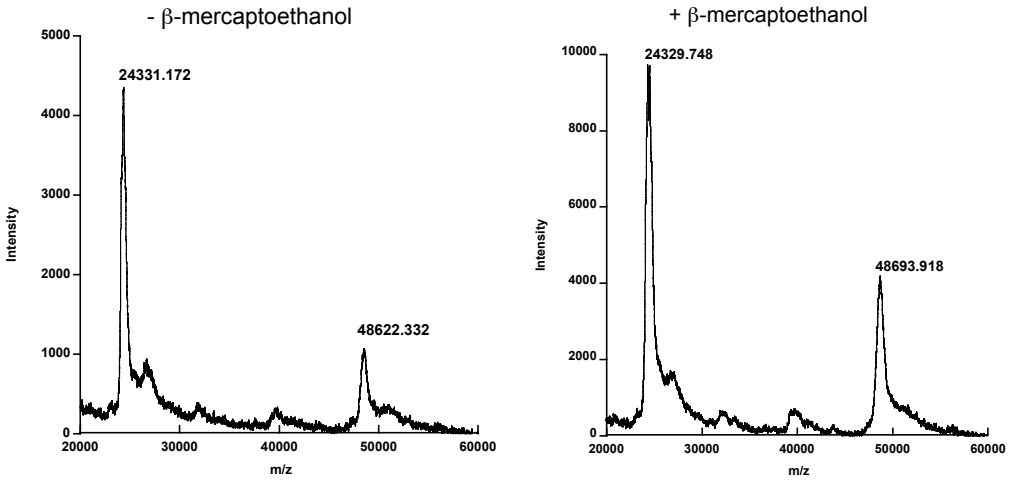


Figure 3.5. MALDI-TOF mass spectrometry gives evidence that recombinant anthrax-P4H exists in both a monomeric and dimeric form. Dialysis of anthrax-P4H into 50 mM Tris, pH 7.4, was performed overnight to eliminate recombinant anthrax-P4H of β-mercaptoethanol. MALDI-TOF MS shows that recombinant anthrax-P4H exists in the equilibrium of dimer/monomer in the presence and absence of 5 mM β-mercaptoethanol. Predicted monomer: m/z = 24,330; dimer: m/z = 48,600

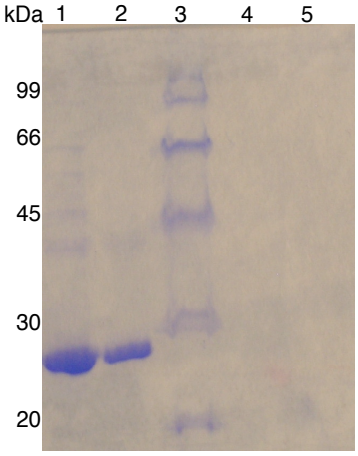


Figure 3.6. SDS-PAGE of dialysis experiment to determine monomer-dimer equilibrium constant. Lane 1) 1 mg/ml anthrax-P4H dimer, Lane 2) 0.5 mg/ml anthrax-P4H dimer, Lane 3) Molecular weight marker, Lane 4) 1 mg/ml anthrax-P4H monomer, Lane 5) 0.5 mg/ml anthrax-P4H monomer.

3.3.4 Preparation of C53S and C53A Mutations of Anthrax-P4H

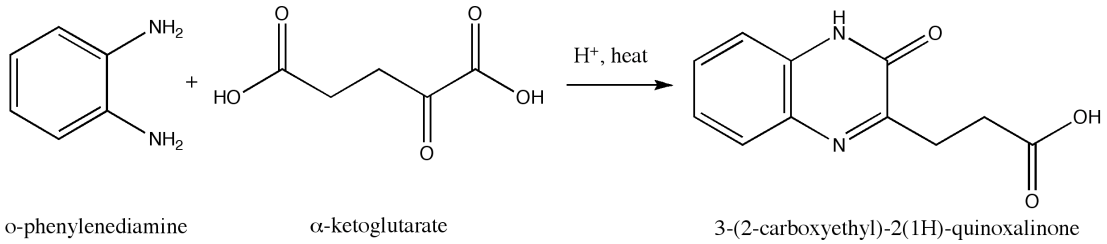
Since the predicted amino acid sequence of anthrax-P4H contains a single cysteine residue, Cys53, we made the C53S and C53A mutants to probe the role of Cys53 in dimer formation. The C53S mutant protein did not fold correctly and was expressed as an insoluble inclusion body. The C53A mutant protein was isolated as a soluble protein and exhibited activity comparable with the wild-type, when tested with the uncoupling and $(GPP)_{10}$ reactions. We also ran the protein on a gel filtration column to determine if the size of anthrax-P4H reduced with the mutation. Neither the gel filtration nor activity changed for C53A compared with wild type. These results suggest that the subunit interaction in the wild-type dimer is not covalent, similar to the results seen with the MALDI-TOF MS data.

3.3.5 Anthrax-P4H Activity Monitored by Fluorometric Detection

Assays for αKG/Fe(II)-oxygenase activities have been reported based on the release of $^{14}CO_2$ from radiolabeled-αKG[18] and the dabsyl-labeled substrate used in the HPLC assay for mammalian-P4Hs.[128] The derivatization of unreacted αKG with OPD is also used to monitor αKG consumption, where the product can be quantitated by fluorescence spectroscopy.[126] The reaction scheme is shown in Scheme 3.1.

The greater the consumption of aKG in the P4H reaction, the lower the amount of αKG available to react with OPD and the lower fluorescence

intensity (FI) will be. Therefore, we expected to see low FI values if anthrax-P4H was active in the presence potential substrates. A standard curve was run to determine the sensitivity and linear response of the fluorescence assay, which was determined to be 0.1 - 0.75 mM αKG (Figure 3.7). Initial activity assays showed increased αKG consumption in the presence of both poly(L-proline) and anthrax-P4H (Figure 3.8).



o-phenylenediamine α-ketoglutarate 3-(2-carboxyethyl)-2(1H)-quinoxalinone

Scheme 3.1. Reaction of OPD with the α-ketoacid motif of 2OG to form 3-(2-carboxyethyl)-2(1*H*)-quinoxalinone.

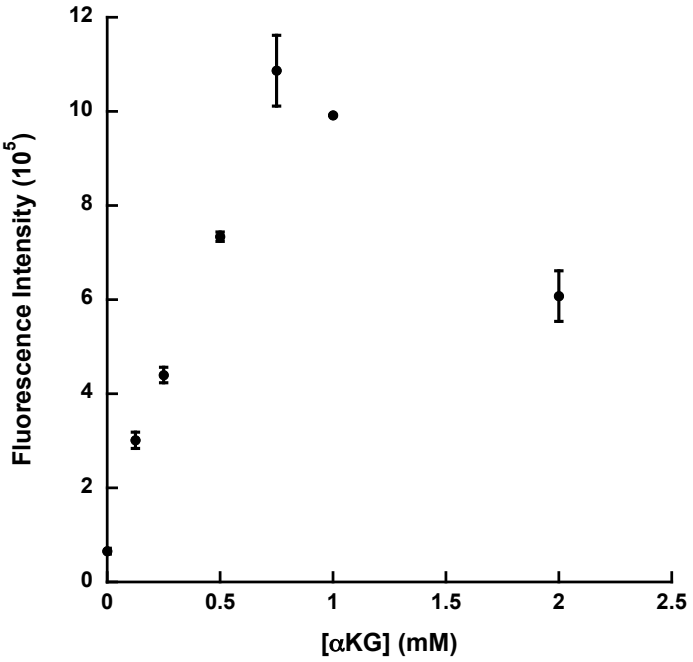


Figure 3.7. αKG standard curve employing fluorescent assay. Samples were run in the presence of 10 mM Fe(II), 1 mM DTT, 50 mM Tris, pH 7.4. Reaction is linear to 0.75 mM αKG.

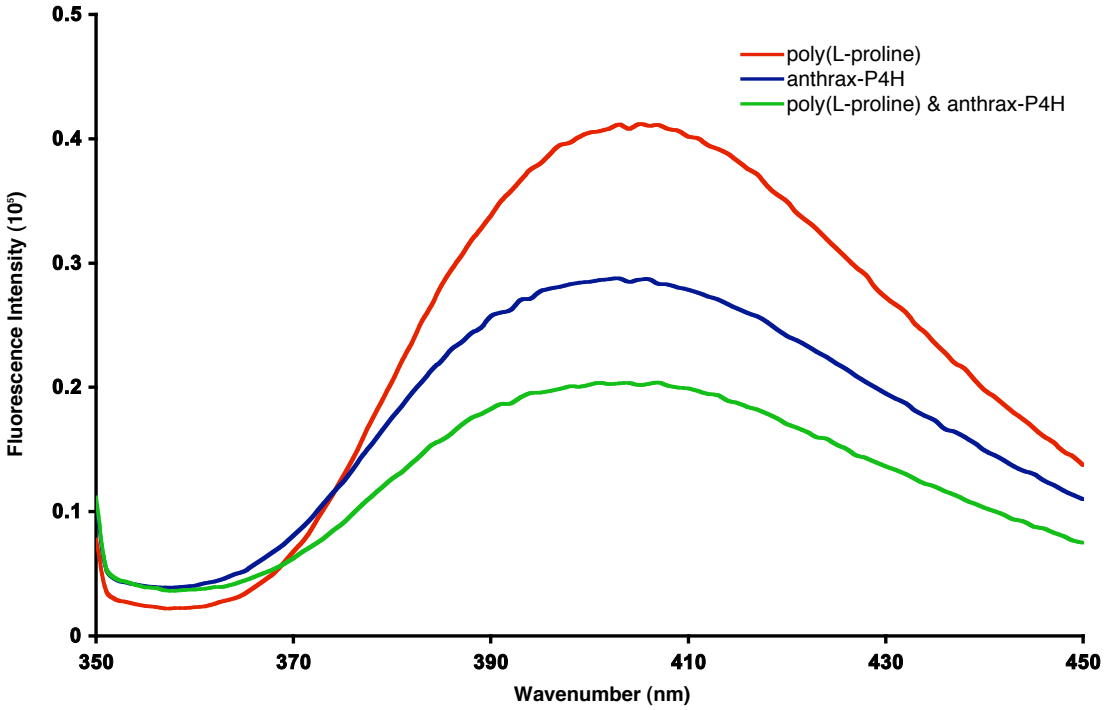


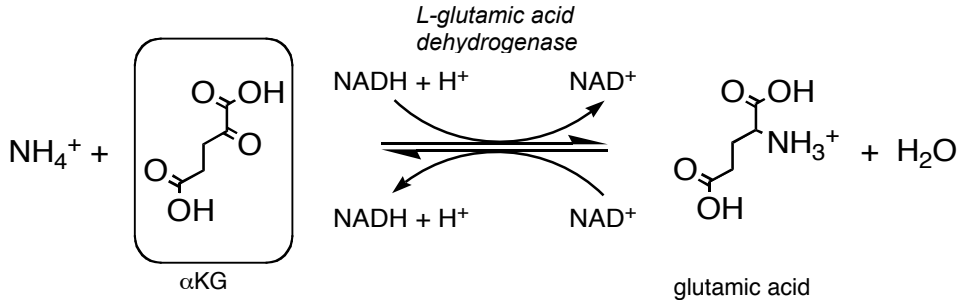
Figure 3.8. Initial activity assay of 10 μM anthrax-P4H with 1 mg/mL poly-(L-proline) in the presence of 750 μM αKG, 750 μM DTT, 10 μM Fe(II), 50 mM Tris, pH 7.5. The reaction mixture was incubated at 37 °C for 5 minutes, and quenched with 0.25 M HCl, and reacted with 5 mg/mL OPD at 95 °C for 10 minutes. The excitation wavelength was 340 nm and emission wavelength 420 nm. The red line corresponds to 1 mg/mL poly(L-proline), blue line is 10 μM anthrax-P4H, and the green line is the total reaction with both anthrax-P4H and poly(L-proline).

The uncoupled reaction of anthrax-P4H with ascorbate could not be used with this assay because the 1,2-dicarbonyl moiety of dehydroascorbate interferes with 1,2-diamine of OPD and therefore alters the fluorescence detection. The uncoupling reaction is used to monitor anthrax-P4H activity because the native substrate of anthrax-P4H is currently not known. Therefore, further methods for assaying the anthrax-P4H activity were

investigated and a coupled assay using UV-visible spectroscopy was developed.

3.3.6 L-Glutamic Dehydrogenase for an Endpoint Assay for αKG/Fe(II)-Oxygenases

We have developed a new endpoint assay using the NADH-dependent reductive amination of αKG by L-glutamic dehydrogenase (Scheme 3.2), where the reaction is driven to completion by the addition of a large excess of NADH (80 μM) and NH_4Cl (100 mM).



Scheme 3.2. Reductive amination of αKG by L-glutamic dehydrogenase used as an endpoint assay for the activity of anthrax-P4H.

Figure 3.9 shows the standard curve where a linear response for NADH consumption can be achieved even when the residual αKG in the assay solution is down to 5 μM. This is a convenient, rapid, and specific method to quantify the remaining αKG in solution at the end of the reaction.

Since procollagen-P4H and other αKG/Fe(II)-oxygenases are known to catalyze the oxidation of ascorbic acid coupled to the oxidative decarboxylation of αKG, i.e. uncoupling (Scheme 1.1 B), we used the

uncoupling reaction as an alternative assay to confirm that recombinant anthrax-P4H has activity consistent with an αKG/Fe(II)-oxygenase. This reaction serves to protect human-P4H from inactivation.[22] After 30 minutes incubation of 20 μM recombinant anthrax-P4H in a buffer containing 40 μM Fe(II), 100 μM ascorbate and 750 μM αKG, 150 μM αKG is consumed. The amount of αKG consumed is dependent on available amounts of ascorbate and anthrax-P4H in the assay mixture. In control experiments in the absence of Fe(II), no consumption of αKG was observed. These results demonstrate that anthrax-P4H has activity consistent with an αKG/Fe(II)-oxygenase.

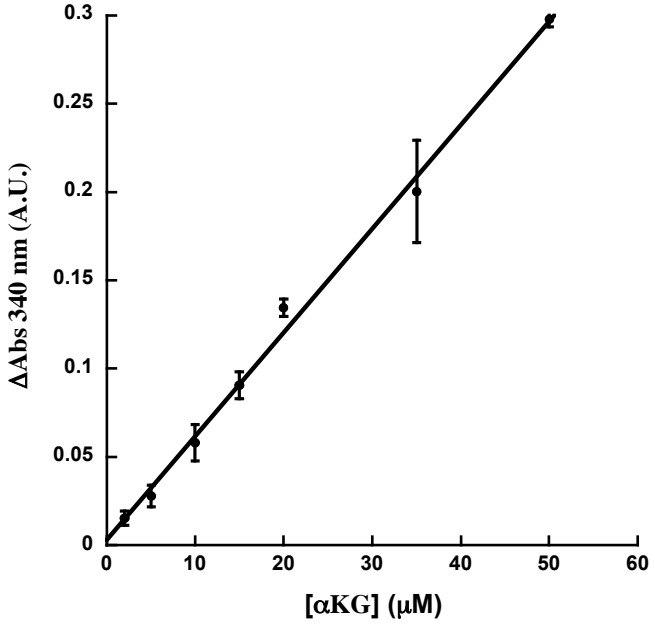


Figure 3.9. A standard curve for the quantification of αKG by L-glutamic dehydrogenase. Aliquots of αKG (2 - 50 μM) were added to 50 mM KPi buffer (pH 7.0) containing NH_4Cl (100 mM), NADH (80 *μ*M), 0.3 mg L-glutamic dehydrogenase, $Fe(NH_4)_2(SO_4)_2$ (10 μM), (+)-sodium *L*-ascorbate (500 μM) and EDTA (100 μM) to a total volume of 800 μL.

3.3.7 O_2 Electrode Assay for Anthrax-P4H Activity

Initial O_2 electrode experiments were performed to optimize the uncoupled reaction. Variation of anthrax-P4H concentration from 5 μM - 20 μM showed that high levels of enzyme (10 μM and higher) produce fast rates, which are too high to see major differences in the presence of potential substrates. 5 μM anthrax-P4H was found to produce a good background rate for monitoring the uncoupled reaction.

The dependence of anthrax-P4H on Fe(II) was determined for the uncoupling reaction. Fe(II)/αKG-oxygenases require one Fe(II) molecule per monomer.[31,129,130] The Fe(II)/anthrax-P4H binary complex was added to an assay mixture containing 750 μM αKG and 50 μM ascorbate to a final concentration of 10 μM anthrax-P4H. We observed a linear dependence of the initial rate on the concentration of Fe(II) up to a stoichiometric amount (Figure 3.10). No inhibition occurred up to three molar equivalents of Fe(II).

The O_2 concentration dependency of anthrax-P4H activity was examined. Figure 3.11 shows a trace following the consumption of O_2 in a solution containing 750 μM of αKG and 50 μM of ascorbate at pH 7.4. When 20 μM Fe(II) is added to the solution, there was small background of O_2 consumption, but when 5 μM of anthrax-P4H was present, a clear inflection point was detected and O_2 was consumed at an initial rate of 5 mol of αKG/mol P4H/min. The total amount of O_2 consumed is 46 μM, which

corresponds to the total available ascorbate in the reaction mixture (background control reaction in the absence of ascorbate consumed 8 μM). The trace plateaued after the ascorbate in the reaction mixture was all consumed.

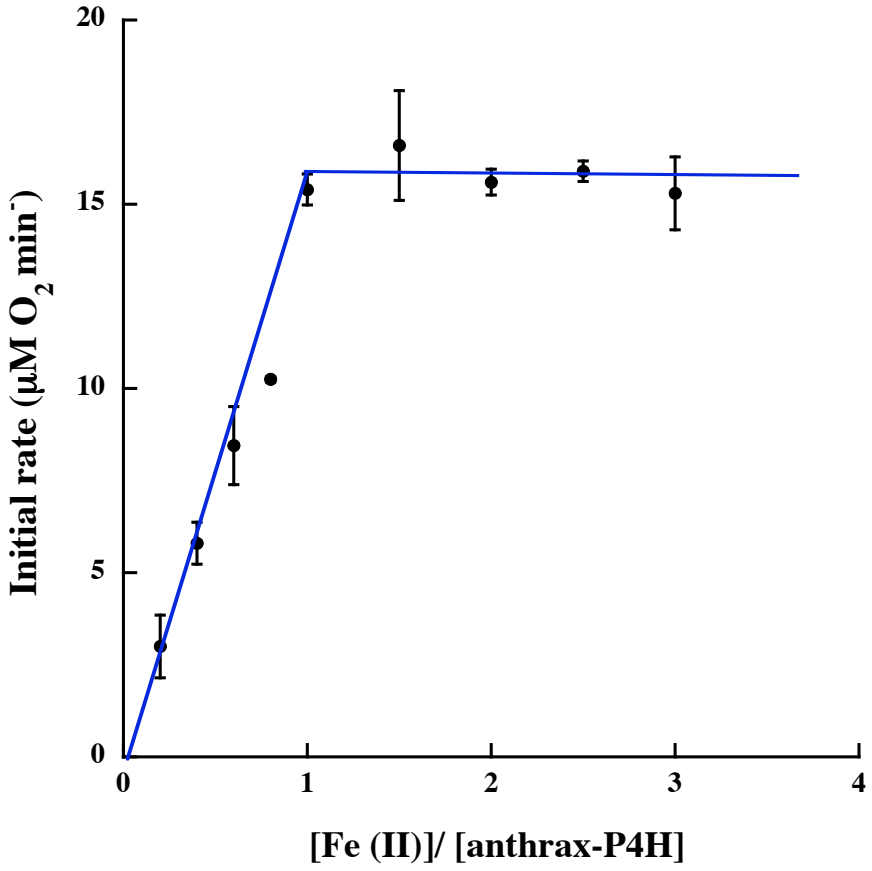


Figure 3.10. Effect of Fe(II) concentration on initial rates of uncoupling reaction catalyzed by anthrax-P4H. Initial rates were measured with an O_2 sensitive Clarke-type electrode. The assay solution contained 10 μM anthrax-P4H, 750 μM αKG and 50 μM (+)-sodium L-ascorbate.

We have measured K_m values for Fe(II), ascorbate, and αKG. These results are summarized in Table 3.1 along with those reported for other P4Hs discussed in this chapter. The k_{cat} for this reaction was determined to be 1.9 ± 0.3 min^{-1}. When either αKG or Fe(II) is eliminated from the solution, only background O_2 consumption was observed.

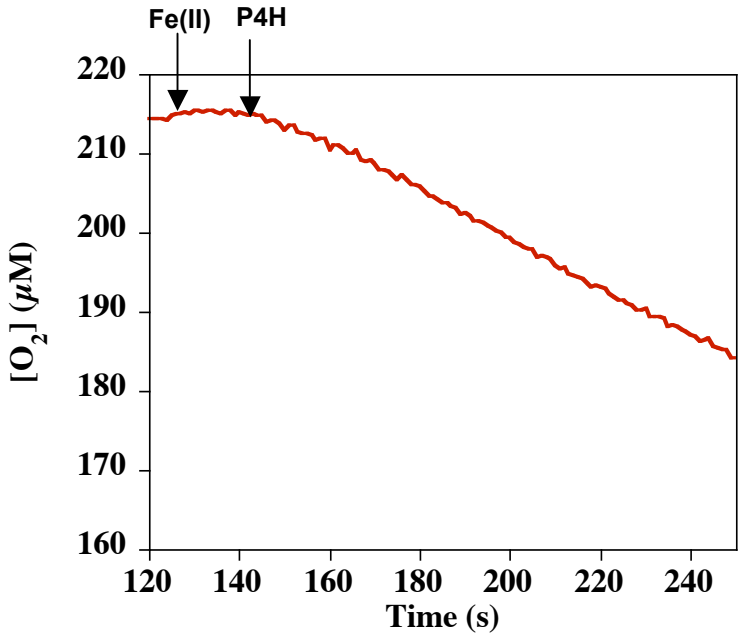


Figure 3.11. An O_2-sensitive Clarke-type electrode trace showing the consumption of O_2 by anthrax-P4H. The reaction was initiated by the sequential addition of 20 μM $Fe(NH_4)_2(SO_4)_2$ and 5 μM anthrax-P4H into the reaction chamber containing 50 μM (+)-sodium L-ascorbate and 750 μM αKG in 50 mM Tris buffer (pH 7.4).

	K_m (μM)					
Cosubstrate	Anthrax-P4H[a]	Human-P4H-1 [b]	PBCV-P4H[c]	Cr-P4H-1 [d]	At-P4H-1[e]	At-P4H-2 [f]
αKG	27.4 ± 15.4	22	20	250	130	170
Fe(II)	8.8 ± 0.9	4	0.4	30	16	5
Ascorbate	15.6 ± 2.5	300	300	20	300	300
$(GlyProPro)_{10}$	9.1 ± 1.3	18	2900	>1500	60	2800

[a]This work; [b] Ref 37;[c] Ref 73; [d] Ref 72; [e] Ref 69; [f] Ref 70

Table 3.1. K_m values of anthrax-P4H, human-P4H-1, PBCV1-P4H, *C. reinhardtii* P4H-1 (Cr-P4H-1), *A. thaliana* P4H-1 and -2 (At-P4H-1 and At-P4H-2) for cosubstrates and $(GlyProPro)_{10}$

3.3.8 pH-Dependency of Uncoupled Reaction Catalyzed by Anthrax-P4H

The effect of pH on initial rates of the uncoupling reaction was measured between pH 5.0 and 9.0. Anthrax-P4H exhibited a bell-shaped pH dependence of the initial rate for the uncoupling reaction with the optimum around 7.0 (Figure 3.12). Two pK_a values of 6.92 ± 0.18 and 7.17 ± 0.18 were determined by non-linear least square fit analysis using Equation 2 (Experimental Methods).

Fe(II) was also replaced with metals Mn(II) and Zn(II) and tested for uncoupled activity. In the presence of either Mn(II) or Zn(II) at the same concentration as Fe(II), activity was abolished.

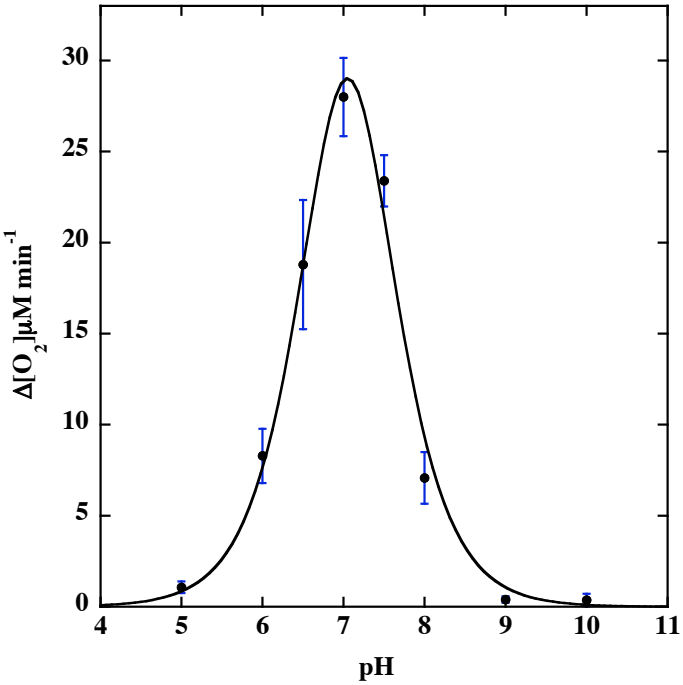


Figure 3.12. pH dependency of initial rates of uncoupling reaction catalyzed by anthrax-P4H. The assay solution contained 5 μM anthrax-P4H, 750 μM αKG, 25 μM $Fe(NH_4)_4(SO_4)_2$ and 50 μM (+)-sodium L-ascorbate.

3.3.9 Anaerobic UV-Vis Spectroscopy Monitoring Anthrax-P4H Binding

The amino acid sequence and catalytic activity of anthrax-P4H suggested that it is an αKG/Fe(II)-oxygenase. Once the Fe(II) is bound to the facial triad of these enzymes, the next step in the catalytic cycle is the binding of αKG to Fe(II) where the formation of the ternary complex can be followed by the appearance of a charge transfer band at ~ 500 nm in the visible region. The extinction coefficient of this transition is generally weak (~ 200 $M^{-1}cm^{-1}$), but since we were able to obtain multiple milligram quantities of recombinant anthrax-P4H, we were able to use visible spectroscopy to probe this initial step in the characterization of anthrax-P4H.

Figure 3.13 shows the visible spectrum of Fe(II)-reconstituted anthrax P4H before and after the addition of one equivalent of αKG under anaerobic conditions. Upon addition of αKG, a band appeared in the visible region, as shown by the red line in Figure 3.13, with a λ_{max} at ~ 480 nm and 510 nm. This band is not seen in the absence of the protein.

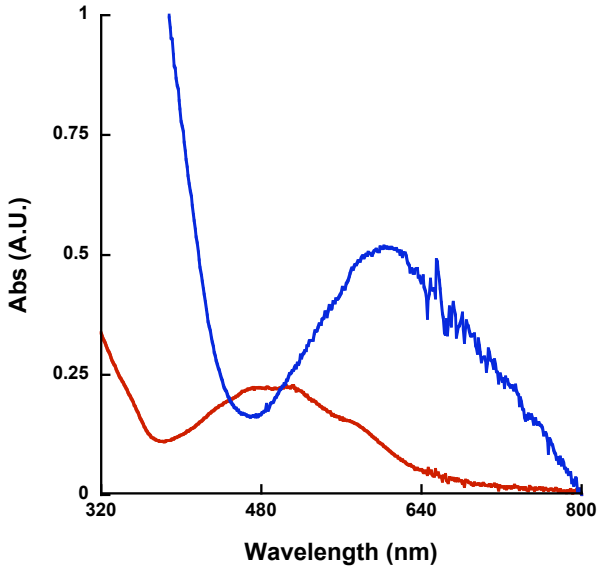


Figure 3.13. A UV-vis spectrum of the Fe(II)/αKG/anthrax-P4H ternary complex under anaerobic conditions (red line). The final concentrations are 1mM $Fe(NH_4)_2(SO_4)_2$, 1mM αKG and 1 mM anthrax-P4H in 50 mM Tris buffer (pH 7.4). Upon exposure to air, a species with the broad absorption with λ_{max} at ~ 600 nm builds up within ten minutes (blue line).

The spectrum is similar to αKG to Fe(II) charge transfer bands seen for TauD and alkylation repair homolog 1 (AlkB).[32,131] The λ_{max} of anthrax-P4H is somewhat blue-shifted compared to those enzymes. The extinction coefficient at 510 nm is 250 $M^{-1}cm^{-1}$, which is consistent with that observed for TauD.[32] These results support the assignment of anthrax-P4H being an αKG/Fe(II)-

oxygenase.

When air was introduced into the solution of the αKG/Fe(II)/anthrax-P4H ternary complex, the solution turned blue (the blue line in Figure 3.13). The resulting spectrum has a broad peak centered at 615 nm with an extinction coefficient of 500 $M^{-1}cm^{-1}$. This spectral change is analogous to that seen in the αKG/Fe(II)-oxygenase TfdA.[132] In TfdA, the blue color is caused by hydroxylation of a Trp residue (W112)[132] where the resulting hydroxyl group coordinates to Fe(III). The hydroxylation occurs concomitant with decarboxylation of αKG and the resulting enzyme is inactivated. It is possible that similar chemistry is occurring with anthrax-P4H and there are several strictly conserved Trp, Tyr, and Phe residues near to the putative iron-binding residues determined by mutagenesis in human-P4H. It is possible that similar chemistry is responsible for the inactivation of human-P4H.

Additional characterization of anthrax-P4H was achieved by titration of O_2 free αKG stock solution into an anaerobically prepared sample of the protein:Fe(II) binary complex. The progress of the formation of the αKG/Fe(II)/anthrax-P4H ternary complex was monitored by UV-vis spectroscopy (Figure 3.14 A). One equivalent of αKG is required for full activity as determined through non-linear least squares fitting using Equation 3 (Experimental Methods). The plot of the background-corrected change in absorbance at 510 nm (λ_{max} of the ternary complex) as a function of αKG

concentration (Figure 3.14 B) was used to determine a dissociation constant (K_d) for the ternary complex to be 54 ± 25 μM. Elimination of absorbance at 510 nm was observed with the substitution of Zn(II) to αKG:anthrax-P4H complex.

The binding of $(Gly\text{-}Pro\text{-}Pro)_{10}$ peptide to the anthrax-P4H/Fe(II)/αKG ternary complex showed no detectable change in the spectrum under the conditions tested. The binding of inhibitors N-oxaloglycine, ciclopirox, and DMHP also showed no spectra shifts upon binding.

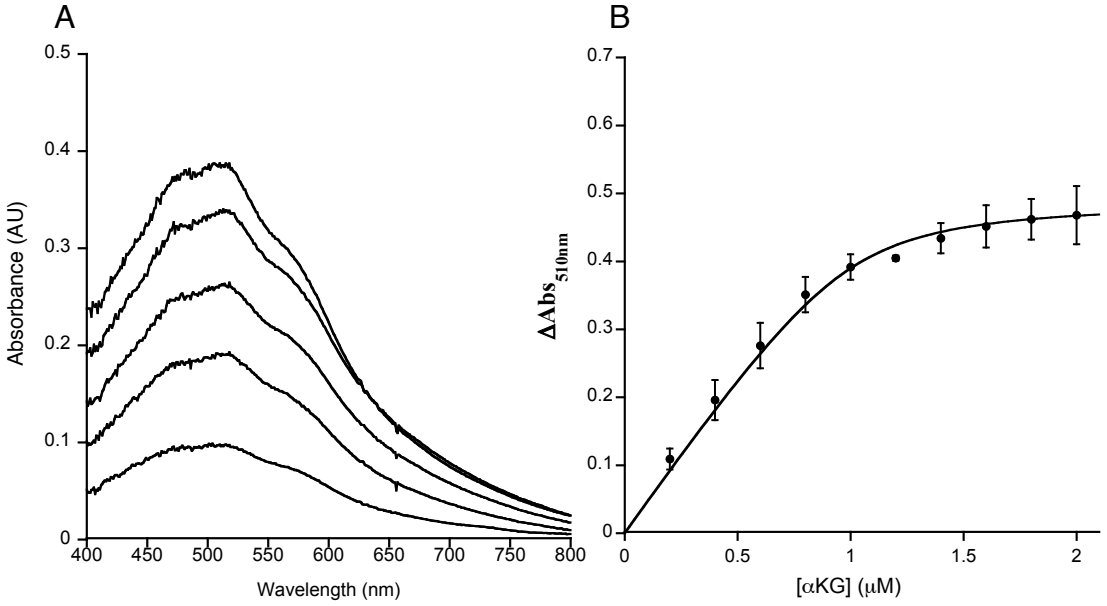


Figure 3.14. UV-vis spectroscopic titration of the Fe(II)/anthrax-P4H binary complex with αKG under anaerobic conditions. The binary complex was prepared by incubating 1 mM each of Fe(II) and anthrax-P4H in the presence of 2 mM sodium dithionite. A) The spectra correspond to addition of 0.2 mM, 0.4 mM, 0.6 mM, 0.8 mM and 1.0 mM αKG (bottom to top). B) A plot of the absorbance change at 510 nm versus concentration of αKG. The solid line is a curve-fit using the following Equation 1 (see Experimental Procedures).

3.3.10 Recombinant *Bacillus* (Collagen-Like Protein) from *anthracis* (BclA)

The *bcla* gene cloned in pQE-30 expression vector, which contains a N-terminal His-tag, was obtained from Dr. George Stewart (University of Missouri-Columbia), and transformed into the M15 (Qiagen) *E.coli* competent cells. Expression of the 35,000 Da BclA protein was induced with 1mM IPTG when the OD_{600} reached 0.6, and grown at 37 °C for 1.5 hours. BclA polypeptide runs higher on SDS-PAGE because its monomers form triple helical structures. The soluble fractions were analyzed by SDS-PAGE. Initial attempts to express and purify rBclA resulted in low yield.

rBclA is hydrophobic and forms stable triple helices, which make the polypeptide difficult to solubilize. Triton X-100 was added at 0.1 % levels to increase solubility of the protein and the amount of rBclA present after lysis was greatly improved. Soluble portions of the lysate showed greater expression levels (Figure 3.15 lane 2). Purification was performed with a Ni-NTA column as previously discussed with the His-tagged anthrax-P4H (Figure 3.15 lane 3) followed by size exclusion column (Figure 3.15 lane 4) with the addition of 0.1% Triton X-100 to all of the buffers.

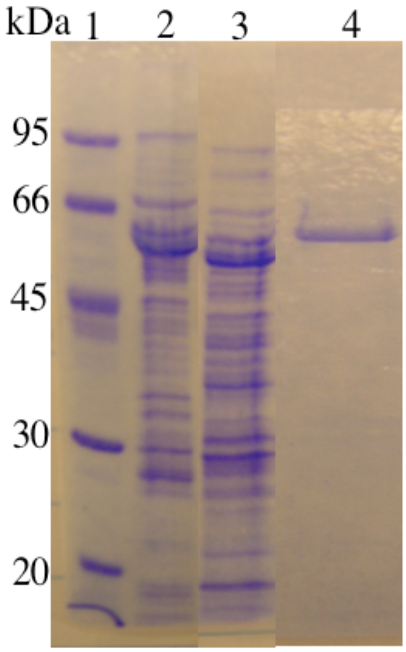


Figure 3.15. Recombinant *Bacillus anthracis* collagen-like polypeptide in the presence of Triton X-100 SDS-PAGE gel. Lane 1) Molecular weight marker. Lane 2) The lysate of recombinant BclA with Triton X-100. Lane 3) Recombinant BclA polypeptide after purification with Ni-NTA affinity column. Lane 4) is purified recombinant BclA after gel filtration chromatography.

3.3.11 Substrate Screening for Recombinant Anthrax-P4H

3.3.11.1 Screening Using the Fluorometric Assay

BclA and poly(L-proline) were assessed as potential substrates using the fluorometric assay. Though the results were difficult to reproduce, initial testing with both BclA and poly(L-proline) showed greater activity than the uncoupled reaction. Figure 3.16 shows a decrease in fluorescence with an increase in BclA concentration in the presence of 375 μM αKG, 375 μM DTT, 20 μM Fe(II), 50 mM Tris, pH 7.4, and 20 μM anthrax-P4H. The green line

represents activity seen with 0.5 mg/mL BclA, and the red line represents 5 mg/mL BclA activity. Poly(L-proline) (yellow line) was also tested and showed greater activity compared with uncoupled reaction, but less than seen in the presence of BclA. These results lead us to believe that BclA is potentially a substrate for anthrax-P4H. Additional activity assays were used to confirm this hypothesis.

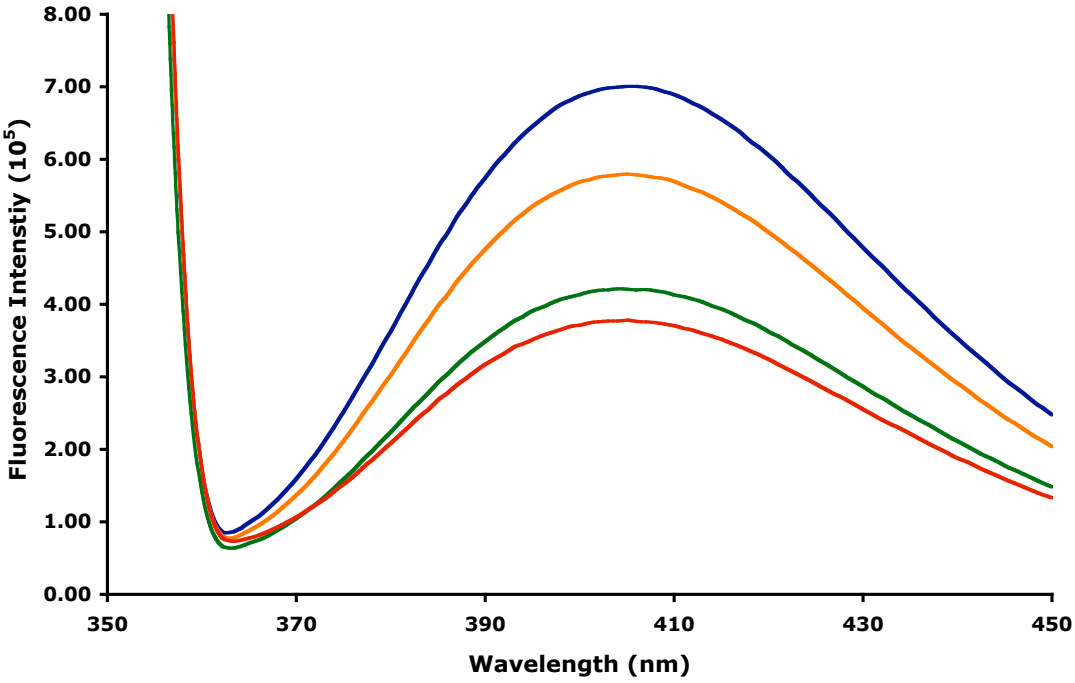


Figure 3.16. Substrate screening using fluorescence assay. Assays were run in the presence of 375 μM αKG, 375 μM DTT, 20 μM Fe(II), 50 mM Tris, pH 7.4 and 20 μM anthrax-P4H. Blue line represents no substrate, orange line represents 1 mg/mL poly(L-proline), green line represents 0.5 mg/mL BclA and red line corresponds to 5 mg/mL BclA.

3.3.11.2 αKG Consumption Assay Utilizing L-Glutamic Dehydrogenase

Potential substrates were screened for enzyme activity using the L-glutamic dehydrogenase assay, which measures αKG consumption. BclA and $(GPT)_5$ were screened for activity in the presence of 750 μM αKG, 1 μM DTT 40 μM Fe(II), 20 μM anthrax-P4H, and 50 mM Tris, pH 7.4. In addition, $(GPT)_5$ was run in the presence of 750 μM ascorbate. BclA concentrations varying from 7 μM - 56 μM and $(GPT)_5$ concentrations from 5 μM - 31 μM were run in triplicate (Table 3.2), and normalized against the uncoupled reaction. An increase of αKG consumption was seen with increasing substrate concentrations. As shown by the error values in Table 3.2, it was difficult to obtain consistent results from run to run as seen with the fluorescence assay.

	αKG Consumed
7 μM BclA	6 μM ± 2
14 μM BclA	16 μM ± 12
29 μM BclA	32 μM ± 3
43 μM BclA	83 μM ± 29
56 μM BclA	58 μM ± 43
5 μM $(GPT)_5$	71 μM ± 1.5
10 μM $(GPT)_5$	85.5 μM ± 15
20 μM $(GPT)_5$	139 μM ± 45
31 μM $(GPT)_5$	157 μM ± 39

Table 3.2. αKG consumption by anthrax-P4H reaction. BclA tests were run in the presence of 750 μM αKG, 750 μM DTT, 40 μM Fe(II), 20 μM anthrax-P4H, 50 mM Tris, pH 7.4. $(GPT)_5$ was run with the same conditions, with the addition of 50 μM ascorbate.

We also screened $(GPP)_{10}$ as a potential substrate of anthrax-P4H. Though *B. anthracis* does not contain high numbers of $(GPP)_{10}$ repeat sequences, this is the common polypeptide used when studying the human form of P4H. The K_m of ascorbate was determined to be 15 μM, which is considerably lower than the value for human, viral and plant forms of P4H of 300 μM (Table 3.1). The tight binding of ascorbate in anthrax-P4H leads us to believe that ascorbate is out competing substrate binding, and we began screening alternate reducing agents to use in activity assays.

Glutathione is a biological reductant that is ubiquitous in the body and responsible for eliminating oxidative stress. We determined that in the presence of glutathione, anthrax-P4H can turnover $(GPP)_{10}$. We used this system to study the time course of the anthrax-P4H reaction using the L-glutamic dehydrogenase assay. We incubated 20 μM of the enzyme at various time points (0 - 60 min) in the presence of 2 mM glutathione, 100 μM αKG, 100 μM $(GPP)_{10}$, 200 μM Fe(II), and 50 mM Tris pH 7.8. Figure 3.17 shows a plot of the consumption of αKG versus time. This assay indicates that the turnover of $(GPP)_{10}$ is slow in the presence of anthrax-P4H with only 66 μM αKG out of 100 μM αKG being consumed after 1 hour.

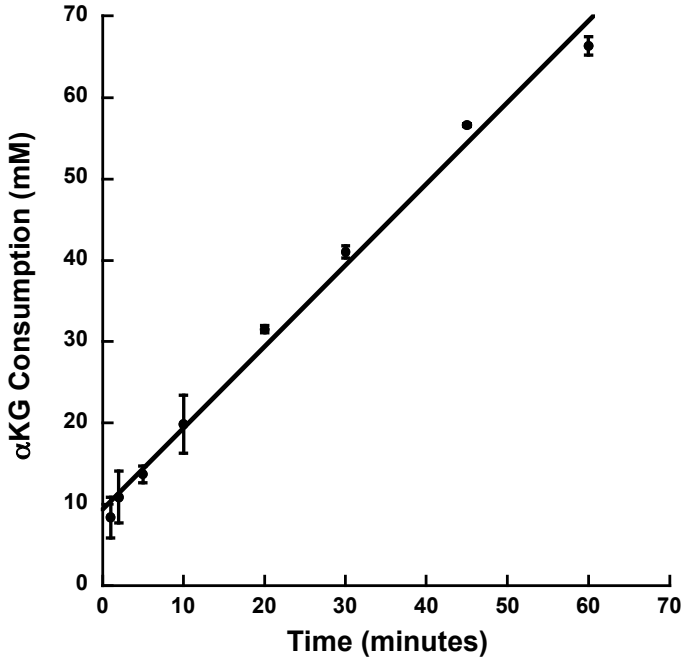


Figure 3.17. The consumption of αKG over time by anthrax-P4H coupled assay. Incubation of 20 μM anthrax-P4H at various time points (0 - 60 min) in the presence of 2 mM glutathione, 100 μM αKG, 100 μM $(GPP)_{10}$, 200 μM Fe(II) and 50 mM Tris pH 7.8.

3.3.11.3 Substrate Screening of Anthrax-P4H with O_2 Electrode

Substrate screening continued with the use of the O_2 electrode, which allows for the analysis on O_2 consumption in a real time manner. Both BclA and $(GPT)_5$ were screened in order to reproduce the results seen with the L-glutamic dehydrogenase coupled assay. The hydrophobicity of the substrates resulted in a large background reaction. Triton X–100 was added to the reaction mixture at a final concentration of 0.1% to improve solubility, but did not show any effect in reducing the noise level. Testing performed showed

minimal activity of BclA compared with the uncoupled reaction and $(GPT)_5$ was seen to inhibit O_2 consumption (Table 3.3).

	$\mu M\ O_2\ min^{-1}$
100 μM $(GPP)_{10}$	9.2 ± 1.6
100 μM $(GPT)_5$	-4.2 ± 0.6
10 μM BclA	2.5 ± 2.2
100 μM peptide 1	3.2 ± 1.1
100 μM peptide 2	3.4 ± 1.3
100 μM peptide 3	0.1 ± 2.0

Table 3.3. O_2 consumption in the presence of substrate. Samples were run in 750 μM αKG, 50 μM glutathione, 25 μM Fe(II), 5 μM anthrax-P4H, 50 mM Tris, pH 7.4.

Experiments were performed using the synthetic peptides of collagen-like repeats from the *B. anthracis* genome: peptide 1, peptide 2, and peptide 3. Screening with the synthetic peptides showed promising results for peptide 1 and peptide 2, but peptide 3 did not produce activity over uncoupling (Table 3.3). Interestingly, both peptides 1 and 2 contain GPP and GPT repeats in their sequences, similar to collagens in humans and *B. anthracis*. Peptide 3 contains a single GPQ and GPA sequence, which is found in bacterial collagen structural motifs.[80,81]

We tested $(GPP)_{10}$ for activity and were able to obtain a K_m of 9 ± 1.3 μM (Figure 3.18). It is intriguing that the K_m for $(Gly\text{-}Pro\text{-}Pro)_{10}$ for anthrax-P4H is almost identical to that reported for human-P4H (Table 3.1). These

results suggest that anthrax-P4H can indeed bind $(Gly\text{-}Pro\text{-}Pro)_{10}$ as tightly as human-P4H-1. k_{cat} was determined to be 2.7 ± 0.3 min^{-1} and the enzyme inactivates within 3 turnovers. We only see the rate increase in the presence of all the components in the reaction mixture, which shows that this activity depends on anthrax-P4H, Fe(II), and αKG.

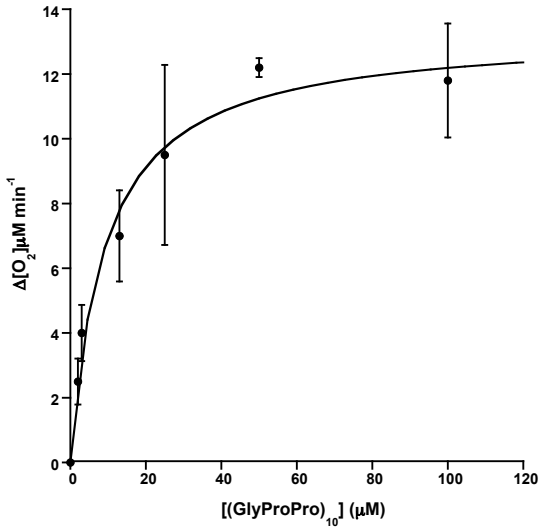


Figure 3.18. Kinetics for the oxidation of $(Gly\text{-}Pro\text{-}Pro)_{10}$ by anthrax-P4H fits to the Michaelis-Menten equation. The assay solution contained 5 μM anthrax-P4H, 750 μM αKG, 20 μM $Fe(NH_4)_2(SO_4)_2$ and 50 μM glutathione.

These experiments did produce some significant leads in identifying the substrate for anthrax-P4H. However, we have not been able to detect hydroxyproline formation from any of these reactions using the double derivatization method discussed in Chapter 2. Further study is required to maximize turnover as well as determine the native substrate for anthrax-P4H. Native substrate determination will be performed through proteomic

approaches using antibodies specific for hydroxyproline (Gemacbio, France) and BclA (Dr. Stewart).

3.4 Discussion:

3.4.1 Anthrax-P4H as an Improved System to Study Human-P4H

As discussed in Chapter 2, prokaryotes were not believed to produce Hyp and therefore not thought to contain P4H proteins. However, the genome of *B. anthracis* as well as *B. cereus* and *B. thuringenesis* all contain putative P4Hs according to National Center for Biotechnology Information-BLAST database. Therefore we sought to clone and express a recombinant collagen-P4H from *B. anthracis*, anthrax-P4H, and biochemically characterize the protein as a potential model system to study human-P4H. Improved spectroscopic and structural knowledge can aide in new strategies for the design of potential therapeutic agents that target fibrosis. We have successfully cloned, expressed, and purified the first recombinant P4H from a prokaryote. Initial characterization studies show anthrax-P4H exhibits both spectroscopic and catalytic properties characteristic of an αKG/Fe(II)-oxygenase.

Anthrax-P4H contains 30 % identity with the C-terminal end of the α subunit of human-P4H. This is approximately the same identity as the viral- and plant-P4Hs demonstrate with human-P4H. Anthrax-P4H is a highly soluble, monodispersed protein that can be purified to > 90 % with a two

column purification. Due to the complexity and restricted solubility of the $\alpha_2\beta_2$ human-P4H, structural determination of the entire human form involved in collagen synthesis has not been determined. A crystal structure of the catalytic domain of human HIF-P4H-2 (pdb 2G1M) has been determined as well as a monomeric algal-P4H (pdb 2JIG) from *C. reinhardtii*.[90,133] Human HIF-P4H-2 has only 14 % identity to human collagen-P4H and was determined to form a separate structural subgroup distinct from collagen-P4Hs.[90,133] Algal-P4H is the only crystal structure of a collagen-P4H currently known. However, it is monomeric and does not have substrate specificity similar to human-P4H.[72,90] We have determined the crystal structure of anthrax-P4H, which will be discussed in detail in Chapter 4. Together both the biochemical characterization and the crystal structure can serve to improve the development and design of specific fibrosis inhibitors targeting human-P4H.

3.4.2 Anthrax-P4H can Exist as an α_2 Homodimer

Through gel filtration and MALDI-TOF analysis, recombinant anthrax-P4H was found to be a homodimer. This is the first known P4H that can exist as a homodimer. As stated previously, the human form exists as an $\alpha_2\beta_2$ heterotetramer. The major form of P4H in *Caenorhabditis elgans* (*C. elgans*) is a α(I)α(II)B_2 tetramer, but α(I)β and α(II)β heterodimers can also be assembled at much lower efficiency.[62] This is the only known P4H that

contains two distinct catalytic subunits and is species-specific to the nematode.[62] Plant (At-P4H-1 and At-P4H-2) and viral (PBCV-1-P4H) forms are monomeric, α.[69,70,73]

Gel filtration experiments did show anthrax-P4H primarily as a homodimer, however, MALDI-TOF MS experiments showed anthrax-P4H predominately in the monomeric state. This may be because the MS experiments are run under harsh conditions that can denature and dissociate the dimer form during analysis.[134,135] The matrix used was a sinapinic acid and ammonium citrate complex adjusted to a pH of 7.0. This matrix has been used previously to characterize protein quaternary structure as the pH of 7.0 stabilizes the protein during MS analysis.[136] The dialysis experiments resulted in almost exclusively dimer form of the enzyme. We believe equilibrium is occurring between the monomeric and dimeric forms of the enzyme.

Site-directed mutagenesis experiments were performed to determine if the dimer is stabilized through a disulfide bond. The α subunit of human-P4H contains five cysteine residues (Cys150, Cys276, Cys293, Cys486, Cys511) which have been shown through site directed mutagenesis studies to be essential in complete tetramer assembly.[122,123] Two intrachain disulfide bonds are formed between Cys276 - Cys293 and Cys486 - Cys511 in the α-subunit.[122,123] To date there is no evidence for the role of Cys150, but mutation to Ser did cause minor structural changes.[123] Anthrax-P4H contains

only one cysteine residue in its sequence, C53. Mutagenesis studies on this cysteine indicate it is not contributing to the dimerization. This supports the theory that the dimer is formed through other forces such as hydrogen bonding or Van der Waals forces (non-covalent dimer). X-Ray crystallography experiments lead to a clearer understanding of the interactions occurring at the dimer interface and will be addressed in Chapter 4.

3.4.3 Anthrax-P4H Undergoes Uncoupling Reaction with Ascorbate

While the rate of the uncoupling reaction (oxidation of ascorbate) by anthrax-P4H is very similar to that reported for human-P4H-1 (1.9 min^{-1} for anthrax-P4H versus 4 min^{-1} for human-P4H-1),[23] the K_m for ascorbate for anthrax-P4H is an order of magnitude smaller compared to human-P4H-1 and is similar to that for Cr-P4H-1 (see Table 3.1).

The K_m for αKG for anthrax-P4H is very similar to that of human-P4H-1 and PBCV1-P4H, but 20 to 30-fold lower than At-P4H-1, At-P4H-2, and Cr-P4H-1.[37,69,72,73] Comparison of the structures of anthrax-P4H and Cr-P4H-1 shows that the flexible loop between β-strands II-III of the core fold in Cr-P4H-1 is $\sim$ 4 Å longer than anthrax-P4H. The difference in length may cause the loop in anthrax-P4H not to be as flexible towards conformational change and leave the active site of anthrax-P4H more exposed compared than that of Cr-P4H-1. The entrance to the active site cavity of the two structures also show a

difference in size, where the Cr-P4H-1 cavity is 2.3 Å smaller in diameter than anthrax-P4H (See Chapter 4). This may explain why αKG is more accessible to the active site cavity of anthrax-P4H than that of Cr-P4H-1. The K_m for Fe(II) for anthrax-P4H is also comparable to that of human-P4H-1 and At-P4H-2 (see Table 3.1).[37,70]

We determined the pH-dependency of the initial rate of the uncoupling reaction and obtained a bell-shaped curve with a pH optimum at 7.0 (Figure 3.12). From the nonlinear least-square fit analysis, we have determined two pK_a values, 6.92 ± 0.18 and 7.17 ± 0.18, respectively. This suggests that there are at least two ionizable groups involved in the catalytic reaction. In the active site of the Cr-P4H-1, two His residues can be found and those residues are conserved in anthrax-P4H, namely His127 and His193 as part of the catalytic triad. Asp129 also as the part of the triad can be one of the ionizable groups, when it is involved in a hydrogen-bonding network. These residues correspond to His412, Asp414 and His483 in human-P4H-1 which have been identified as the Fe(II) binding site residues by the site-directed mutagenesis study,[40] where H412E, D414N and H483E mutants resulted in complete loss of activity.[40]

Anthrax-P4H takes ascorbate as a presumed surrogate substrate as has been observed for human-P4H.[23,137] Ascorbate can act as an alternate oxygen acceptor in the catalytic cycle and is consumed stoichiometrically in

the absence of substrate.[23] P4Hs are known to require ascorbate as a co-substrate that is essential for turnover and protects the enzyme from oxidative damage leading to inactivation.[22] The mechanism for the inactivation of human-P4H is not well understood due to the limited amount of recombinant protein that is not suitable for spectroscopic study. It has been shown in other αKG/Fe(II)-oxygenases that the inactivation in the absence of substrate is due to the self-hydroxylation of a Phe, Tyr or Trp residue in the active site, where the blue chromophore is assigned as a charge transfer complex of hydroxylated product to Fe(III).[138-140] This self-hydroxylation occurs concomitantly with decarboxylation of αKG and resulting enzyme is inactivated. In the absence of ascorbate, the catalytic cycle of anthrax-P4H lead to inactivation coupled to the formation of a blue chromophore with λ_{max} at ~ 600 nm (Figure 3.13). The absorption maximum seen in anthrax-P4H is very close to that seen when the hydroxyl group of Tyr (hydroxylated-Phe) and hydroxyl-Trp coordinate to Fe(III).[138-140] There are several strictly conserved Phe, Trp, and Tyr residues near the iron-binding site of anthrax-P4H as well as in human-P4H-1. It is possible that similar chemistry is responsible for the inactivation of anthrax and human-P4H-1.

3.4.4 Anthrax-P4H Exhibits Spectroscopic Characteristics Similar to other αKG-Fe Oxygenases

Anthrax-P4H exhibits both spectroscopic and catalytic properties characteristics of an αKG/Fe(II)-oxygenase. The Fe(II)-αKG binary complex

of αKG/Fe(II)-oxygenases have a distinct visible spectrum, which can be monitored under anaerobic conditions.

DAOCS was the first αKG-Fe(II)-oxgenase to have both its Fe(II) and Fe(II)/αKG-bound crystal structures published.[141] These results confirmed the octahedral coordination of Fe(II) surrounded by the His-Asp/Glu-His facial triad and water molecules. The binding of αKG to the ferrous ion showed the displacement of two water molecules and bidentate binding of αKG.[141] Reaction intermediates have been characterized spectroscopically through UV-Vis, CD, and MCD spectroscopy studies.[26,29-31]

A characteristic spectra at 500 nm is correlated to the MLCT absorption transitions under anaerobic conditions is widely believed to represent the bidentate binding of αKG to Fe(II)-enzyme complex. Krebs and Bollinger obtained this characteristic spectra at 520 nm with viral-P4H, with a ε_{520} = 250 M^{-1} cm^{-1}, which compares closely to that obtained with anthrax-P4H.[124] The shape of our peaks showed two maxima, λ_{max} ~ 480 nm and 508 nm, and we obtained the same extinction coefficient. The titration of αKG into the anthrax-P4H complex resulted in a K_D = 54 ± 25 μM, which compares with viral-P4H's K_D = 27 ± 6 μM

Upon exposure to air, the solution turns blue, and the absorption spectra shifts from λ_{max} ~ 508 nm to λ_{max} ~ 615 nm with an extinction coefficient of 500 M^{-1} cm^{-1}. Similar spectra and blue chromophore formation

have been seen in the αKG-Fe(II) oxygenase TfDa, which exhibits a shift from a weak transition at λ_{max} ~ 530 nm to the strong, broad signal at λ_{max} ~ 580 nm.[132] Through EPR data, the blue chomophore in TfDa was shown to arise from a high-spin Fe(III) complex resulting from the hydroxylation of an active site Trp residue as previously discussed.

Attempts were made to examine the effect of the binding of $(Gly\text{-}Pro\text{-}Pro)_{10}$ to the UV-vis spectrum of the αKG/anthrax-P4H/Fe(II) ternary complex. However, no spectral changes were observed. UV-Vis spectra of viral-P4H in the presence of substrate $(Pro\text{-}Ala\text{-}Pro\text{-}Lys)_3$ also resulted in no noticeable spectral changes.[124] A small spectral shift was seen upon the binding of taurine to TauD from 530 nm (ε_{530} = 140 M^{-1} cm^{-1}) to 520 nm (ε_{520} = 40 M^{-1} cm^{-1}), with more distinctive features at 470 nm and 570 nm.[31] Hausinger and coworkers were able to use 5 mM taurine in their studies, and we were only able to use 1 mM $(Gly\text{-}Pro\text{-}Pro)_{10}$ in our experiments due to the solubility of the peptide substrate. This may be another factor in the absence of a shift with the binding of $(Gly\text{-}Pro\text{-}Pro)_{10}$ to anthrax-P4H.

3.4.5 Potential Substrates for Anthrax-P4H

As stated previously, prokaryotes including *B. anthracis*, do not contain the $(Gly\text{-}Pro\text{-}Pro)_n$ repeats involved in collagen production in humans. However, BclA protein has been found to contain Gly-Pro-Thr repeats.[98] Turnbough and coworkers found the Thr residues to have extensive

glycosylation,[96] and recombinant BclA forms trimers which are sensitive to collagenase and exhibit CD spectrum similar to that of the collagen triple-helix.[99] Because of its collagen-like properties, we believe BclA may act as a substrate for anthrax-P4H. Therefore, we obtained a recombinant form of BclA from Dr. Stewart, University of Missouri, and were able to obtain soluble protein in low levels. The hydrophobic nature and tightly folded trimeric structure of the recombinant BclA resulted in a inefficient reaction with anthrax-P4H.

Establishing anthrax-P4H as an effective model system for human-P4H requires that the enzyme hydroxylate collagen substrates. The Limburg lab has developed an recombinant system for $(Gly\text{-}Pro\text{-}Pro)_{10}$ expressed as a fusion protein in *E. coli*. This procedure allows for the purification of milligram quantities of substrate for assays at a fraction of the cost and time as that with peptide synthesis. $(Gly\text{-}Pro\text{-}Pro)_{10}$ was tested for activity with anthrax-P4H in the presence of ascorbate, Fe(II), αKG, and O_2. It is known that some αKG/Fe(II) enzymes can use co-substrates other than ascorbate, such as DTT, to support catalysis. Of these thiol-containing reagents, glutathione, a reductant that is present in bacteria, supported limited turnover whose rate increased in the presence of $(Gly\text{-}Pro\text{-}Pro)_{10}$, an *in vitro* substrate for human-P4H. On the other hand, other proline-substrates such as poly(L-proline), an *in vitro* substrate for PBCV-P4H, At-P4H-1 and At-P4H-2, and Cr-P4H-1 and

a synthetic $(Gly\text{-}Pro\text{-}Thr)_5$ peptide mimicking the repeat in BclA showed no effect. *O*-glycosylation of Thr in the $(Gly\text{-}Pro\text{-}Thr)_5$ may be necessary for substrate recognition by anthrax-P4H as seen in native BclA. It is intriguing that the K_m for $(Gly\text{-}Pro\text{-}Pro)_{10}$ for anthrax-P4H is almost identical to that reported for human-P4H (Table 3.1). These results suggest that anthrax-P4H can indeed bind $(Gly\text{-}Pro\text{-}Pro)_{10}$ as tightly as human-P4H-1.

It is not yet clear if the increase in turnover represents formation of Hyp as we have not to date detected any using standard HPLC protocols. Further, the rate of turnover is much slower than what has been reported for other P4Hs (2 min^{-1} vs > 300 min^{-1}). One possible explanation is that $(Gly\text{-}Pro\text{-}Pro)_{10}$ binds to the anthrax-P4H and promotes the uncoupled turnover of glutathione. A similar phenomenon has been reported for the interaction of poly(L-proline) with human-P4H where the former acts as an inhibitor for $(Gly\text{-}Pro\text{-}Pro)_{10}$ but promotes the uncoupled oxidation of ascorbate two-fold.[23,142]

The similar K_m between human and anthrax-P4H for $(Gly\text{-}Pro\text{-}Pro)_{10}$ is somewhat surprising as anthrax-P4H lacks the proposed PBD of human-P4H. Removal or mutation of the human-PBD greatly raises the K_m for substrate in human-P4H-I.[117] However, At-P4H-1 will react with $(Gly\text{-}Pro\text{-}Pro)_{10}$ with a K_m of 60 μM, where the relative activity is ~ 1/30 of that of poly(L-proline), the best substrate tested for that enzyme.[69] It was not determined if $(Gly\text{-}Pro\text{-}Pro)_{10}$ was hydroxylated.[69] It remains unknown whether $(Gly\text{-}Pro\text{-}Pro)_{10}$ is

hydroxylated by At-P4H-2.

The active site structure of Cr-P4H-1 revealed an extended shallow groove near the active site that is proposed to serve as a collagen-like peptide binding/recognition.[90] The residues that are shown to be essential in substrate binding in Cr-P4H-1 by site-directed mutagenesis studies are strictly conserved in anthrax-P4H, human-P4H-1, At-P4H-1 and At-P4H-2 (Figure 3.1). It is possible that collagen-like substrates are primarily recognized by the groove near the active site by P4Hs. In human-P4H-1, it is most likely that the additional PBD is located spatially close to the groove and functions in concert with the groove to bind procollagen.

3.5 Conclusions and Future Directions

The attempt to establish an improved model system of human-P4H to conduct structural and mechanistic study was initiated through the cloning and expression of the first example of a bacterial P4H. A putative P4H from *Bacillus anthracis* was cloned, expressed, and purified in *E. coli* with yields of ~10 mg per liter of cell culture. Characterization studies showed enzyme activity dependence on Fe(II), αKG, and O_2, which fits its classification as an αKG/Fe(II)-oxygenase. Anthrax-P4H was found to adopt both monomeric and dimeric forms dependent on conditions and types of analysis. MALDI-TOF MS and site-directed mutagenesis studies showed that homodimer formation is stabilized through forces other than a disulfide bond. This is the

first P4H found to form a homodimer.

Purified, highly soluble anthrax-P4H was obtained and tested for activity. We developed an assay utilizing the enzyme L-glutamic dehydrogenase's dependence on αKG, and monitored activity using an O_2 electrode. We were able to measure consumption of both αKG and O_2 and obtain leads for potential substrates. BclA protein was expressed and purified as a potential substrate and gave promising but irreproducible results. We are currently screening additional synthetic peptides as possible substrates for anthrax-P4H. Once enzymatic turnover is optimized, further characterization of the mechanistic intermediates will be determined. This includes stopped-flow UV-vis, Mössbauer, and rapid-freeze quenching EPR spectroscopic analysis.

Anthrax-P4H shows apparent tight-binding of the collagen-like substrate $(Gly\text{-}Pro\text{-}Pro)_{10}$ with the same affinity as human-P4H-1 as well as the surrogate substrate ascorbate. The structure of anthrax-P4H may be structurally related to the α_2 subunit of human-P4H-1, and may help understand the assembly of the α_2 subunit and interactions between the two α domains.

Chapter 4: X-Ray Crystal Structure Determination of a Proyl-4-Hydroxylase-Like Protein from *Bacillus anthracis*

4.1 Introduction

C-P4Hs have long been targets for the development of selective inhibitors against diseases such as fibrosis.[17] Scar formation and fibrotic tissue development are natural healing responses after injury.[143] However, in some situations the production of fibrotic tissue accumulation is excessive and can lead to tumor formation. The organs most commonly affected are the liver, lungs, kidneys, and skin.[143] Human-P4H has gained much attention in the development of a specific inhibitor as a potential therapeutic targeting fibrosis, as it catalyzes the rate-limiting step in the biosynthesis of collagen. In particular, Human-P4H-1 is the most abundant form of C-P4H in the body, and therefore most involved in fibrosis formation after injury.[14]

There are three tissue specific isoenzymes of C-P4H in humans, with amino acid differences among the three located in the α subunit.[14] The sequence identity among human-P4H-1 and -2 is 65 % and the identity among -1 and -3 and -2 and -3 is 35 % and 37 % repectively.[14] The development of selective inhibitors for human-P4H-1 is difficult because the isoenzymes are so closely related, with the C-terminus having the highest percentage of identity. αKG analogs have been shown to successfully inhibit C-P4Hs *in vitro*. 2',4'- and 2',5'-Dicarboylate pyridines inhibit human-P4H-1 with K_i values of 2 μM and 0.8 μM respectively.[7] However, these inhibitors

have low permeability *in vivo* and the development of specific inhibitors is needed to combat the delivery problem as well as specificity and selectivity.

The major difference in the isoenzymes of C-P4H is in the PBD.[17,127] Studies have shown two regions in the PBD are responsible for poly(L-proline) (inhibitor) binding corresponding to residues 162 - 192 and 231 - 240 in human-P4H-1.[127] Site-directed mutagenesis studies strongly suggested that Ile182 and Tyr223 (human-P4H-1 numbering) and Glu180 and Gln231 (human-P4H-2 numbering) are involved in peptide binding where mutations E180I and Q231Y in human-P4H-2 lowered the K_m of $(GPP)_{10}$ to values similar to human-P4H-1.[127]

The crystal structure of the full length $\alpha_2\beta_2$ human-P4H-1 has not been determined. The α subunit of human-P4H-1 is insoluble in the absence of PDI and does not remain properly folded in the ER. This has hindered attempts to crystallize the α_2 subunit of human-P4H-1. An improved expression and purification system for recombinant human-P4H-1 is required to obtain the complete human-P4H-1 structure. We have successfully expressed a recombinant form of human-P4H-1 in high purity and at yields of 5 mg/L. We are currently screening crystallization conditions for structure determination. The structural information can aid in developing selective inhibitors as potential therapeutics for fibrosis.

The crystal structure of the isolated PBD from the α-subunit of human-

P4H-1 has been determined to 2.3 Å resolution.[117] The structure consists of five α-helices and belongs to the family of tetratricopeptide-repeat (TPR) domains that are involved in many protein-protein interactions.[144,145] A groove in the PBD containing eight tyrosine residues has been found to be essential in substrate recognition through site-directed mutagenesis experiments.[117] The crystal structure of a PDI from yeast has been determined to 2.4 Å.[46]

The crystal structures of a monomeric C-P4H from *C. reinhardtii* (Cr-P4H-1) complexed with Zn(II) and the αKG analog 2,4-dicarboxylate pyridine have been reported.[90] The jellyroll motif characteristic of αKG/Fe(II)-oxygenases was confirmed. Cr-P4H-1 lacks the separate PBD but instead, a putative peptide-binding groove with high plasticity has been proposed. Analyzing the structures of Cr-P4H-1 and anthrax-P4H can also provide insights into the differing substrate specificity and affinity for αKG of the two enzymes. The structure of HIF-P4H2 has also been determined, but was found to belong to a structural subgroup distinct from C-P4Hs.[90,133]

We report here in the X-ray structure determination of the recombinant anthrax-P4H. We have shown that the anthrax-P4H binds $(Gly\text{-}Pro\text{-}Pro)_{10}$ with a similar K_m values as that of human-P4H-1 (Chapter 3), and its ability to form homodimers may provide insight into the α domain interactions of human-P4H-1 and its catalytic property, in particular the recognition mode of the collagen-like peptide.

4.2 Materials and Methods

4.2.1 Expression and Purification of Selenomethionine (SeMet) Substituted Anthrax-P4H

To produce SeMet protein, the native *anthrax-p4h* plasmid was transformed into methionine-requiring auxotrophic *E. coli* strain B384(DE3) (Novagen). A 100 mL starter culture was grown overnight at 37 °C in SelenoMet Medium (Molecular Dimensions Ltd.) containing 100 μg/L ampicillin and 30 μg/L L-SeMet (Anatrace). A 10 mL of starter culture was used to inoculate 1 L of SelenoMet Medium containing 100 μg/L ampicillin and 30 μg/L L-SeMet. Cells were induced at an OD_{600} of 0.6 by addition of 200 μM IPTG as described for the native protein (Chapter 3). Cell growth was maintained at 37 ˚C for 12 hours after induction. After 12 hours the cells were harvested by centrifugation similar to native anthrax-P4H as described in Chapter 3.

The SeMet-anthrax-P4H was purified using the method described for the native form of the enzyme (Chapter 3), and the overall protein yield of the SeMet-anthrax-P4H was 12 mg/L, comparable to that of the native form (Chapter 3). MALDI-TOF mass spectrometry analysis determined a mean incorporation of 3 selenium atoms per anthrax-P4H monomer, where four methionines are predicted in the deduced amino acid sequence including the N-terminus.

4.2.2 Crystallization

Initial crystallization screening was performed by the hanging-drop vapor-diffusion method using commercially available sparse matrix screening kits (Hampton Research and Emerald Biosystems). The purified SeMet-anthrax-P4H was concentrated to 24 mg/mL in 50 mM Tris buffer (pH 7.4) containing 150 mM KCl, and 5 mM β-mercaptoethanol. Equal volumes of protein and reservoir solution (1 μL + 1 μL) were mixed and equilibrated against the 750 μL reservoir solution at 293 K. Initial crystals were obtained with 10% PEG-8000, 0.1 M MES pH 6.0, 0.2 M $Zn(OAc)_2$ from the Emerald Biosystems Wizard II kit. These crystals were not of diffraction quality and further optimization was attempted. Because the initial mother liquor solution contained $Zn(OAc)_2$, co-crystallization of all the solutions in the presence of 1 mM Zn(II) was performed. However, we have not been able to obtain diffractable crystals with these screening kits.

Additional sparse-matrix screens were performed, and rhombic crystals of varying sizes were obtained upon equilibration with 16% (w/v) PEG 8000, 40 mM potassium phosphate (monobasic) and 20% (v/v) glycerol from the Emerald Biosystems Wizard III kit. The formation of crystals was found to be pH dependent, and crystal growth was optimized at pH range 4.0 – 4.2.

Co-crystallization experiments were performed in an anaerobic glovebox (COY Laboratories) in the presence of Fe(II), αKG, $(GlyProPro)_{10}$ or ascorbate under the same mother liquor conditions as apo-anthrax-P4H.

4.2.3 Data Collection

X-ray diffraction data from an apo-anthrax-P4H crystal was collected in-house (KU-Protein Structure Lab) on an R-AXIS IV++ detector with Cu $K\alpha$ X-rays generated by a Rigaku RU-H3RHB rotating-anode generator and focused using an Osmic confocal optical system (Rigaku, Japan). The crystal was flash-cooled by direct immersion into liquid nitrogen. No additional cryoprotectant was necessary due to the high content of PEG-8000 and glycerol in the well solution. A total of 113º of data were collected (0.5º oscillation, 4.5 min exposure per oscillation) and the data was indexed, scaled, and integrated using the CrystalClear software package. The data collection statistics are summarized in Table 4.1.

A three-wavelength anomalous dataset of the SeMet-anthrax-P4H was collected at beamline BL9-2 at the Stanford Synchrotron Radiation Laboratory (SSRL) to 1.4 Å resolution, and was solved using multi-wavelength anomalous dispersion (MAD) phasing.[146] The SeMet crystals were harvested as described for the apo-anthrax-P4H, loaded into a sample cassette designed for the Stanford Automated Mounting (SAM) system, and screened for diffraction quality.[147] The wavelengths were determined by the

fluorescence absorption spectrum near the selenium L_{III} absorption edge. The data set was collected at remote, peak, and inflection wavelengths (0.912 Å, 0.9793 Å, 0.912 Å). Several MAD data sets were collected from separate crystals and the data set producing the best overall figure of merit for the experimental phases was used for structure determination.

A total of 180º of data were collected (1º oscillation, 2 sec exposure per oscillation). The data was processed with the program MOSFLM,[148] and scaled with SCALA of the CCP4 program suite.[149] The space group was determined as $P2_1$ with cell dimensions a = 41.38 Å, b = 63.80 Å, c = 98.50 Å, β = 98.74°. The data collection statistics are summarized in Table 4.2 (See Results section).

Data collection of Fe(II)/αKG, Fe(II)/αKG/$(GPP)_{10}$, and Fe(II)/αKG/ascorbate crystals were collected to 1.39 Å, 1.65 Å, and 1.40 Å respectively at the SSRL. The data collection statistics are summarized in Table 4.3 (See Results section).

4.2.4 Structure Determination and Validation

A data set from the SeMet-anthrax-P4H data collection was used to solve the structure of apo anthrax-P4H. Shelx[150] was used to calculate the heavy atom substructure. Phases were calculated using SOLVE,[151] which was followed by electron-density modification using RESOLVE.[152] The model was further rebuilt manually using the graphical software COOT[153], and

crystallographic refinement was performed with the programs REFMAC5[154] and Arp/wArp 6.1[155,156] of the CCP4 suite.[157] The structure refinement statistics are summarized in Table 4.2.

The structure was validated using Procheck[158], WHATIF,[159] and Molprobity.[160] The geometry statistics are summarized in Table 4.2 (See Results section). All molecular figures were produced with PyMOL.[161]

4.2.5 Site-Directed Mutagenesis to Probe Self-Hydroxylation Reaction

As discussed briefly in Chapter 3, members of the Fe(II)/αKG-oxygenases have been shown to undergo enzyme inactivation resulting in the self-hydroxylation of a Phe, Tyr or Trp residue near the Fe(II) binding site in the absence of substrate. Site-directed point mutations were performed using the QuikChange method (Stratagene). Primers were designed using Primer-X (http://www.bioinformatics.org/primerx/). For F178Y, forward primer: 5'-GAAAGGGAATGGCAGTATACTATGAGTATTTCTATCAAGACC-3' and reverse primer: 5'-GGTCTTGATAGAAATACTCATAGTATACTGCCATTCCCTTTC-3' were used. For F178A, forward primer: 5'-GAAAGGGAATGGCAGTATACGCTGAGTATTTCTATCAAGACC-3' and reverse primer: 5'-GGTCTTGATAGAAATACTCAGCGTATACTGCCATTCCCTTTC-3' were used. The resulting plasmids were fully sequenced and the point mutations were confirmed. The mutant proteins were expressed and purified as described for the WT.

4.2.6 Anaerobic Spectroscopy Monitoring Anthrax-P4H Self-Oxidation

Samples were prepared as discussed previously in Chapter 3. Briefly, 100 μL of 1 mM enzyme (F178Y or F178A) in 50 mM Tris buffer (pH 7.4) containing 2 mM sodium dithionite was placed in a septum-rubber sealed 50 μL cuvette and was degassed by gently purging O_2-free Ar for 30 min. Stoichiometric amounts of anaerobic solutions of Fe(II) and αKG were added to the cuvette and the absorbance was measured. After 30 minutes, the cuvette was opened to air and the spectra were monitored over time up to 2 hours for spectral changes.

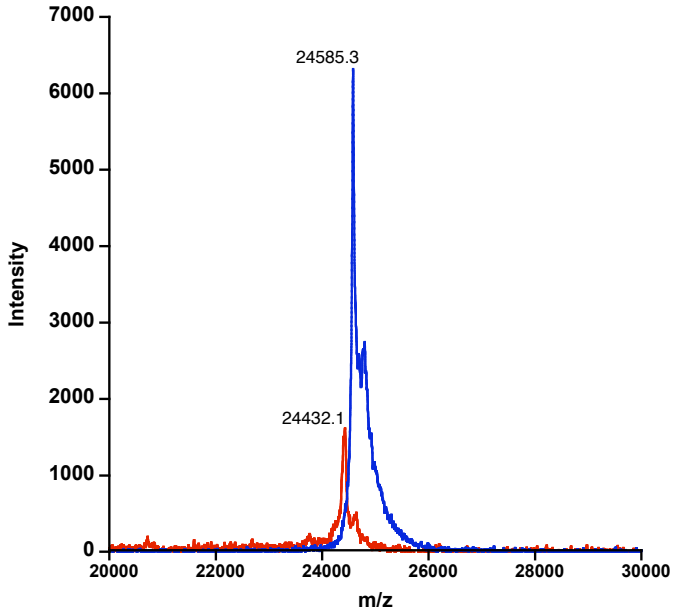
4.3 Results

4.3.1 SeMet Incorporation in Anthrax-P4H

Attempts to perform molecular replacement using the search model of Cr-P4H-1 (2V4A, 2JIJ, or 2JIG), and the single wavelength anthrax-P4H data were unsuccessful. We therefore proceeded to use MAD methods to solve the structure. SeMet-anthrax-P4H could be easily prepared in large amounts using auxotrophic *E. coli* strain 834(DE3). SeMet incorporation was confirmed by MALDI-TOF MS (Figure 4.1 A). Purified samples of native anthrax-P4H and SeMet-anthrax-P4H resulted in the average m/z values of 24,432 and 24,585 respectively. The additional mass of 153 corresponds to the difference obtained when selenium replaces sulfur in three of the methionine residues. The fluorescence scan of the SeMet-anthraxP4H showed a strong

signal at the selenium absorption edge (Figure 4.1 B) and was used to determine the peak, edge, and remote wavelengths used in the MAD data collection.

A



B

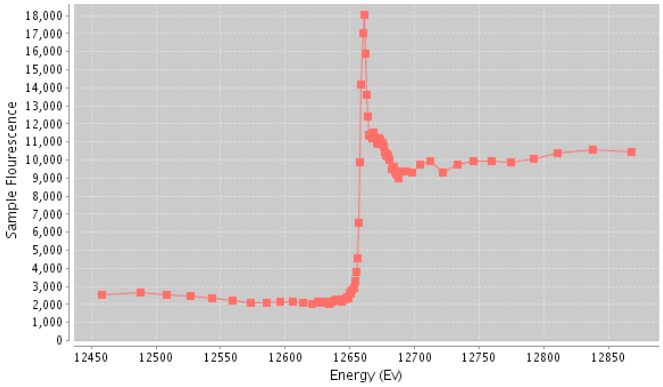


Figure 4.1. SeMet incorporation in recombinant anthrax-P4H. A) MALDI-TOF mass spectrometry of native (red) and SeMet-substituted (blue) anthrax-P4H. B) Fluorescence scan of SeMet anthrax-P4H crystal at SSRL to determine optimal energies for MAD data collection.

4.3.2 Crystallization of Anthrax-P4H

Initial sparse matrix screening resulted in needle shaped urchin crystals (Figure 4.2 A) and we were not able to optimize the conditions to produce diffraction quality crystals. We were able to obtain three-dimensional crystals for both native and SeMet-anthrax-P4H with further screening (Figure 4.2 B).[162] We have found that a subtle pH difference between stock solutions was the controlling factor to obtain diffraction quality suitable crystals. The optimum crystals were produced in the buffers between pH 4.0 - 4.2, and crystals grew to a typical size of 0.8 x 0.5 x 0.1 mm over 1 to 2 weeks.[162]

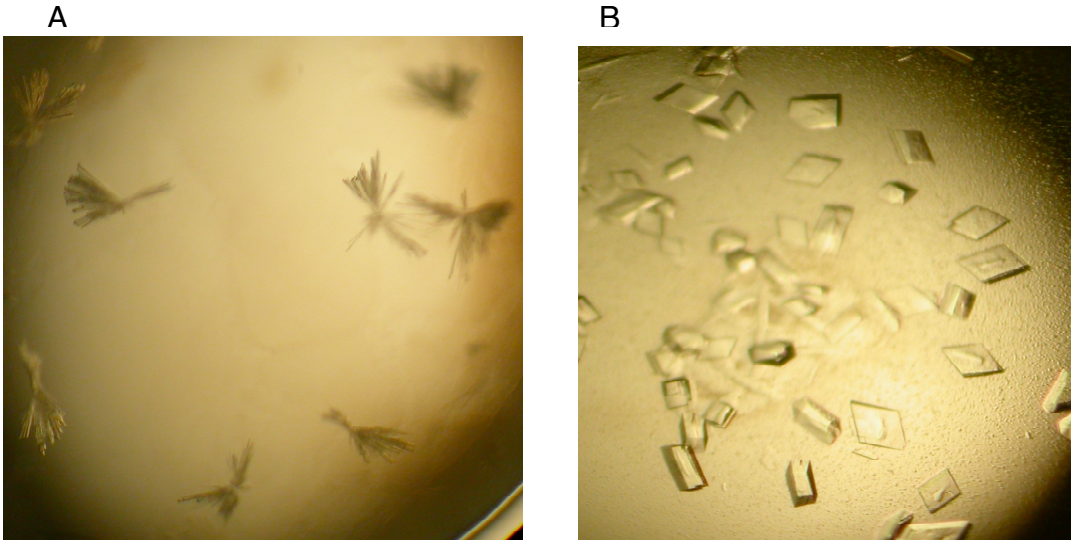


Figure 4.2. A) Initial crystals produced from Emerald Biosystems kit Wizard I, solution # 7 made up of 10 % (w/v) PEG-8000, 0.1 M MES pH 6.0, and 0.2 M $Zn(OAc)_2$. B) Crystals produced from Emerald Biosystems kit wizard III, solution # 32 made up of 16 % (w/v) PEG-8000, 0.04 M potassium phosphate (monobasic), and 20 % (v/v) glycerol. A pH of 4.0 - 4.2 is required for crystallization seen in B.

4.3.3. Data Collection and Structure Determination/Validation

Initial native anthrax-P4H crystals grew in the orthorhombic space group *P222* with unit cell parameters a = 63.76 Å, b= 80.12 Å, c = 98.63 Å, and diffracted to 2.1 Å. The presence of two protein molecules in the asymmetric unit gave a calculated Matthews coefficient V_M of 2.57 $Å^3$/Da, which corresponds to a solvent content of 52%. The V_M was within the range commonly observed for protein crystals.[163] Data collection is summarized in Table 4.1

Space Group	P222 Orthorhombic
Resolution (Å)	16.04 – 2.10 (2.17 – 2.10)
Unit-cell parameters (Å, °)	63.76 80.12 98.63 90.00 90.00 90.00
No. of measurements	96761
No. of unique reflections	28619
Redundancy	3.38 (3.39)
Completeness (%)	95.0 (93.2)
Rmerge (%)	14.6 (35.5)
Reduced ChiSquared	0.97 (1.01)
Average I/σ(I)	6.0 (3.4)

$$R_{merge} = \sum_{hkl} \sum_{i} \left| I_i(hkl) - \langle I(hkl) \rangle \right| / \sum_{hkl} \sum_{i} I_i(hkl)$$

Table 4.1. Native anthrax-P4H data collection summary from Rigaku home source (University of Kansas-Protein Structure Lab). Values in parentheses indicate highest resolution shell.

SeMet-anthrax-P4H crystals grew in the monoclinic space group $P2_1$ with unit cell parameters a = 41.38 Å, b = 63.80 Å, c = 98.50 Å, β = 98.73 °, and diffracted to 1.4 Å at three wavelengths around the selenium edge.[162] The data set was collected at remote, peak, and inflection wavelengths (0.912

Å, 0.9793 Å, 0.9795 Å). Figure 4.3 is a diffraction pattern representative of the data collection at the SSRL. After repeated cycles of refinement and model building the crystallographic R/R_{free} factors are 17.9%/19.9% (Table 4.2). Average B-factors of the structure are 12.5 for the main chain and 13.4 for the side chain. Analysis of the structure with Procheck shows that 91 % of the residues reside in the mostly favored region of the Ramachandran plot as displayed in Table 4.2. Because the SeMet dataset was collected to 1.4 Å, we used it to solve the structure that will be the basis of the following discussion.

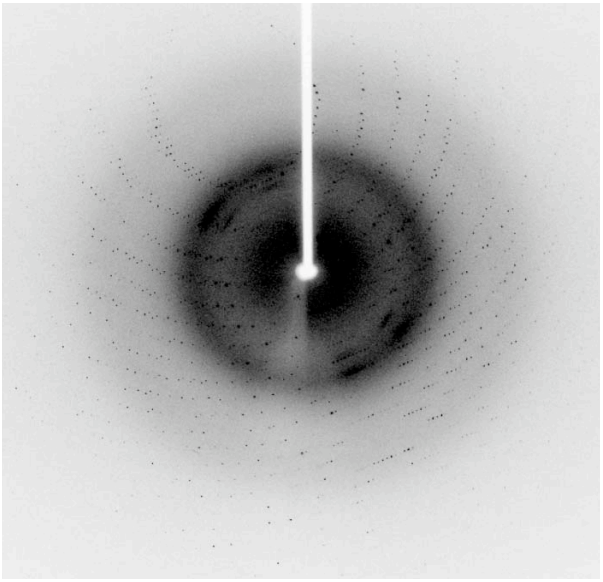


Figure 4.3. Diffraction pattern of SeMet-anthrax-P4H obtained on beamline BL9-2 at the SSRL. The resolution of this data set was 1.4 Å. The detector edge was set at 1.36 Å.

Crystals containing cofactors and substrates also grew in the $P2_1$ space group with unit cell parameters close in value to the SeMet-anthrax-P4H crystals (Table 4.3). Analysis of the data showed no electron density consistent with the incorporation of Fe(II), αKG, $(GPP)_{10}$ or ascorbate into the crystals.

	Se peak	Se inflection	Se remote
Data Collection			
Space Group	$P2_1$	$P2_1$	$P2_1$
Wavelength (Å)	0.9793	0.9795	0.912
Resolution (Å)	53.38 – 1.40 (1.44 – 1.40)	53.38 – 1.40 (1.44 – 1.40)	53.38 – 1.40 (1.44 – 1.40)
Unit-cell parameters (Å, °)	a = 41.38, b = 63.80, c = 98.50, β = 98.74	a = 41.38, b = 63.80, c = 98.50, β = 98.73	a = 41.38, b = 63.80, c = 98.50, β = 98.80
No. of measurements	374798 (27211)	374659 (27152)	376690 (27404)
No. of unique reflections	99723 (7385)	99692 (7369)	99677 (7368)
Redundancy	3.8 (3.7)	3.8 (3.7)	3.8 (3.7)
Completeness (%)	100 (100)	100 (100)	100 (100)
Rmerge (%)	4.4 (21.9)	5.0 (22.2)	4.1 (18.8)
Average I/σ(I)	16.7 (4.8)	14.9 (4.3)	18.1 (5.9)
Refinement			
Resolution	29.9 - 1.4		
R (%)	17.9		
Rfree (%)	19.9		
Number of Reflections	94685		
Number of atoms			
Protein	3233		
Waters	383		
Glycerol	16		
Geometry Statistics			
rmsd in bond length (Å)	0.018		
rmsd in bond angle (°)	1.5		
Average B-factor (Å^2)			
Main/side chain	12.5/13.4		
Waters	25.1		
Ligands	21.4		
Ramachandran plot (Procheck)			
Favored (%)	90.5		
Allowed (%)	9.5		

$$R_{merge} = \sum_{hkl} \sum_{i} \left| I_i(hkl) - \langle I(hkl) \rangle \right| / \sum_{hkl} \sum_{i} I_i(hkl)$$

Table 4.2. Data collection, refinement and validation statistics for anthrax-P4H structure determination. Values in parentheses indicate the highest resolution shell.

	Fe(II)/αKG	Fe(II)/αKG/$(GPP)_{10}$	Fe(II)/αKG/ascorbate
	41.34,63.37,98.12	41.21,63.06,98.60	41.38,64.07,98.63
Unit Cell	98.87°	99.03°	98.69°
Completeness	99.4% (99.9%)	95.8 % (92.5 %)	99.8 % (99.8 %)
Multiplicity	2.7 (2.6)	3.7 (3.6)	2.5 (2.5)
Rsym	0.050 (0.318)	0.062 (0.406)	0.069 (0.388)
I/0	14.5 (3.3)	14.0 (2.9)	10.3 (2.5)
Resolution	31.11 - 1.39 (1.43 - 1.39)	25.62 - 1.65 (1.69 -1.65)	32.49 - 1.40 (1.44 - 1.40)

$$R_{merge} = \sum_{hkl} \sum_{i} \left| I_i(hkl) - \langle I(hkl) \rangle \right| / \sum_{hkl} \sum_{i} I_i(hkl)$$

Table 4.3. Data Collection of anthrax-P4H crystals co-crystallized with cofactors and substrates $(GPP)_{10}$ or ascorbate, SSRL. Values in parentheses indicate highest resolution shell.

4.3.4 Dimeric Structure of Anthrax-P4H

Anthrax-P4H crystallizes with two molecules in the asymmetric unit (Figure 4.4 A). Molecule A and B are similar with an RMSD value on Cα atoms of 0.334 Å. Within the dimer, monomer A and monomer B are arranged with intermolecular contacts between β2 and β9 (V) of the neighboring molecule. Residues Gln31-Glu38 from one molecule reacts with Arg171 - Leu164 from the other and visa versa. The symmetry of the structure forms two groups of these interactions. The orientation of the dimer arranges the active sites of each monomer directed toward the exterior of the dimer.

The dimer interface is stabilized through both hydrogen bonding and Van der Waals (VdW) interactions. Eight total hydrogen bonds (4 from each interaction) from the peptide backbone (Figure 4.5 B, magenta lines) stabilize

the dimer over a distance of 2.86 - 2.96 Å. These stabilizing forces generate an extensive network of β-sheets going from one monomer to the next. There are a total of seventeen VdW interactions stabilizing molecule A and B over a distance of 3.5 - 4.1 Å (Figure 4.5 B, orange lines). There are eight total intermolecular VdW interactions stabilizing each interface, and one interaction between the two Phe36 residues interacting with one another at 3.8 Å (Figure 4.5 B, interior view). There are three water molecules present in the dimer interface.

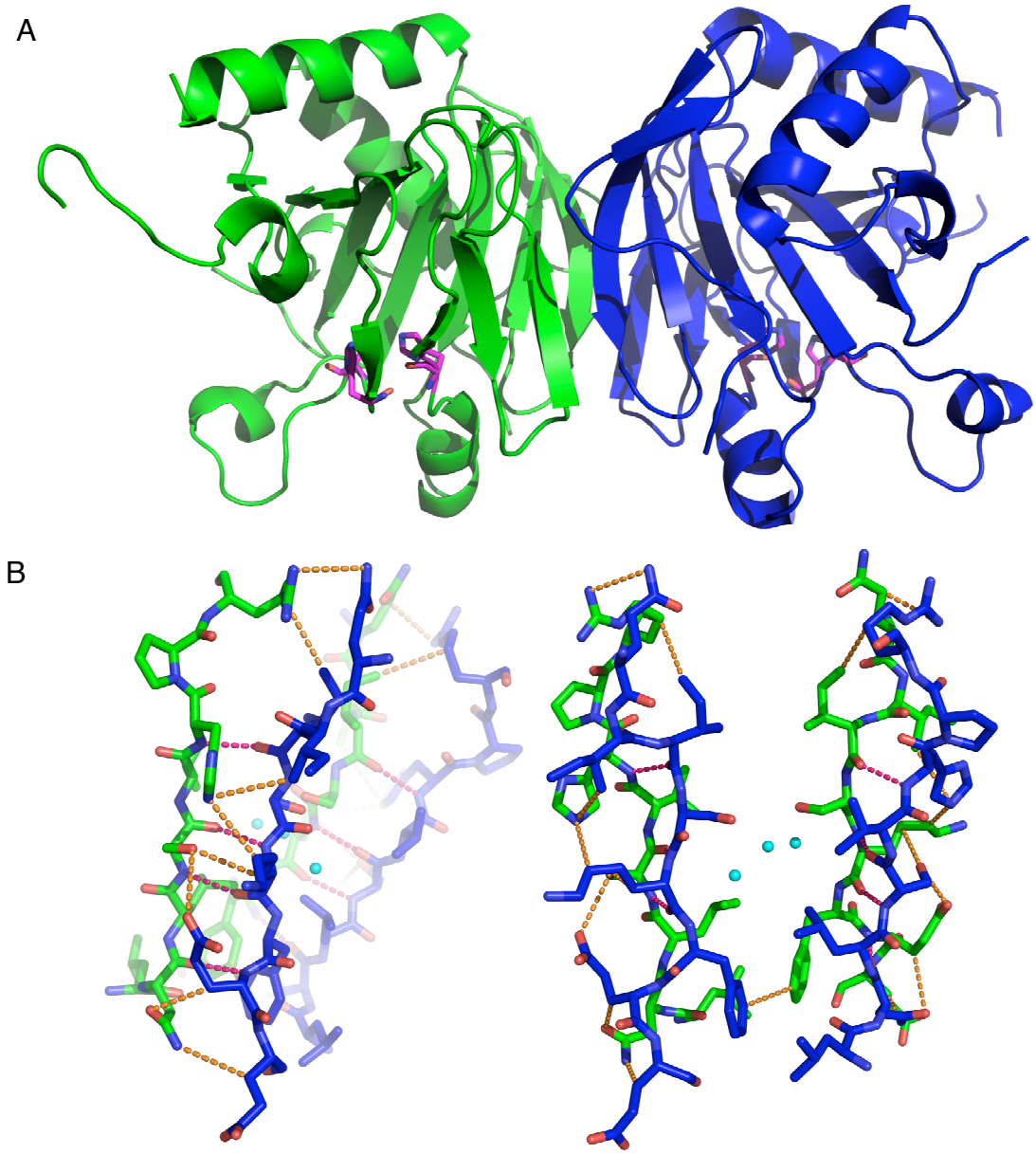


Figure 4.4. Oligomerization of anthrax-P4H. A) Ribbon diagram of asymmetric unit. Each monomer A is displayed in green and monomer B is displayed in blue. Proposed active site residues are displayed in magenta. B) Close up view of the hydrogen bonding (orange lines) and VdW forces (magenta lines) occurring at the dimer: dimer interface. Front view from the face of one interaction and interior view between the two interactions. Three water molecules present in the dimer interface are colored cyan.

4.3.5 Overall Structure of Anthrax-P4H

Anthrax-P4H is composed of 6 α-helices and 12 β-strands per monomer, with eight (β6-β14) of the the β-sheets forming the DSBH motif characteristic with αKG/Fe(II)-oxygenases (Figure 4.5).[24] β-strands located in the DSBH are labeled I-VIII for consistent nomenclature. There are three β-strands (β1 - β3) at the N-terminus that extends the major sheet of the DSBH conformation. The major sheet of the DSBH motif is comprised of β-strands I, VIII, III, VI (Figure 4.5). A small β-strand of two residues (β4) extends the opposite end of the major sheet. The minor sheet of the DSBH motif is made up of β-strands II, VII, IV, V (Figure 4.5). There are three α-helices and three 3_{10} helices located throughout the structure. The N-terminus of anthrax-P4H consists of two α-helices and two 3_{10} helices, that contribute to stabilizing the backside of the major sheet of the DSBH motif (Figure 4.5), as commonly seen among the αKG/Fe(II)-oxygenases.[24] α5 (3_{10} helix) is located on the solvent exposed loop between II/III and most likely stabilizes Asp127 and His129 of the iron binding site. α6 is also in a loop created by VI/VII providing stability for His193 in the iron binding site. Both of these helices give additional stability to the overall core fold of anthrax-P4H.

The N-terminal portions (residues Met1 - Lys11) of both molecule A and B had poor electron density that could not be modeled. In addition, there is one disordered loop region present in both monomers, which yielded low

electron density. This region corresponds to residues Ala65 - Ser71 in molecule A, and Ser67 - Arg73 in molecule B.

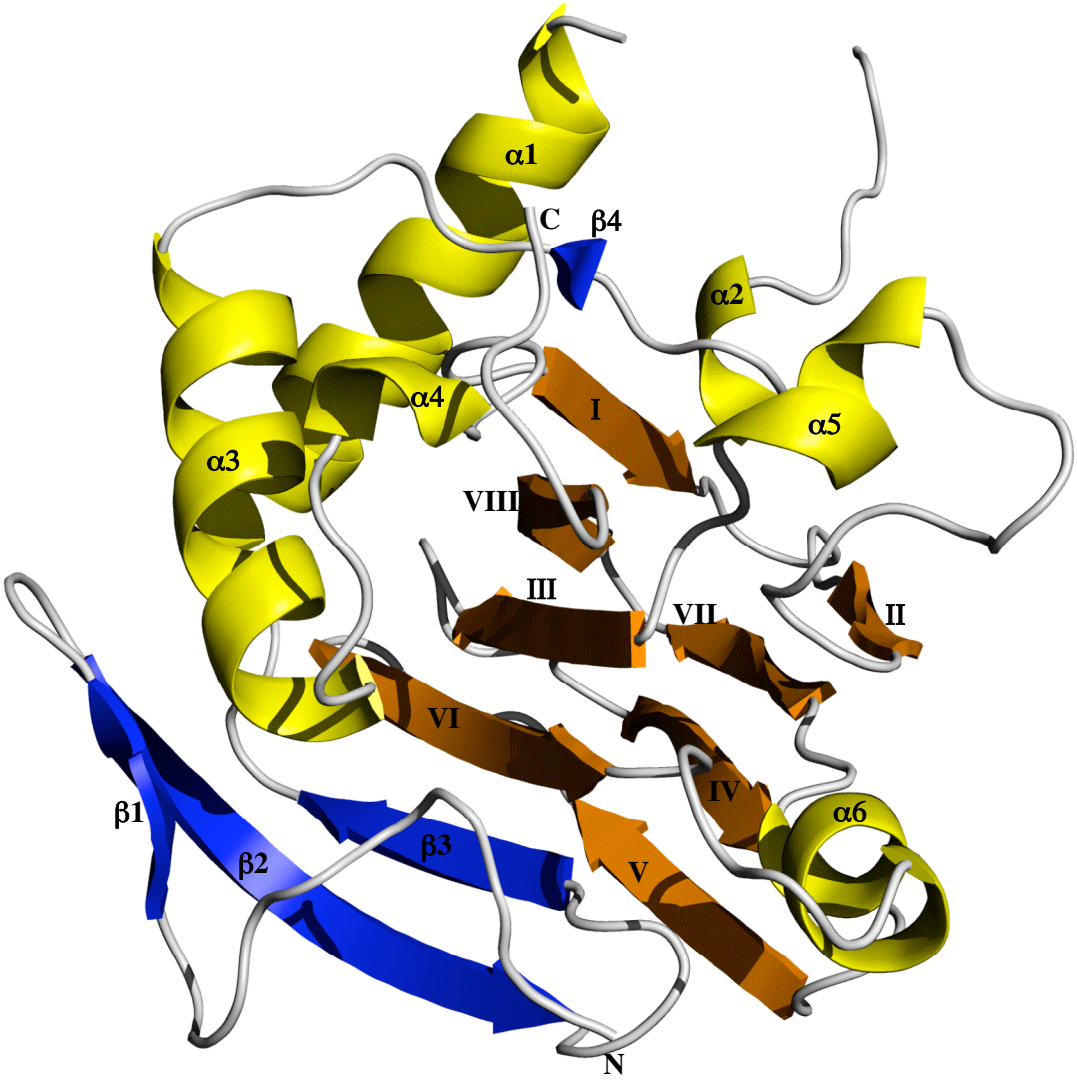


Figure 4.5. Overall view of the crystal structure of anthrax-P4H monomer B. Ribbon representation of anthrax-P4H. α helices are colored yellow, β-strands in the DSBH core fold are colored orange, and additional β-strands are colored blue. The secondary structure is numbered sequentially and Roman numerals indicate the eight β-strands of the DSBH core fold.

4.3.6 Active Site of Anthrax-P4H

The putative Fe(II) binding site is located within the minor β-sheet in an open end of the DSBH core (Figure 4.4) as seen in other αKG/Fe(II)-oxygenases.[24] Figure 4.6 depicts the three residues of the canonical H^1-X-D/E-X_n-H^2 facial triad motif, His127, Asp129, and His193, that comprise the putative Fe(II) binding site. The H^1-X-D of the motif is located at the end of the second strand of the DSBH (II) motif, and the distal histidine (H^2) is located on the seventh strand of the DSBH (VII) motif. Electron density consistent with a glycerol molecule is detected at this site in both molecule A and B (Figure 4.6). The Fe(II) binding site is conserved throughout the members of αKG/Fe(II)-oxygenases.[24] Asp129 forms further hydrogen bonds with Thr145 (2.7 Å) of βIII, Trp209 (3.1 Å) of βVIII, and glycerol (2.7 Å) for greater stability (Figure 4.7).

In each monomer the proposed αKG binding site contains a water molecule (A:Wat2, B:Wat592). Comparison with structures of other αKG/Fe(II)-oxygenases indicates that in anthrax-P4H, binding of the 5-carboxylate of αKG to the Fe (II) atom complex involves side-chains Tyr118 (3.0 Å), Tyr124 (3.1 Å), Thr159 (3.0 Å), and Lys203 (3.1 Å) (Figure 4.7). Lys203 is a conserved basic residue corresponding to Lys493 in human type I C-P4H, which through site-directed mutagenesis studies has been shown to be essential in binding αKG.[40] Residue Arg211, which is proposed to be

involved in αKG binding stability, is not located in the active site (Figure 4.7). This residue corresponds to His501 in human-P4H-1 and is proposed to undergo hydrogen bonding to one of the oxygens in the oxygen-bridged Fe(IV)-peroxo intermediate, assist in cleavage and the formation of the Fe(IV)=oxo intermediate.[25,40]

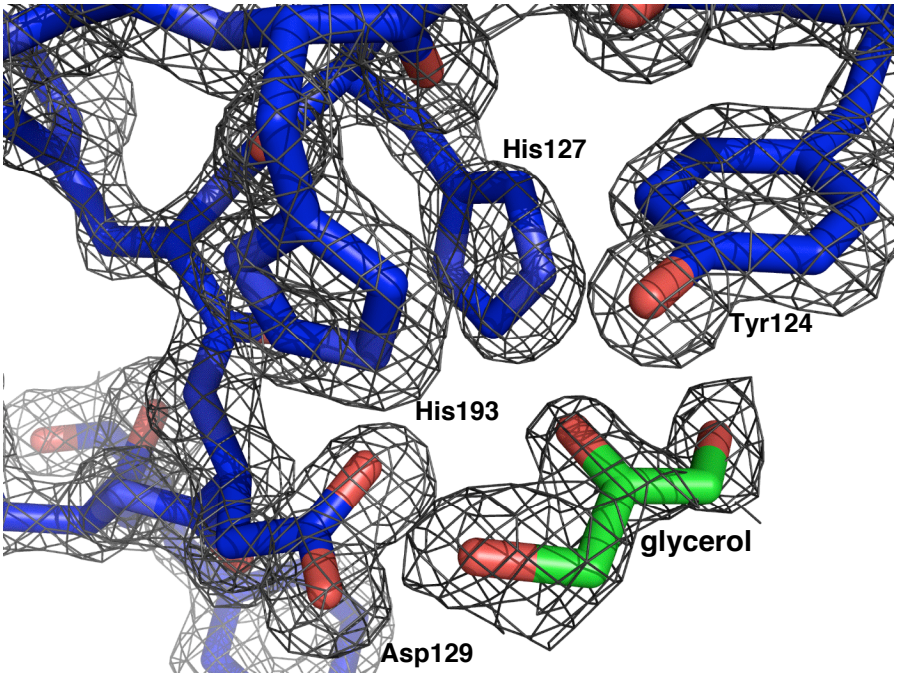


Figure 4.6. The 2Fo-Fc electron density map, contoured to 1σ, of canonical H^1-X-D/E-X_n-H^2 facial triad motif of anthrax-P4H. Electron density consistent with a glycerol molecule (green) occupies the proposed Fe(II) binding site.

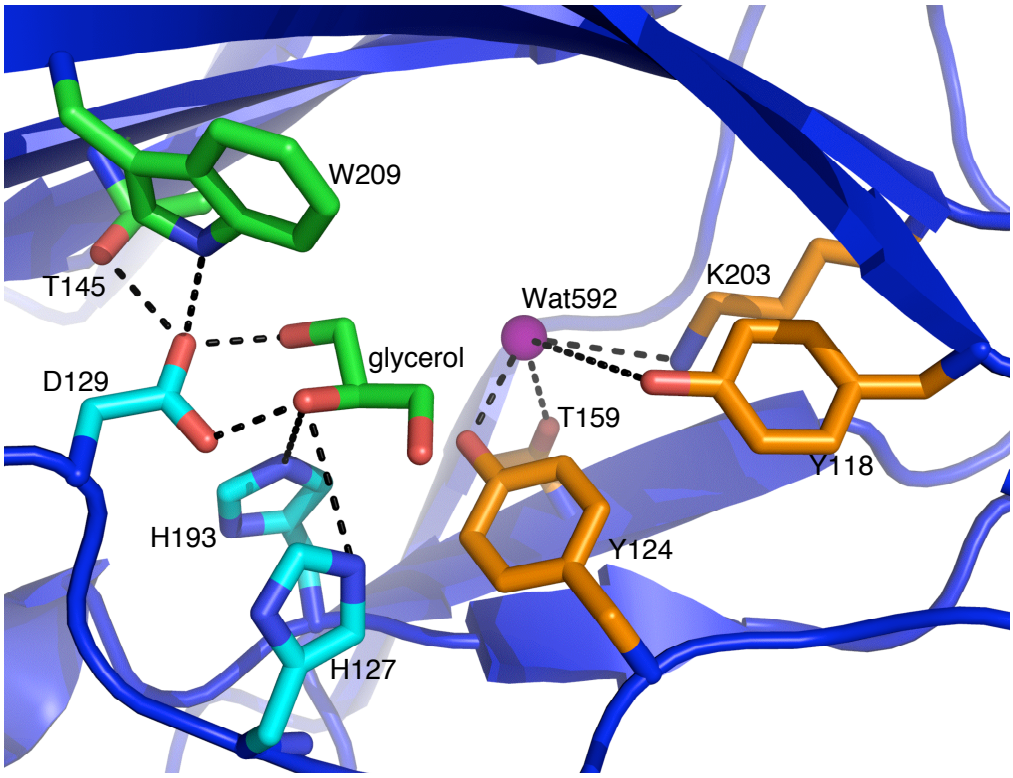


Figure 4.7. Proposed active site of anthrax-P4H. Residues His127, Asp129, and His193 involved in Fe(II) binding are displayed in cyan. Residues Thr145, Trp209, and the glycerol molecule are involved in a hydrogen-bonding network responsible for stabilizing Asp129 and are shown in green. Glycerol occupies the proposed Fe(II) binding site. A water molecule (Wat592) proposed in the putative αKG binding site is labeled and colored magenta. Residues Tyr118, Tyr124, Thr159, and Lys203 proposed in the binding of the 5-carboxylate moiety of αKG are shown in orange.

4.3.7 Self-Hydroxylation of Active Site Residues Leads to Inactivation of the Enzyme

In the absence of ascorbate, the catalytic cycle of anthrax-P4H leads to inactivation coupled to the formation of a blue chromophore with λ_{max} at ~ 600 nm (Figure 3.13) as discussed in Chapter 3. In members of this superfamily of enzymes including TauD and AlkB, it has been shown that the inactivation is due to the self-hydroxylation of a Phe, Tyr, or Trp residue in the active site, where the blue chromophore is assigned as a charge transfer complex of hydroxylated residue to Fe(III).[132,139,140] The self-hydroxylation occurs concomitantly with decarboxylation of αKG and the resulting enzyme is inactivated. Based on the active site structure of anthrax-P4H, we set out to identify potential residues that may be hydroxylated through this mechanism.

In AlkB, Trp178 has been shown to undergo self-hydroxylation by UV-vis spectroscopy and ESI-MS.[164] By superimposing the active site structure of anthrax-P4H with that of AlkB (PDB: 2FD8), Phe178 (anthrax-P4H) in the active site is the most likely candidate to undergo self-hydroxylation (Figure 4.8 A). We also performed structural overlays of anthrax-P4H with TauD (PDB: 1OS7) as shown in Figure 4.8 B. Tyr73 in TauD has been shown to undergo hydroxylation by LC-MS/MS and Resonance raman spectroscopy.[139] However, Tyr73 in TauD does not align well with Trp203 in the active site of anthrax-P4H. We therefore decided to prepare mutations of residue Phe178 to determine whether this residue undergoes self-hydroxylation.

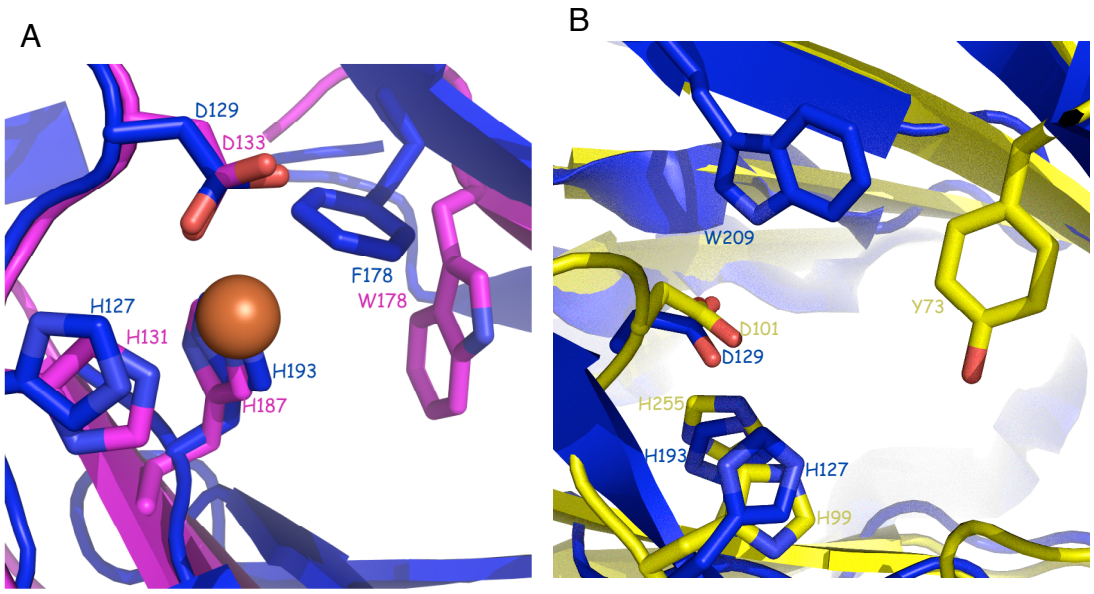


Figure 4.8. Superimposition of anthrax-P4H (blue) active site structure with αKG/Fe(II)-oxygenases, which undergo self-hydroxylation in response to enzyme inactivation. A) Trp178 in AlkB (PDB:2FD8) (magenta) is the residue responsible for self-hydroxylation during inactivation. The structures overlay well and we have proposed F178 in anthrax-P4H undergoes self-hydroxylation. The bound iron present in overlay is from the AlkB structure and is colored orange. B) Tyr73 undergoes hydroxylation in TauD (PDB:1OS7) (yellow), but does not overlay with Trp209 of anthrax-P4H.

F178Y and F178A were expressed and purified in the same fashion as WT. Both mutants showed lower O_2 consumption when compared to WT. F178Y had 54 % of the activity of WT and F178A had 10 % of the activity of WT. F178A was found to be not stable and precipitation was observed after several days at 4 °C. Thus this enzyme was characterized immediately after protein purification.

The extent of self-hydroxylation was examined by UV-vis spectroscopy. The αKG/Fe(II)/anthrax-P4H ternary complex was prepared under anaerobic conditions as described previously (Chapter 3). The

absorbance of the αKG/Fe(II)/enzyme ternary complex was the same for both mutants compared with WT λ_{max} = 510 nm. However, when O_2 was introduced, the mutants behaved differently from WT (Figure 4.9). The λ_{max} of F178Y is 530 nm, or 70 nm blue-shifted compared to WT. In contrast F178A showed no spectral change after being exposed to air (data not shown due to precipitation).

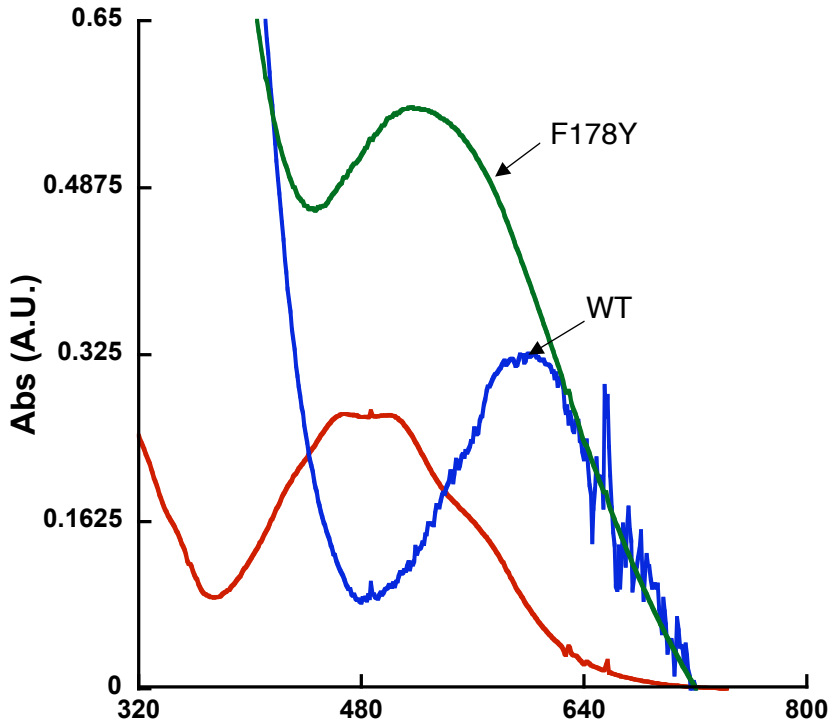


Figure 4.9. A UV-vis spectrum of the Fe(II)/αKG/anthrax-P4H WT ternary complex under anaerobic conditions (red line). The final concentrations are 1mM $Fe(NH_4)_2(SO_4)_2$, 1mM αKG and 1 mM enzyme (WT, F178Y or F178A) in 50 mM Tris buffer (pH 7.4). All samples showed the same spectra, characteristic of the ternary complex. Exposure to air results in a difference between the WT and mutants. WT (blue) displays λ_{max} = 615 nm, whereas F178Y (green) has λ_{max} = 530 nm. F178A showed no spectral change when exposed to air.

4.3.8 Putative Peptide Binding Groove

The human-P4H isoenzymes are the only P4H proteins that contain the PBD, which is the major domain distinguishing each of the forms.[165] The binding mode of a collagen-like peptide to C-P4H has not been defined. This may be due to the complexity of the polypeptide substrates that P4Hs use compared with small molecule substrates of other members in αKG/Fe(II)-oxygenases.

The structure of Cr-P4H-1 complexed with Zn(II) and the αKG analog 2',4'-dicarboxylate pyridine has been solved and revealed an extended shallow groove lined by two flexible loops proposed in substrate binding. Point mutation of residues in the groove (Arg93, Ser95, Glu127, Tyr140, Arg161 and His245 in Cr-P4H-1 numbering) to Ala resulted in complete loss of activity. Therefore these residues are proposed to serve in collagen-like peptide recognition/binding.[90] The corresponding residues are strictly conserved in anthrax-P4H (Arg79, Ser81, Glu111, Tyr124, Arg142, anthrax-P4H numbering). In anthrax-P4H a shallow groove is formed from Phe85 of β5 to Y124 of βII, which extends across the face of the molecule. Figure 4.10 illustrates the shape of the groove and the residues in the groove. The groove is centered near the active site at residue Trp209. Trp209 is stabilized on the solvent side by a salt bridge made up of Glu111/Arg142. Above Trp209, residues shape the groove from I (Y118 for example) and II (Y124 for

example). Below Trp209, Arg211 (VIII) further stabilizes the salt bridge formed by Glu111/Arg142.

The base of the groove is formed by Ser81 and Phe131 and extends to large loops creating a distinct valley in which a large substrate such as $(GlyProPro)_{10}$ can bind. The groove is shaped by two distinct "horns" that form either side of the valley. The top of one horn is made up of Asp74, which is located close to the unmodeled portion of the structure after α1. Arg136, which is located on α5, occupies the top of the other horn. These "horns" lend shape to a total valley that is 25.5 Å wide and 27.7 Å deep.

The assignment of the peptide binding groove is putative and based off of the results of the Cr-P4H-1 structure.[90] However, many of the residues are located relatively close to the active site and may be necessary for catalysis but not for substrate binding. Further studies such as structure determination of the collagen-like peptide bound form of anthrax-P4H are necessary to evaluate the assignment of the peptide-binding groove.

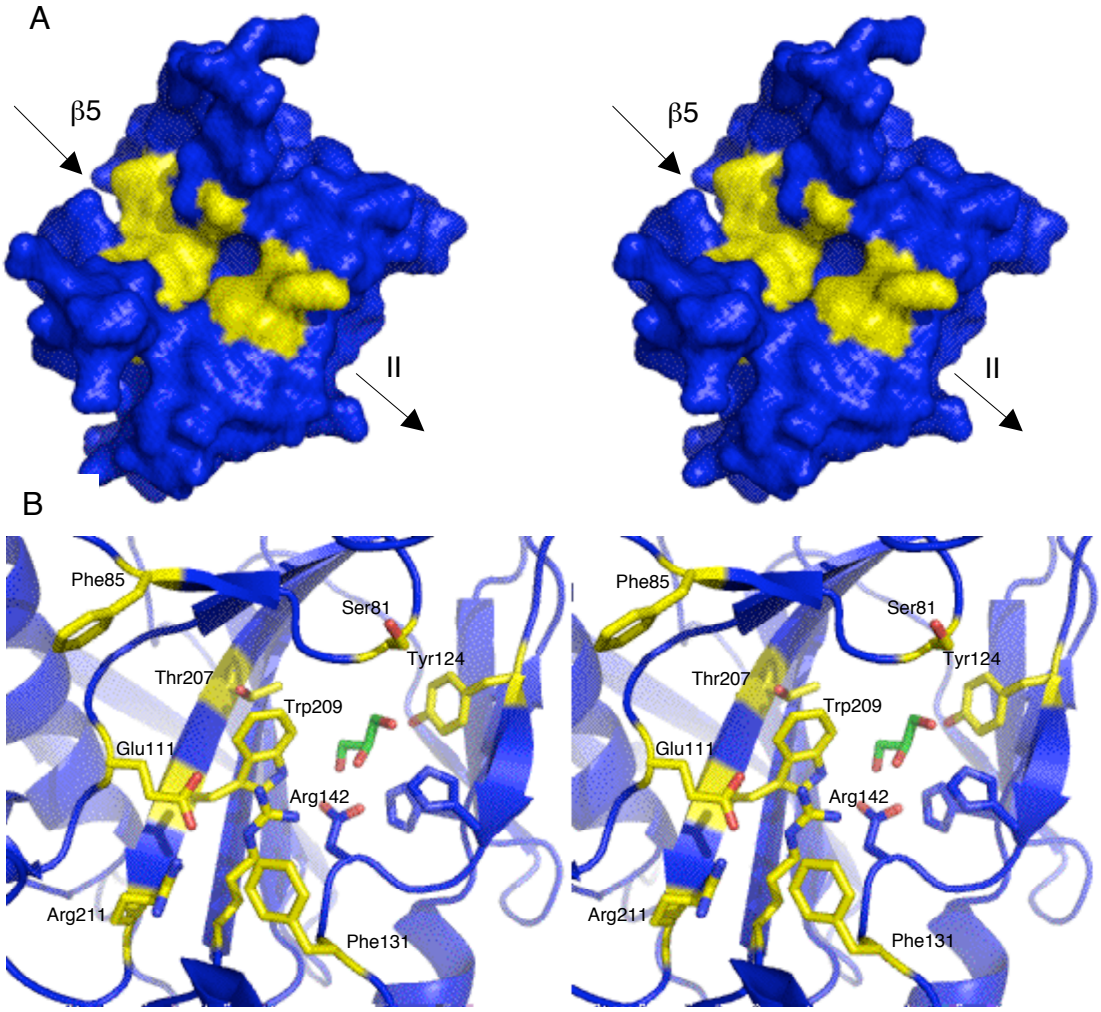


Figure 4.10. Stereoview diagram of putative collagen-like peptide binding groove of anthrax-P4H. Extends from β5 - II with Trp209 at the center of the groove near the active site. A) Surface representation of residues in the groove colored yellow. B) Representation of the residues in the groove labeled and colored yellow.

4.4 Discussion

4.4.1 SeMet Incorporation and Crystallization

A protein crystal structure is solved using both direct measurements such as intensity and position (h,k,l) of the diffracted rays and by indirect measurements such as the phase of the structure. Both intensity and phase give information about the atom arrangement in the crystal lattice. The phase provides information on the electron density wave nature of the unit cell. The direct phase information is lost in the data collection process and must be determined by indirect methods. If a structure of a homologous protein is known then molecular replacement can be used. This method translates and rotates the unknown structure to obtain the phases from an already solved crystal structure.[166] However, this only works when the proteins are structurally homologous. Otherwise alternate methods are required.

The structure of anthrax-P4H was determined using MAD methods. Molecular replacement could not be used at the time because no other collagen-P4H crystal structures were solved. Since then the structure of an Cr-P4H-1 has been solved (PDB: 2JIG, 2JIJ, and 2V4A). However, attempts to solve anthrax-P4H using the data set of the Cr-P4H-1 have been unsuccessful.

MAD uses the properties of anomalous scatters to indicate phases. Anomalous scatterers are heavy atoms, which absorb X-rays at a wavelength

near the element's absorption edge.[166] These atoms absorb energy at different wavelengths compared with the native protein and the difference can be used to determine the phase. MAD data collection must be performed at a synchrotron radiation source as data must be collected at 3 different wavelengths.[166]

SeMet was used as the anomalous scatterer for anthrax-P4H MAD phasing experiment. SeMet was easily incorporated into the protein with the use of auxotrophic *E. coli* and substituting SeMet in the growing media for Se. We were successfully able to incorporate 3 Se atoms out of the 4 present in anthrax-P4H sequence as determined from MS analysis. The native and SeMet anthrax-P4H crystals were not isomorphous, meaning they did not share the same unit cell parameters or space group. Therefore, we used only the MAD data set to solve the structure of anthrax-P4H. The high-resolution data set was collected at the SSRL and the structure was solved to 1.4 Å with R/R_{free} values of 17.9/19.9.

The crystallization of anthrax-P4H was performed with hanging-drop diffusion method. Three-dimensional crystals were obtained after a one time screening with Wizard III kit from Emerald Biosystems. The crystals were birefringent, single, and large enough to collect data. However, attempts to remake the solution did not produce the same crystal morphology. A systematic screening of each of the solutions revealed that a change in the

pH of the precipitant solution between the original batch of solution and subsequent batches. Lowering pH of the mother liquor solution to 4.0 resulted in diffraction quality, highly reproducible crystal formation.

4.4.2 Oligomerization

Members of collagen P4Hs exist in various oligomeric forms, which have been determined experimentally. Studies on human-P4H-1 have shown the enzyme to form an $\alpha_2\beta_2$ heterotetramer. Plant and viral forms of C-P4H are monomeric. Nematode forms C-P4H from *C. elegans* have been seen as heterodimers experimentally.[36]

Little is known about the α_2 dimer interaction of human-P4H-1 as there is no crystal structure for the catalytic subunit. Through site-directed mutagenesis studies, the α_2 dimer has been shown to be stabilized by two intrachain disulfide bonds (C486-C511 and C276-C293).[122,123] Though these disulfide bonds exist between the α subunits, they are also required for tetramer assembly.[123] There may be other interactions at work stabilizing the α_2 dimer in human-P4H-1. In addition to the two covalent bonds, human-P4H-1 α_2 subunit may contain a network of β-sheets stabilized with VdW forces and backbone hydrogen bonding, similar to anthrax-P4H. Site-directed mutagenesis experiments can aid in identifying the residues responsible for the key stability of the dimer of human-P4H-1.

Members of the αKG/Fe(II)-oxygenase family of enzymes exist in a wide range of oligomeric forms. Human phytanoyl-CoA 2-hydroxylase[167] (PAHX) crystallizes as a monomer whereas 1-aminocyclopropane-1-carboxylic acid oxidase (ACCO) from *P. hybrida* crystallizes as a tetramer.[168] Among some enzymes in the αKG/Fe(II) oxygenases, the orientation of the C-terminus can play a role in the oligomerization and crystallization. In DAOCS, the C-terminus peptide inhibited the co-crystallization with substrate by forming a lid over the active site cavity.[141] The mutation of the last four residues was shown to abolish activity and is believed to be involved in substrate recognition/binding.[141] In ACCO[168] and factor Inhibiting HIF-1[169] (FIH), the quaternary structure is formed by C-terminus peptide locking of the monomers. However, other enzymes such as alkylsulfatase (AtsK)[170] and proline 3-hydroxylase[171] (P3H) have quaternary structures that are stabilized by other forces such as hydrophobic or electrostatic interactions respectively. The latter description applies to anthrax-P4H where the C-terminal peptide has no role in the dimerization.

4.4.3 Overall Structure

The crystal structure of isopenicillin N synthase (IPNS) was the first structure to be solved in the family of non heme Fe(II)-oxygenases, as IPNS does not use αKG as a cofactor.[172] Though Fe(II)/αKG-dependent oxygenases catalyze wide range of reactions, a characteristic double

stranded β-helix core fold was discovered and conserved in all the structures determined so far.[24,90,173] We have determined the crystal structure of anthrax-P4H in its apo form to 1.4 Å. The closest structural homolog of anthrax-P4H is the SeMet structure of a monomeric Cr-P4H-1 (PDB: 2V4A), with a Z-score of 22.6 using the DALI server. These structures have a RMSD of 1.68 Å for their Cα chains. Comparison with HIF-P4H2 involved in hypoxic response yielded a score of 13.6 and RMSD of 2.53 Å.

Anthrax-P4H forms a β-sandwich consisting of 4 anti-parallel β-sheets, also known as the jellyroll core fold motif. The sandwich is joined through a "hairpin" loop which connects the fourth (IV) and fifth strands (V) of the DSBH domain.[24] In anthrax-P4H this hairpin loop is a tight turn/loop as seen with some members of αKG/Fe(II)-oxygenases such as DAOCS (Figure 4.5).[141] The Cr-P4H-1 adopts an extended conformation in its' IV-V loop similar to the structure of TauD.[174] The second β-strand in the DSBH (II) motif of anthrax-P4H is short, but in an ordered conformation, unlike some oxygenases where it is missing altogether (Figure 4.5).[167] Therefore all eight β-strands of the DSBH motif are seen in this structure, and stabilized through internal hydrophobic interactions.[24]

Anthrax-P4H and Cr-P4H-1 share similar structural characteristics (Figure 4.11). These similarities include the DSBH motif, β3 at the N-terminus, α1, α2, and 3_{10} helix (2) also at the N-terminus, and α3 located in

between β10 (V) and β11 (VI) in the DSBH. Both structures have no secondary structure after β12(VIII) of DSBH, resulting in similar C-terminal ends.

There are distinct structural differences between anthrax-P4H and Cr-P4H-1 as well (Figure 4.11). Anthrax-P4H has an additional β-strand (β1) at its' N-terminus, and β2 is considerably longer than with the algal structure. This may be due to insufficient electron density in the algal structure at the N-terminus, as the first 28 residues in the sequence are not modeled in. In addition, there are 2 additional β-strands (β4-β5 algal numbering) in the algal ternary structure that corresponds to an extended loop and missing electron density in the anthrax-P4H structure (Ala65 - Ser71 in molecule B and Ser67-Arg73 in molecule B). Instead anthrax-P4H forms a 3_{10} helix not present in the algal form. This short helix causes β4 to be shorter in the anthrax-P4H structure compared to the algal structure. Loop β8-β9 (IV-V) has a long loop conformation with a 3_{10} helix present in Cr-P4H-1 forming a large loop insertion composed of 20 residues compared with the anthrax-P4H structure, which contains a short, tight turn of 5 residues. This is the loop that connects the major and minor sheets of the DSBH. In some αKG/Fe(II) oxygenases this loop is long and flexible as seen with Cr-P4H-1 and in others in exhibits a tight hairpin turn as seen in anthrax-P4H.[24]

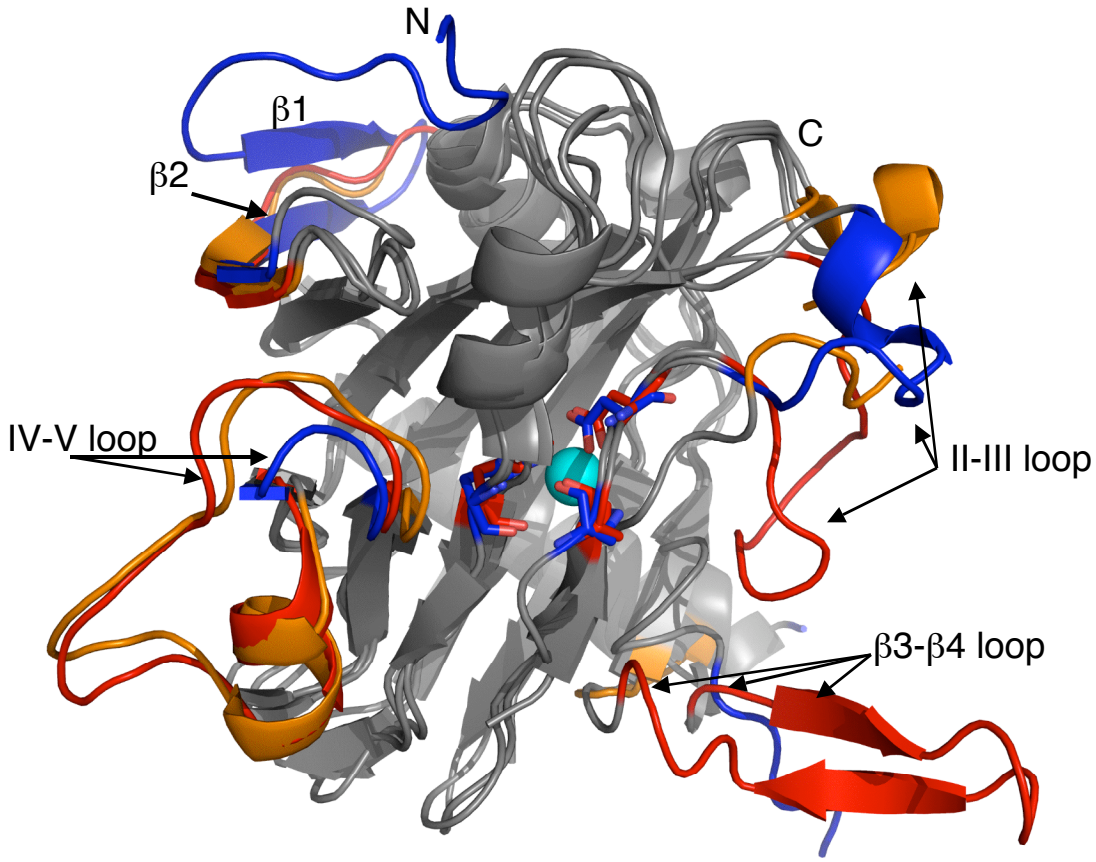


Figure 4.11. Structural overlay of anthrax-P4H with Cr-P4H-1 SeMet(2V4A) and ternary (2JIG) structures. The portion of the st ructure with high similarity is colored grey. The flexible loops displaying key difference in the structures are colored in blue for anthrax-P4H, orange for SeMet algal-P4H (2V4A) and red for ternary Cr-P4H-1 (2JIG). The cyan sphere represents a Zn atom at th e active site of the ternary structure of Cr-P4H-1. The structural overlay was performed in COOT.

The ternary form of Cr-P4H-1 structure (2JIG) shows high plasticity at loop II-III that undergoes a 19 Å conformational change with binding of Zn(II) and an αKG analog, which causes the loop to reside in a "closed" conformation.[90] However comparison of the anthrax-P4H structure with the SeMet Cr-P4H-1 structure (2V4A) revealed a similar conformation of this loop between the two structures, and the presence of an additional 3_{10} helix that is

not present in the ternary structure (2JIG) (Figure 4.11). Both Cr-P4H-1 and anthrax-P4H structures are present in the "open" conformation, where the central cavity is exposed to the solvent. The loop in anthrax-P4H is 4 residues shorter, which may result in less flexibility and conformational change upon binding of the cofactors. Crystals of the Fe(II) and αKG bound form are currently being screened. Except for the differences in these flexible loop regions, the rest of the structures are conserved.

4.4.4 The Active Site Structure of Anthrax-P4H

The active site of αKG/Fe(II)-oxygenases is conserved among members of the family of enzymes and is located on the minor β-sheet of the DSBH motif.[24] One of the roles of the DSBH motif is to provide a rigid scaffold for the Fe(II) binding site. Therefore, we can compare the active site of anthrax-P4H with crystal structures containing a metal and an αKG analog to examine putative interactions at the anthrax-P4H active site.

The spatial arrangement of the active site residues of αKG/Fe(II)-oxygenases is not affected by binding of Fe(II) or 2',4'-dicarboxylate pyridine, and therefore the amino acid residues align nicely with those of the Cr-P4H-1 ternary structure (Figure 4.12). Both the H-X-D/E-X_n-H facial triad and the Lys residue responsible for stabilizing the C5 position of the αKG are conserved. The Tyr124 in anthrax -P4H and corresponding Tyr140 in Cr-P4H-1, do show different orientations. In the algal structure this residue

undergoes a conformational change upon binding of 2',4'-dicaroxylate pyridine.[90] In the ligand free and Zn(II) only binary (yellow residues) algal structures Tyr140 is flipped toward the bulk solvent of the protein, away from the active site, while the algal structure including both Zn(II) and 2',4'-dicaroxylate pyridine (red residues) shows this residue stacking with His143 of the facial triad (Figure 4.12).[90] In the anthrax-P4H structure (blue residues), Tyr124 also points toward the active site and potentially stabilizes the bound αKG by hydrogen bonding with its C1 carboxyl group. This Tyr residue is conserved throughout all P4Hs and may contribute to the differences observed in K_m for αKG for anthrax-P4H and Cr-P4H-1.

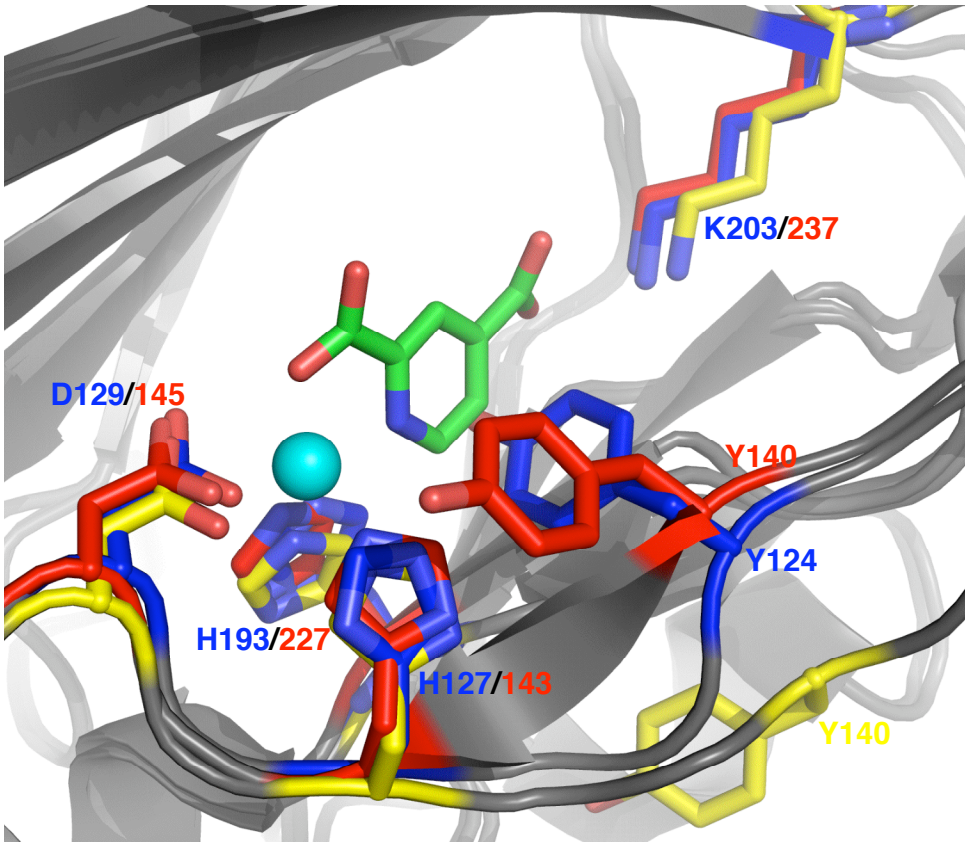


Figure 4.12. Structural overlay of the active sites of anthrax-P4H (blue), binary complex of Cr-P4H-1 (yellow) (2JIJ molecule B), and ternary complex of Cr-P4H-1 (red) (2JIG molecule A). The active site coordination does not change the facial triad or Lys residue identified as binding the C5 position of αKG. Tyr124 (anthrax) and Tyr140 (algal) are spatially different between all three structures and in anthrax-P4H may interact directly with the C1 carboxy of αKG. The numbering of the Cr-P4H-1 binary and ternary structures are the same and displayed in red for simplicity, except for Tyr140 where it is labeled in both red and yellow. The Zn(II) atom (cyan sphere) and 2',4'-dicarboxylate pyridine (green) are present from the ternary Cr-P4H-1 structure.

4.4.5 Self-Hydroxylation of Residues Leads to Enzyme Inactivation

We have shown through UV-vis spectroscopic study of F178Y and F178A that Phe178 likely undergoes self-hydroxylation to inactivate anthrax-P4H. In the WT, Phe178 is most likely hydroxylated to form a tyrosinate-to-Fe(III) MLCT band as seen in 4-hydroxyphenylpyruvate dioxygenase (HPPD).[140] The F178Y mutation blue-shifted the spectra compared with WT (~ 600 nm) to a λ_{max} of ~ 530 nm (Figure 4.9). TauD has a similar spectra in the presence of Fe(II) and αKG with a λ_{max} of 550 nm.[175] Tyr73 has been assigned as the residue that undergoes self-hydroxylation to form a catechol. Similarly, we propose that a catechol is forming in F178Y. The F178A mutant showed no spectral change upon exposure of the Fe(II)/αKG/anthrax-P4H ternary product to air. The lack of any spectral feature supports the suggestion that F178 undergoes the self-hydroxylation in the active site.

Contrary evidence to the results obtained in anthrax-P4H is seen in top down mass spectrometry experiments performed on viral-P4H in which four sites of oxidation were detected.[176] This technique does not require proteolytic digestion of the protein and uses intact protein for its analysis. They were not able to identify exact residues, but only areas in the sequence undergoing oxidation. MS/MS data identified the four areas of oxidation as R28 - V31 (22 %), E95 - F107 (28 %), K216 (34 %), and V125 - D150 (~4 %).[176] These samples were examined in the presence of excess substrates

and the residues identified do not correlate well with previous results obtained from other members of the αKG/Fe(II)-oxygenases.[177] Two of the four sites contain aromatic residues (F107 and Y138), but four sites of oxidation have not been observed in previous reports. Therefore, the mechanism of self-hydroxylation in anthrax-P4H needs to be further investigated.

To determine if Phe178 is the residue undergoing self-hydroxylation, the system can be examined with traditional proteolytic digest followed by LC/MS. Comparison of the samples with and without ascorbate will assist in identifying the residue responsible for inactivation, as ascorbate protects the enzyme from inactivation. The crystal structure of the inactivated form of anthrax-P4H could also be determined to identify the residue that undergoes self-hydroxylation.

4.4.6 Peptide-Binding Domain of Anthrax-P4H

In addition to proline hydroxylation involved in hypoxia, HIFα undergoes hydroxylation of an asparagine residue (Asn801) in the C-terminal activation domain (CAD) of HIFα.[178] This enzyme was identified as HIF asparagine hydroxylase (FIH) and provides a secondary oxygen-dependent mechanism whereby HIFα transcription is prevented.[178] FIH is a member of the αKG/Fe(II)-oxygenase family and binds to the polypeptide CAD domain of HIFα. The polypeptide nature of the CAD substrate for FIH is similar to the collagen-P4H substrate of collagen. The crystal structure in the presence of

Fe(II), αKG, and CAD polypeptide fragments has been determined to 2.25 Å (Figure 4.13).[179] Two binding sites were determined in the structure. Site 1 is a groove that adopts an extended conformation containing 10 hydrogen bonds.[179] Site 2 is on the surface of FIH and is stabilized by 2 hydrogen bonds and appears to have weaker binding as determined from kinetic analysis.[179] The polypeptide had to be at least a 20-mer in order to obtain a

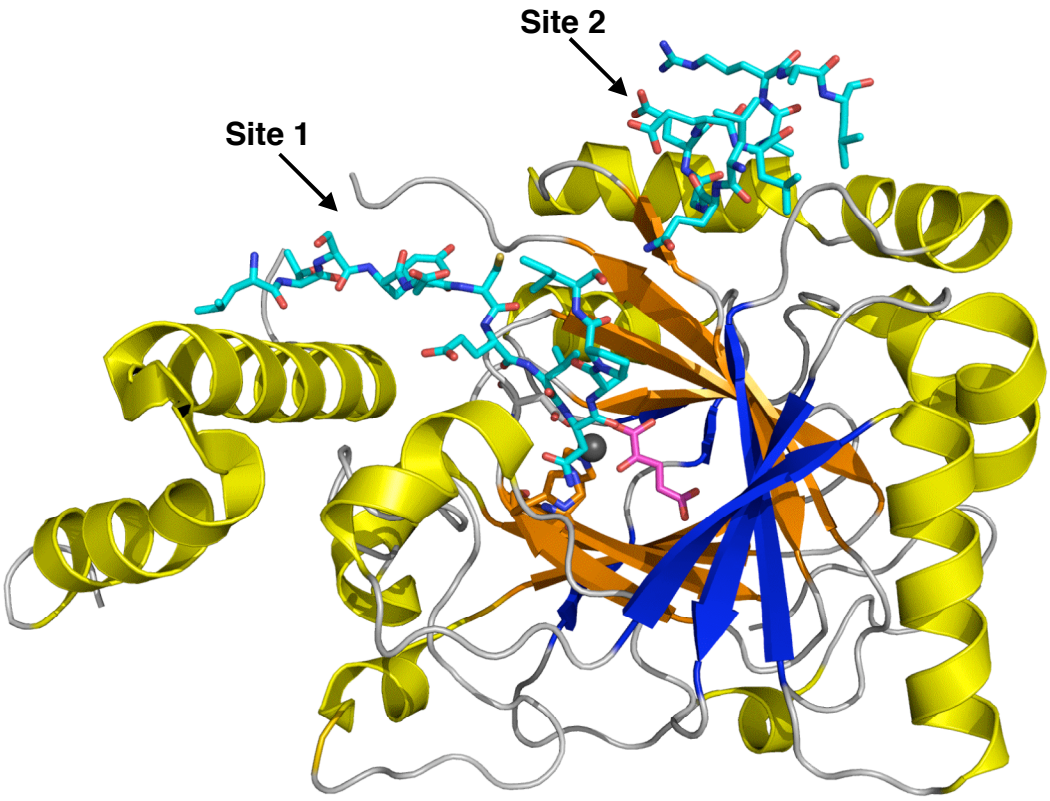


Figure 4.13. Monomeric structure of FIH (PDB: 1H2L). The α helices are colored yellow, β-strands in the DSBH core fold are colored orange, and additional β-strands are colored blue. The CAD peptide fragments are shown in stick representation in cyan, the αKG is shown in s tick representation in magenta, and the Fe(II) atom is represented as a grey sphere. The two sites of substrate binding are labeled.

co-crystal structure. Attempts to crystallize FIH with smaller substrates were unsuccessful. The structure revealed an extended loop conformation of the CAD fragments at site 1 as compared with the α-helical conformation these same residues adopt when binding to the 1st transcriptional adaptor zinc-binding domain (TAZ1).[180,181] The conformational flexibility of the CAD fragments in FIH reveal an induced fit binding process.[179]

Currently, there is no crystal structure of human-P4H-1 solved mainly due to the poor solubility of the α subunit. The available expression systems for recombinant human-P4H-1 are also low yielding (~ 2 mg/L) and do not produce pure protein.[88,89] The crystal structure of the human-P4H-1 PBD made up of residues Gly138 - Ser244 has been determined to 2.3 Å (PDB: 1TJC), and is composed of 5 antiparallel α-helices (Figure 4.14).[117] The PBD belongs to a tetratricopeptide repeat domains involved in protein-protein interactions,[144,145] and contains two TRP repeat domains and a solvating helix. A groove containing eight tyrosine residues is the most unique feature of the structure and site-directed mutagenesis studies have identified three tyrosine residues as highly important for peptide binding.[117] It is believed that the peptide substrate binds in the deep groove in the concave surface of the domain. The alanine scanning of all eight tyrosine residues suggested that these residues have no role in the tetramer formation.[117] However, no further

analysis has been performed to determine if additional residues in the α subunit are also involved in substrate binding.

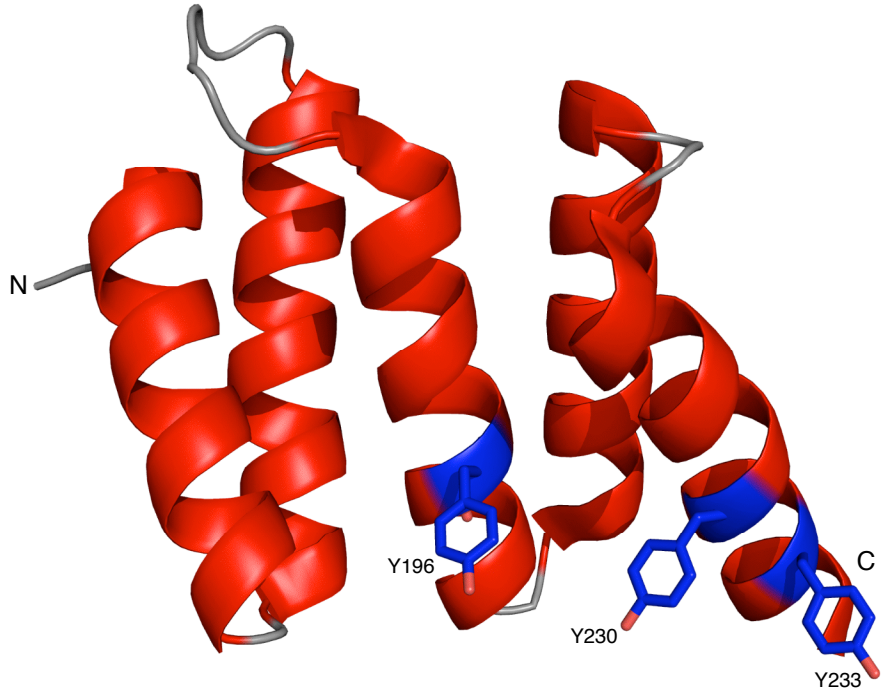


Figure 4.14 Crystal structure of one TRP domain of human-P4H-1 peptide-binding domain (PBD) (PDB: 1TJC). The 5 antiparallel α-helices are colored red, and the three tyrosine residues proposed to be involved in substrate recognition are displayed as sticks and colored blue.

A crystal structure of the collagen-binding domain (151-318) of the *Staphylococcus aureus* adhesin protein (Cna) has been determined to 2 Å (PDB: 1AMX), where the collagen binding interface of the domain is built along a groove on a concave β-sheet and has significant structural and chemical complimentarily to the triple helix of collagen.[103] The Cna binding of collagen can be used as a general model for collagen binding, and may provide insight on the interactions of the PBD and active site of human-P4H-

1. A subsequent structural study of Cna apo-protein (31-344, N1 and N2 subdomains) and Cna complexed with a synthetic collagen-like peptide revealed that the peptide penetrates through a spherical hole formed by the two subdomains and the N1-N2 linker.[182] A "collagen hug" model has been proposed for the binding of collagen to Cna in which a long linker is present to wrap the collagen toward the binding site (Figure 4.15).[182] It is possible that a similar event may be occurring in human-P4H-1 where the PBD is acting as the "linker" to draw the collagen substrate toward the binding site closer to the

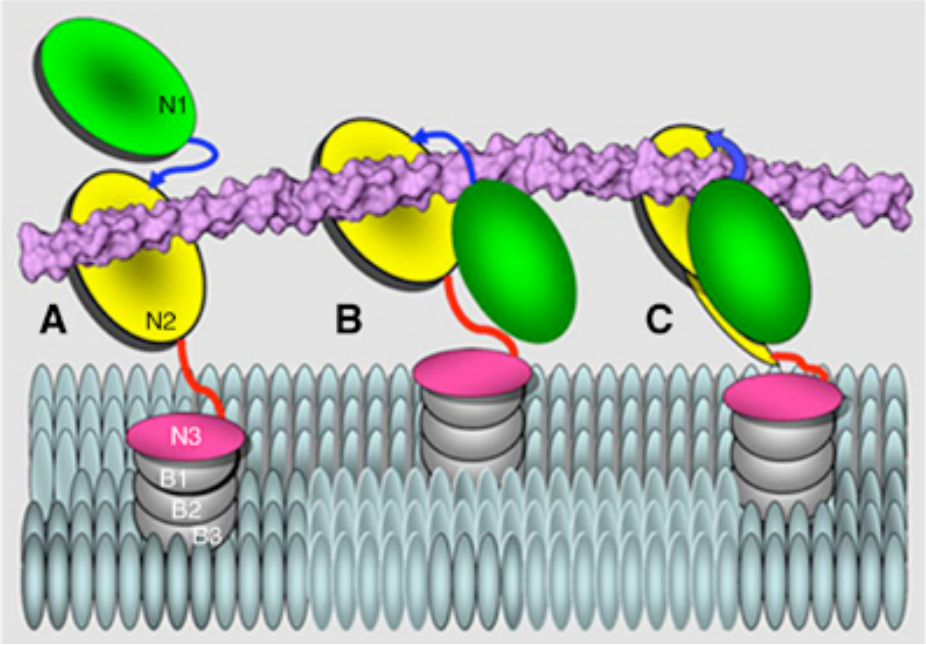


Figure 4.15 Proposed "Collagen Hug" model shown in cartoon representation. A) The collagen triple helix associates with the groove in the N2 domain (yellow). B) The collagen is wrapped by the linker (blue) between the N1 and N2 domains and the N1 domain (green). C) The N1 domain interacts with the N2 domain through hydrophobic interactions and the C-terminal latch (red) interacts with the N1 domain to secure the collagen triple helix in place. Figure was reprinted with permission from Macmillan Publishers Ltd: the Embo Journal, Zong et.al. 2005.[180]

active site. This would locate the PBD spatially close to the groove and function concertedly with the groove to bind procollagen. It should be noted that there is no sequence or structural similarity between Cna and procollagen-P4H.

The structure of Cr-P4H-1 reveals a groove proposed to be essential in substrate binding.[90] Comparison of the algal and anthrax-P4H structures show that the grooves are very similar as shown by Figure 4.16 A and B. Both grooves extend from the β5 strand to the II strand close to the active site. Anthrax-P4H has a much more prominent groove that forms a deep valley of ~ 28 Å flanked by residues Asp74 and Arg136 (Figure 4.16 C). The algal structure forms a much more shallow valley from Lys102 to Glu156, which is 11.40 Å wide and 6.9 Å at its deepest and 4.45 Å at its most shallow (Figure 4.16 D). The deep valley in anthrax-P4H may play a greater role in recognizing collagen-like substrates and help to guide the substrate toward the active site. We have just begun to conduct docking studies and co-crystallization experiments to address these questions.

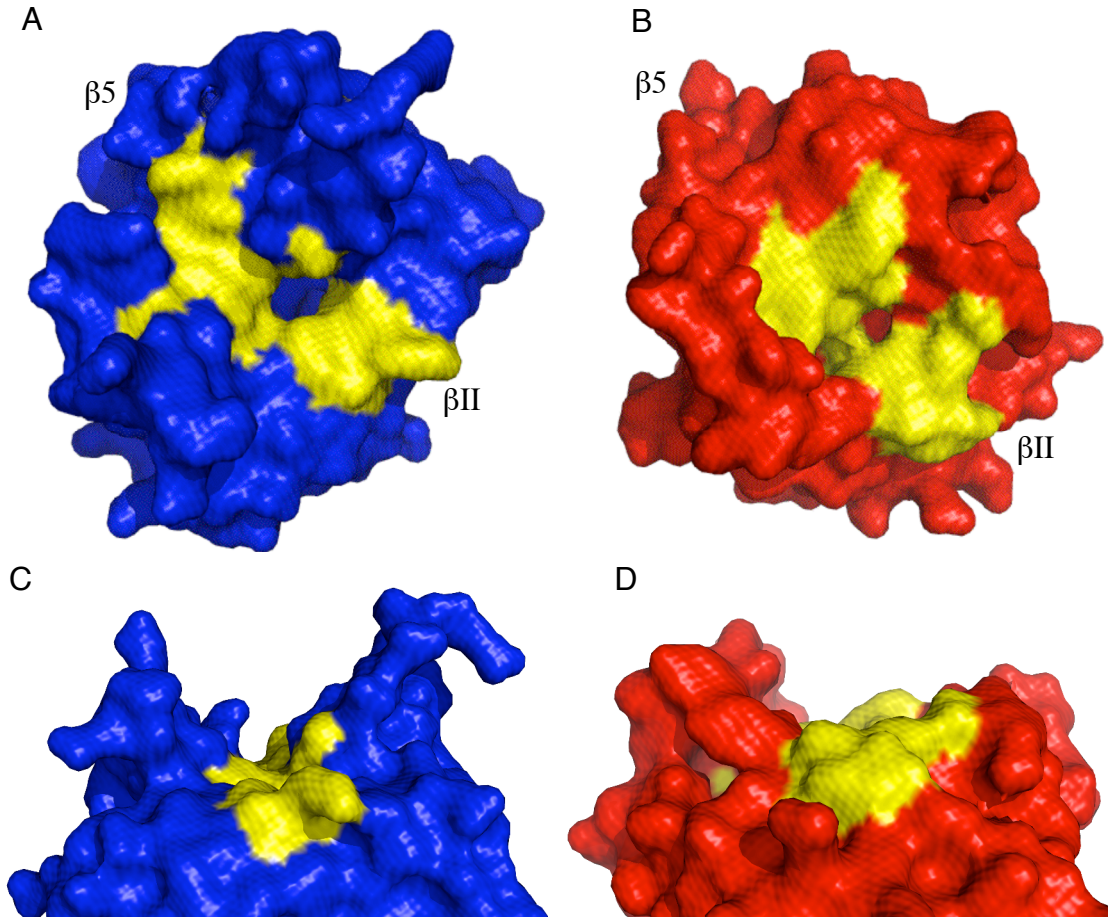


Figure 4.16 Comparison of the putative peptide-substrate binding groove of A) anthrax-P4H (blue) and B) Cr-P4H-1 (red). The putative peptide binding groove is identified by yellow in both structures and extends from β5 (Phe85/Trp99) to βII (Tyr124/Tyr140) (anthrax/algal). Comparison of valley formed around the peptide binding groove for C) anthrax-P4H (blue) and D) Cr-P4H-1 (red).

4.5 Conclusions and Future Directions

We have determined the crystal structure of apo anthrax-P4H, which is the first identified bacterial P4H, to 1.4 Å. MAD phasing was used to determine the crystal structure with SeMet substituted anthrax-P4H. Anthrax-P4H crystallized with 2 molecules per asymmetric unit. This is the first putative homodimeric P4H that has been identified. The dimer is stabilized through hydrogen bonding of the peptide backbone and van der Waals interactions in the dimer interface.

Anthrax-P4H contains the jellyroll core fold motif common among αKG/Fe(II)-oxygenases. It contains all 8 β-strands, which make up a major and minor β-sheet. The structure of anthrax-P4H contains fewer flexible loops when compared with the structures of the other known C-P4H, Cr-P4H-1.

The anthrax-P4H active site contains the characteristic H-X-D/E-X_n-H facial triad and Lys residue responsible for binding the C5 position of the αKG as defined from site-directed mutagenesis studies on human-P4H-1. Tyr124 is also located in the active site and may play a role in αKG binding. Its orientation in anthrax-P4H differs from that in all of the algal structures and may form a direct interaction with αKG.

We have suggested that Phe178 may be responsible for enzyme inactivation due to self-hydroxylation mechanism. F178Y and F178A were

purified and showed distinct UV-vis spectral changes upon exposure to O_2 compared to the WT enzyme. Crystallization and MS studies should be used to confirm this theory.

Anthrax-P4H and Cr-P4H-1 both have putative peptide-binding grooves. This was surprising as the substrate specificity of the two differs greatly with collagen substrate, $(GlyProPro)_{10}$. We aim to determine the difference in substrate specificity between anthrax-P4H and Cr-P4H-1. The proposed peptide-binding groove in anthrax-P4H consists of a deeper valley, which may help with collagen-like substrate recognition. The structure of anthrax-P4H can provide insights in the catalytic properties as well as dimerization of a subunits of human-P4H-1 and aide in the development of more selective therapeutics to inhibit fibrosis. The co-crystal structure of anthrax-P4H and $(GlyProPro)_{10}$ can provide essential information in this goal and is under current examination.

Chapter 5: Additional P4H Systems for Biochemical Characterization

5.1 Introduction

In order to gain further insight into the mechanism of P4H catalysis and their substrate specificity, the recombinant forms of two viral P4Hs were prepared in addition to anthrax-P4H. The genomes of both PBCV-1 and MT325 contain putative *p4h* genes. PBCV-1 and MT325 are both large, icosahedral, plaque-forming viruses that infect and replicate in differing unicellular green algae, Chlorella.[183] Members of the Chlorella viruses are genetically diverse but similar in morphology and infect hosts in both marine and fresh water. PBCV-1 infects NC64A Chlorella, which resides in fresh water isolated from the United States, Brazil, Argentina, Australia, Israel, and Italy.[184] Pbi Chlorella, which is infected by the MT325 virus, was initially isolated in fresh water from Europe, but has been found in regions of higher altitudes in Australia, Canada, and northern United States.[184] The viruses from these Chlorella do not affect one another.

PBCV-1 is the prototypic virus currently used in chlorovirus research to study gene expression, proteomics, and virus-host relationship.[183] The genome of PBCV-1 contains a 621 bp putative *p4h* gene. A recombinant form of PBCV-P4H was expressed in *E. coli* and purified to homogeneity by Eriksson and coworkers.[73] The biological function and native substrate of PBCV-P4H is currently unknown. Proline-rich peptides deduced from the

PBCV-1 genome were screened as potential substrates in the P4H activity assay. Among those tested, $(Pro\text{-}Ala\text{-}Pro\text{-}Lys)_{10}$ had a K_m of 20 μM and showed the highest activity among the proline-rich peptides tested.[73] This substrate, and the others tested including $(Ser\text{-}Pro\text{-}Lys\text{-}Pro\text{-}Pro)_5$ and $(Pro\text{-}Glu\text{-}Pro\text{-}Pro\text{-}Ala)_5$ are distinct from both animal ($(Gly\text{-}Pro\text{-}Pro)_n$ of collagen) and plant ($(Ala\text{-}Thr\text{-}Pro\text{-}Pro\text{-}Pro\text{-}Val)_3$) substrates. The activity of PBCV-P4H was assayed with both $(Gly\text{-}Pro\text{-}Pro)_{10}$ and poly(L-proline) substrates and K_m values of 2900 μM and 500 μM were determined, respectively.[73] A recent spectroscopic study of PBCV-P4H identified one of the proposed mechanistic intermediates, the high spin Fe(IV)=O complex, as discussed in Chapter 3.[124]

Another *p4h* gene was cloned from the genome of virus MT325. MT325 has the smallest chlorella viral genome sequenced to date and has not been extensively studied.[185] Characterizing MT325-P4H has not been previously performed and can add further insight into the role of the P4H enzyme in lower order organisms.

To develop potential models for human-P4H-1, chimeric-P4Hs were examined (Figure 5.1). The chimeras were designed to contain both the human-P4H-1 PBD and the catalytic domain of PBCV-P4H in attempt to increase the solubility from that observed for the catalytic core of human-P4H-1.

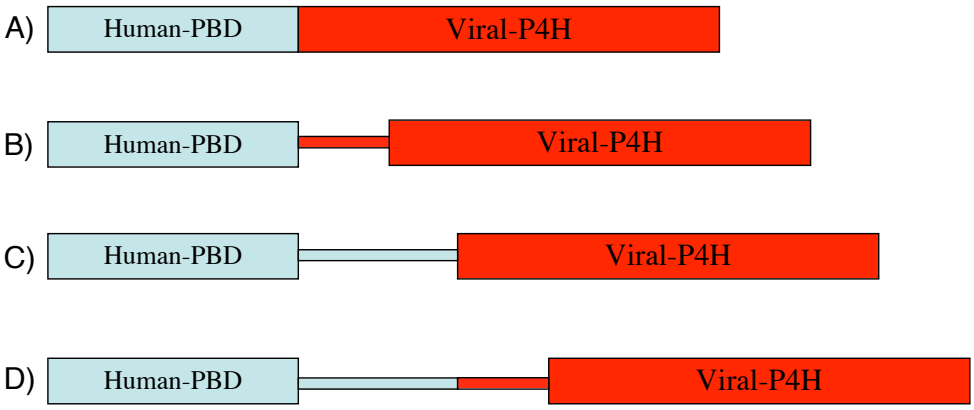


Figure 5.1 Schematic depiction of the four kinds of chimera proteins. Blue blocks represent human-PBD domain and human-linker peptide. Red blocks represent catalytic domain of viral-P4 and viral-linker peptide.

5.2 Materials and Methods

5.2.1 Cloning of Truncated and Full-Length Forms of *pbcv-p4h* Gene

Initial cloning of a 621 bp *pbcv-p4h* gene was reproduced from previously published methods by Erikkson et al.[73] The genomic PBCV DNA (GenBank accession number; U42580) was obtained from Dr. Van Etten at the University of Nebraska. The gene of interest was amplified using the PCR primers 5'-CGCGCATATGGAGGGGTTTGAAACCAGCGAT-3' and 5'-CGCGCTCGAGTCATTTAACAGCACGGATCCATT-3', incorporating NdeI and XhoI restriction sites, respectively (underlined). The resulting PCR product was digested and ligated into NdeI/XhoI digested pET15b (Novagen) expression vector, in frame with the 5' end His-tag coding sequence. This PCR product codes for a 106 amino acid protein corresponding to residues

Glu36-Lys242 without the putative signal peptide at the N-terminus.[73] The DNA of the purified plasmid was sequenced in the forward and reverse direction at the DNA sequencing facility of University of California-Berkeley.

A full-length 684 bp *pbcv-p4h* gene was also amplified from genomic DNA (GenBank™ accession number U42580) by PCR using primers 5'-CAT GCCATGGAAACCCGAGAATACTTGTTG-3' and 5'-CCCTCGAGTTAGATC CATTGGTTCGCGATG-3' containing NcoI and XhoI restriction enzyme sites, respectively (underlined). The PCR product was first ligated into pGEMT-Easy (Promega) cloning vector. The purified plasmid was sequenced in the forward and reverse direction, and then subcloned into NcoI/XhoI digested pET15b.

5.2.2 Cloning of Viral *mt325-p4h* Gene

A 672 bp open reading frame (ORF) encoding for the *mt325-p4h* gene was amplified from genomic DNA (GenBank™ accession number DQ491001) using the PCR primers 5'-CATGCCATGGATGGAAAATCCGAGTATATTC-3' and 5'-CCCTCGAGTTAAATAGCTCGAATCCATTG-3' containing NcoI and XhoI restriction enzyme sites, respectively (underlined). The *mt325* genomic DNA was obtained from Dr. Van Etten at the University of Nebraska. The resulting purified PCR product was ligated into a pGEMT-Easy cloning vector and sequenced in the forward and reverse direction. The sequence verified

plasmid was digested with NcoI and XhoI and the purified insert was ligated into the digested pET-15b.

5.2.3 Expression of Truncated PBCV-P4H and MT325-P4H

The truncated form of PBCV-P4H (Glu36-Lys242) without the N-terminal putative signal peptide was expressed as previously reported by Eriksson, et.al.[73] Briefly, purified expression plasmid of the truncated gene was transformed into BL21(DE)3 (Invitrogen) *E. coli*. Fifteen mL of LB-amp starter cultures were used to inoculate 2.8 L shaker flasks containing 1.5 L LB-amp media. Cells were grown at 37 ºC until the OD_{600} reached 0.6 at which time the temperature was reduced to 28 ºC for 30 minutes. The protein expression was induced by addition of IPTG to a final concentration of 800 mM. After induction, cells were incubated at 28 ºC for 3 hours, harvested by centrifugation at 8,000 rpm at 4 ºC for 20 minutes, and stored frozen at - 80 ºC until ready for use.

Cells were thawed on ice, resuspended in 60 mM imidazole, 0.5 M NaCl and 20 mM Tris, pH 7.9 containing protease inhibitors (1 mM PMSF, 1 mM leupeptin, 1 mM antipain, 1 mM pepstatin), lysozyme, and DNAse and RNAse. Cells were disrupted by ultrasonication and the lysate was centrifuged at 20,000 rpm at 4 ºC for 1 hour.

Recombinant truncated PBCV-P4H was purified with HisTrap HP (GE Biosciences) Ni^{+2} Sepharose column as described previously in Chapter 3.

The fractions were analyzed with SDS-PAGE and those containing pure protein were pooled, concentrated using a 10,000 MWCO membrane in an Amicon stirred cell concentrator (Millipore) and desalted with a HiTrap desalting column (GE Biosciences) to remove imidazole. These partially purified fractions were purified on Superdex 16/60 Size Exclusion column fitted to an FPLC (Akta). Truncated PBCV-P4H was eluted with 50 mM KPi, 150 mM KCl, pH 7.0. Fractions containing the truncated PBCV-P4H (> 99% purity) by SDS-PAGE were combined, concentrated to ~ 40 mg/mL, made 20% v/v with glycerol and stored frozen at - 80˚C. Protein concentration was determined with Bradford Assay (Pierce) using BSA as the protein standard. A yield of 3.5 mg of protein was isolated per liter of culture.

MT325-P4H expression plasmid was transformed into BL21(DE3) *E. coli* cells for expression. Pilot expression of MT325-P4H was performed in 500 mL LB-amp media. Cells were grown at 37 ºC until the OD_{600} reached 0.6, and the protein expression was induced by addition of IPTG to a final concentration of 200 μM. After induction, cells were maintained at 37 ºC for 3 hours and then treated with Bug Buster (Novagen) containing Benzoase Nuclease for expression analysis. Soluble and insoluble samples were separated and subjected to SDS-PAGE analysis.

5.2.4 Cloning of *chimera-p4h* Constructs

5.2.4.1 Chimera1: *human-1-pbd* with Truncated *pbcv-p4h*

The NcoI/NdeI digest of 321 bp PCR product-encoding human-1-PDB was then ligated into the truncated *pbcv*/pET-15b plasmid between restriction sites NcoI and NdeI. This construct has no linker nucleotides between the *human-1-pdb* and *pbcv-p4h* gene (Figure 5.1 A).

5.2.4.2 Chimera 2: *human-p4h-1 pbd* with Full-Length *pbcv-p4h*

Initially, the 321 bp gene encoding the PBD (Gly154-Ser260 residues) of human-P4H-1 was cloned from the vector containing the full-length $\alpha_2\beta_2$ *human-p4h-1* (prepared by Dr. Hirakawa in the Limburg Lab). PCR primers 5'-CATGCCATGGGAGTGAAACACAAATC-3' and 5'-GGAATTCCATATGAG ACTTATTGACATC-3' were used to clone out the *human-1-pdb*. The PCR product was digested with NcoI/NdeI and the purified insert was ligated into the full-length *pbcv-p4h*/pET-15b clone between restriction sites NcoI and NdeI. This construct has a linker of 64 bp nucleotide derived from *pbcv-p4h* between the *human-1-pdb* and *pbcv-p4h* gene (Figure 5.1 B).

5.2.4.3 Chimera 3: *human-1-pbd* with Human-Linker Sequence and Truncated *pbcv-p4h*

Cloning of a 561 bp gene fragment encoding the PDB domain (Gly154-His340) and extra nucleotide linker from the *human-p4h*-1 sequence was performed using PCR primers 5'-CATGCCATGGGAGTGAAACACAAATC-3' and 5'-GGAATTCCATATGATGGAAGCGAATAATA-3'. The PCR product

was digested with NcoI and NdeI, and the purified product was inserted into the truncated *pbcv*/pET-15b clone between restriction sites NcoI and NdeI. This construct contains a linker of 81 bp nucleotides derived from human-*p4h*-1 gene between the *human-1-pdb* and *pbcv-p4h* gene (Figure 5.1 C).

5.2.4.4 Chimera 4: *human-1-pdb* with Human-Linker Sequence and Full-Length *pbcv-p4h*

The 561 bp gene fragment containing *human-1-pdb* and human-linker sequence described above was inserted at NcoI and NdeI into the full-length pbcv/pET15b construct. This chimera contains both the 81 bp linker originated from human-*p4h-1* and the 64 bp linker originated from *pbcv-p4h* to anneal the human-*pdb*-1 and *pbcv-p4h* genes to achieve maximum separation between the domains (Figure 5.1 D).

5.2.5 Expression of Four Chimera-P4H Proteins

Plasmids for expression of chimeras 1 and 2 were transformed into BL21(DE3) *E. coli* cells and initial pilot expressions were tested. Single colonies of each chimera were grown in LB-amp media to an OD_{600} of 0.6 at which time the protein expression was induced by addition of IPTG to a final concentration of 200 μM. 1 mL aliquot of cell culture was taken every hour post induction, centrifuged, supernatant decanted, and cell pellets were stored frozen at - 20 ºC. Lysis of the cells was performed by standard Bug Buster protocol with Benzonase Nuclease. The soluble fraction was analyzed by SDS-PAGE.

No protein expression was seen with the BL21(DE3) *E. coli* cells therefore all chimera constructs were transformed into *E. coli* cell lines BL21(DE3)*plyseS* and BL21Origami(DE3) (Invitrogen). The BL21(DE3)plyseS cells provide a tighter control of protein expression by expressing T7 lysozyme, which reduces the basal expression of the protein of interest. BL21Origami(DE3) cells contain mutation in both the thioredoxin reductase and glutathione reductase genes to enhance disulfide bond formation. Cells were grown, the protein expression was induced by addition of IPTG and the cell culture was collected every hour as described for the protein expression using the BL21(DE3) cells. Cells were lysed using BugBuster and the protein expression was analyzed by SDS-PAGE. Protein production was seen with BL21*plyseS* cells, however only in the insoluble fraction. Cell growth time was extended from 3 hours to overnight and the growth temperature was decreased from 37 ºC to 20 ºC in attempt to obtain soluble protein.

5.3 Results

5.3.1 Cloning and Expression of Truncated and Full-Length *pbcv-p4h*

We have cloned the 621 bp truncated form of *pbcv-p4h* coding for Glu36–Lys242 without the N-terminal secretion signal peptide as reported by Eriksson and coworkers[73] (Figure 5.2 A). The recombinant protein was purified using a Ni^{2+} sepharose affinity column followed by a Superdex sizing column (Figure 5.2 B).

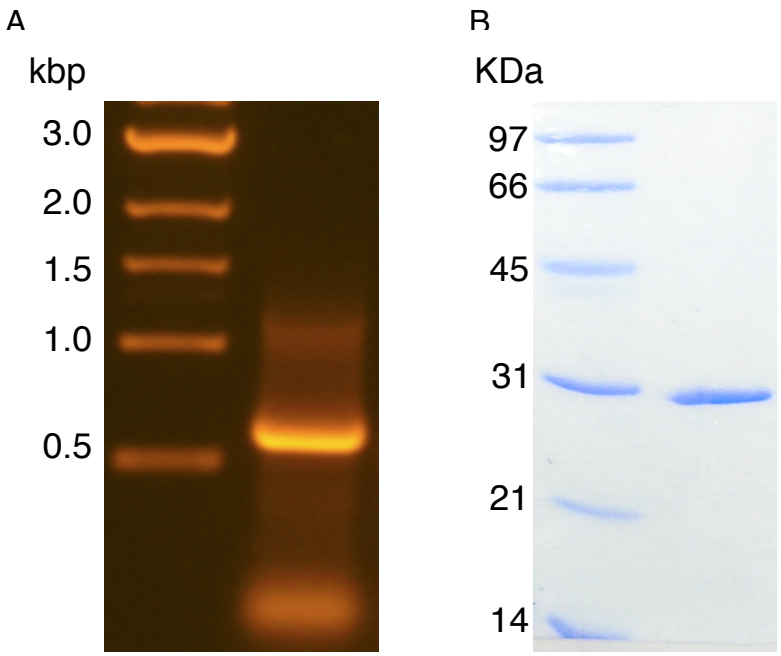


Figure 5.2 Cloning and expression/purification of truncated PBCV-P4H. A) Agarose DNA gel of 621 bp pbcv-p4h gene from PCR. B) Purified PBCV-P4H protein.

Although PBCV-P4H was initially isolated as a soluble protein as reported,[73] the protein was found to be not stable and precipitation was observed upon final concentration and under storage at 4 °C. The protein yield was lower compared to that of anthrax-P4H, 6 mg/L vs. 10mg/L, respectively. No protein yield was reported for the PBCV-P4H. We had previously seen precipitation of a recombinant anthrax-P4H when it was expressed with the N-terminal His-tag. Therefore, we have cloned the 684 bp full-length *pbcv-p4h* in the absence of the His-tag coding sequence. We have not pursued the expression and characterization of protein product yet.

5.3.2 Cloning and Expression of Viral *mt325-p4h*

A BLAST search identified a 672 bp ORF of a putative *p4h* gene in the genome of virus MT325. We PCR-cloned the *mt325-p4h* gene, and the recombinant protein was expressed in BL21(DE3) *E. coli* and found to be soluble (Figure 5.3).

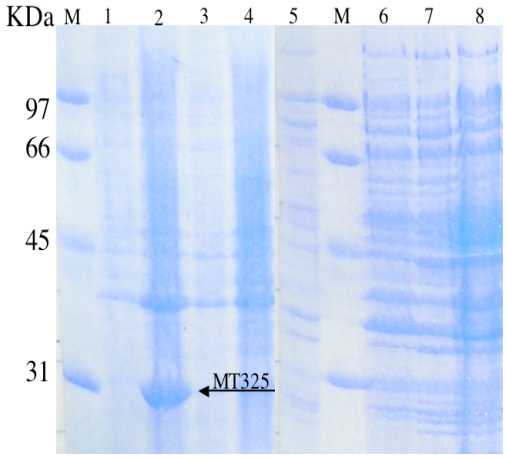


Figure 5.3. Pilot expression of recombinant MT325-P4H. Lane 1) soluble fraction of cell lysate from pre-induction, 2) soluble fraction of cell lysates after 3 hour post-induction, 3) soluble fraction of cell lysate from uninduced cells, 4) soluble fraction of cell lysate uninduced at 3 hour, 5) insoluble fraction of cell lysate from pre-induction, 6) insoluble fraction of cell lysate from 3 hour post-induction, 7) insoluble fraction of cell lysate from uninduced cells, 8 insoluble fraction of cell lysate from uninduced cells at 3 hours uninduced.

5.3.3 Cloning of *human-1-pbd* and *pbcv-p4h* Chimera Constructs

The expression constructs for four chimeric P4Hs containing both the human-*pbd* and gene fragment encoding the catalytic core of PBCV-P4H were prepared. Chimera 1 construct is 924 bp and does not contain a linker

between the *pbd* and *pbcv-p4h* genes (Figure 5.1 A). We used the gene fragment lacking the putative signal peptide coding sequence of *pbcv-p4h* in this clone. This corresponds to the gene fragment encoding the C-terminal end of the α subunit and the P4H catalytic core of human-*p4h*.

We incorporated linker sequences on three of the chimera constructs in attempt to help proper protein folding. Chimera 2 is a 1005 bp gene containing the human-*pbd* and full-length *pbcv-p4h* (Figure 5.1 B). This construct gives a 64 bp linker between the *pbd* and the corresponding gene encoding the catalytic core of *pbcv-p4h*. Chimera 3 contains an 81 bp linker from the human-*p4h*-1 downstream of the *pbd* to the *pbcv-p4h*, making up the 1185 bp gene (Figure 5.1 C). The linker sequence was derived from the sequence between the *pbd* and the gene encoding that catalytic core of the a-subunit of human-*p4h*-1. Chimera 4, a 1245 bp gene, is a blend of Chimera 1 and 3 incorporating both the viral and human linkers that form a 145 bp linker (Figure 5.1 D).

5.3.4. Expression of Human-1-PDB and PBCV-P4H Chimeras

Pilot expression was tried on each of the *chimera-p4h* clones. Screening of the host cells and cell growth and induction conditions were performed. The length for incubation time after induction was screened from 3 hours post induction to overnight, and growth temperatures were varied from 37 ºC to 20 ºC. None of the chimera-P4H proteins were found in the

soluble protein after cell lysis (Figure 5.4 A). Chimera 3 was expressed as an inclusion body in BL21plyseS *E. coli* cells, overnight at 20 °C (Figure 5.4 B). We have not screened the conditions to solubilize the inclusion body yet.

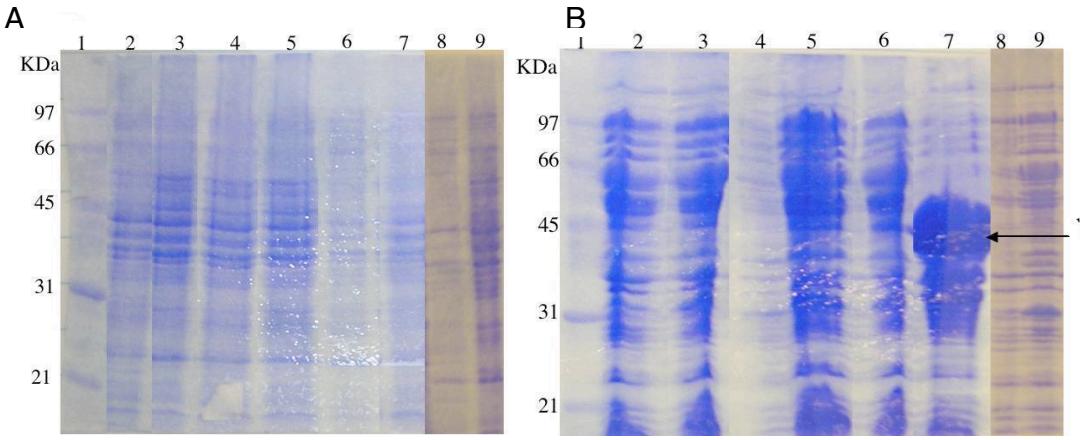


Figure 5.4. Pilot expression of four Chimeric P4H proteins in BL21plyseS *E. coli* cells. A) Soluble fractions of cell lysate: uninduced and induced, Chimera 1 (lanes 2 and 3), Chimera 2 (lanes 4 and 5), Chimera 3 (lanes 6 and 7), Chimera 4 (lanes 8 and 9). B) Insoluble fractions: uninduced and induced, Chimera 1 (lanes 2 and 3), Chimera 2 (lanes 4 and 5), Chimera 3 (lanes 6 and 7) indicated by arrow, Chimera 4 (8,9).

5.4 Discussion

5.4.1 PBCV- and MT325-P4H Model systems

In addition to anthrax-P4H, several other P4H systems have been developed and studied in our laboratory. These include two different viral-P4Hs and four chimera-P4Hs, containing both the human-1 PBD and catalytic core of the monomeric PBCV-P4H.

Further insight into the mechanism of P4H catalysis and substrate specificity can be gained through comparative study on the two viral forms.

The function of P4H in PBCV-1 is currently unknown. Identifying the role of PBCV-P4H may help explain the role P4Hs have structurally *in vivo* in lower organisms and may correlate with its role in anthrax-P4H. Prior to the characterization of PBCV-P4H, it was thought that viruses and bacteria did not produce Hyp. However, this not the case as Hyp has been detected in viruses as well as *B. anthracis* (chapter 2), and collagen-like repeat sequences have been identified in both viral and bacterial genomes. The biological role of collagen-like peptides in chlorella viruses is not well understood. In *B.anthracis*, BclA is not responsible for the toxicity nor the spore formation but it seems to contribute to the hydrophobicity of the surface of the spore, potentially have roles in protecting spore from host cell defense.[98,186,187]

5.4.2 The Role of PBD can be Determined with a Chimera-P4H

The PBD of human-P4H-1 is composed of ~100 residues (Gly138 - Ser244) and is present only in the type (I) human-P4H.[14,127] This domain is distinct structurally from previously identified proline-rich-peptide binding proteins, which mainly consist of β-strands and forms a TRP protein domain discussed in Chapter 4.[188] Research in the Limburg lab has continued searching for the best model system to understand the structure/function relationship of human-P4H-1 in order to develop more selective inhibitors to target fibrosis.

A model system that incorporates the PBD functions of human-P4H to the catalytic core of P4H is desirable as none of the available recombinant forms of P4H contain this domain. However, anthrax-P4H and At-P4H-1 bind $(GPP)_{10}$ with similar affinity as human-P4H-1 (18 μM), 9 μM, and 60 μM, respectively.[37,69] The structure of Cr-P4H-1 identified a shallow groove proposed to be essential for peptide binding. Point mutations identified residues Arg93, Ser95, Glu127, Tyr140, Arg161, and His245 (Cr-P4H-1 numbering) as involved in substrate binding/recognition.[90] These residues are all conserved in human-P4H-1 and correspond to residues Arg362, Ser364, Glu356, Tyr409, Arg432, and His501 (human-P4H-1 numbering). His501 has been shown to be essential in αKG binding in human-P4H-1 and therefore most likely does not play a role in substrate recognition.[40] Additional mutations (Trp99, Gln130, and Trp243 Cr-P4H-1 numbering) in the groove also showed an increase in the K_m of poly(L-proline).[90] These residues are all also conserved in human-P4H-1 (Trp385, Gln406, and Trp499 human-P4H-1 numbering). This suggests that there may be alternative roles of the PBD that have not yet been determined. A high yielding, soluble expression system for human-P4H-1 in *E. coli* has successfully been developed in our laboratory as previously discussed. Comparative mechanistic studies on the recombinant forms of human-P4H-1 and anthrax-P4H, in particular to define the substrate recognition, should provide insights into the role of PDB domain.

An alternative approach for defining the role of PDB is the development of chimeric P4Hs. The expression of four chimeric forms of the PBD and PBCV-P4H have been examined (Figure 5.1). Chimeras were expected to be more soluble than the α_2 domain of human-P4H-1. However, we have not been able to obtain soluble protein yet. Chimera 3, which contains a 27 amino acid residues linker, yielded inclusion bodies. We have not screened conditions to solubilize the inclusion body.

We have now characterized the biochemical and structural properties of the recombinant anthrax-P4H (Chapter 2-4), where we have shown that the anthrax-P4H is α_2 dimer and exhibits the similar catalytic properties as human-P4H-1. Anthrax-P4H may serve as a better catalytic component for the chimera proteins than PBCV-P4H. The crystal structure of anthrax-P4H is also known and shows the N-terminus is accessible (Chapter 3) and in an orientation where PBD can attach.

5.5 Conclusions

Attempts have been made to prepare alternate forms of P4Hs in order to conduct comparative structure/function studies. Two viral P4Hs, one from PBCV-1 and the other from MT325, have been expressed. Future goals of these proteins include optimizing conditions to obtain soluble MT325-P4H. Identification of the native substrates of both PBCV- and MT325-P4H is also desired to help ascertain the role of P4H in these organisms *in vivo*. The

constructs of four chimeric P4Hs expression system were made, however none produced soluble proteins. Chimeric proteins incorporating anthrax-P4H as the catalytic subunit may be more fruitful as its it binds $(GPP)_{10}$ with a similar affinity as human-P4H-1.

6.1 Final Conclusions

The current available system for a recombinant human-P4H expression gives low yield and low purity enzyme, which is not sufficient for spectroscopic studies or structure determination. Both spectroscopic and structural knowledge can aide in new strategies for the design of potential therapeutic agents that target fibrosis. We proposed to characterize collagen-P4H from *B. anthracis* (anthrax-P4H) as a structural and spectroscopic model for human-P4H. However, we initially had to determine if ΔSterne *B. anthracis* produced Hyp as it is common belief that prokaryotes do not stabilize their collagens through Hyp.

B. anthracis contains BclA which forms a collagen-like triple helix structure and contains (GPT)n repeats. The Thr residues have been shown to be glycosylated, but we believe Pro is undergoing hydroxylation at the Xaa position as seen in P4Hs from lower organisms such as hydrothermal vent worms and Cr-P4H-1. Hydroxylation at the Xaa position has been shown to further stabilize model peptides with glycosylation at the Yaa position. We were able to grow ΔSterne *B. anthracis* cells and spores and isolate their proteins by varying solubility. Double derivatization and HPLC detected the presence of Hyp in both *B. anthracis* cells and spores. This is the first evidence of Hyp in *B. anthracis*. Our results show that there is a significant amount of Hyp associated with sporulation in *B. anthracis.* This strongly

suggests that a P4H-like protein that can catalyze the formation of Hyp exists in *B. anthracis*.

BLAST search analysis identified a collagen-P4H protein in the *B. anthracis* genome. We have cloned and expressed the first known bacterial P4H as a recombinant protein in *E. coli* (anthrax-P4H). Anthrax-P4H is expressed in high yield, highly soluble, and is monodispersed. Through gel filtration experiments, anthrax-P4H is predicted to be a homodimer, and site-directed mutagenesis studies show that it is a non-covalent dimer. Dialysis experiments and MALDI-TOF MS data support anthrax-P4H as a homodimer. This is the first known P4H present as a homodimer.

Characterization studies showed enzyme activity dependence on Fe(II), αKG, and O_2, which fits its classification of an αKG/Fe(II)-oxygenase. Anthrax-P4H binds $(GPP)_{10}$ with the same affinity as human-P4H. Anthrax-P4H is an improved model for human-P4H, and can aide in the development of improved, selective inhibitor design.

We have crystallized, collected data, and solved the structure of the apo form of anthrax-P4H to 1.4 Å. This is the second structure of a Collagen-P4H. The structure of anthrax-P4H has the DSBH or "jellyroll" core fold characteristic of the αKG/Fe(II) oxygenases. The active site also has the characteristic H^1-X-D/E-X_n-H^2 facial triad motif. It has been shown that self-hydroxylation of active site residues leads to inactivation of enzyme. The

inactivation is due to the self-hydroxylation of a Phe, Tyr or Trp residue in the active site, where the blue chromophore is assigned as a charge transfer complex of hydroxylated product to Fe(III). The self-hydroxylation occurs concomitantly with decarboxylation of αKG and resulting enzyme is inactivated. Structural overlays of anthrax-P4H with TauD and AlkB we proposed that F178 in the active site is the most likely candidate that undergoes self-hydroxylation. Spectral shifts were seen with F178A and F18Y mutations compared to WT, providing evidence that F178 is responsible for the self-hydroxylation.

A putative peptide-binding groove has been identified in algal-P4H, and is also conserved in anthax-P4H. Anthrax-P4H has been shown to catalyze $(Gly\text{-}Pro\text{-}Pro)_{10}$ with similar K_m values as human C-P4H, and its homodimeric property can provide insights into the α domain interactions of human C-P4H and catalytic property. We aim to determine the difference in substrate specificity between anthrax-P4H and algal-P4H, with their structures that are so similar.

Alternate P4H systems have also been developed in the laboratory. These include viral systems, PBCV- and MT325-P4H. Little is known about the function of these proteins *in vivo* or their native substrates. Chimera P4H proteins incorporating the PBD of human-P4H-1 and catalytic domain of

PBCV-P4H were also developed. The chimera P4Hs can be used to establish the role of the PBD in human-P4H peptide recognition.

The biochemical characterization and structure determination of anthrax-P4H has established new avenues for inhibitor development. Current work in the Limburg lab is focusing on the development of inhibitors targeting the substrate specificity of the PBD and the Fe(II) binding site of the catalytic core. Future work includes identifying the residues that play a key role in catalysis and inactivation, most importantly the residues conserved between anthrax-P4H and human-P4H-1. The identification of mechanistic intermediates is also proposed to fully understand the P4H mechanism to aid in the long-term goal of the development of selective fibrosis inhibitors.

References:

(1) Thompson, R. *FDA Consumer* **1985**, *v19*, p34(3).

(2) Myllyharju, J., & Kivirikko, K.I. *Trends Genet.* **2004**, *20*, 33-43.

(3) Brodsky, B.; Shah, N. K. *FASEB J.* **1995**, *9*, 1537-1546.

(4) Kivirikko, K. I.; Myllyla, R. *Methods Enzymol.* **1983**, *82*, 245-304.

(5) Prockop, J. D. *Ann. Rev. Biochem.* **1995**, *64*, 403.

(6) Vranka, J. A.; Sakai, L. Y.; Bachinger, H. P. *J. Biol. Chem.* **2004**, M312807200.

(7) Koivunen, P.; Salo, K. E. H.; Myllyharju, J.; Ruddocks, L. W. *J. Biol. Chem.* **2005**, *280*, 5227-5235.

(8) Kivirikko, K. I.; Pihlajaniemi, T. *Adv. Enzymol. Related Areas Mol. Biol.* **1998**, *72*, 325-398.

(9) Jimenez, S.; Harsh, M.; Rosenbloom, J. *Biochem. Biophy. Res. Commun.* **1973**, *52*, 106-114.

(10) Berg, R. A.; Prockop, D. J. *Biochem. Biophy. Res. Commun.* **1973**, *52*.

(11) Bella, J.; Eaton, M.; Brodsky, B.; Berman, H. M. *Science* **1994**, *266*, 75-81.

(12) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. *Nature* **1998**, *392*, 666-667.

(13) Kivirikko, K. I. *Ann. Med.* **1993**, *25*, 113-126.

(14) Myllyharju, J. *Ann. Med* **2008**, *Apr 23*, 1-16.

(15) Kisseleva, T.; Brenner, D. A. *Exp. Biol. Med.* **2008**, *233*, 109-122.

(16) Sivakumar, P.; Das, A. *Inflammation Res.* **2008**, *57*, 410-418.

(17) Myllyharju, J.; Kivirikko, K. I. *Ann. Med.* **2001**, *33*, 7-21.

(18) Rhoades, R. E.; Udenfriend, S. *Pro. Natl. Acad. Sci.* **1968**, *60*, 1473-1478.

(19) Kivirikko, K. I.; Prockop, D. J. *J. Biol. Chem.* **1967**, *242*, 4007-4012.

(20) Hausinger, R. P. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 21-68.

(21) Kivirikko, K. I.; Prockop, D. J. *Pro. Natl. Acad. Sci.* **1967**, *57*, 782-789.

(22) de Jong, L.; Albracht, S. P. J.; Kemp, A. *Biochim. Biophys. Acta* **1982**, *704*, 326-332.

(23) Myllyla, R., Majamma, K., Gunzler, V., Hanauske-Able, H. M., & Kivirikko, K. *J. Biol. Chem* **1984**, *259*, 5403-2405.

(24) Clifton, I. J., McDonough, M.A., Ehrismann, D., Kershaw, N.J., Granatino, N., & Schofield, C.J. *J. Inorg. Biochem.* **2006**, *100*, 644-669.

(25) Hanauske-Abel, H. M.; Gunzler, V. *J. Theor. Biol* **1982**, *94*, 421-455.

(26) Neidig, M. L.; Solomon, E. I. *Chem. Commun.* **2005**, 5843-5863.

(27) Cosper, N. J.; Stålhandske, C. M. V.; Saari, R. E.; Hausinger, R. P.; Scott, R. A. *J. Biol. Inorg. Chem.* **1999**, *4*, 122-129.

(28) Lloyd, M. D.; Lee, H.-J.; Harlos, K.; Zhang, Z.-H.; Baldwin, J. E.; Schofield, C. J.; Charnock, J. M.; Garner, C. D.; Hara, T.; Terwisscha van

Scheltinga, A. C.; ValegÂrd, K.; Viklund, J. A. C.; Hajdu, J.; Andersson, I.; Danielsson, k.; Bhikhabhai, R. *J. Mol. Biol.* **1999**, *287*, 943-960.
(29) Pavel, E. G. Z., J., Busby, R.W, Gunsior, M., Townsend, C.A., Solomon, E.I. *J. Am. Chem. Soc.* **1998**, *120*, 743-753.
(30) Purpero, V., and Moran,G.R. *J.Biol.Inorg. Chem.* **2007**, *12*, 587-601.
(31) Ryle, M. J.; Padmakumr, R.; Hausinger, R. P. *Biochemistry* **1999**, *38*, 15278-15286.
(32) Price, J. C.; Barr, E. W.; Tirupati, B.; Bollinger, J. M., Jr.; Krebs, C. *Biochemistry* **2003**, *42*, 7497-7508.
(33) Proshlyakov, D. A.; Henshaw, T. F.; Monterosso, G. R.; Ryle, M. J.; Hausinger, R. P. *J. Am. Chem. Soc.* **2004**, *126*, 1022-1023.
(34) Riggs-Gelasco, P. J.; Price, J. C.; Guyer, R. B.; Brehm, J. H.; Barr, E. W.; Jr. Bollinger, J. M.; Krebs, C. *J. Am. Chem. Soc.* **2004**, *126*, 8108-8109.
(35) Kivirikko, K. I.; Myllyharju, J. *Matrix Biol.* **1998**, *16*, 357-368.
(36) Myllyharju, J. *Matrix Biol.* **2003**, *22*, 15-24.
(37) Annunen, P.; Helaakoski, T.; Myllyharju, J.; Veijola, J.; Pihlajaniemi, T.; Kivirikko, K. I. *J. Biol. Chem.* **1997**, *272*, 17342-17348.
(38) Kukkola, L.; Hieta, R.; Kivirikko, K. I.; Myllyharju, J. *J. Biol. Chem.* **2003**, *278*, 47685-47693.
(39) Myllyharju, J. *Ann. Med.* **2008**, *Apr 23*, 1-16.
(40) Myllyharju, J.; Kivirikko, K. I. *Embo Journal* **1997**, *16*, 1173-1180.
(41) Kivirikko, K. I.; Myllyla, R.; Pihlajaniemi, T. *FASEB J.* **1989**, *3*, 1609-1617.
(42) Nissi, R.; Autio-Harmainen, H.; Marttila, P.; Sormunen, R.; Kivirikko, K. I. *J. Histochem. Cytochem.* **2001**, *49*, 1143-1154.
(43) Van Den Diepstraten, C.; Papay, K.; Bolender, Z.; Brown, A.; Pickering, J. G. *Circulation* **2003**, *108*, 508-511.
(44) Wilkinson, B.; Gilbert, H. F. *Biochim. Biophys. Acta* **2004**, *1699*, 35-44.
(45) Wetterau, J. R.; Combs, K. A.; Spinner, S. N.; Joiner, B. J. *J. Biol. Chem.* **1990**, *265*, 9801-9807.
(46) Tian, G.; Xiang, S.; Noiva, R.; Lennarz, W. J.; Schindelin, H. *Cell* **2006**, *124*, 61-73.
(47) Pirneskoski, A.; Ruddock, L. W.; Klappa, P.; Freedman, R. B.; Kivirikko, K. I.; Koivunen, P. *J. Biol. Chem.* **2001**, *276*, 11287-11293.
(48) Koivunen, P.; Salo, K. E. H.; Myllyharju, J.; Ruddocks, L. W. *J. Biol. Chem.* **2005**, *280*, 5227-5235.
(49) Vuori, K.; Pihlajaniemi, T.; Myllyla, R.; Kivirikko, K. I. *EMBO J* **1992**, *11*, 4213-4217.
(50) Gunzler, V.; Brocks, D.; Henke, S.; Myllyla, R.; Geiger, R.; Kivirikko, K. I. *J. Biol. Chem.* **1988**, *263*, 19498-19504.
(51) Baader, E.; Tschank, G.; Baringhaus, K. H.; Burghard, H.; Gunzler, V. *Biochem. J.* **1994**, *300*, 525-530.

(52) Cunliffe, C. J.; Frankline, T. J.; Hales, N. J.; Hill, G. B. *J. Med. Chem.* **1992**, *35*, 2652-2658.
(53) Majamaa, K.; Hanauske-Abel, H. M.; Gunzler, V.; Kivirikko, K. I. *Eur. J. Biochem.* **1984**, *138*, 239-245.
(54) Majamaa, K.; Turpeenniemi-Hujanen, T. M.; Latipaa, P.; Gunzler, V.; Hanauske-Abel, H. M.; Hassinen, I. E.; Kivirikko, K. I. *Biochem. J.* **1985**, *229*, 127-133.
(55) Majamaa, K.; Gunzler, V.; Hanauske-Abel, H. M.; Myllyla, R.; Kivirikko, K. I. *J. Biol. Chem.* **1986**, *261*, 7819-7823.
(56) Ivan, M.; Kondo, K.; Yang, H.; Kim, W.; Valiando, J.; Ohh, M.; Salic, A.; Asara, J. M.; Lane, W. S.; Kaelin, W. G., Jr. *Science* **2001**, *292*, 464-468.
(57) Jaakkola, P.; Mole, D. R.; Tian, Y.-M.; Wilson, M. I.; Gielbert, J.; Gaskell, S. J.; Kriegsheim, A. v.; Hebestreit, H. F.; Mukherji, M.; Schofield, C. J.; Maxwell, P. H.; Pugh, C. W.; Ratcliffe, P. J. *Science* **2001**, *292*, 468-472.
(58) Bruick, R. K.; McKnight, S. L. *Science* **2001**, *294*, 1337-1340.
(59) Veijola, J.; Koivunen, P.; Annunen, P.; Pihlajaniemi, T.; Kivirikko, K. I. *J. Biol. Chem.* **1994**, *269*, 26746-26753.
(60) Friedman, L.; Higgin, J. J.; Moulder, G.; Barstead, R.; Raines, R. T.; Kimble, J. *Pro. Natl. Acad. Sci.* **2000**, *97*, 4736-4741.
(61) Winter, A. D.; Page, A. P. *Mol. Cell. Biol.* **2000**, *20*, 4084-4093.
(62) Myllyharju, J.; Kukkola, L.; Winter, A. D.; Page, A. P. *J. Biol. Chem.* **2002**, *277*, 29187-29196.
(63) Kivirikko, K. I.; Myllyla, R. *Methods Enzymol.* **1982**, *82*, 245-304.
(64) Riihimaa, P.; Nissi, R.; Page, A. P.; Winter, A. D.; Keskiaho, K.; Kivirikko, K. I.; Myllyharju, J. *J. Biol. Chem.* **2002**, *277*, 18238-18243.
(65) Abrams, E. W.; Andrew, D. J. *Mech. Dev.* **2002**, *112*, 165-171.
(66) Annunen, P.; Koivunen, P.; Kivirikko, K. I. *J. Biol. Chem.* **1999**, *274*, 6790-6796.
(67) Yasothornsrikul, S.; Davis, W. J.; Cramer, G.; Kimbrell, D. A.; Dearolf, C. R. *Gene* **1997**, *198*, 17-25.
(68) Kieliszewski, M. J.; Shpak, E. *CMLS* **2001**, *58*, 1386-1398.
(69) Hieta, R.; Myllyharju, J. *J. Biol. Chem* **2002**, *277*, 23965-23971.
(70) Tiainen, P.; Myllyharju, J.; Koivunen, P. *J. Biol. Chem.* **2005**, *280*, 1142-1148.
(71) Ferris, P. J.; Woessner, J. P.; Waffenschmidt, S.; Kilz, S.; Drees, J.; Goodenough, U. W. *Biochemistry* **2001**, *40*, 2978-2987.
(72) Keskiaho, K., Hieta, R., Sormunen, R., and Myllyharju, J. *Plant Cell* **2007**, *19*, 256-269.
(73) Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., & Kivirikko, K.I. *J. Biol. Chem* **1999**, *274*, 22131-22134.
(74) Smith, M. C. M.; Burns, N.; Sayers, J. R.; Sorrell, J. A.; Casjens, S. R.; Hendrix, R. W. *Science* **1998**, *279*, 1834.

(75) Xu, Y.; Keene, D. R.; Bujnicki, J. M.; Hook, M.; Lukomski, S. *J. Biol. Chem* **2002**, *277*, 27312-27318.
(76) Lawrence, C. C., Sobey, W.J., Field, R.A., Baldwin, J.E., Schofield, C.J. *Biochem. J.* **1996**, *313*, 185-191.
(77) Shibasaki, T.; Mori, H.; Chiba, S.; Ozaki, A. *Appl. Environ. Microbiol.* **1999**, *65*, 4028-4031.
(78) Engel, J.; Bachinger, H. P. *Proc. Indian Acad. Sci (Chem. Sci.)* **1999**, *111*, 81-86.
(79) Charalambous, B. M.; Keen, J. N.; McPherson, M. J. *Embo Journal* **1988**, *7*, 2903-2909.
(80) Rasmussen, M.; Jacobsson, M.; Bjorck, L. *J. Biol. Chem* **2003**, *278*, 32313-32316.
(81) Mohs, A.; Silva, T.; Yoshida, T.; Amin, R.; Lukomski, S.; Inouye, M.; Brodsky, B. *J. Biol. Chem.* **2007**, *282*, 29757-29765.
(82) Rety, S.; Salamitou, S.; Garcia-Verdugo, I.; Hulmes, D. J. S.; Le Hegarat, F.; Chaby, R.; Lewit-Bentley, A. *J. Biol. Chem.* **2005**, *280*, 43073-43078.
(83) Daubenspeck, J. M.; Zeng, H. D.; Chen, P.; Dong, S. L.; Steichen, C. T.; Krishna, N. R.; Pritchard, D. G.; Turnbough, C. L. *J. Biol. Chem* **2004**, *279*, 30945-30953.
(84) Bann, J. G.; Peyton, D. H.; Bachinger, H. P. *FEBS Letters* **2000**, *473*, 237-240.
(85) Bann, J. G.; Bachinger, H. P. *J. Biol. Chem* **2000**, *275*, 24466-24469.
(86) Mann, K.; Mechling, D. E.; Bachinger, H. P.; Eckerskorn, C.; Gaill, F.; Timpl, R. *J. Mol. Biol.* **1996**, *261*, 255-266.
(87) Adams, E.; Lamon, M. *J. Biol. Chem* **1977**, *252*, 7591-7597.
(88) Neubauer, A.; Neubauer, P.; Myllyharju, J. *Matrix Biol.* **2005**, *24*, 59-68.
(89) Kersteen, E. A., Higgin, J.A., Raines, R.A. *Protein Expression Purif.* **2004**, *38*, 279-291.
(90) Koski, M. K., Hieta, R., Bollner, C., Kivirikko, K., Myllyharju, J., & Wierenga, R.K. *J. Biol. Chem* **2007**, *282*, 37112-37123.
(91) Mock, M.; Fouet, A. *Annu. Rev. Microbiol.* **2001**, *55*, 647-671.
(92) Dixon, T. C.; Meselson, M.; Guillemin, J.; Hanna, P. C. *New Engl. J. Med* **1999**, *341*, 815-826.
(93) Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Friedlander, A. M.; Hauer, J.; McDade, J.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Tonat, K. *JAMA* **1999**, *281*, 1735-1745.
(94) Liu, C. Q.; Nuttall, S. D.; Tran, H.; Wilkins, M.; Streltsov, V. A.; Alderton, M. R. *Biotechnol. Bioeng.* **2008**, *99*, 774-782.
(95) Levine, S. M.; Tang, Y.; Pei, Z. *Rev. Med. Microbiol.* **2005**, *16*, 125-133.

(96) Steichen, C.; Chen, P.; Kearney, J. F.; Turnbough, C. L. *J. Bacteriol.* **2003**, *185*, 1903-1910.
(97) Redmond, C.; Baillie, L. W. J.; Hibbs, S.; Moir, A. J. G.; Moir, A. *Microbiology-Sgm* **2004**, *150*, 355-363.
(98) Sylvestre, P., Couture-Tosi, E., and Mock, M. *Mol. Micro.* **2002**, *45*, 169-178.
(99) Boydston, J. A.; Chen, P.; Steichen, C. T.; Turnbough, C. L. *J. Bacteriol.* **2005**, *187*, 5310-5317.
(100) Gaill, F.; Hunt, S. *Rev. Aquat. Sci.* **1991**, *4*, 107-137.
(101) Waller, L. N., Stump, M.J., Fox, K.F., Harley, W.M., Fox, A., Stewart, G.C., Shahgholi, M. *J. Bacteriol.* **2005**, *187*, 4592-4597.
(102) Thompson, B. M.; Waller, L. N.; Fox, K. F.; Fox, A.; Stewart, G. C. *J. Bacteriol.* **2007**, *189*, 6704-6713.
(103) Symersky, J.; Patti, J. M.; Carson, M.; House-Pompeo, K.; Teale, M.; Moore, D.; Jin, L.; Schneider, A.; DeLucas, L. J.; Hook, M.; Narayana, S. V. L. *Nat. Struct. Biol.* **1997**, *4*, 833-838.
(104) Zong, Y.; Xu, Y.; Liang, X.; Keene, D. R.; Hook, A.; Gurusiddappa, S.; Hook, M.; Narayana, S. V. L. *Embo Journal* **2005**, *24*, 4224-4236.
(105) Nicholson, W. L.; Setlow, P. In *Molecular Biological Methods for Bacillus*; Harwood, C. R., Cutting, S. M., Eds.; John Wiley & Sons: New York, 1990, p 391- 450.
(106) Ikeda, M.; Sorimachi, K.; Akimoto, K.; Yasamura, Y. *J. Chromatogr. Biomed.* **1993**, *621*, 133-138.
(107) Boot-Handford, R. P.; Tuckwell, D. S. *BioEssays* **2003**, *25*, 142-151.
(108) Sylvestre, P.; Couture-Tosi, E.; Mock, M. *J. Bacteriol.* **2003**, *185*, 1555-1563.
(109) Kaule, G.; Timpl, R.; Gaill, F.; Gunzler, V. *Matrix Biol.* **1998**, *17*, 205-212.
(110) Gaill, F., Mann,K., Wiedemanna,H., Jurgen,E., Timpl,R. *J. Mol. Biol.* **1995**, *246*, 284-294.
(111) Goldstein, A.; Adams, E. *J. Biol. Chem* **1968**, *243*, 3550-3552.
(112) Goldstein, A.; Adams, E. *J. Biol. Chem* **1970**, *245*, 5478-5483.
(113) Josse, J.; Harrington, W. F. *J. Mol. Biol.* **1964**, *9*, 269-287.
(114) Eberhardt, E. S.; Jr. Panasik, N.; Raines, R. T. *J. Am. Chem. Soc.* **1996**, *118*, 12261-12266.
(115) Peterson, J. D.; Umayam, L. A.; Dickinson, T.; Hickey, E. K.; White, O. *Nucleic Acids Res.* **2001**, *29*, 123-125.
(116) Pihlajaniemi, T.; Helaakoski, T.; Tasanen, K.; Myllyla, R.; Huhtala, M. L.; Koivu, J.; Kivirikko, K. I. *Embo Journal* **1987**, *6*, 643-649.
(117) Pekkala, M.; Hieta, R.; Bergmann, U.; Kivirikko, K. I.; Wierenga, R. K.; Myllyharju, J. *J. Biol. Chem.* **2004**, *279*, 52255-52261.
(118) Kivirikko, K. I.; Myllyharju, J. *Matrix Biology* **1998**, *16*, 357-368.

(119) Engel, J. *Science* **1997**, *277*, 1785-1786.
(120) Vuorela, A.; Myllyharju, J.; Nissi, R.; Pihlajaniemi, T.; Kivirikko, K. I. *Embo Journal* **1997**, *16*, 6702-6712.
(121) Vuori, K., Pihlajaniemi, T., Marttila, M. and Kivirikko, K. I. *Pro. Natl. Acad. Sci.* **1992**, *89*, 7467-7470.
(122) John, D. C. A.; Bulleid, N. J. *Biochemistry* **1994**, *33*, 14018-14025.
(123) Lamberg, A.; Pihlajaniemi, T.; Kivirikko, K. I. *J. Biol. Chem.* **1995**, *270*, 9926-9931.
(124) Hoffart, L. M., Barr, E.W., Guyer, R.B., Jr. Bollinger, J.M., Krebs, C. *Pro. Natl. Acad. Sci.* **2006**, *103*, 14738-14743.
(125) Bollinger, J. M., Jr.; Price, J. C.; Hoffart, L. M.; Barr, E. W.; Krebs, C. *Eur. J. Inorg. Chem.* **2005**, 4245-4254.
(126) McNeill, L. A.; Bethge, L.; Hewitson, K. S.; Schofield, C. J. *Anal. Biochem.* **2005**, *336*, 125-131.
(127) Myllyharju, J.; Kivirikko, K. I. *Embo Journal* **1999**, *18*, 306-312.
(128) Kersteen, E. A., Higgin, J. A., Raines, R. A. *Protein Expres. Purif.* **2004**, *38*, 279-291.
(129) Fukumori, F.; Hausinger, R. P. *J. Biol. Chem* **1993**, *268*, 24311-24317.
(130) Tuderman, L.; Myllyla, R.; Kivirikko, K. I. *Eur. J. Biochem.* **1977**, *80*, 341-348.
(131) Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. *Nature* **2002**, *419*, 174-178.
(132) Liu, A.; Ho, R. Y. N.; Que, L.; Ryle, M. J.; Phinney, B. S.; Hausinger, R. P. *J. Am. Chem. Soc.* **2001**, *123*, 5126-5127.
(133) McDonough, M. A.; Li, V.; Flashman, E.; Chowdhury, R.; Mohr, C.; Lienard, B. M. R.; Zondlo, J.; Oldham, N. J.; Clifton, I. J.; Lewis, J.; McNeill, L. A.; Kurzeja, R. J. M.; Hewitson, K. S.; Yang, E.; Jordan, S.; Syed, R. S.; Schofield, C. J. *Pro. Natl. Acad. Sci.* **2006**, *103*, 9814-9819.
(134) Farmer, T. B.; Caprioli, R. M. *J. Mass Spectrom.* **1998**, *33*, 697-704.
(135) Tayeh, M. A.; Madigan, M. T. *J. Bacteriol.* **1987**, *169*, 4196-4202.
(136) Song, F. *J. Am. Soc. Mass. Spectrom.* **2007**, *18*, 1286-1290.
(137) Counts, D. F.; Cardinale, G. J.; Udenfriend, S. *Pro. Natl. Acad. Sci.* **1978**, *75*, 2145-2149.
(138) Liu, A.; Ho, R. Y. N.; Que, L., Jr.; Ryle, M. J.; Phinney, B. S.; Hausinger, R. P. *J. Am. Chem. Soc.* **2001**, *123*, 5126-5127.
(139) Koehntop, K. D.; Marimanikkuppam, S.; Ryle, M. J.; Hausinger, R. P.; Que, L., Jr. *J. Biol. Inorg. Chem.* **2006**, *11*, 63-72.
(140) Bradley, F. C.; Lindstedt, S.; Lipscomb, J. D.; Que, L., Jr.; Roe, A. L.; Rundgren, M. *J. Biol. Chem* **1986**, *261*, 11693-11696.
(141) Valegard, K.; van Scheltinga, A. C. T.; Lloyd, M. D.; Hara, T.; Ramaswamy, S.; Perrakis, A.; Thompson, A.; Lee, H. J.; Baldwin, J. E.; Schofield, C. J.; Hajdu, J.; Andersson, I. *Nature* **1998**, *394*, 805-809.

(142) Kivirikko, K. I.; Myllyla, R. *Ann. N.Y. Acad. Sci.* **1985**, *460*, 187-201.
(143) Prockop, D. J.; Kivirikko, K. I. *Annu. Rev. Biochem.* **1995**, *64*, 403-34.
(144) Main, E. R. G.; Xiong, Y.; Cocco, M. J.; D'Andrea, L.; Regan, L. *Structure* **2003**, *11*, 497-508.
(145) Das, A.; Cohen, P.; Barford, D. *EMBO J* **1998**, *17*, 1192-1199.
(146) Hendrickson, W. A. *Science* **1991**, *254*, 51-58.
(147) Cohen, A. E., Ellis, P. J., Miller, M. D., Deacon, A. M. & Phizackerley, R. P. *J. Appl. Cryst.* **2002**, *35*, 720-726.
(148) Leslie, A. G. *Acta Cryst.* **1999**, *D55*, 1696-1702.
(149) Collaborative Computational Project, N. *Acta Cryst.* **1994**, *D50*, 760-763.
(150) Sheldrick, G. M. *Acta Cryst.* **2008**, *A64*, 112-122.
(151) Terwilliger, T. C., & Berendzen, J. *Acta Cryst.* **1999**, *D55*, 849-861.
(152) Terwilliger, T. C. *Acta Cryst.* **2000**, *D56*, 965-972.
(153) Emsley, P., Cowtan, K. *Acta Cryst.* **2004**, *D60*, 2126-2132.
(154) Murshudov, G. N., Vagin, A.A., and Dodson, E.J. *Acta Cryst.* **1997**, *D53*, 240-255.
(155) Perrakis, A.; Sixma, T. K.; Wilson, K. S.; Lamzin, V. S. *Acta Cryst. D.* **1997**, *53*, 448-455.
(156) Perrakis, A.; Morris, R. M.; Lamzin, V. S. *Nat. Struct. Biol.* **1999**, *6*, 458-463.
(157) Collaborative Computational Project, N. *Acta Cryst.* **1994**, *D50*, 760-763.
(158) Laskowski R A, M. M. W., Moss D S & Thornton J M *J. Appl. Cryst.* **1993**, *26*.
(159) Vriend, G. *J. Mol. Graph.* **1990**, *8*.
(160) Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. *Proteins* **2003**, *50*.
(161) Delano, W. L.; Scientific, D., Ed. San Carlos, CA, USA, 2002.
(162) Miller, M. A.; Scott, E. E.; Limburg, J. *Acta Cryst.* **2008**, F*64*, 788-791.
(163) Matthews, B. W. *J. Mol. Biol.* **1968**, *33*, 491-497.
(164) Henshaw, T. F.; Feig, M.; Hausinger, R. P. *J. Inorg. Biochem.* **2004**, *98*, 856-861.
(165) Myllyharju, J. *Top. Curr. Chem* **2005**, *247*, 115-147.
(166) Rhodes, G. *Crystallography Made Crystal Clear: A Guide for users of macromolecular models*; Third Edition ed.; Elsevier/Academic Press: San Diego, 2006.
(167) McDonough, M. A.; Kavanagh, K. L.; Butler, D.; Searls, T.; Oppermann, U.; Schofield, C. J. *J. Biol. Chem.* **2005**, *280*, 41101-41110.
(168) Zhang, Z., Ren, J.S., CLifton, I.J., Schofield, C.J. *Chem. Biol* **2004**, *11*, 1383-1394.

(169) Lee, C.; Kim, S. J.; Jeong, D. G.; Lee, S. M.; Ryu, S. E. *J. Biol. Chem.* **2003**, *278*, 7558-7563.
(170) Muller, I.; Kahnert, A.; Pape, T.; Sheldrick, G. M.; Meyer-Klauck, W.; Dierks, T.; Kertesz, M.; Uson, I. *Biochemistry* **2004**, *43*, 3075-3088.
(171) Clifton, I. J.; Hsueh, L. C.; Baldwin, J. E.; Harlos, K.; Schofield, C. J. *Eur. J. Biochem.* **2001**, *268*, 6625-6636.
(172) Roach, P. L.; Clifton, I. J.; Fulop, V.; Harlos, K.; Barton, G. J.; Hajdu, J.; Andersson, I.; Schofield, C. J.; Baldwin, J. E. *Nature* **1995**, *375*, 700-704.
(173) You, Z.; Omura, S.; Ikeda, H.; Cane, D. E.; Jogl, G. *J. Biol. Chem.* **2007**, *282*, 36552-36560.
(174) Elkins, J. M.; Ryle, M. J.; Clifton, I. J.; Hotopp, J. C. D.; Lloyd, J. S.; Burzlaff, N. I.; Baldwin, J. E.; Hausinger, R. P.; Roach, P. L. *Biochemistry* **2002**, *41*, 5185-5192.
(175) Ryle, M. J.; Liu, A.; Muthukumaran, R. B.; Ho, R. Y. N.; Koehntop, K. D.; McCracken, J.; Jr. Que, L.; Hausinger, R. P. *Biochemistry* **2003**, *42*, 1854-1862.
(176) Ge, Y.; Lawhorn, B. G.; Elnaggar, M.; Sze, S. K.; Begley, T. P.; Mclafferty, F. L. *Protein Sci.* **2003**, *12*.
(177) Farquhar, E. R.; Koehntop, K. D.; Emerson, J. P.; Que, L. *Biochem. Biophy. Res. Commun.* **2005**, *338*, 230-239.
(178) Schofield, C. J., & Ratcliffe, P.J. *Biochem. Biophy. Res. Commun.* **2005**, *338*, 617-626.
(179) Elkins, J. M.; Hewitson, K. S.; McNeill, L. A.; Seibel, J. F.; Schlemminger, I.; Pughl, C. W.; Ratcliffe, P. J.; Schofield, C. J. *J. Biol. Chem.* **2003**, *278*, 1802-1806.
(180) Freedman, S. J.; Sun, Z.-Y. J.; Poy, F.; Kung, A. L.; Livingston, D. M.; Wagner, G.; Eck, M. J. *PNAS* **2002**, *99*, 5367-5372.
(181) Dames, S. A.; Martinez-Yamout, M.; De Guzman, R. N.; Dyson, H. J.; Wright, P. E. *Pro. Nat. Acad. Sci.* **2002**, *99*, 5271-5276.
(182) Zong, Y.; Xu, Y.; Liang, X.; Keene, D. R.; Hook, A.; Gurusiddappa, S.; Hook, M.; Narayana, S. V. *EMBO J* **2005**, *24*, 4224-36.
(183) Yamada, T.; Onimatsu, H.; Van Etten, J. L.; Karl, M.; Aaron, J. S. In *Advances in Virus Research*; Academic Press: 2006; Vol. Volume 66, p 293-336.
(184) Kang, M.; Dunigan, D. D.; Van Etten, J. L. *Mol. Plant Pathol.* **2005**, *6*, 213-224.
(185) Fitzgerald, L. A.; Graves, M. V.; Li, X.; Feldblyum, T.; Hartigan, J.; Van Etten, J. L. *Virology* **2007**, *358*, 459-471.
(186) Bozue, J.; Cote, C. K.; Moody, K. L.; Welkos, S. L. *Infect. Immun.* **2007**, *75*, 508-511.
(187) Brahmbhatt, T. N.; Janes, B. K.; Stibitz, E. S.; Darnell, S. C.; Sanz, P.; Rasmussen, S. B.; O'Brien, A. D. *Infect. Immun.* **2007**, *75*, 5233-5239.

(188) Hieta, R.; Kukkola, L.; Permi, P.; Pirila, P.; Kivirikko, K. I.; Kilpelainen, I.; Myllyharju, J. *J. Biol. Chem.* **2003**, *278*, 34966-34974.