

SUBSTRATE RECOGNITION BY PROLYL 4-HYDROXYLASE

by

Kelly Lynn Gorres

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
(Biochemistry)

at the
UNIVERSITY OF WISCONSIN–MADISON

2009

A dissertation entitled

Substrate Recognition by Prolyl 4-Hydroxylase

submitted to the Graduate School of the
University of Wisconsin-Madison
in partial fulfillment of the requirements for the
degree of Doctor of Philosophy

by

Kelly Lynn Gorres




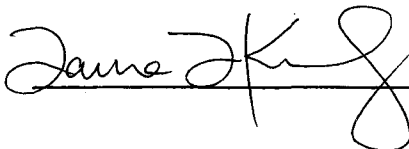
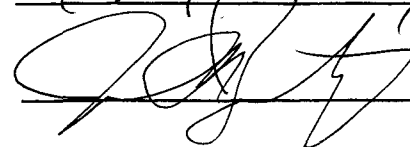
Date of Final Oral Examination: August 19, 2009

Month & Year Degree to be awarded: December

May

August 2009

Approval Signatures of Dissertation Committee

Signature, Dean of Graduate School

 / MW

SUBSTRATE RECOGNITION BY PROLYL 4-HYDROXYLASE

Kelly L. Gorres

Under the supervision of Professor Ronald T. Raines

At the University of Wisconsin–Madison

The post-translational hydroxylation of proline is found in a growing number of proteins, but is most widely known in collagen, a structural protein in animals. The conversion of (2*S*)-proline (Pro) to (2*S*,4*R*)-hydroxyproline (Hyp) stabilizes the triple-helical structure of collagen. Hydroxylation of proline residues is catalyzed by the enzyme prolyl 4-hydroxylase (P4H). P4H utilizes iron(II), molecular oxygen, and α -ketoglutarate as co-substrates to hydroxylate proline residues in a polypeptide substrate. This dissertation focuses on how P4H recognizes its substrates.

P4H catalyzes the hydroxylation of proline residues within Xaa–Pro–Gly sequences, where Xaa is any amino acid and Gly is glycine. The amino acids surrounding Pro influence the peptide structure and hydroxylation of Pro, but the conformation of Pro recognized by P4H was unknown. CHAPTER TWO describes the use of proline analogues to determine that P4H prefers the proline C $^{\gamma}$ -endo ring pucker conformation. This study proposes a mechanism by which P4H avoids product inhibition. This information will facilitate future design of inhibitors of the enzyme.

The study of enzymes, including P4H, relies heavily on the ability to assay the activity of the enzyme quickly, easily, and accurately. Because P4H uses a number of co-substrates and the natural substrate, collagen, is so large, no efficient assay of catalytic turnover by prolyl hydroxylation had been developed. In CHAPTER THREE, we describe the design and use of an

assay that is both continuous and direct by utilizing (2*S*,4*S*)-4-fluoroproline. This assay is demonstrated to be useful for both mechanistic studies and the screening of inhibitors.

P4H binds iron(II) in its active site via two histidines and one aspartate residue, termed a 2-His-1-carboxylate motif. This motif is common to many iron(II) dioxygenases similar to P4H. A class of related enzymes catalyzes halogenation reactions, rather than hydroxylations. These halogenases contain two histidine iron-binding residues, but lack the aspartate. Instead, a halide replaces the carboxylate. CHAPTER FOUR describes the conversion of P4H to mimic the halogenase active site and the analysis of the orientation of the iron-binding residues in P4H. These experiments demonstrate that the native P4H active site is resistant to alteration and critical for any catalytic activity.

Acknowledgements

I am grateful for the support I received from many people throughout graduate school. First, I would like to thank my advisor, Ron Raines. Ron provided many opportunities, and his communication skills and creativity served as a great example. Conversations with Ron were always fruitful. Perhaps most importantly, Ron cultivated a lab environment that is collaborative and energetic.

I am grateful for the many members of the Raines lab, past and present, who provided support and advice that has been important to my research. Dr. Betsy Kersteen was a tremendous mentor and invaluable in helping me get started in the lab. I appreciate the help from those on the collagen and related projects, Matt Shoulders, Amit Choudhary, Frank Kotch, Nicky McElfresh, and Greg Jakubczak. It was a pleasure working with Khian Hong Pua, a talented undergraduate student, who helped to collect data that appears in Chapter Four. I thank Kim Dickson, Steve Fuchs, Brian Miller, Bryan Smith, Eugene Lee, Jeremy Johnson, Jeet Kalia, Annie Tam, Eddie Myers, Luke Lavis, Joe Binder, Cindy Chao, Greg Ellis, Rex Watkins and the many other Raines lab members for contributive discussions. Rebecca Turcotte and Tom Rutkoski were exceptional colleagues and friends, along with other members of the Biochemistry department, Bryan Becklund, Todd Gruber, Alison Meyer, and Summer Raines.

The help I received from Dr. Gary Case at the Peptide Synthesis Facility at the University of Wisconsin Biotechnology Center is much appreciated. He has been a great source of advice and assistance, along with Dr. Darrell McCaslin at the Biophysics Instrumentation Facility. I am grateful for the opportunities provided by the Delta program for teaching and learning, the financial assistance provided by the NIH Chemistry–Biology Interface Training Program, and

the internship experience with Drs. Sharyn Su and Bruce Odegaard at Beckman Coulter during my graduate training. Professors Carpenter, Soderberg, Wyckoff, Cole, Gooch, and Togeas at the University of Minnesota, Morris contributed to my education prior to graduate school.

None of this work would have been possible without the love and support of my family and friends. Drs. Kelly Linn, Andrea Van Wyk, and Sarah Wynia-Smith were fantastic roommates, and are wonderful friends. My parents Jim and Deb Gorres, and sisters Brittany and Shannon, have always been there for me. I am so lucky and blessed to have such a wonderful family. Finally, I am so happy to have met and married John May—he is the best.

Table of Contents

Abstract.....	i
Acknowledgements	iii
Table of Contents	v
List of Figures	ix
List of Tables and Schemes.....	xii
List of Abbreviations.....	xiii
Chapter 1 Substrate Diversity of Prolyl 4-Hydroxylases	1
1.1 Abstract.....	2
1.2 Introduction.....	3
1.3 Prolyl 4-Hydroxylase (P4H)	6
1.3.1 Collagen	6
1.3.2 Collagen-domain-containing proteins.....	16
1.3.3 Elastin	18
1.3.4 Prion protein.....	19
1.3.5 Conotoxins	20
1.3.6 Argonaute 2.....	21
1.4 Prolyl Hydroxylase Domain Protein (PHD)	22
1.4.1 Hypoxia inducible factor α	22

1.4.2 Large subunit of RNA polymerase II.....	26
1.4.3 I κ B kinase- β	27
1.4.4 Activating transcription factor 4	28
1.4.5 Iron regulatory protein 2	29
1.5 Plant and algal prolyl 4-hydroxylases	30
1.6 Prolyl 4-hydroxylases in microorganisms	32
1.6.1 Protozoa	32
1.6.2 Bacteria	33
1.6.3 Virus.....	35
1.7 Protein structure.....	35
1.8 Conclusions.....	36
 Chapter 2 Conformational Preferences of Substrates for Human Prolyl 4-Hydroxylase	 48
2.1 Abstract.....	49
2.2 Introduction.....	50
2.3 Experimental Procedures	52
2.4 Results.....	57
2.5 Discussion	60
2.6 Conclusions.....	64
 Chapter 3 Direct and Continuous Assay for Prolyl 4-Hydroxylase	 76
3.1 Abstract.....	77
3.2 Introduction.....	78

3.3 Materials and methods	81
3.4 Results.....	82
3.5 Discussion	84
3.6 Conclusions.....	85
 Chapter 4 Stringency of the 2-His-1-Asp Active-Site Motif in Prolyl 4-Hydroxylase	 93
4.1 Abstract	94
4.2 Introduction.....	95
4.3 Materials and methods	97
4.4 Results.....	99
4.5 Discussion	101
 Chapter 5 Future Directions.....	 111
5.1 Three-dimensional structure of P4H.....	112
5.2 P4H recognition of prolyl ring pucker	114
5.3 P4H Xaa position substrate preference	114
5.4 Inhibitors of P4H	116
 Appendix 1 Structure and Function of <i>Bacillus subtilis</i> YphP, a Novel Prokaryotic Protein Disulfide Isomerase with a CXC Catalytic Motif.....	 118
A1.1 Abstract	119
A1.2 Introduction.....	120

	viii
A1.3 Materials and methods	122
A1.4 Results and discussion	126
A1.5 Conclusions.....	133
Appendix 2 Towards Triple Helices in <i>Escherichia coli</i>	146
A2.1 Abstract	147
A2.2 Introduction.....	148
A2.3 Materials and methods	150
A2.4 Results.....	153
A2.5 Discussion	157
Appendix 3 Delta Program: Integrating Research, Teaching, and Learning....	169
A3.1 Description of the Delta Program	170
A3.2 All (content) is not lost	171
A3.2.1 Abstract	171
A3.2.2 Introduction.....	172
A3.2.3 Project design.....	173
A3.2.4 Results and Discussion.....	174
A3.2.5 Lessons learned	177
A3.2.6 Conclusions.....	178
References	182

List of Figures

Figure 1.1	Reaction catalyzed by prolyl 4-hydroxylase.....	40
Figure 1.2	Crystallographic structure of the collagen triple helix.....	41
Figure 1.3	Putative mechanism for prolyl hydroxylation by P4H.	42
Figure 1.4	Surfactant proteins A and D.....	43
Figure 1.5	Hypoxia sensing pathway.	44
Figure 1.6	Role of Hyp in oxygen sensing.	45
Figure 1.7	Three-dimensional structures of prolyl 4-hydroxylases.	46
Figure 1.8	Structures of peptide substrate-bound prolyl 4-hydroxylases.....	47
Figure 2.1	Reaction catalyzed by prolyl 4-hydroxylase.....	69
Figure 2.2	Structures of proline and its analogues studied herein, in their predominant ring pucker.....	70
Figure 2.3	Comparison of PEG-Gly-Tyr-Yaa-GlyOEt peptides.....	71
Figure 2.4	Peptides containing Flp, Kep, or Mtp do not inhibit Hyp formation from Pro by P4H.	72
Figure 2.5	Characterization of the products for the turnover of PEG-Gly-Tyr-flp-GlyOEt by P4H.	73
Figure 2.6	Characterization of the products for the turnover of PEG-Gly-Tyr-Thp-GlyOEt by P4H.	74
Figure 2.7	Putative model for substrate recognition by P4H.	75
Figure 3.1	Reaction catalyzed by prolyl 4-hydroxylase when a Pro-containing peptide or a flp-containing peptide is the substrate.	88
Figure 3.2	Fluoride ion-detection assay for P4H.	89

Figure 3.3	Dependence of the rate of fluoride-ion release from PEG-Gly-Tyr-flp-GlyOEt (1.0 mM) on the concentration of P4H.	90
Figure 3.4	Catalysis of fluoride-ion release from PEG-Gly-Tyr-flp-GlyOEt by P4H.	91
Figure 3.5	Catalysis of fluoride ion-release from PEG-Gly-Tyr-flp-GlyOEt (0.20 mM) by a P4H variant or in the presence of a small-molecule inhibitor.	92
Figure 4.1	Active sites of α -ketoglutarate-dependent iron(II) dioxygenases.	106
Figure 4.2	Reaction catalyzed by P4H.	107
Figure 4.3	Proposed reaction mechanisms for the hydroxylase P4H (A) and halogenase SyrB2 (B).	109
Figure 4.4	Tolerance of P4H activity to salts.	110
Figure A1.1	FRET-based assay for disulfide bond isomerization.	137
Figure A1.2	Ribbons diagrams of YphP.	139
Figure A1.3	Structure of the catalytic loop of YphP.	140
Figure A1.4	Sequence alignment of representative members of the DUF1094 family.	142
Figure A1.5	Reduction potential of YphP.	143
Figure A1.6	A diagram of the proposed reaction mechanism catalyzed by YphP and its homologues.	144
Figure A1.7	Structures of known oxidized CXC motifs and the reduced CXC motif of YphP.	145
Figure A2.1	Schematic of the lambda repressor (cI) and NcI-fusion protein control of gene expression.	161
Figure A2.2	Structure of foldon.	162
Figure A2.3	Map of plasmid pNcI-(PPG) _n	163

Figure A2.4	Purification of NcI-(PPG) ₇	164
Figure A2.5	Dependence of triple helix formation on temperature.	165
Figure A2.6	NcI-foldon represses <i>lacZ</i> expression.....	166
Figure A2.7	Oligomerization of NcI-foldon variants.....	168
Figure A3.1	Non-technical responses increased from pre-course to post-course	181

List of Tables and Schemes

Table 1.1	The mid-point of thermal stability (T_m) values for collagen-like peptides that vary in the Yaa position.	38
Table 1.2	Summary of prolyl and proline hydroxylases, their substrates and the proteins that contain hydroxyproline and the effect/result of hydroxylation.....	39
Table 2.1	Comparison of PEG-Gly-Tyr-Yaa-GlyOEt as substrates or inhibitors for P4H.	66
Table 2.2	Comparison of [^{14}C]CO ₂ release from [$1\text{-}^{14}\text{C}$]α-ketoglutarate in P4H assays with PEG-Gly-Tyr-Yaa-GlyOEt as the substrat	67
Scheme 2.1	68
Table 3.1	Comparison of the turnover of PEG-Gly-Tyr-flp-GlyOEt by P4H (90 nM) determined by using the fluoride detection assay or the HPLC-based assay	87
Table 4.1	Characterization of P4H variants for hydroxylase and halogenase activity.	105
Table A1.1	Crystallographic data	135
Table A1.2	Values of k_{cat}/K_M for catalysis of disulfide bond isomerization.....	136
Table A2.1	Values of m/z of purified NcI-(PPG) _n proteins.....	160
Table A3.1	Number of students that gave technical and non-technical responses.	179
Table A3.2	The technical knowledge increased over the BME 601 course.	180

List of Abbreviations

Ago 2	argonaute 2
Ala (A)	alanine
Asp (D)	aspartic acid
ATF4	activating transcription factor 4
At-P4H	P4H from <i>Arabidopsis thaliana</i>
AtsK	alkylsulfatase from <i>Pseudomonas putida</i>
β -gal	β -galactosidase
Boc	<i>tert</i> -butyloxycarbonyl
cI	lambda repressor
c.p.s.	counts per second
Cr-P4H	P4H from <i>Chlamydomonas reinhardtii</i>
CRTAP	cartilage-associated protein
CS	clavamate synthase
CXC	Cys-Xaa-Cys
CXXC	Cys-Xaa-Xaa-Cys
CypB	cyclophilin B
DAOCS	deacetoxycephalosporin C synthase
DdP4H	P4H from <i>Dictyostelium discoideum</i>
DMF	dimethylformamide
DMSO	dimethylsulfoxide

dns-sTI	dansyl scrambled tachyplesin I
DTNB	5,5'-dithiobis(2-nitrobenzoic acid)
DTT	D,L-dithiothreitol
E°	reduction potential
$E^{\circ'}$	standard reduction potential
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ESI-MS	electrospray ionization mass spectrometry
ε	extinction coefficient
F	fluorescence emission intensity
F_0	initial fluorescence emission intensity
FIH-1	factor-inhibiting hypoxia-inducible factor 1
Flp	4(<i>R</i>)-fluoro-L-proline or (2 <i>S</i> ,4 <i>R</i>)-4-fluoroproline
flp	4(<i>S</i>)-fluoro-L-proline or (2 <i>S</i> ,4 <i>S</i>)-4-fluoroproline
F_{\max}	maximal fluorescence emission intensity
Fmoc	fluorenylmethoxycarbonyl
FPLC	fast performance liquid chromatography
FRET	fluorescence resonance energy transfer
F	Faraday constant
Glu (E)	glutamic acid
Gly (G)	glycine
<i>gor</i>	glutathione reductase gene
Gros1	growth suppressor 1

GSH	reduced glutathione
GSSG	oxidized glutathione
HCl	hydrochloric acid
HIF	hypoxia inducible factor
His (H)	histidine
HP	hibernation protein
HPLC	high-performance liquid chromatography
HRE	hypoxia response element
HRGP	hydroxyproline-rich glycoprotein
Hyp	4(<i>R</i>)-hydroxy-L-proline or (2 <i>S</i> ,4 <i>R</i>)-4-hydroxyproline
hyp	4(<i>S</i>)-hydroxy-L-proline or (2 <i>S</i> ,4 <i>S</i>)-4-hydroxyproline
3-Hyp	(2 <i>S</i> ,3 <i>S</i>)-3-hydroxyproline
Ig	immunoglobulin
IKK	inhibitory κ B kinase
IPTG	isopropyl-1-thio- β -D-galactoside
IRP2	iron regulatory protein 2
K_a	acid dissociation constant
k_{cat}	first-order enzymatic rate constant
K_d	equilibrium dissociation constant
kDa	kilodalton
Kep	(2 <i>S</i>)-4-keto-L-proline
K_{eq}	equilibrium constant
K_m	Michaelis constant

<i>lacZ</i>	gene encoding β -galactosidase
LB	Luria–Bertani
Leu (L)	leucine
MALDI–TOF	matrix-assisted laser desorption/ionization time of flight
MBL	mannan-binding lectin
MCPBA	3-chloroperoxybenzoic acid
MS	mass spectrometry
Mtp	3,5-methano-L-proline or 2-azabicyclo[2.1.1]hexane-3-carboxylic acid
MW	molecular weight
NaCl	sodium chloride
NcI	N-terminal domain of lambda repressor
NMR	nuclear magnetic resonance
OD	optical density
ODDD	oxygen-dependent degradation domain
OI	osteogenesis imperfecta
P3H	prolyl 3-hydroxylase
P4H	collagen prolyl 4-hydroxylase
P4H α (I)	isoform I of the α subunit of human collagen prolyl 4-hydroxylase
P4H-TM	P4H containing a transmembrane domain
PBCV-1	<i>Paramecium bursaria</i> chlorella virus-1
PDI	protein disulfide isomerase
PEG	poly(ethylene glycol)
PHD	prolyl hydroxylase domain protein

pK_a	log of the acid dissociation constant
Pro (P)	(2 <i>S</i>)-proline
PrP ^C	cellular prion protein
PrP ^{Sc}	scrapie prion protein
pVHL	von Hippel Lindau protein
RNAi	RNA interference
Rbp1	RNA polymerase II large subunit
RISC	RNA-induced silencing complex
SDS	sodium dodecyl sulfate
Ser (S)	serine
siRNA	small interfering RNA
SP-A	lung surfactant protein A
SP-D	lung surfactant protein D
STEM	science, technology, engineering, math
SyrB2	<i>Pseudomonas syringae</i> syringomycin biosynthesis enzyme
TauD	taurine/ α -ketoglutarate dioxygenase
TB	Terrific Broth
TEMPO	2,2,6,6-tetramethylpiperidine-1-oxyl
TFA	trifluoroacetic acid
TfdA	2,4-dichlorophenoxyacetate/ α -ketoglutarate dioxygenase
Thp	(2 <i>S</i>)-4-thiaproline
Thp(O)	(2 <i>S</i> ,4 <i>R</i>)-4-thiaoxoproline
thp(O)	(2 <i>S</i> ,4 <i>S</i>)-4-thiaoxoproline

Thp(O,O)	(2S)-4-thiadioxoproline
<i>T</i>	temperature
<i>T_m</i>	temperature at the midpoint of the thermal denaturation curve
TPR	tetratricopeptide repeat
Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride
Trx	thioredoxin
<i>trxB</i>	thioredoxin reductase gene
UV	ultraviolet
Val (V)	valine
VCB	pVHL-elonginC-elonginB
<i>V_{max}</i>	maximal velocity
VOSTS	Views on Science-Technology-Society
Xaa	any amino acid
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
Yaa	any amino acid

Chapter One

Introduction:

Substrate Diversity of Prolyl 4-Hydroxylases

1.1 Abstract

Posttranslational modifications affect the biochemical properties of proteins, which can cause profound changes in protein function. Proline can be posttranslationally modified by the addition of a hydroxyl group, most often at the 4-position of the pyrrolidine ring, resulting in (2*S*,4*R*)-4-hydroxy-L-proline (Hyp). Hyp is most well known for its role in stabilizing the triple-helical structure of collagen. Hyp is also found in proteins with collagen-like domains, as well as elastin, conotoxins, and argonaute 2. Prolyl 4-hydroxylation in these proteins is catalyzed by the enzyme prolyl 4-hydroxylase (P4H), which is located in the endoplasmic reticulum. A cytosolic prolyl hydroxylase domain protein (PHD) catalyzes the prolyl 4-hydroxylation of the hypoxia inducible factor α (HIF α), which plays a role in oxygen sensing. PHD is also proposed to hydroxylate Inhibitory κ B kinase (IKK) and RNA polymerase II. Prolyl 4-hydroxylases are also found in plants, and a bacterial and a viral peptidyl prolyl 4-hydroxylase have recently been reported. The protein substrates for the prolyl 4-hydroxylases are incredibly diverse beyond the prolyl residue that is hydroxylated. Additionally, the biological consequences of prolyl hydroxylation vary widely, including alterations in protein conformation, protein–protein interactions, and further modification. Here, we discuss collagen and more recently discovered proteins containing Hyp, the role of Hyp in these proteins, and the prolyl hydroxylases that catalyze this modification.

1.2 Introduction

Proteins are biological macromolecules that perform a variety of roles in catalysis, structure, signal transduction, immune response, transport, and movement. Proteins are comprised of a chain of amino acids joined by peptide bonds. In most organisms, there are twenty so-called proteinogenic amino acids that are building blocks for synthesis of polypeptides by the ribosome. An immense number of proteins can be formed from the possible combinations of the twenty natural amino acids, yet an even greater diversity of chemical functionalities and structures is obtained by altering proteins after ribosomal synthesis, termed posttranslational modifications.

Chemical modification of proteins occurs both cotranslationally and posttranslationally (Walsh et al., 2005). Modifications influence protein function and behavior by altering characteristics such as conformation, stability, protein–protein interactions, and the potential for further modification. Some modifications are permanent, while others are transient. A protein can be modified by cleavage of the polypeptide backbone and removal of a portion of the protein. Other modifications covalently append molecules or functional groups onto the protein. These additions range in size from an entire protein, carbohydrate, or lipid, as in ubiquitinylation, glycosylation, and farnesylation, to relatively small modifications of individual amino acids that involve only a few atoms, such as disulfide bond formation, phosphorylation, acetylation, methylation, sulfonation, and carboxylation.

Here, we focus on the posttranslational hydroxylation of proline residues. The hydroxylation of proline residues, the conversion of a C–H bond into a C–OH bond, adds only 16 atomic mass units. Hydroxylation most commonly occurs at the 4-position (C^{γ}) of the prolyl pyrrolidine ring, producing (2*S*,4*R*)-4-hydroxy-L-proline (Hyp), also termed *trans*-4-hydroxyproline (Figure 1.1). The insertion of an electronegative atom alters the stereoelectronic properties of the amino acid,

as well as creates the potential for hydrogen bonding. The biological consequences of prolyl hydroxylation are diverse. This modification can have effects on the structure of the protein being modified, protein–protein interactions, and the reactivity of the protein.

Hydroxylation of proline residues occurs in a wide variety of proteins and in organisms from across the kingdoms of life. Prolyl hydroxylation is perhaps most widely known in collagen. Prolyl 4-hydroxylation in collagen is catalyzed by the enzyme prolyl 4-hydroxylase (P4H). P4H hydroxylates collagen in the endoplasmic reticulum prior to secretion. In collagen, the addition of the electron-withdrawing hydroxyl group alters the structure of the polypeptide resulting in stabilization of the collagen triple helix. In addition to collagen, other secreted proteins that contain collagen-like domains are hydroxylated by P4H, such as acetylcholinesterase and many proteins involved in innate immunity. P4H also catalyzes the hydroxylation of proteins that do not resemble collagen, such as another structural protein, elastin, the conotoxin peptides produced by snails, prion protein, and the argonaute 2 protein of the RNA-induced silencing complex.

Prolyl hydroxylation also takes place in the cytoplasm of eukaryotic cells, and is catalyzed by the prolyl hydroxylase domain protein (PHD). Cytoplasmic prolyl hydroxylation is part of the mechanism by which metazoans sense oxygen through the hypoxia inducible factor (HIF). Rather than playing a role in protein structure, as in collagen, Hyp in HIF α creates a hydrogen bond donor that instigates a protein–protein interaction with the von Hippel Lindau protein (pVHL) complex resulting in HIF α degradation. PHD is also proposed to hydroxylate RNA polymerase II, inhibitory κ B kinase β (IKK β), and possibly activating transcription factor 4 and iron regulatory protein 2. Prolyl hydroxylation affects the stability and activity of these proteins, though the mechanisms are not well defined.

Though predominantly studied in animals, prolyl hydroxylases are also found in organisms from the other kingdoms of life. Plants contain prolyl 4-hydroxylases that produce Hyp in cell wall proteins and proteins involved in defense. In plants, unlike in animals, prolyl hydroxylation generates a substrate for further posttranslational modifications, as sugars are added to the hydroxyl group of Hyp in plants. In addition to being found in plants, Hyp is found and glycosylated in the Skp1 protein of the soil amoeba *Dictyostelium*. Hyp is also made by bacteria, although Hyp found in bacterial nonribosomal peptide antibiotics is produced by proline 4-hydroxylase, which hydroxylates free proline rather than prolyl residues within polypeptides. Nevertheless, a bacterial peptidyl prolyl hydroxylase has recently been reported in *Bacillus anthracis*. A viral prolyl 4-hydroxylase has also been reported, though the biological function of the bacterial and viral prolyl hydroxylases is unknown.

All of the prolyl hydroxylases are a part of a larger class of enzymes, the α -ketoglutarate-dependent, iron(II) dioxygenases. These enzymes utilize iron, α -ketoglutarate, and molecular oxygen to catalyze the oxidation of nonreactive molecules, including the hydroxylation of an alkyl prolyl substrate (Figure 1.1). The α -ketoglutarate-dependent, iron(II) dioxygenases share a catalytic triad, and are thought to use a similar mechanism, but do not share much overall sequence identity. Herein, the similarities and differences among the prolyl hydroxylases are described in the context of the variety of proteins that are hydroxylated and their substrate recognition elements for prolyl hydroxylation. Proteins containing hydroxyproline are continually being discovered, and additional prolyl hydroxylase enzymes perhaps will be discovered.

1.3 Prolyl 4-Hydroxylase (P4H)

1.3.1 Collagen

Proline hydroxylation is perhaps most commonly discussed in collagen. Collagen is the most abundant protein in animals, and it is the major structural protein of connective tissues in mammals. Collagens are characterized by a repeating amino acid sequence in which every third amino acid is glycine, (Xaa–Yaa–Gly)_n. The strands within the most abundant type of collagen, Type I, each contain approximately 300 Xaa–Yaa–Gly triplets. Proline (Pro) is the amino acid most commonly found in the Xaa position, and in the Yaa position most often is Hyp. The Pro–Hyp–Gly sequence occurs in 10.5% of collagen triplets (Ramshaw et al., 1998).

Hyp stabilizes the collagen triple helix

Collagen forms a characteristic triple-helical tertiary structure, shown in Figure 1.2 (Bella et al., 1994; Berisio et al., 2001; Kramer et al., 1999; Nagarajan et al., 1999). The triple helix consists of three poly(L-proline)-like chains in a left-handed supercoil. The presence of Hyp is required for collagen stability at physiological temperatures, demonstrated by the difference in melting temperature (T_m , the temperature at the midpoint of the thermal transition) between a fully hydroxylated type I collagen of 43 °C and an unhydroxylated form of only 27 °C (Berg and Prockop, 1973c). Stabilization of the triple helix by the presence of Hyp has been studied extensively using peptide mimics of collagen (Holmgren et al., 1998; Sakakibara et al., 1973). Table 1.1 shows the T_m values for collagen peptides containing various amino acids.

Mechanisms by which Hyp stabilizes the collagen triple helix have been reviewed recently (Shoulders and Raines, 2009). The prevailing hypothesis for many years was that Hyp mediated

hydrogen bonding interactions along with an extensive network of water molecules that would stabilize the triple helix. X-ray crystallographic studies provided evidence for this model, showing that the triple helix is highly hydrated and that the hydroxyl group of Hyp binds to a carbonyl on the main chain of a neighboring strand through one or two water molecules (Bella et al., 1994; Kramer et al., 1998). Structural studies of (Pro–Pro–Gly)₁₀ and (Pro–Hyp–Gly)₁₀, however, presented a less substantial hydration network, and the same number of water molecules was detected in the structure with Pro in place of Hyp (Nagarajan et al., 1998; Nagarajan et al., 1999). Furthermore, Hyp stabilizes the triple helix even in anhydrous conditions (Engel et al., 1977; Holmgren et al., 1998), and (2*S*,4*S*)-4-hydroxyproline (hyp) does not stabilize the triple helix (Takeuchi and Prockop, 1969). These findings suggest that the stability of the Hyp-containing helix is not due to water-mediated hydrogen bonds.

An alternative hypothesis probed the electron-withdrawing property of the hydroxyl group. Hyp analogues in which the hydroxyl group is replaced with a fluorine atom were incorporated into collagen mimic peptides. Organic fluorine has a low tendency to form hydrogen bonds, but fluorine is an electronegative element. Peptides containing (2*S*,4*R*)-4-fluoroproline (Flp) in the Yaa position, (Pro–Flp–Gly)₁₀, assemble into triple helices, and the T_m of these helices is 91 °C, which is ~20 °C greater than that of the (Pro–Hyp–Gly)₁₀ peptide (T_m of 69 °C; Table 1.1; (Holmgren et al., 1998)). These results suggest that the stabilizing effect of Flp (and Hyp) is due to an inductive effect generated from the electron-withdrawing substituent on the proline ring, rather than hydrogen bonding (Bretscher et al., 2001; Holmgren et al., 1999; Jenkins and Raines, 2002). The effect depends on the stereochemistry, as (2*S*,4*S*)-4-fluoroproline (flp) in place of Hyp does not stabilize the triple helix (Bretscher et al., 2001).

The effectiveness of each amino acid at stabilizing the triple helix can be assessed by the conformation induced by the amino acid compared to the conformational preferences of collagen. The conformations discussed here are the pucker of the pyrrolidine ring of proline and the isomerization of the Xaa–Pro amide bond. Analysis of the crystal structure of (Pro–Pro–Gly)₁₀ revealed that residues in the Yaa position prefer torsion angles [$\phi = (-60.1 \pm 3.6)^\circ$ and $\psi = (152.4 \pm 2.6)^\circ$] that correspond to the C^γ-*exo* (up) ring pucker conformation. Residues in the Xaa position, however, prefer torsion angles [$\phi = (-74.5 \pm 2.9)^\circ$ and $\psi = (164.3 \pm 4.3)^\circ$] that correspond to the C^γ-*endo* (down) ring pucker conformation (Berisio et al., 2002). Substitutions on the pyrrolidine ring affect the ring pucker conformation. Stereoelectronic substitutions influence the ring pucker via the *gauche* effect. Hyp and Flp prefer the C^γ-*exo* conformation, whereas Pro and flp favor the C^γ-*endo* pucker (Bretscher et al., 2001; DeRider et al., 2002). The presence of an amino acid that favors the C^γ-*endo* pucker in the Xaa position and the C^γ-*exo* pucker in the Yaa position typically stabilize the collagen triple helix. The ring pucker, and hence collagen stability, can also be directed by steric substitutions (Shoulders et al., 2006).

Substitutions on the pyrrolidine ring also influence the *trans*:*cis* ratio of the peptidyl-prolyl amide bond. Proline is unique among amino acids in that its amide bond is found in the *cis* conformation in proteins, whereas all other amino acids are found exclusively in the *trans* conformation. The amide bonds found in collagen are all in the *trans* conformation, and inhibition of peptidyl-prolyl *cis*–*trans* isomerase decreases collagen production (Bächinger, 1987; Steinmann et al., 1991). Electronegative substitutions at the 4*R* position, as in Hyp and Flp, increase the percentage of amide bonds in the *trans* conformation (Bretscher et al., 2001). This occurs because the C^γ-*exo* ring pucker, enforced by the *gauche* effect, provides backbone

angles that allow an $n \rightarrow \pi^*$ interaction between the oxygen of an amide bond and the carbonyl of the subsequent amide bond when the amide bond is in the *trans* conformation. Proline derivatives with electronegative substitutions at the 4S position, such as flp, prefer the C^γ -*endo* ring pucker, which does not stabilize the *trans* peptide bond, and thus there is a higher percentage of amide bonds in the *cis* conformation. The stereoelectronic preorganization of the ϕ , ψ , and ω dihedral angles explains the mechanism by which the collagen triple helix is stabilized by the hydroxylation of Pro in the Yaa position. This seemingly small change has global effects and is crucial for the tertiary structure of collagen.

Hyp is formed in collagen by prolyl 4-hydroxylase

The biosynthesis of fibrillar collagens entails a series of posttranslational modifications. The hydroxylation of specific proline residues is one of the first modifications that takes place in the lumen of the endoplasmic reticulum. This hydroxylation is catalyzed by prolyl 4-hydroxylase (P4H; EC 1.14.11.2). P4H activity is critical for proper folding of collagen, and P4H activity is necessary for the viability of the nematode *Caenorhabditis elegans* (Myllyharju et al., 2002; Winter and Page, 2000) and the mouse *Mus musculus* (Holster et al., 2007).

Mammalian P4H is an $\alpha_2\beta_2$ tetramer (Berg and Prockop, 1973a; Koivu and Myllyla, 1986; Nietfeld et al., 1981)) in which the 59 kDa α subunit (P4H α) contains the peptide substrate binding domain and the enzymic active site (Helaakoski et al., 1989; Hutton et al., 1966b; Prockop and Juva, 1965). Three isoforms of the P4H α subunit, α (I), α (II), and α (III), have been identified in vertebrates, with α (I) being the most prevalent (Annunen et al., 1997; Helaakoski et al., 1995; Helaakoski et al., 1989; Kukkola et al., 2003; Myllyharju et al., 2002). All of the

isoforms associate in the $\alpha_2\beta_2$ tetrameric form. Most of the conserved amino acids occur in the C-terminal region where the catalytic residues of the active site are located. P4Hs from other animals, such as the nematode *Caenorhabditis elegans* and the fly *Drosophila melanogaster* have been characterized. The prolyl 4-hydroxylases from *C. elegans*, PHY-1 and PHY-2, can assemble with a single β subunit to form dimers or as a mixed PHY-1/PHY-2/ β_2 tetramer (Myllyharju and Kivirikko, 2004).

The β subunit functions independently as protein disulfide isomerase (PDI; 55 kDa) (Koivu and Myllyla, 1986; Pihlajaniemi et al., 1987). As a P4H subunit, PDI retains the enzyme in the lumen of the ER through the KDEL retention signal and keeps the α subunit in a soluble and active form (Vuori et al., 1992b; Vuori et al., 1992c). In the absence of PDI, the α subunit is insoluble and cannot be refolded in vitro (Nietfeld and Kemp, 1981). Recombinant P4H tetramers have been produced by coexpression of the α subunit and PDI in mammalian, plant, insect, and yeast cells, as well as more recent bacterial expression systems (Kersteen et al., 2004; Neubauer et al., 2005).

P4H is a member of the non-heme iron(II), α -ketoglutarate-dependent dioxygenase family. Fe(II), α -ketoglutarate, and O_2 are required for activity (Hutton and Udenfriend, 1966). During the reaction, α -ketoglutarate is oxidatively decarboxylated to produce succinate and CO_2 (Figure 1.1; (Rhoads and Udenfriend, 1968)). The proposed mechanism for prolyl hydroxylation by P4H is depicted in Figure 1.3. This mechanism is based on studies of related enzymes in the family of α -ketoglutarate-dependent non-heme iron(II) oxygenases (reviewed in (Costas et al., 2004)). The reaction occurs in two main stages. The first stage involves the formation of the highly reactive Fe(IV)=O species, which then goes on to hydroxylate proline in the second stage.

The reactive iron-oxo intermediate catalyzes prolyl hydroxylation by abstracting the pro-4R hydrogen atom, which forms a radical, followed by oxygen rebound resulting in Hyp formation (Groves and McClusky, 1976).

Ascorbate, the ionized form of L-ascorbic acid, is not required for turnover of P4H, and the enzyme can go through a number of turnovers without ascorbate before loss of activity (Myllyla et al., 1978; Nietfeld and Kemp, 1981). Ascorbate, however, is able to rescue the enzyme when the iron is oxidized and thus inactivated. Ascorbate is able to rescue P4H by reducing the iron(III) to the active iron(II) state (de Jong et al., 1982; de Jong and Kemp, 1984). P4H can catalyze the decarboxylation of α -ketoglutarate without hydroxylation of proline, leading to an uncoupling of α -ketoglutarate and peptide substrate turnover (Counts et al., 1978; Rao and Adams, 1978). The uncoupled reaction leads to inactivation of the enzyme, and under these conditions ascorbate is used stoichiometrically (Myllyla et al., 1984).

Recognition of proline residues that are hydroxylated by P4H

P4H catalyzes hydroxylation of Pro residues in the Yaa position of the repeating Xaa–Yaa–Gly units of collagen strands (Hutton et al., 1967). The hydroxylation reaction is performed on individual procollagen chains and is prevented by the triple-helical conformation (Berg and Prockop, 1973b). Free proline amino acids are not hydroxylated by P4H (Cardinale and Udenfriend, 1974). The minimum substrate required for hydroxylation is an Xaa–Pro–Gly tripeptide, and Pro is the preferred residue in the Xaa position, although hydroxylation can occur at lower rates with a variety of amino acids at this position (Kivirikko et al., 1972; Rapaka et al., 1978).

P4H α interacts with substrates in two sites, the peptide-substrate-binding domain and the active site. The peptide-substrate-binding domain binds to poly(L-proline) type II structures. Poly(L-proline) is not hydroxylated by P4H, though it does bind to the enzyme and is a competitive inhibitor of enzymatic activity (Prockop and Kivirikko, 1969). The K_m for peptide substrates decreases with increasing peptide length (Lamberg et al., 1995; Myllyharju and Kivirikko, 1997). The three-dimensional structure of the peptide binding domain (Phe144–Ser244) was solved by X-ray crystallography (Pekkala et al., 2004). The α -helical structure forms a concave, “bowl-like surface” that contains a number of hydrophobic amino acids that likely form the peptide substrate binding site. The three-dimensional structure of the entire P4H tetramer has not been elucidated. Recently, the structure of yeast PDI was solved (Tian et al., 2006), but provided little insight into how mammalian PDI might associate with P4H α to form a tetramer.

P4H α also contains the catalytic active site. The iron is bound in the active site by two histidine residues and an aspartate residue. This catalytic triad 2-His-1-carboxylate motif is common to the α -ketoglutarate, iron(II) dioxygenases. This domain also has a positively charged residue that binds the carboxylate of α -ketoglutarate. No structural information is available for the catalytic site of P4H, though a model for the binding of peptide substrate to the P4H active site has been proposed. Structural, spectral and computational analyses of Pro–Gly sequences in substrate peptides and proteins suggested that adoption of a β -turn conformation is required for recognition by P4H (Atreya and Ananthanarayanan, 1991; Brahmachari and Ananthanarayanan, 1979; Chopra and Ananthanarayanan, 1982; Rapaka et al., 1978). The β -turn structure forms the structural requirement for binding and catalysis in the active site, and longer substrates having a polyproline II type helical structure add to the binding interaction by making contacts within the peptide-substrate-binding domain. Hydroxylation of the proline residues then results in a “straightening” of that turn, which allows the collagen triple helix to form. In those studies the amino acids surrounding the Pro–Gly sequence were varied to influence the conformation, and it was assumed that the peptide bond was in the *trans* conformation. More recent work describes P4H recognition of the conformation of the proline ring itself, and perhaps the *cis* conformation of the peptide bond. Using peptide substrates containing proline analogues that vary in ring pucker preference, it was determined the P4H recognizes the C $^{\gamma}$ -*endo* ring pucker. The proline analogues determined to be P4H substrates also favored the *cis* peptide bond. Upon hydroxylation, the switch to the C $^{\gamma}$ -*exo* ring pucker and *trans* peptide bond could provide a mechanism to prevent product inhibition for P4H (Gorres et al., 2008). This work is the subject of Chapter 2.

Given that collagen is a polymeric substrate, the question arises as to whether P4H is a processive enzyme. The strength of binding of the collagen peptide to P4H depends greatly on the length of the polypeptide, where the K_m values decrease with increasing peptide length. Yet, no evidence has been presented that consecutive proline residues are hydroxylated by P4H. In peptide fragments derived from collagen, proline residues in the Yaa position are hydroxylated incompletely (Bornstein, 1967a; Bornstein, 1967b). In (Pro-Pro-Gly)₅ collagen model peptides, the proline residues in the 3rd and 4th triplets were hydroxylated preferentially (Kivirikko et al., 1971). P4H tetramers contain two α subunits, each containing an active site. There was evidence that one substrate peptide, if of sufficient length, could interact with both binding sites. This was described as processivity or synergy, (de Waal and de Jong, 1988), and later specified as processive *binding* of the substrate, not necessarily processive hydroxylation (de Jong et al., 1991). That model relies on multiple binding sites, two P4H α subunits per enzyme tetramer. Studies of P4H dimers with only one catalytic subunit from *C. elegans* (Kukkola et al., 2004) and the peptide-substrate-binding domain human P4H α (I) amino acids 138–244 show that the strength of binding is still dependent on peptide length, providing evidence against a processive mode of binding. Mutation in one P4H α active site does, however, decrease the activity of the other active site, even though the binding of iron and α -ketoglutarate are unaffected (Pekkala et al., 2004).

Prolyl 3-hydroxylation

Collagen also contains (2*S*,3*S*)-3-hydroxyproline (3-Hyp), though 3-Hyp is much less abundant than 4-Hyp (Rhodes and Miller, 1978). 3-Hyp is more prevalent in the Type IV

collagen of basement membranes, which contains 10–15 3-Hyp residues, than in Type I and II fibrillar collagens, each having a single 3-Hyp residue. 3-Hyp is formed from Pro in the Xaa position of Xaa–Hyp–Gly sequences (Gryder et al., 1975). The role of 3-Hyp in collagen is not well defined, but is known to destabilize the triple helix of a collagen model peptide (Jenkins et al., 2003). It has been hypothesized that 3-Hyp adjusts the stability of basement membrane collagen in order for the meshwork structure to form. Destabilization could also play a role in protein binding.

3-Hyp is formed by prolyl 3-hydroxylases (P3H; EC 1.14.11.7). Three isoforms of P3H have been identified in vertebrates, and all contain an ER-retention signal. Like P4H, P3Hs also require iron, α -ketoglutarate, O₂, and ascorbate for activity. P3Hs contain the conserved catalytic residues and do not hydroxylate triple-helical collagen. P3H1 was found to be homologous to mammalian leprecan or growth suppressor 1 (Gros1), and to form a tight complex with cartilage-associated protein (CRTAP) and cyclophilin B (CypB) (Vranka et al., 2004). Lack of 3-Hyp in Type I and II collagens leads to an osteogenesis imperfecta (OI)-like disease, demonstrated by a *Crtap* knock-out mouse (Morello et al., 2006) and human mutations in the gene encoding P3H1 (Cabral et al., 2007). The P3H1/CRTAP/CypB complex has also been shown to have chaperone activity (Ishikawa et al., 2009). P3H2 hydroxylated peptides derived from Type IV collagen more efficiently than Type I peptides, and was localized to tissues rich in basement membrane (Tiainen et al., 2008). The effect of prolyl 3-hydroxylation on basement membrane collagens remains unknown.

Lysine hydroxylation

Posttranslational modifications of collagen also include the hydroxylation of lysine residues by lysyl hydroxylase (E.C. 1.14.11.4). The resulting hydroxylysine residues can be oxidized by lysyl oxidase and form covalent cross-links among collagen strands, or the hydroxylysine can be modified further by a galactosyltransferase and a glucosyltransferase (Myllyla et al., 2007). Like P4H, lysyl hydroxylase is an α -ketoglutarate-dependent, iron(II) dioxygenase. Lysine residues in the Yaa position of the (Xaa–Yaa–Gly)_n collagenous sequence, as well as some lysines in the terminal, non-helical regions, are hydroxylated. When recognizing its substrate, lysyl hydroxylase does not hydroxylate triple-helical collagen, and the binding of substrate is influenced by the surrounding amino acids (Risteli et al., 2004).

1.3.2 Collagen-domain-containing proteins

The Xaa–Yaa–Gly amino acid sequence characteristic of collagen is found in other proteins as well. These sequences form triple helices, though the triple-helical domains are usually much shorter than is found in collagen. The prolyl residues within these sequences are still hydroxylated. Collagen-domain-containing proteins have globular domains attached to the triple helix that have distinct functions. The triple-helical domains can act as a spacer between globular domains, as oligomerization domains, or as a binding site for interacting partners.

The asymmetric form of acetylcholinesterase is comprised of catalytic subunits attached to a collagenous, triple-helical domain (Rosenbloom and Cywinski, 1976a). Acetylcholinesterase hydrolyzes the neurotransmitter acetylcholine, which ends signal transmission at neuromuscular junctions. The collagenous domain, like collagen, is comprised of Xaa–Yaa–Gly triplets that contain Hyp. This collagen-like tail domain is responsible for attaching acetylcholinesterase to

the basal lamina via heparin sulfate proteoglycans (Deprez et al., 2003). Mutations that prevent the triple helix from forming result in an absence of acetylcholinesterase at the neuromuscular junction and lead to abnormal functioning (reviewed in (Aldunate et al., 2004)).

The complement protein C1q also contains a collagenous domain. C1q is a subunit of C1 that functions in the complement pathway of the innate immune response. The globular head domains of C1q bind antigen-bound immunoglobulins (Igs), as well as ligands associated with pathogens. The globular domains are held together by an oligomerization domain that is comprised of collagen-like triple helices comprised of an Xaa–Yaa–Gly amino acid sequence containing Hyp (Porter and Reid, 1978). Six trimers of C1q are linked by a collagen microfibril that produces an overall ‘bunch of tulips’ structure for C1q. Individual C1q globular domains bind weakly to IgG and IgM, but oligomerization increases the strength of the interaction with clusters of IgG. The hydroxylation of prolyl residues is critical, as C1q secretion and function is decreased in the presence of the iron chelator α,α -dipyridyl and 3,4-dehydroproline, which are known to inhibit P4H and triple helix formation (Mocharla et al., 1987; Muller et al., 1978).

Collectins are a class of proteins that contain a lectin domain, in addition to collagenous domains (van de Wetering et al., 2004). Also involved in the innate immune response, the lectin domains bind carbohydrates on the cell surface of pathogens. The collagen-like triple-helical domains perform a number of functions. Oligomerization of the lectin domains is accomplished by the collagenous domains. Binding of a single lectin domain to its carbohydrate ligand is weak without multivalency. Inhibition of prolyl hydroxylation, and thus triple-helix formation, in the collectin mannan-binding lectin (MBL) prevents proper oligomerization (Ma et al., 1997). The collagenous domain also dictates the shape and spacing of the lectin domains. The triple helix acts as a spacer in lung surfactant protein D (SP-D), whereas SP-A and MBL have interruptions

in the (Xaa–Yaa–Gly) triple-helical sequence which results in bending of the collagenous domain (Figure 1.4; (Voss et al., 1988)).

The collagenous domains also play a role in the effector function of the collectins and C1q. When a lectin domain binds a ligand, the proteases that bind the collagenous region are activated and lead to initiation of the complement pathway. The collagenous domain can also bind to cell-surface receptors, which can elicit many responses, such as phagocytosis, chemotaxis, coagulation and regulation of the adaptive immune response (Kishore et al., 2006).

Collagens, the collagen-domain-containing proteins discussed thus far, and the hibernation proteins HP-20, 25, and 7 that also contain collagenous domains (Takamatsu et al., 1993) are all secreted proteins. Additionally, there are integral membrane proteins with collagenous domains (Franzke et al., 2005). The macrophage scavenger receptors were the first collagenous membrane proteins to be discovered. The collagen-like domain contains positively charged residues that bind a wide range of negatively charged ligands, including oxidized low-density lipoprotein. Ligand binding can lead to endocytosis, phagocytosis, or mediate adhesion.

1.3.3 Elastin

Elastin is a structural protein that provides elasticity in connective tissues. Elasticity is especially important for blood vessels and lung tissues, which have a high elastin content. The amino-acid composition of elastin is rich in proline and glycine, like that of collagen, but elastin does not have glycine as every third residue, nor does it have a triple-helical structure. Elastin is rich in alanine and valine, and a prototypical elastin sequence is Val–Pro–Gly–Val–Gly, and peptides comprised of (Val–Pro–Gly–Val–Gly)_n repeats are substrates for P4H (Bhatnagar et al., 1978). Hydroxylation of prolyl residues at the 4-position also occurs in elastin by the collagen

P4H, although there is less Hyp in elastin than in collagen (Rosenbloom and Cywinski, 1976a). Hydroxyproline is not required for the synthesis and secretion of elastin (Rosenbloom and Cywinski, 1976b). The role of Hyp in elastin, though, remains unknown. It has been proposed that Hyp in elastin is coincidentally formed in the ER because P4H is located there and the same cells make elastin and collagen. Accumulation of elastin, however, is affected by levels of ascorbic acid. Cell cultures grown in the presence of ascorbate produce overhydroxylated elastin, but less insoluble and cross-linked elastin. It appears the levels of Hyp influence elastin fibril formation (Dunn and Franzblau, 1982).

1.3.4 Prion protein

The conversion of the cellular prion protein (PrP^{C}) to a partially protease-resistant, aggregated scrapie form (PrP^{Sc}) is characteristic of the prion neurodegenerative disorders. The C-terminal portion of PrP is mostly α -helical in the PrP^{C} and changes to all β -sheet in the PrP^{Sc} form. The physiological function(s) of PrP remains unknown, although roles in antiapoptosis, antioxidation, sensing and transport of copper or other metals, neuronal development, differentiation and maintenance, and even with the immune system have been proposed (Marc et al., 2007). The proposed functions are based on interactions between PrP and metals, other proteins, or nucleic acids. The majority of these interactions occur within the N-terminus of PrP. The N-terminal domain of PrP is unstructured, but contains distinct regions of nonapeptide repeats and octapeptide repeats. A portion of the N-terminus also has a polyproline II type helical structure. This region contains a Pro–Gly sequence found to be hydroxylated in PrP produced in mammalian cell culture and from the brains of scrapie-infected mice (Gill et al., 2000). A peptide derived from amino acids 37–53 was found to be hydroxylated in vitro by purified human P4H

(K.L. Gorres, E.S. Eberhardt, R.T. Raines, unpublished results). The biological consequence of this modification is unknown. It is possible that hydroxylation results in structural changes within PrP, or alters the metal–protein or protein–protein interactions required for the normal function of PrP^C or the conversion and transmission of PrP^{Sc}.

1.3.5 Conotoxins

Cone snails (genus *Conus*) produce venomous peptides that often target ion channels in the nervous system. These peptide toxins, known as conotoxins, are translated by the ribosome, and are highly cross-linked by disulfide bonds. Conotoxins also contain a great number of posttranslational modifications, including prolyl hydroxylation (Buczek et al., 2005). Hyp has been identified in a number of conotoxins, and the hydroxylation seems to be specific because some peptides contain both proline and hydroxyproline. Hyp is found in amino acid sequences that are different than the Pro–Gly sequence hydroxylated in collagen, and there is no obvious consensus sequence among the conotoxins that is hydroxylated. No prolyl hydroxylase from *Conus* has been identified and characterized.

Hyp in conotoxins affects peptide folding, structure and biological activity. An NMR structure of the O10P conomarphin variant (where O = Hyp) revealed structural differences compared to the native peptide (Huang and Du, 2009). A study of peptides from each of the μ -, α -, and ω -conotoxin families showed the effect of Hyp varied (Lopez-Vera et al., 2008). Removal of all three Hyp hydroxyl groups in the μ -GIIIA conotoxin slightly increased the rate of folding, but greatly decreased its biological activity. In these and other peptides Hyp mediates the conotoxin peptide–protein interaction. The α -conotoxins do not contain Hyp, however the

Pro to Hyp mutation in α -ImI and α -GI increased the rate of folding and reduced the bioactivity. Hyp in the α -conotoxins interrupts peptide–protein interaction. Hyp has no effect on biological activity in ω -MVIIC conotoxin. The yield of folded peptide and rate of folding, however, were improved. And so, the role of Hyp in conotoxins can be either to stabilize structure or enhance molecular recognition. Because conotoxins are highly modified, it may also be possible that one modification, such as Hyp, alters other modifications.

1.3.6 Argonaute 2

RNA interference (RNAi), gene silencing resulting from double-stranded RNA, is performed in cells by RNA-induced silencing complexes (RISCs). RISCs are comprised of small interfering RNAs (siRNAs) and proteins from the Argonaute family. Argonaute 2 (Ago2) cleaves target mRNA (Liu et al., 2004). Ago2 was found to interact with collagen P4H, and Hyp was found at Pro700 in Ago2 via mass spectrometric analysis (Qi et al., 2008). Pro700 is located within a Pro–Gly sequence that is typically hydroxylated in collagen. Other proline residues within Pro–Gly sequences of Ago2 were not hydroxylated, suggesting specificity. Although Ago2 is mostly located in the cytoplasm, there is evidence for some Ago2 in the ER, where collagen P4H is located. Furthermore, the prolyl hydroxylase domain proteins (see below) did not hydroxylate Ago in vitro.

The result of hydroxylation of Ago2 at Pro700 is increased protein stability (Qi et al., 2008). The P700A variant of Ago2 has less conformational stability than does the wild-type enzyme, and the half-life of Ago2 is decreased in cells when P4H is inhibited. Hydroxylation was not required for the catalytic activity of Ago2 or for siRNA binding. Hydroxylation could alter the cellular localization of Ago2. The mechanism of hydroxyproline stabilization of Ago2 remains to

be determined. The degradation of Ago2 appears to be proteasome-mediated, but what is the role of Hyp? Does Hyp stabilize the structure of Ago2, as it does for collagen? Does Hyp promote the binding of another protein that stabilizes Ago2, or does the absence of Hyp allow recognition of Ago2 and its direction to the proteasome, as in the degradation of HIF (see below)? It will be interesting to learn whether prolyl hydroxylation affects other posttranslational modifications of Ago2, such as ubiquitination.

1.4 Prolyl Hydroxylase Domain Protein (PHD)

1.4.1 Hypoxia inducible factor α

The detection of molecular oxygen (O_2) levels and homeostasis in animals are controlled through the hypoxia-inducible transcription factors (HIFs). HIFs direct the transcription of >100 genes through regulatory hypoxia response elements (HRE) (Ke and Costa, 2006). HIF-regulated genes are involved in cell proliferation, angiogenesis, erythropoiesis, and metabolism. The principal HIF, HIF-1, is comprised of two subunits, HIF-1 α and HIF-1 β , which are both constitutively expressed. The level of HIF-1 α , however, is regulated by the availability of oxygen. As shown in Figure 1.5, under normal oxygen conditions, HIF-1 α is polyubiquitinated and degraded rapidly by the proteasome. During hypoxia, HIF-1 α is not degraded, translocates to the nucleus, and dimerizes with HIF-1 β to form the active transcription factor.

The concentration of oxygen is sensed by oxygenase enzymes that catalyze the posttranslational hydroxylation of HIF- α (reviewed in (Chowdhury et al., 2008; Kaelin and Ratcliffe, 2008)). Under normal oxygen conditions (normoxia), prolyl hydroxylase domain proteins (PHDs) hydroxylate two highly conserved proline residues (Pro402 and Pro564) in HIF-

1 α . These proline residues are located within the oxygen-dependent degradation domain (ODDD). The presence of Hyp within the ODDD of HIF-1 α is recognized by the von Hippel-Lindau tumor suppressor protein (pVHL), which is a component of a ubiquitin–protein E3 ligase complex along with elonginB, elonginC, cul2, and rbx1. Upon hydroxylation, HIF-1 α is recognized by the ubiquitin E3 ligase, polyubiquitinated, and directed to the proteasome for degradation (Figure 1.5). Under hypoxic conditions PHD activity is decreased due to its use of O₂ as a cosubstrate. Without proline hydroxylation, HIF-1 α is not directed for degradation. HIF-1 α can then dimerize with HIF-1 β to activate transcription.

The interaction between HIF-1 α and the pVHL–elonginC–elonginB (VCB) complex is controlled by prolyl hydroxylation. A 20 amino acid peptide derived from HIF-1 α that encompasses Pro564 can be hydroxylated by HIF-P4H and then recognized by pVHL. The mode of the interaction was revealed by the three-dimensional structure of the VCB complex co-crystallized with a HIF-1 α peptide containing Hyp. The HIF-1 α peptide and the pyrrolidine ring of Hyp form contacts with hydrophobic areas of pVHL. The hydroxyl group of Hyp564 in the HIF-1 α peptide forms hydrogen bonds with the hydroxyl group of Ser111 and the imidazole group of His115 in pVHL (Figure 1.6; (Hon et al., 2002; Min et al., 2002)). The presence of Hyp in a peptide fragment of HIF-1 α increases its affinity for the VCB complex approximately 1000-fold. The dissociation constant of the complex between the HIF-1 α peptide containing Hyp and pVHL is $K_d = (0.22 \pm 0.04) \mu\text{M}$, whereas an interaction was not detected with the peptide containing Pro.

HIF Prolyl hydroxylases

There are three isoforms of prolyl hydroxylase domain proteins (PHDs), PHD1–3, also known as HIF-P4Hs (HPHs) 3–1, and EGLNs 2, 1 and 3. The PHDs are like collagen P4H in that they require iron(II), α -ketoglutarate, and molecular oxygen for catalytic activity, and the PHDs have the 2-His-1-Asp iron-binding motif (Bruick and McKnight, 2001). PHDs also use ascorbate as a cosubstrate, and are thought to go through a similar mechanism. PHDs, however, are distinct from the P4H involved in collagen biosynthesis—PHDs are cytosolic enzymes. The apparent K_m value for O_2 for PHDs is higher than collagen P4H and is greater than the oxygen concentration in tissues, which allows the enzyme activity to correspond to oxygen concentration throughout the biological range (Ehrismann et al., 2007; Hirsila et al., 2003).

The two prolyl residues in HIF α hydroxylated by PHDs are located in LXXLAP motifs. The preferences for the N-terminal oxygen-dependent degradation domain (ODDD) versus the C-terminal ODDD vary among the HIF α and PHD isoforms. Collagen P4H cannot hydroxylate the LXXLAP motif in HIF α (Jaakkola et al., 2001). Recognition of the sequence by PHDs, however, is quite flexible, though the alanine residue is the strictest requirement (Li et al., 2004). The minimum peptide length is eight amino acids, but peptides of 19–20 residues are hydroxylated much more efficiently. There is no evidence for secondary structural requirements within the HIF α peptide for PHD recognition. The proline residue is an absolute requirement in the HIF α peptide, even for binding to PHD. When proline is replaced with the proline analogues 3,4-dehydropyrroline or L-azetidine-2-carboxylic acid, the rate of uncoupled α -ketoglutarate decarboxylation increases. Other proline analogues, (2*S*,4*S*)-4-hydroxypyrroline (hyp) and (2*S*,4*S*)-4-fluoropyrroline (flp), are substrates for PHD when incorporated in to peptides derived from

HIF-1 α (Loenarz et al., 2009). The structure of a HIF α -derived peptide bound to PHD2 revealed the substrate prolyl residue in a C $^{\gamma}$ -endo ring pucker conformation (Figure 1.8) (Chowdhury et al., 2009). These results indicate that PHDs recognize the prolyl ring pucker in a manner similar to P4H (Gorres et al., 2008).

Asparagine hydroxylation

In addition to prolyl hydroxylation, oxygen sensing and control of HIF-1 α is also accomplished by factor-inhibiting hypoxia-inducible factor 1 (FIH) asparaginyl hydroxylase. FIH is also a member of the α -ketoglutarate-dependent, iron(II) dioxygenase family. Under normoxic conditions, FIH hydroxylates Asn803 in HIF-1 α , which prevents an interaction with the transcriptional coactivator p300 (Lancaster et al., 2004). Hypoxia diminishes asparagine hydroxylation, allowing the coactivators to interact with HIF-1 α and activate transcription. Three-dimensional structures have confirmed the 2-His-1-carboxylate active site motif in FIH (Dann et al., 2002; Elkins et al., 2003; Lee et al., 2003). Like P4H, FIH has two sites for interacting with its peptide substrate. In addition to the active site there is a second peptide-binding site that is important for tight binding. Compared to PHDs, FIH recognizes an even longer portion of HIF-1 α , nearly all of a 35 amino acid peptide is needed for substantial activity (Koivunen et al., 2004). The asparagine that is hydroxylated is critical for FIH binding, and the preceding valine is also important, though many other neighboring residues can be substituted without effect (Linke et al., 2004).

Transmembrane prolyl 4-hydroxylase

A recently discovered prolyl 4-hydroxylase, PH-4 or P4H-TM, contains a transmembrane domain near its N-terminus (Koivunen et al., 2007; Oehme et al., 2002). P4H-TM is associated with the membrane of the endoplasmic reticulum. By comparison of amino acid sequences, P4H-TM is related more closely to the catalytic C-terminal region of collagen P4H than to the PHDs, though P4H-TM does not show any sequence similarity to the N-terminal peptide-substrate-binding domain of P4H. P4H-TM, however, decreases transcriptional activation by HIF-1 α . In vitro, P4H-TM hydroxylates HIF-1 α and does not hydroxylate collagen, even though the active site resides in the ER lumen. P4H-TM expression is induced under hypoxic conditions in cell culture, although its cellular location does not change. How the active site of P4H-TM inside the ER can act upon a typically cytosolic protein and the role of the cellular localization of P4H-TM is not known. It is possible that P4H-TM has a specialized function in regulating HIF-1 α . Alternatively, HIF-1 α might not be the primary substrate, and P4H-TM could be active in other pathways.

1.4.2 Large subunit of RNA polymerase II

The RNA polymerase II complex, responsible for transcribing DNA into mRNA, transitions from transcription initiation to elongation through phosphorylation of the C-terminal domain of the large subunit Rpb1. In response to UV irradiation or oxidative stress, hyperphosphorylated Rpb1 is bound by pVHL and ubiquitinated. Ubiquitylation of Rpb1 does not lead to its degradation. Binding of pVHL is dependent on the hyperphosphorylation of the C-terminal domain of Rpb1 and the hydroxylation of Pro1465 (Kuznetsova et al., 2003).

Rpb1 shares some amino acid sequence similarity with HIF-1 α , including a LXXLAP motif, suggesting the involvement of a PHD enzyme rather than the collagen P4Hs. PHD1 was found to be the main catalyst of Rpb1 prolyl hydroxylation (Mikhaylova et al., 2008). Surprisingly, PHD2 inhibited hydroxylation of Pro1465 and phosphorylation. The role of prolyl 4-hydroxylation in Rpb1, as in HIF-1 α , is to recruit pVHL. Rpb1 is translocated from the soluble fraction to the chromatin-engaged fraction by pVHL under conditions of oxidative stress. The PHDs were also found in the chromatin fraction. The consequences of Rpb1 hydroxylation and pVHL binding within a cell have not yet been determined. Regulation of Rpb1, and thus RNA polymerase, by pVHL could be involved in transcription elongation that alters gene expression during stresses that result in DNA damage.

1.4.3 I κ B kinase- β

NF κ B is a transcription factor involved in fundamental aspects of the innate immune response and inflammation, and is also important for the progression of tumor development. Hypoxia has been shown to activate NF κ B. The link between oxygen sensing and NF κ B appears to be prolyl hydroxylation by the same PHD that is crucial to oxygen sensing and HIF response. Inhibition of PHD, particularly PHD1, by either small-molecule inhibitors or siRNA resulted in NF κ B activation (Cummins et al., 2006). Conversely, overexpression of PHD1 under normal oxygen conditions caused a decrease in NF κ B activity. PHD does not, however, act directly on NF κ B.

NF κ B is controlled by a cascade of inhibitory proteins. NF κ B is sequestered in the cytosol by its interaction with Inhibitory κ B (I κ B). Phosphorylation of I κ B by I κ B kinase- β (IKK β) leads to

the ubiquitinylation and degradation of I κ B, which exposes the nuclear localization signal of NF κ B. IKK β contains a conserved LXXLAP motif, which is the same sequence that is required for hydroxylation in HIF α . When the proline in the LXXLAP motif in IKK β is replaced (as in the P191A variant), NF κ B is no longer induced by hypoxia. Hydroxylation of these substrates remains to be confirmed by mass spectrometry.

The mechanism by which prolyl hydroxylation inhibits phosphorylation by IKK β could be by altering the protein conformation, as in collagen, or interrupting a protein–protein interaction, as is the case for HIF. It is proposed that hydroxylation at Pro191 changes the conformation of the activation loop, making the kinase inactive. Alternatively, hydroxylation could disrupt the binding of the substrate. Hydroxylation could also induce the binding of another protein, possibly pVHL, which would block IKK β from being phosphorylated and activated.

1.4.4 Activating transcription factor 4 (ATF-4)

There is also evidence for prolyl hydroxylation-dependent degradation of activating transcription factor 4 (ATF-4; (Koditz et al., 2007)). ATF-4 was found to interact with PHD3, but not PHD1 or PHD2. Like HIF-1 α , ATF-4 was stabilized by PHD inhibitors, hypoxia, and proteasome inhibitors. The interaction was mapped to a portion of the zipper II domain, which contains five proline residues, although none are within an LXXLAP motif. ATF-4 variants lacking this region or all five proline residues were more stable than wild-type ATF-4. Individual proline variants, however, did not have the same effect. The combinations of Hyp residues required for protein stabilization has not been determined. ATF-4, incubated under appropriate conditions for prolyl hydroxylation, did not interact with pVHL. Rather, degradation of ATF-4

was found to be dependent on the SCF^{βTrCP} ubiquitin ligase (Lassot et al., 2001). Whether prolyl hydroxylation is required for this interaction, or an interaction with another ubiquitin ligase or some adaptor protein, or whether the hydroxylation has a structural effect remains to be determined.

1.4.5 Iron regulatory protein 2 (IRP2)

Iron regulatory protein 2 (IRP2) senses iron in mammalian cells and postranscriptionally regulates expression of genes containing iron-responsive elements. IRP2 is stable and binds RNA at low iron concentration or under hypoxia, but is degraded by the proteasome at higher iron levels (Hanson et al., 1999). This control of IRP2 is similar to that of HIF α . The iron-mediated degradation of IRP2 is blocked by the presence of an α -ketoglutarate analogue that is a known inhibitor of α -ketoglutarate-dependent prolyl hydroxylases (Hanson et al., 2003; Wang et al., 2004). This evidence, along with the iron and oxygen dependence, suggests the involvement of an α -ketoglutarate-dependent dioxygenase. The residue(s) modified has not been identified, though the C-terminal domain has been identified as necessary for iron-dependent degradation (Wang et al., 2008). This region contains a number of proline residues that could be modified by a prolyl hydroxylase, though none are located in a LXXLAP motif typical for PHD hydroxylation. Hydroxylation by another α -ketoglutarate-dependent dioxygenase, such as asparaginyl hydroxylase, also cannot be ruled out.

1.5 Plant and algal prolyl 4-hydroxylases

Prolyl 4-hydroxylation occurs in a number of proteins in plants and algae. Peptides containing Hyp are part of systemin defense mechanisms (Pearce et al., 2009; Ryan and Pearce, 2003), and Hyp is found in some secreted and vacuolar proteins (Shimizu et al., 2005). Hyp is very abundant in a large class of proteins, termed hydroxyproline-rich glycoproteins (HPRGs) in which 15–25% of the amino acids are Hyp. HPRGs are the major proteinaceous components of the cell walls in higher plants and green algae. In addition to functions in cell wall self-assembly and rigidity, HPRGs also play roles in plant growth, development, cell–cell interactions, and cellular communication (Wu et al., 2001). The HPRGs are subgrouped by the type of amino acids in characteristic repetitive amino acid sequences. The extensins typically contain a Ser–Hyp₄ motif, the repetitive proline-rich proteins have variations of pentapeptide repeats containing Hyp and little Ser, the arabinoglactan proteins contain Hyp alternating with other amino acids, and other HRGPs have contiguous Hyp residues (Kieliszewski and Lamport, 1994; Kieliszewski and Shpak, 2001).

Some Hyp residues in plants and algae are modified further by the addition of oligoarabinose or arabinogalactan. Glycosylation of Hyp has not been found in animals. The extent and type of *O*-glycosylation is predicted by the Hyp contiguity hypothesis, where glycosylation correlates with the location and context of Hyp residues (Kieliszewski, 2001). Where Hyp residues are contiguous in the amino acid sequence, arabinosylation is predominant, whereas arabinogalactans are added to clustered, non-contiguous Hyp residues.

Plant prolyl 4-hydroxylases have been cloned from *Arabidopsis thaliana* (Hieta and Myllyharju, 2002; Tiainen et al., 2005), *Nicotiana tabacum* (Yuasa et al., 2005), and the green alga *Chlamydomonas reinhardtii* (Keskiäho et al., 2007). Prolyl 4-hydroxylases in plants, like in

animals, require Fe(II), oxygen, ascorbate and α -ketoglutarate. In general, plant P4Hs are smaller in size, ~30–60 kD, compared to collagen P4H. Plant and algal P4Hs are soluble monomers, and the three-dimensional structure of the *C. reinhardtii* P4H has been solved (Koski et al., 2007). The 2-His-1-carboxylate iron-binding residues and overall structure are consistent with what is known for P4H and PHD. The Cr-P4H-1, however, seems to be more similar to P4H in that it contains a polyproline-binding domain. An N-terminal transmembrane domain was identified in a P4H from *N. tabacum*, and is predicted by sequence analysis in other plant P4Hs. This membrane-bound P4H localizes to the ER and Golgi (Yuasa et al., 2005).

The plant prolyl 4-hydroxylases differ substantially from the animal enzymes in their substrate specificity. P4Hs isolated from plants can hydroxylate polyproline, which is a competitive inhibitor of the collagen P4Hs. Peptides that mimic collagen, (Xaa-Pro-Gly)_n, are hydroxylated by some plant P4Hs, though generally not very efficiently (Kaska et al., 1987; Tanaka et al., 1981). The *A. thaliana* At-P4H-1 enzyme does hydroxylate collagen-like peptides, as well as a peptide derived from HIF α with only one proline residue, suggesting a polyproline II type structure might not be required for hydroxylation. The At-P4H-2 enzyme, however, did not efficiently hydroxylate either the collagen-like peptide or the HIF α peptide. Both At-P4H-1 and At-P4H-2 did hydroxylate peptides representing the plant proline-rich proteins, arabinogalactan protein and extensin.

Despite differences in the amino acid sequence of native substrates for plant and animal prolyl 4-hydroxylases, the recognition of the prolyl residue through its ring pucker seems to be related. P4H and PHD prefer substrates containing proline analogues that favor the C $^{\gamma}$ -endo ring pucker, and do not bind Hyp-containing peptides that favor the C $^{\gamma}$ -exo ring pucker (Gorres et al.,

2008; Loenarz et al., 2009). Similarly, the structure of the algal P4H, Cr-P4H-1, complexed with a (Pro-Ser)₅ peptide substrate revealed the Pro in the active site in the C^γ-endo ring pucker (Figure 1.8) (Koski et al., 2009). In fact, Tyr140 in the Cr-P4H-1 active site would prevent Hyp from binding. Upon prolyl hydroxylation, the change in ring pucker from C^γ-endo to C^γ-exo could cause the tyrosine residue to alter its conformation, leading to opening of the active site to allow product release.

1.6 Prolyl 4-hydroxylases in microorganisms

1.6.1 Protozoa

Skp1 is a eukaryotic protein that is a subunit in a number of multi-subunit complexes, though most well studied as an adaptor in the SCF (Skp1-cullin-F box protein) E3 ubiquitin ligase complex. Uniquely in the amoeba *Dictyostelium*, commonly referred to as slime mold, Skp1 is glycosylated on a hydroxylated proline residue (West et al., 2004). Pro143 is hydroxylated, likely forming (2*S*,4*R*)-4-hydroxyproline. A pentasaccharide is then added to Hyp by five glycosyltransferases. Glycosylation of hydroxyproline is common in secreted plant cell wall proteins. Skp1, however, is a cytoplasmic and nuclear protein, and so hydroxylation and glycosylation of Skp1 takes place in the cytosol, rather than in the organelles of the secretory pathway.

A gene encoding P4H from *Dictyostelium discoideum*, *phyA*, has been cloned and characterized (van der Wel et al., 2005). Activity of recombinant *Dictyostelium* P4H1, DdP4H1, required α -ketoglutarate, ascorbate, and O₂, and activity decreased in the presence of iron chelators. Recombinant Skp1 was used as the protein substrate, but a peptide derived from Skp1 was not hydroxylated. DdP4H1 was found to be a soluble and cytoplasmic protein. The *phyA*

gene for DdP4H1 contains the conserved 2-His-1-Asp iron-binding residues and is related more closely to the PHDs than the P4Hs. The hydroxylated proline in Skp1 is not, however, within an LXXLAP motif. Like PHDs, DdP4H1 appears to sense oxygen in order to regulate *Dictyostelium* development (West et al., 2007).

BLAST searches identified P4H-like sequences in the genomes of other eukaryotic microorganisms, such as the diatom *Thalassiosira pseudonana* and the oomycete *Phytophthora sojae*. Interestingly, these genes are predicted to be bifunctional, encoding the first glycosyltransferase in the pathway in addition to a P4H. The P4H/glycosyltransferase pathway might also exist in *Toxoplasma gondii*, the causative agent of toxoplasmosis (West et al., 2006).

1.6.2 Bacteria

Hyp is also found in bacterial antibiotic peptides. These peptides are synthesized by enzymatic pathways rather than by the ribosome. These non-ribosomal peptides often contain a high percentage of non-natural and modified amino acids, including Hyp. Like in animals and plants, Hyp is formed by stereospecific hydroxylation at the 4*R*-position (Baldwin et al., 1993), and the hydroxyl oxygen of Hyp comes from O₂ (Diegelmann et al., 1969). In addition to Hyp, other isomers of hydroxyproline and other proline modifications occur in bacteria. Hyp and (2*S*)-4-ketoproline (Kep) are found in actinomycins produced by *Streptomyces antibioticus* (Katz et al., 1962), the two diastereomers of (2*S*)-3-hydroxyproline are found in telomycin (Sheehan et al., 1968), (2*R*,4*R*)-hydroxyproline is found in etamycin (Katz et al., 1979), and (2*S*,4*S*)-hydroxyproline is found in microcolin A (Koehn et al., 1992). Pro is the precursor to all the different forms of hydroxyproline in bacteria. A major difference from all other organisms,

however, is that bacterial Hyp is produced from free proline instead of peptidyl-proline (Adefarati et al., 1991).

The enzymes that catalyze the hydroxylation of free proline are identified as the proline hydroxylases, differentiated from the prolyl hydroxylases that hydroxylate peptidyl-proline. A proline 4-hydroxylase and a proline 3-hydroxylase have been purified from *Streptomyces*. Proline 4-hydroxylase forms Hyp in the production of etamycin (Lawrence et al., 1996), and proline 3-hydroxylase catalyzes the formation of (2*S*,3*S*)-hydroxyproline (Mori et al., 1997). A proline 4-hydroxylase converting (2*S*)-proline to (2*S*,4*S*)-hydroxyproline has also been discovered recently (Hara and Kino, 2009). These proline hydroxylases are thought to be members of the non-heme iron(II) dioxygenase family, like P4H and PHD. They also require iron, α -ketoglutarate and O₂. The proline hydroxylases seem to show less substrate specificity than P4H in that proline analogues (2*S*)-3,4-dehydropyrolidine and L-pipecolic acid are substrates (Baldwin et al., 1994). A three-dimensional structure of proline 3-hydroxylase confirmed the 2-His-1-Asp iron-binding residues in the active site (Clifton et al., 2001). The structure also revealed a number of charged amino acids in the active site that could be important for binding the amino and carboxyl groups of the proline substrate, which presumably is different than in peptidyl-prolyl hydroxylases.

Although proline hydroxylation in bacteria occurs mainly on free (2*S*)-proline, a bacterial peptidyl-prolyl hydroxylase has recently been reported. A P4H from *Bacillus anthracis*, designated anthrax-P4H, was found to be homodimeric and iron(II)-, α -ketoglutarate-, and O₂-dependent (Miller et al., 2008). Unlike other bacterial hydroxylases that hydroxylate free proline, anthrax-P4H binds the collagen-like peptide (Gly-Pro-Pro)₁₀ with a similar K_M value as P4H. The physiological substrate and role of anthrax-P4H remains to be elucidated.

1.6.3 Virus

An enzyme catalyzing prolyl hydroxylation has also been identified in the eukaryotic algal virus *Paramecium bursaria* chlorella virus-1 (PBCV-1) (Eriksson et al., 1999). The PBCV-1 prolyl 4-hydroxylase sequence shows similarity to the C-terminal region of the catalytic subunit of P4H. PBCV-1 P4H is a monomer, and can hydroxylate collagen-like peptides, as well as polyproline, the typical plant P4H substrate. The viral genome contains open reading frames for proteins with proline-rich repeats, and peptides containing these (Pro-Ala-Pro-Lys)_n proline-rich sequences are hydroxylated by the viral P4H. The natural viral substrate and the function of hydroxylation are unknown.

1.7 Protein structure

All prolyl and proline hydroxylases are members of a family of enzymes that utilize iron(II), α -ketoglutarate, and molecular oxygen, and most show increased activity in the presence of ascorbate. Studies on a number of α -ketoglutarate-dependent iron(II) dioxygenases have revealed a common iron-binding motif that includes two His residues and one Asp/Glu residue (Schofield and Zhang, 1999). The exception to the 2-His-1-carboxylate motif is the recently discovered halogenases, which catalyze the addition of a halide ion instead of a hydroxyl moiety (Blasiak et al., 2006; Wong et al., 2009). In the halogenases, the Asp/Glu is replaced by an alanine residue and a chloride ion. All of these enzymes also contain a conserved basic residue, an arginine or lysine, that interacts with the carboxylate of α -ketoglutarate. Overall, these enzymes show low sequence homology, but do share a common three-dimensional structural

fold. Within a β barrel jelly roll motif the 2-His-1-carboxylate motif in the active sites are all positioned in a similar manner, illustrated in Figure 1.7 with the structures of PHD2 (McDonough et al., 2006), proline 3-hydroxylase (Clifton et al., 2001), Cr-P4H-1 (Koski et al., 2007), FIH-1 (Dann et al., 2002; Elkins et al., 2003; Lee et al., 2003), and the halogenase SyrB2 (Blasiak et al., 2006). Outside the jelly roll motif, the structures among this family of enzymes vary due to the wide variety of substrates that each enzyme binds.

1.8 Conclusions

The prolyl hydroxylases and proteins containing Hyp are summarized in Table 1.2. The protein substrates are quite diverse, with the proline being the only overall commonality. The P4Hs involved in collagen biosynthesis mainly recognize the characteristic (XPG)_n collagen sequence, but proline residues preceding glycine (PG) in non-collagenous proteins are also hydroxylated. The hydroxylated prolines in conotoxins, though, do not seem to be in any consensus sequence. The LXXLAP motif is a common substrate for PHDs, as in HIF α , RNA pol II Rpb1, and IKK β , but proteins without this motif also might prove to be substrates for PHDs. Plant P4Hs hydroxylate sequences rich in proline residues with a variety of repetitive motifs, and bacteria are unique in that they hydroxylate proline as a free amino acid.

The function of Hyp also varies greatly. In collagen, Hyp plays a structural role preorganizing the collagen strands to stabilize the triple-helical structure. Hyp can also act as a recognition motif for protein–protein interactions that can lead to a variety of consequences. Hyp allows pVHL recognition of HIF α that leads to protein degradation, conotoxin binding to target ion channels, and bacterial nonribosomal peptide antibacterial activity. Prolyl hydroxylation

inhibits the enzymatic activity of IKK β which may be caused by a change in protein conformation or another protein binding. In plants and algae, Hyp is abundant and provides a substrate for the addition of sugars that have many functions on the cell surface. Intriguingly, the biological consequences of the presence of Hyp in place of Pro in a number of proteins are yet to be determined. Moreover, it is likely that prolyl hydroxylation, a small posttranslational modification, will be discovered in additional proteins and organisms.

Table 1.1

The mid-point of thermal stability (T_m) values for collagen-like peptides that vary in the Yaa position.

Peptide (Pro-Yaa-Gly) _n	T_m (°C)
(Pro-Flp-Gly) ₁₀	91 ^a
(Pro-Hyp-Gly) ₁₀	61–69 ^a
(Pro-Pro-Gly) ₁₀	31–41 ^a
(Pro-hyp-Gly) ₁₀	No helix ^b
(Pro-Flp-Gly) ₇	45 ^c
(Pro-Hyp-Gly) ₇	36 ^c
(Pro-Pro-Gly) ₇	No helix ^d
(Pro-flp-Gly) ₇	No helix ^c

^a Ref. (Holmgren et al., 1999)

^b Ref. (Inouye et al., 1976)

^c Ref. (Bretscher et al., 2001)

^d Ref. (Hodges and Raines, 2005)

Table 1.2

Summary of prolyl and proline hydroxylases, the proteins that contain hydroxyproline, the amino-acid sequence in which Hyp occurs, and the function of hydroxylation.

Enzyme	Substrate	Sequence	Function
P4H	Collagen	(XPG) _n	Structure
	Collagen-domain proteins	(XPG) _n	Structure
	Elastin	PG	?
	Prion protein	PG	?
	Conotoxin	No consensus	Structure and Activity
	Ago2	PG	Protein stability
P3H	Collagen	(PHypG) _n	?
PHD	HIF α	LXXLAP	Protein interaction
	RNA pol II Rpb1	LXXLAP	Protein interaction
	I κ B kinase β	LXXLAP	Activity
	ATF 4	Not LXXLAP	Protein stability
	IRP	Not LXXLAP	Protein stability
PH-4 ER	HIF-1 α	LXXLAP	?
Plant and Algal P4H	HPGPs	Poly(L-proline) and Proline-rich sequences	Glycosylation
DdP4H1	Skp1	KNDFTPEEEQIRK	Glycosylation
Proline hydroxylases	Peptide antibiotics	Free L-proline	Antibacterial activity
Anthrax-P4H	?	?	?
PBCV P4H	?	?	?

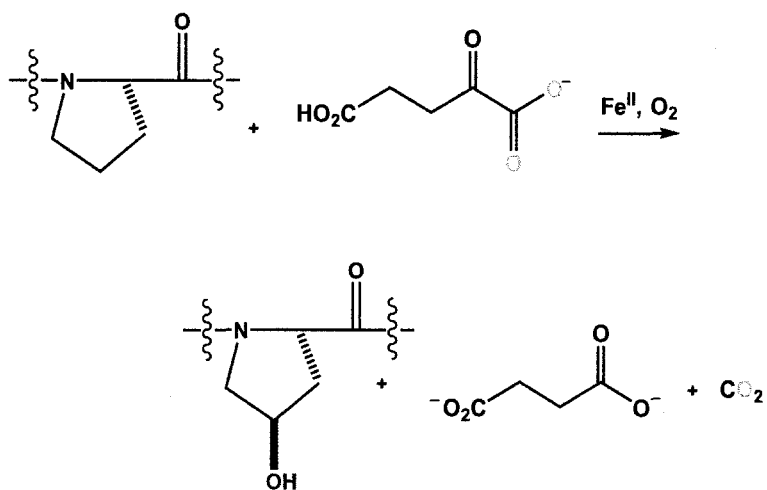


Figure 1.1 Reaction catalyzed by prolyl 4-hydroxylase.

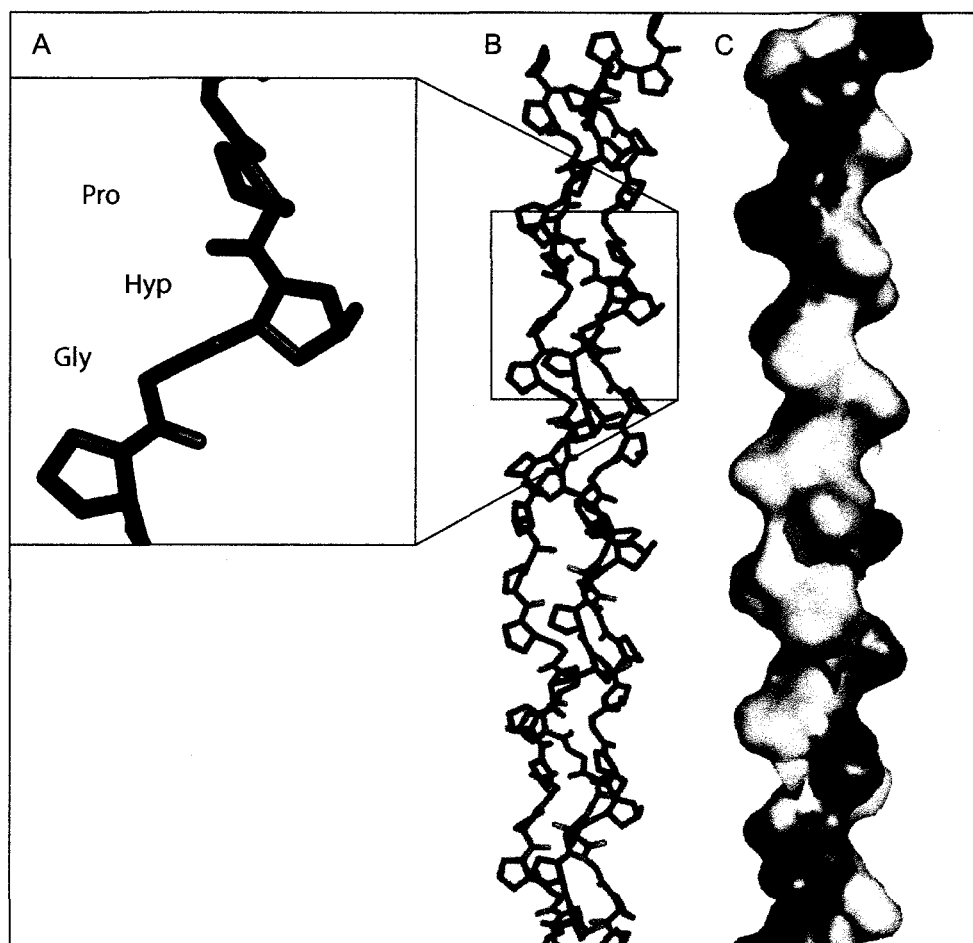


Figure 1.2 Crystallographic structure of the collagen triple helix from the collagen peptide mimic $[(\text{Pro-Hyp-Gly})_4\text{-Pro-Hyp-Ala-(Pro-Hyp-Gly)}_5]_3$ (PDB entry 1CAG; (Bella et al., 1994)). A) A segment of a single strand shows the structure of a Pro-Hyp-Gly sequence. B) A stick representation of the triple helix. In panels A) and B) the stick representation shows atoms colored for carbon (black), nitrogen (blue), and oxygen (red). C) The surface diagram shows each strand indicated with a different color.

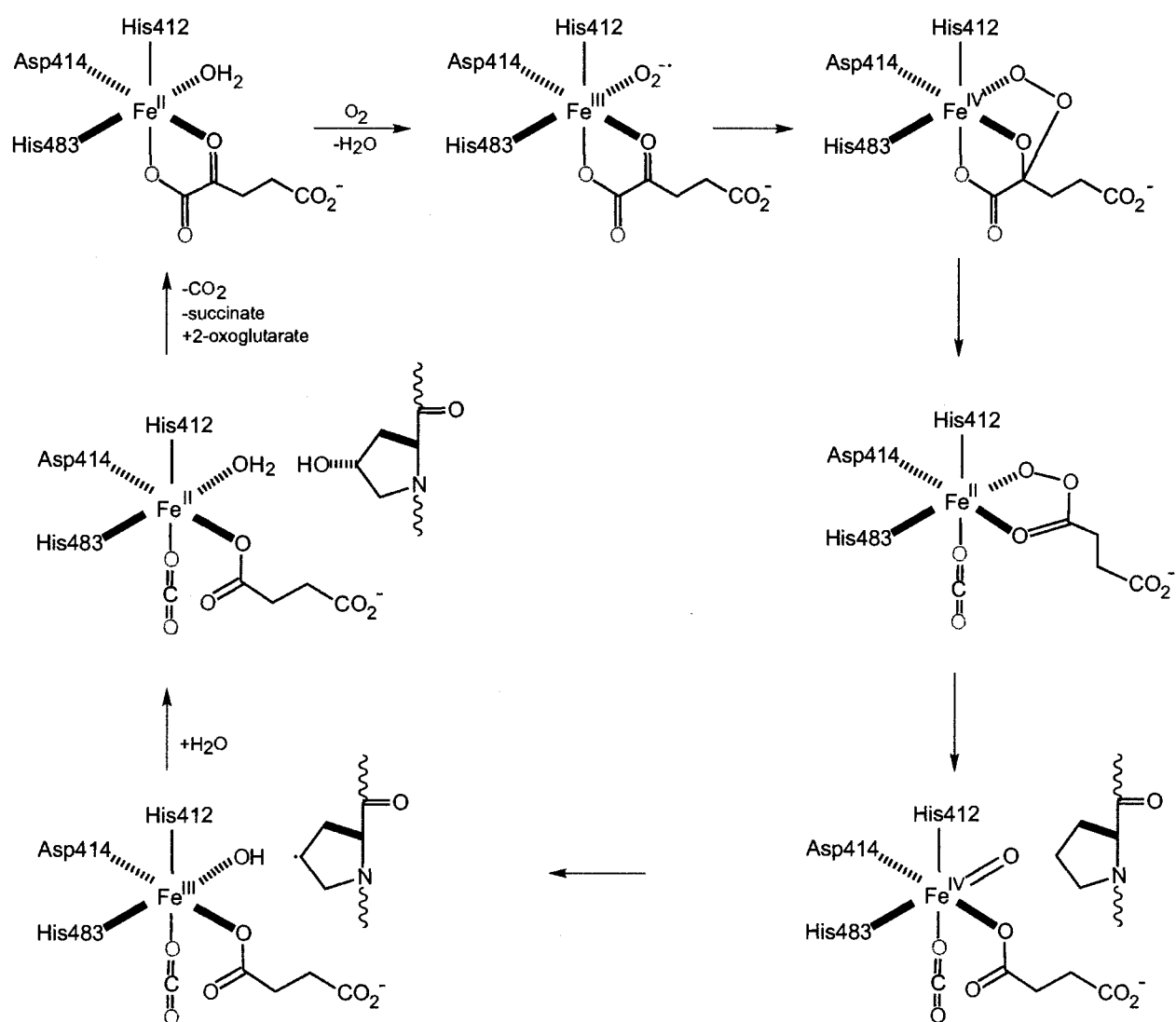


Figure 1.3 Putative mechanism for prolyl hydroxylation by P4H.

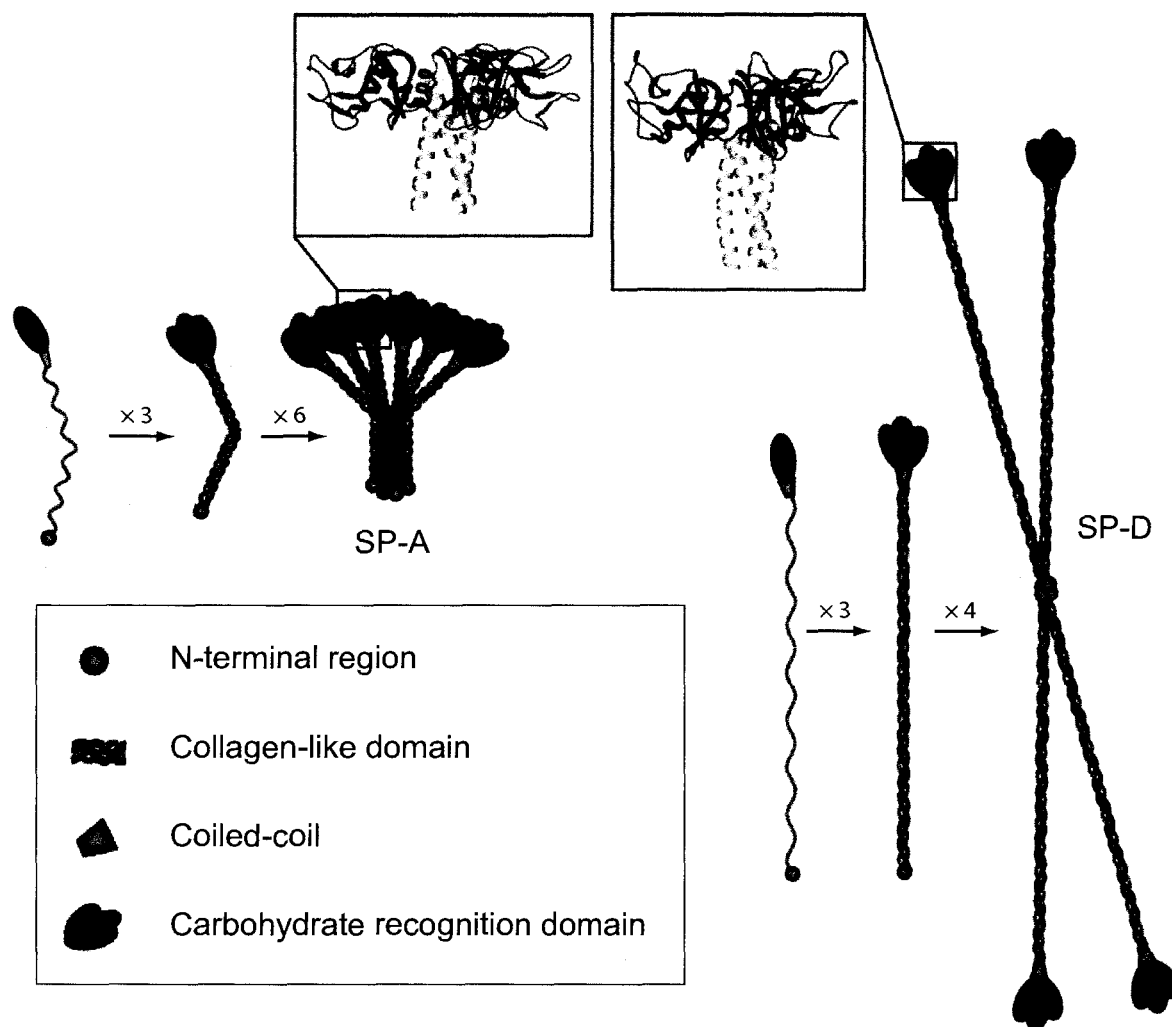


Figure 1.4 Surfactant proteins A and D (SP-A and SP-D). SP-A forms a ‘bunch of tulips’ overall structure composed of 18 proteins with 6 sets of triple helices. SP-D forms from 12 proteins with 4 sets of triple helices. Figure modified from (Kishore et al., 2006).

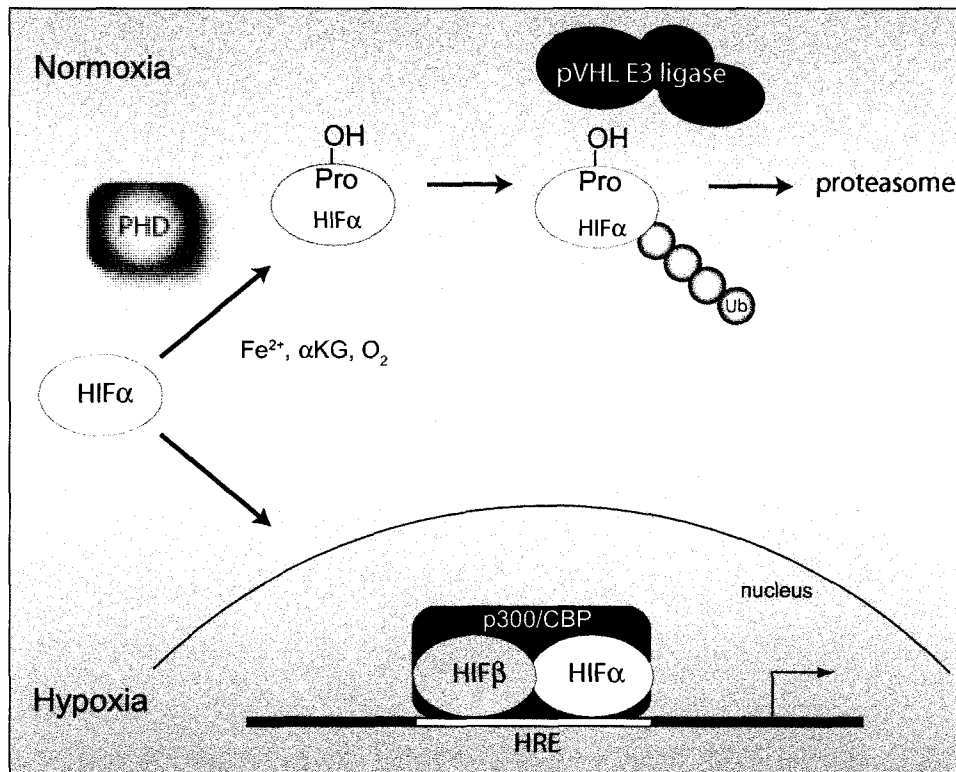


Figure 1.5 Hypoxia sensing pathway. Under normoxic conditions, hypoxia inducible factor-1 α (HIF-1 α) is hydroxylated by prolyl hydroxylase domain-containing proteins (PHDs), and then recognized by pVHL and targeted for degradation by the proteasome. During hypoxia, HIF-1 α is not degraded and translocates to the nucleus to activate the transcription of genes controlled by the hypoxia response element (HRE).

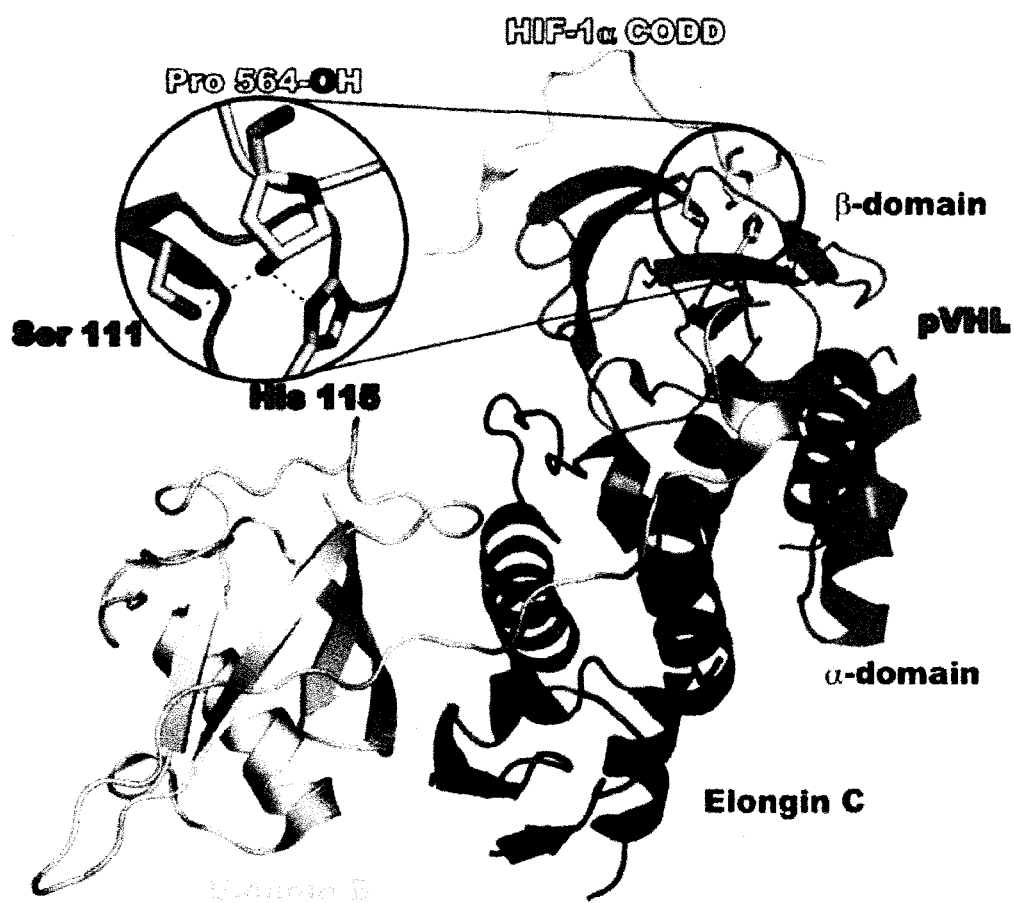


Figure 1.6 Role of Hyp in oxygen sensing. The three-dimensional structure of pVHL–elonginB–elonginC and a HIF-1 α peptide (PDB 1LQB) shows the recognition of HIF-1 α by pVHL (adapted from (Chowdhury et al., 2008)). A hydrogen bond is formed between the hydroxyl group of Hyp564 of HIF-1 α and Ser111 and His115 of pVHL. This interaction directs the degradation of HIF-1 α .

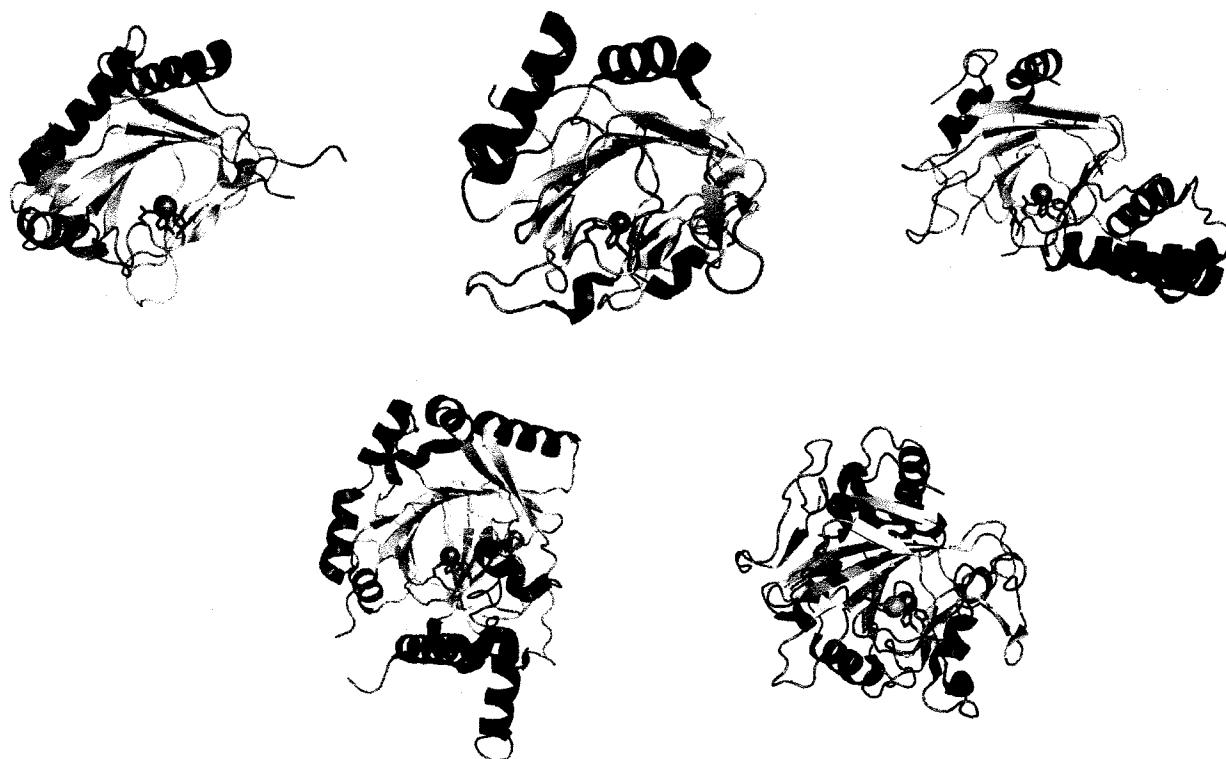


Figure 1.7 Three-dimensional structures of three prolyl 4-hydroxylases. From the top left are PHD2 (2G1M; (McDonough et al., 2006)), Cr-P4H-1 (2JIG; (Koski et al., 2007)), and proline 3-hydroxylase (1E5S; (Clifton et al., 2001)). Two related enzymes are shown on the bottom, from the left asparaginyl hydroxylase FIH (1H2N; (Elkins et al., 2003)), and halogenase SyrB2 (2FCT; (Blasiak et al., 2006)). The proteins are colored by secondary structure with helices in dark red and sheets in tan. The active site iron is shown in orange, or zinc in gray in Cr-P4H-1. The 2-His-1-Asp residues that coordinate the metal are shown in light blue. The chloride in SyrB2 is shown in green.

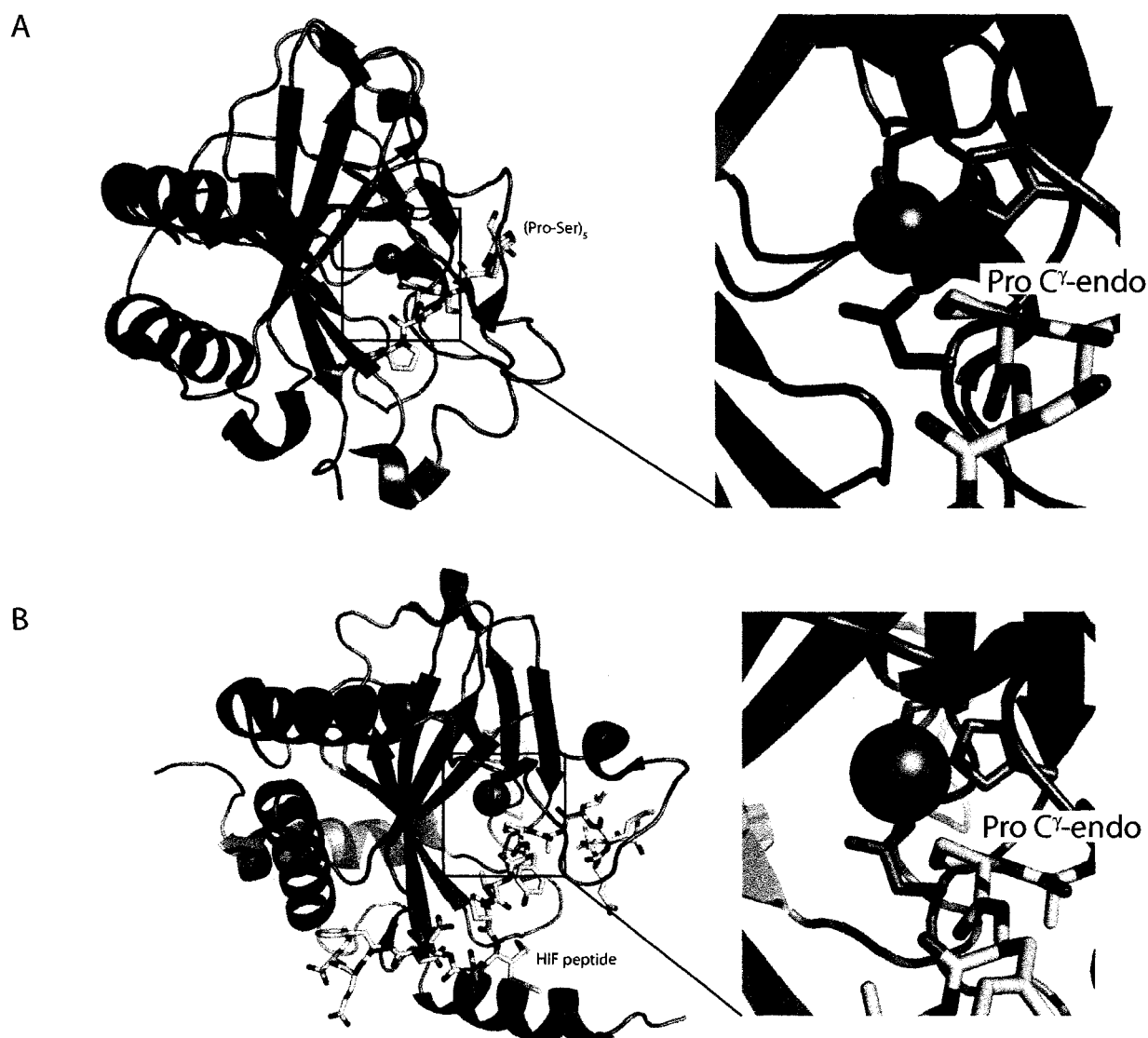


Figure 1.8 Structures of peptide substrate-bound prolyl 4-hydroxylases. The secondary structure of A) Cr-P4H-1 with Zn in the active (pdb 3gze), and B) PHD2 with Mn in the active site (pdb 3hqr) are shown with iron-binding residues in stick representation. The peptide substrates in gray stick representation, (Pro-Ser)₅ for Cr-P4H-1 and a HIF-derived peptide for PHD2, reveal Pro in the C γ -endo ring pucker conformation.

Chapter Two

Conformational Preferences of Substrates for Human Prolyl 4-Hydroxylase

This work was published as

K. L. Gorres, R. Edupuganti, G. R. Krow, and R. T. Raines (2008) *Biochemistry* 47(36):9447-9455.

This work was done in collaboration with Prof. Grant R. Krow. Methanoproline was synthesized by R. Edupuganti and G. R. Krow.

2.1 Abstract

Prolyl 4-hydroxylase (P4H) catalyzes the post-translational hydroxylation of (2*S*)-proline (Pro) residues in procollagen strands. The resulting (2*S*,4*R*)-4-hydroxyproline (Hyp) residues are essential for the folding, secretion, and stability of the collagen triple helix. Even though its product (Hyp) differs from its substrate (Pro) by only a single oxygen atom, no product inhibition has been observed for P4H. Here, we examine the basis for the binding and turnover of substrates by human P4H. Synthetic peptides containing (2*S*,4*R*)-4-fluoroproline (Flp), (2*S*,4*S*)-4-fluoroproline (flp), (2*S*)-4-ketoproline (Kep), (2*S*)-4-thiaproline (Thp), and methanoproline (Mtp) were evaluated as substrates for P4H. Peptides containing Pro, flp, and Thp were found to be excellent substrates for P4H, forming Hyp, Kep, and (2*S*,4*R*)-thiaoxoproline, respectively. Thus, P4H is tolerant to some substitutions on C-4 of the pyrrolidine ring. In contrast, peptides containing Flp, Kep, or Mtp did not even bind to the active site of P4H. Each proline analogue that does bind to P4H is also a substrate, indicating that discrimination occurs at the level of binding rather than turnover. As the iron(IV)-oxo species that forms in the active site of P4H is highly reactive, P4H has an imperative for forming a snug complex with its substrate, and appears to do so. Most notably, those proline analogues with a greater preference for a C^γ-*endo* pucker and *cis* peptide bond were the ones recognized by P4H. As Hyp has a strong preference for C^γ-*exo* pucker and *trans* peptide bond, P4H appears to discriminate against the conformation of proline residues in a manner that diminishes product inhibition during collagen biosynthesis.

2.2 Introduction

Collagens are the major structural proteins in the extracellular matrix. All collagens are comprised of three polypeptide strands that coil together into a right-handed triple helix. Each strand contains a repeating three amino-acid sequence in which every third residue is a glycine (Gly): Xaa–Yaa–Gly. The Xaa position is often (2*S*)-proline (Pro), and the Yaa position is often (2*S*,4*R*)-4-hydroxyproline (Hyp) (Prockop and Kivirikko, 1995). Hyp is produced by the action of prolyl 4-hydroxylase (P4H; EC 1.14.11.2). This post-translational modification is the most prevalent in the kingdom Animalia, where Hyp is more common than seven of the “common” amino-acid residues: Cys, Gln, His, Met, Phe, Trp, and Tyr.¹ The presence of Hyp is crucial for the folding, secretion, and stability of the collagen triple helix under physiological conditions (Berg and Prockop, 1973c; Bulleid et al., 1996; Chopra and Ananthanarayanan, 1982; Jenkins and Raines, 2002), and P4H activity is necessary for the viability of the nematode *Caenorhabditis elegans* (Friedman et al., 2000; Myllyharju et al., 2002) and the mouse *Mus musculus* (Holster et al., 2007). The enzyme-catalyzed hydroxylation of proline residues also plays a role in the sensing of molecular oxygen, though those enzymes are distinct from the P4H that acts on collagen (Kaelin, 2005).

P4H is a tetrameric enzyme comprised of two α subunits and two β subunits. Each α subunit (59 kDa) contains a catalytic domain and a peptide-binding domain (Lamberg et al., 1995). The β subunit (55 kDa) is protein disulfide isomerase (Koivu et al., 1987), which is necessary to keep the α subunit from aggregating (Vuori et al., 1992b; Vuori et al., 1992c) and to retain the enzyme in the endoplasmic reticulum (Kivirikko et al., 1989). The study of P4H has been facilitated by

¹ The abundance of Hyp in animal proteins is <4%, a value calculated from the abundance of collagen amongst animal proteins ($1/3$) and the prevalence of Hyp within the Xaa–Yaa–Gly sequence of collagen ($<38\% \times 1/3$). The abundance of the “common” amino acids is given in ref. (McCaldon and Argos, 1988).

the recent development of recombinant DNA systems for the high-level production of active P4H tetramers in *Escherichia coli* (Kersteen et al., 2004; Neubauer et al., 2005). Still, the three-dimensional structure of P4H remains unknown.

P4H is a non-heme iron(II) dioxygenase (Fox, 1998; Guzman, 1998; Kivirikko and Pihlajaniemi, 1998). This class of enzymes uses α -ketoglutarate and O₂ as co-substrates (Figure 2.1). Hydroxylation of proline is accompanied by the oxidative decarboxylation of α -ketoglutarate to form succinate (Rhoads and Udenfriend, 1968). During the reaction, one atom of molecular oxygen is incorporated into Hyp, and the other into succinate (Cardinale et al., 1971). Uncoupling of the decarboxylation of α -ketoglutarate from substrate hydroxylation results in inactivated P4H, which can be rescued by ascorbate (de Jong et al., 1982; Myllyla et al., 1978; Myllyla et al., 1984). Details of the mechanism by which P4H catalyzes hydroxylation have been proposed based on related α -ketoglutarate-dependent dioxygenases (Costas et al., 2004; Hoffart et al., 2006; Krebs et al., 2007). In the initial steps of catalysis, decarboxylation of α -ketoglutarate produces an iron(IV)-oxo species. This highly reactive species abstracts the proR hydrogen from C-4 of a proline residue in the substrate (Fujita et al., 1964). A hydroxyl group is transferred to the radical intermediate to form the Hyp product.

P4H does not hydroxylate all proline residues. Free proline amino acids are not hydroxylated by P4H (Cardinale and Udenfriend, 1974). Nor is polyproline hydroxylated, though it does bind to the enzyme and is a competitive inhibitor of enzymatic activity (Prockop and Kivirikko, 1969). Hydroxylation occurs only on proline residues in an Xaa-Pro-Gly sequence within a polypeptide. The strength of substrate binding increases with the length of the polypeptide but is obviated by triple-helical structure (Kivirikko et al., 1972).

Computational, structural, and spectral analyses of Pro–Gly sequences in peptides and proteins have shown that adoption of a β -turn conformation is energetically favorable (Atreya and Ananthanarayanan, 1991; Brahmachari and Ananthanarayanan, 1979; Chopra and Ananthanarayanan, 1982; Rapaka et al., 1978). Other conformation preferences for the substrates of prolyl 4-hydroxylase are less clear. Because proline is a secondary amine and prolyl peptide bonds are tertiary amides, the *cis* isomer of prolyl peptide bonds occurs much more frequently than in peptide bonds between non-prolyl residues (Fischer, 2000). The conformation of the prolyl peptide bond is correlated with another conformational feature of proline, its ring pucker (Bretscher et al., 2001; DeRider et al., 2002). The pyrrolidine ring of proline primarily adopts two puckered conformations: *C ^{γ} -endo* and *C ^{γ} -exo* (Figure 2.2).² The *trans/cis* ratio and ring pucker are both influenced by substituents on the pyrrolidine ring that evoke stereoelectronic effects (Raines, 2006).

In this work, we create variations in *trans/cis* ratio and preferred ring pucker by using proline analogues with subtle substitutions at the 4-position of the pyrrolidine ring (Figure 2.2). We then demonstrate that the ability of P4H to bind to a proline analogue and catalyze its oxidation correlates with the *trans/cis* ratio and preferred ring pucker. The results provide new insight into how this essential enzyme recognizes its substrate.

2.3 Experimental Procedures

Materials. Boc-FlpOH and Boc-flpOH were from OmegaChem (Lévis, Québec). Boc-protected amino acids were converted to their Fmoc derivatives via standard methods and used

² The pyrrolidine ring of proline actually prefers two distinct twist, rather than envelope, conformations (Giacovazzo et al., 2002). As *C ^{γ}* experiences a large out-of-plane displacement in these twisted rings, we refer to pyrrolidine ring conformations simply as “*C ^{γ} -exo*” and “*C ^{γ} -endo*”.

without further purification (Hodges and Raines, 2005). Fluorenylmethoxycarbonyl (Fmoc)-mini-PEG-3TM (Fmoc-11-amino-3,6,9-trioxaundecanoic acid) was from Peptides International (Louisville, KY). Fmoc-ThpOH was from Advanced ChemTech (Louisville, KY). All other Fmoc-protected amino acids were from Novabiochem (La Jolla, CA). [1-¹⁴C]α-Ketoglutarate was from American Radiolabeled Chemicals (St. Louis, MO). All other chemicals were of reagent grade or better, and were used without further purification.

Synthesis of Boc-D,L-MtpOH. Boc-D,L-MtpOH was prepared by the route shown in Scheme 1, which is less costly and more convenient than those described previously (Jenkins et al., 2004; Krow et al., 2005). Briefly, carbamate **1** was prepared as in ref. (Krow et al., 2006), and then used in a manner similar to that in ref. (Krow et al., 2002). Specifically, a solution of carbamate **1** (2.0 g, 10.91 mmol) and TMEDA (2.1 mL, 14.18 mmol, 1.3 equiv) in dry ether (50 mL) was cooled to -78 °C, and *s*-BuLi (10.1 mL, 14.18 mmol, 1.3 equiv, 1.4 M solution in cyclohexane) was added dropwise. The resulting solution was stirred for 2 h at -78 °C and then transferred *via* a canula to a solution of distilled DMF (6.3 mL, 81.83 mmol, 7.5 equiv) in dry diethyl ether (30 mL) that had been cooled to -78 °C. The solution was warmed slowly to room temperature, and then quenched with 10 mL saturated NH₄Cl(aq). The ether layer was washed with distilled water (2 × 20 mL) and dried over Na₂SO₄(s). After filtration and concentration, the oil was taken up in 10 mL of dry MeOH, and this solution was cooled to 0 °C. NaBH₄ (2.1 g, 54.56 mmol, 5.0 equiv) was added slowly to the reaction mixture, which was then stirred at 0 °C for 15 min. The reaction mixture was warmed to room temperature, and 10 mL of saturated NH₄Cl(aq) was added slowly, followed by 30 mL of CH₂Cl₂. The aqueous layer was extracted with CH₂Cl₂ (3 × 30 mL). The organic extracts were combined, dried over Na₂SO₄(s), and filtered. The solvent was removed under reduced pressure, and the residue was purified by

chromatography on silica gel (3:2 hexanes/diethyl ether) to furnish 3-CH₂OH (**2**) (609 mg, 26%, $R_f = 0.36$ with 1:1 hexanes/diethyl ether), 1-CH₂OH (**3**) (638 mg, 27%, $R_f = 0.51$), and starting material **1** (190 mg, 10%, $R_f = 0.71$).

Following the procedure in refs. (Jenkins et al., 2004; Krow et al., 2002), 3-CH₂OH (**2**) (600 mg, 2.813 mmol) was dissolved in a solution of CH₂Cl₂ (14 mL), TEMPO (27 mg), saturated NaHCO₃(aq) (11 mL), KBr (54 mg), and Bu₄NCl (67 mg). This solution was cooled to 0 °C, and a solution of NaOCl (14 mmol), saturated NaHCO₃(aq) (6 mL), and brine (12 mL) was added dropwise over 45 min. The reaction mixture was stirred for an additional 1 h, and then warmed to room temperature. The aqueous layer was separated, and the organic layer was washed with 50% w/v NaHCO₃(aq) (3 × 65 mL). The aqueous layers were combined, and washed with CH₂Cl₂ (2 × 25 mL), acidified with dilute HCl solution until the pH was <3, and then extracted with ethyl acetate (5 × 200 mL). The combined organic extracts were dried over Na₂SO₄(s), and the solvent was removed to furnish Boc-D,L-MtpOH (524 mg, 82%) as off-white solid of sufficient purity.

Instrumentation. Measurements of UV and visible absorbance were made with a Cary model 3 spectrophotometer (Varian, Palo Alto, CA). Peptide synthesis was conducted with a Pioneer (PerCeptive Biosystems) or Symphony (Protein Technologies) automated synthesizer at the University of Wisconsin Biotechnology Center. Preparative high-performance liquid chromatography (HPLC) was performed with a system from Waters (Milford, MA) equipped with two 510 pumps and a 486 tunable absorbance detector. Analytical HPLC was performed with a Waters system equipped with two 515 pumps, a 717 plus autosampler, and a 996 photodiode array detector. Fast protein liquid chromatography (FPLC) was performed with an ÄKTA system from Amersham Pharmacia (Piscataway, NJ). Matrix-assisted laser

desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry was performed with a Perkin-Elmer (Wellesley, MA) Voyager MALDI-TOF mass spectrometer at the University of Wisconsin-Madison Biophysics Instrumentation Facility. Scintillation counting was performed with a Wallac 1450 MicroBeta TriLux liquid scintillation counter from Perkin-Elmer (Wellesley, MA).

Production and Purification of P4H. P4H was produced and purified by using procedures reported previously (Kersteen et al., 2004).

Synthesis of Peptides. Peptides were synthesized on a solid support (PEG-PS, Applied Biosystems) by standard Fmoc-protection methods using HATU activation. To increase the water-solubility of the Pro-, Flp-, flp-, Kep-, and Thp-containing peptides, a short PEG chain was added to their N-terminus. The PEG was added by coupling Fmoc-mini-PEG-3TM (Peptides International, Louisville, KY) to the N-terminus of the peptide while on solid support. The N-terminus remained protected with the Fmoc while the peptides were cleaved from the solid support with 8 mL of 95:2.5:2.5 trifluoroacetic acid (TFA)/triisopropylsilane/water, and then washed with CH₂Cl₂ and dried under vacuum. The dried peptide was heated at reflux with SOCl₂ (15-fold molar excess) in ethanol for 2 h. The solvent was removed by rotary evaporation under reduced pressure. The N-terminal Fmoc was then removed by stirring the peptide in 20% v/v piperidine in DMF for 20 min. The peptides were purified by preparatory HPLC with a gradient (10–30% v/v) of acetonitrile in water containing TFA (0.1% v/v) to yield PEG-Gly-Tyr-Yaa-GlyOEt. MALDI-MS *m/z*: [M + Na]⁺ Yaa = Pro (14% yield; calcd 632.3, found 632.6), Flp (6% yield; calcd 650.3; found 650.5), flp (72% yield; calcd 650.3; found 650.6), Kep (1.8% yield; calcd 646.3, found 646.5), Thp (8% yield; calcd 650.3; found 650.7).

Cbz-Gly-Tyr-Mtp-GlyOEt was synthesized by standard Fmoc-protection methods on solid support. The cleavage from solid support and esterification were performed as described for the PEGylated peptides. Purification was done by preparatory HPLC with a gradient (20–40% v/v) of acetonitrile in water. The synthesis was done with racemic Fmoc-D,L-Mtp, and the diastereomers were not separated. Cbz-Gly-Tyr-Mtp-GlyOEt was analyzed by MALDI-MS m/z : $[M + Na]^+$ (calcd 589.2, found 589.5).

Peptide concentrations were determined by absorbance measurement in 6 M guanidine hydrochloride at pH 6.5 using $\epsilon = 1450 \text{ M}^{-1} \text{ cm}^{-1}$ at 276 nm (Gill and von Hippel, 1989).

HPLC-based Assay of Enzymatic Activity. An HPLC-based assay described previously (Kersteen et al., 2004) was used to monitor product formation by P4H. Assays were performed for 5 min at 30 °C in 100 μL of 50 mM Tris-HCl buffer, pH 7.8, containing bovine serum albumin (1 mg/mL), catalase (100 $\mu\text{g/mL}$), dithiothreitol (100 μM), ascorbate (2 mM), FeSO_4 (50 μM), P4H (90 nM), and α -ketoglutarate (500 μM). The tetrapeptide substrate (stock solution in ethanol) was added to initiate the reaction. The reactions were quenched by boiling for 60 s. All assays were performed in triplicate. A reversed-phase analytical Alltima HP C18 AQ column (4.6 \times 250 mm) from Alltech (Deerfield, IL) was used to separate peptides by elution with aqueous acetonitrile (12–30 % v/v in 20 min) containing TFA (0.1% v/v) at 1.0 mL/min. Product formation was quantified by the ratio of product/substrate, as determined by integration of the $A_{214 \text{ nm}}$ by the Millennium32 software from Waters (Millford, MA).

[^{14}C]CO $_2$ -Release Assay for Enzymatic Activity. An alternative means to assess P4H activity is to monitor the release of CO $_2$, which is a product of catalysis (Figure 2.1). Procedures for monitoring the release of [^{14}C]CO $_2$ from [$1\text{-}^{14}\text{C}$] α -ketoglutarate were as described elsewhere (Kivirikko and Myllyla, 1982; Rhoads and Udenfriend, 1968). Concentrations were the same as

above, except for that of P4H (395 nM). All reactions were performed in duplicate and corrected for the rate of decarboxylation in the absence of the peptide substrate.

Reduction and Conversion of Kep. For analysis, the ketone in PEG-Gly-Tyr-Kep-GlyOEt was either reduced to the alcohol with sodium borohydride (10 equiv) for 30 min at room temperature or converted to the oxime via reaction with hydroxylamine (10 equiv) in 250 mM sodium phosphate buffer, pH 5.0, for 1 h at 100 °C.

Chemical Oxidation of Thp. A mixture of PEG-Gly-Tyr-Thp(O)-GlyOEt and PEG-Gly-Tyr-thp(O)-GlyOEt was produced by chemical oxidation of the Thp-containing peptide with either 3-chloroperoxybenzoic acid (MCPBA) (1 equiv) in chloroform for 2.5 h at room temperature or sodium periodate (1.1 equiv) in aqueous methanol (75%) for 30 min at 50 °C (Kanai et al., 2002). A peptide containing Thp(O,O) was produced by reaction of the Thp-containing peptide with MCPBA (10 equiv) in chloroform for 5 h at room temperature.

2.4 Results

Design of Peptide Substrates for P4H. Previously, we reported on an HPLC-based assay for P4H activity using the peptide substrate dansyl-Gly-Phe-Pro-GlyOEt (Kersteen et al., 2004). This peptide is not especially soluble, and is thus not useful for assays at high concentration. To increase its solubility, the dansyl moiety was replaced with a short PEG segment. In addition, Phe was replaced by Tyr to aid in the determination of peptide concentration. PEG-Gly-Tyr-Pro-GlyOEt is a substrate for P4H with $k_{\text{cat}} = 360 \text{ min}^{-1}$, $K_{\text{M}} = 0.58 \text{ mM}$, and $k_{\text{cat}}/K_{\text{M}} = 1.0 \times 10^4 \text{ M}^{-1}\text{s}^{-1}$ (Table 2.1, Figure 2.3). This $k_{\text{cat}}/K_{\text{M}}$ value is 3-fold greater than that for dansyl-Gly-Phe-Pro-GlyOEt (Kersteen et al., 2004).

Flp does not Inhibit P4H. No product formation was detected when Flp was used as a substrate for P4H (data not shown). It is not surprising that P4H cannot turn over Flp because P4H abstracts the proR hydrogen from C-4 of a proline residue, and that hydrogen is replaced with fluorine in Flp. Likewise, Flp does not inhibit the P4H-catalyzed conversion of Pro to Hyp. Under standard P4H reaction conditions as described in the experimental procedures with PEG-Gly-Tyr-Pro-GlyOEt at 0.06 mM, which is an order of magnitude below its K_M value, PEG-Gly-Tyr-Flp-GlyOEt had no measurable effect on hydroxylation of the Pro-containing peptide (Figure 2.4).

flp is a Substrate for P4H. Unlike Flp, flp has a hydrogen in the 4R position of the pyrrolidine ring, and PEG-Gly-Tyr-flp-GlyOEt was determined to be a substrate for P4H (Figures 2.3 and 2.5). The product was identified as 4-ketoproline (Kep) by MALDI-MS (m/z 624.4, calc 624.3), and coeluted with a standard during HPLC (data not shown). Kep was characterized further by reduction of its ketone to a hydroxyl group with sodium borohydride (MALDI-MS m/z 626.6, calc 626.3) and reaction of its ketone with hydroxylamine to form an oxime (MALDI-MS m/z 639.5, calc 639.3) (Figure 2.5). Kep can form by hydroxylation of flp to produce a fluorohydrin, followed by the elimination of a fluoride ion. From triplicate P4H reactions under standard conditions with varying concentrations of PEG-Gly-Tyr-flp-GlyOEt, the value of k_{cat}/K_M was determined to be $(2.7 \pm 0.5) \times 10^3 \text{ M}^{-1}\text{s}^{-1}$. This value is 27% that of the Pro-containing peptide, which has a k_{cat}/K_M value of $(10 \pm 0.6) \times 10^3 \text{ M}^{-1}\text{s}^{-1}$ (Table 2.1, Figure 2.3).

Kep does not Inhibit P4H. Kep, the product of the P4H-catalyzed reaction of flp, was not found to be a substrate for P4H (data not shown). Like Flp, Kep did not even inhibit the P4H-catalyzed conversion of Pro to Hyp (Figure 2.4).

Thp is a Substrate for P4H. Thp has a sulfur atom bearing no hydrogen atoms at the 4-position of the thiazolidine ring. Nonetheless, PEG-Gly-Tyr-Thp-GlyOEt was found to be a substrate for P4H (Figure 2.6). P4H added a single oxygen atom to Thp to form a sulfoxide (Thp(O)) in the peptide product (MALDI-MS m/z 666.6, calc 666.3 $[M + Na]^+$). Thp(O) was also formed, along with thp(O), in the Thp-containing peptide by chemical oxidation using MCPBA or sodium periodate. Oxidation by sodium periodate yielded two products with the same mass according to HPLC and MALDI-MS. Kanai and coworkers reported that this nonenzymatic reaction with a thiopropylpyrrolidine derivative substrate produces an 85:15 mixture of the 4*R* and 4*S* diastereomers (Kanai et al., 2002). HPLC analysis of the sodium periodate reaction compared to the P4H-catalyzed formation of the sulfoxide showed that P4H produces the 4*R* diastereomer exclusively. Chemical oxidation by MCPBA yielded the sulfoxide diastereomers, as well as the sulfone (Thp(O,O); MALDI-MS m/z 682.4, calc 682.3 $[M + Na]^+$), which was not formed by P4H. From triplicate P4H reactions under standard conditions with varying concentrations of PEG-Gly-Tyr-Thp-GlyOEt, the k_{cat}/K_M value was determined to be $(1.2 \pm 0.3) \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, which is 12% that of the Pro-containing peptide (Table 2.1, Figure 2.3).

P4H Decarboxylates [$1\text{-}^{14}\text{C}$] α -Ketoglutarate when Pro, flp or Thp is a Substrate. The HPLC-based assay described above identified flp- and Thp-containing peptides as novel substrates of P4H. These proline analogues, along with the Pro-containing peptide, were also tested in an assay for P4H activity that measures the release of $[^{14}\text{C}]\text{CO}_2$ from $[1\text{-}^{14}\text{C}]\alpha$ -ketoglutarate. The rate of decarboxylation in the absence of the peptide substrate was subtracted from each measurement. Substrate concentrations for assays with the Pro-, flp-, and Thp-containing peptides were 58, 160, and 640 μM , which were 10% of the K_M values determined in the HPLC-based assay. Under these conditions the rate of CO_2 released in the presence of the Pro-

containing peptide was $1.9 \pm 0.7 \mu\text{M}/\text{min}$ (Table 2.2). The rate with flp in the peptide substrate was $0.5 \pm 0.1 \mu\text{M}/\text{min}$, which is 26% that of Pro; the rate with Thp was $0.1 \pm 0.1 \mu\text{M}/\text{min}$, 6% that of Pro. These rates follow closely the trend observed with the HPLC-based assay.

Mtp does not Inhibit P4H. Mtp, the bicyclic proline analogue, was not found to be a substrate for P4H. No product was detected by either HPLC or mass spectrometry (data not shown). Like Flp and Kep, Mtp did not inhibit the P4H-catalyzed conversion of Pro to Hyp (Figure 2.4).

2.5 Discussion

P4H catalyzes an extremely difficult chemical reaction—the hydroxylation of an unactivated methylene group to form a secondary alcohol. The putative mechanism involves the oxidative decarboxylation of α -ketoglutarate, which promotes the formation of a highly reactive iron(IV)-oxo species from molecular oxygen (Costas et al., 2004; Hoffart et al., 2006; Krebs et al., 2007). This iron(IV)-oxo species then abstracts the proR hydrogen from C-4 of a proline residue, replacing it with a hydroxyl group to form Hyp (Fujita et al., 1964).

In previous work, proline analogues with substituents at C-3 and C-5 have been used to probe catalysis by P4H. Peptides containing a racemic mixture of 3-fluoroprolines in the Yaa position were found to be substrates for the enzyme (Tandon et al., 1998). Peptides containing the proline analogues 3-exomethyleneproline (Tandon et al., 1998), 5-oxaproline (Gunzler et al., 1988), and 3,4-dehydropyrolidine (Kerwar et al., 1976; Nolan et al., 1978; Salvador et al., 1976) inhibited P4H activity. In our work, we have focused instead on C-4, which is the carbon that undergoes a change in covalency during the reaction catalyzed by P4H.³

³ The diastereomers of (2S)-4-methylproline have been reported to be neither substrates nor inhibitors of human P4H.

Requirements of Nonnatural Proline Analogues for Turnover by P4H. P4H can elicit the homolytic cleavage of a C–H bond but not a C–F bond. The change from Pro to Flp is conservative from the perspective of sterics, as hydrogen and fluorine have comparable van der Waals radii ($r_{\text{H}} = 1.20 \text{ \AA}$; $r_{\text{F}} = 1.35 \text{ \AA}$ (Pauling, 1960)). A C–F bond ($\Delta H^\circ = 116 \text{ kcal/mol}$) is, however, much stronger than a C–H bond ($\Delta H^\circ = 98 \text{ kcal/mol}$). Accordingly, we were not surprised to learn that P4H cannot turn over Flp. This finding contrasts with reports by others in the 1960s that relied on less direct assays (Gottlieb et al., 1965; Hutton et al., 1968). We did, however, expect P4H to bind to Flp as it does to Pro. Yet, no inhibition of P4H activity by Flp was detectable (Table 2.1).

Unlike Flp, P4H does turn over flp. The stability of a carbon radical with an α -fluoro substituent is $\Delta H = 0.7 \text{ kcal/mol}$ greater than that of an unsubstituted carbon radical, according to homolytic bond dissociation enthalpies of 2-fluoropropane and propane calculated at 25°C (Zhang, 1998). Hence, the hydroxylation of flp should be slightly faster than that of Pro. Of course, other factors contribute to the rate of an enzymatic reaction, and the actual values of k_{cat} are 360 and 250 min^{-1} for Pro and flp, respectively (Table 2.1).

The product of the turnover of flp by P4H is Kep (Figure 2.5). This product has notable utility for future studies of P4H in biological systems. The turnover of flp introduces a functional group, a ketone, with orthogonal reactivity into a protein. This group could serve as a handle for a proteomic analyses of P4H substrates (Chen and Ting, 2005), as flp can be incorporated into proteins by biosynthesis (Kim and Conticello, 2007; Kim et al., 2004; Kim et al., 2005; Renner et al., 2001; Steiner et al., 2008). Furthermore, the turnover of flp not only produces Kep, but also releases fluoride ion, whose detection could provide the basis for a direct, continuous assay of P4H activity.

The sulfide in the Thp-containing peptide is oxidized to a sulfoxide by P4H (Table 2.1; Figure 2.6). Sulfides can be oxidized by other dioxygenases, such as thymine hydroxylase (Thornburg et al., 1993), 4-hydroxyphenylpyruvate dioxygenase (Pascal et al., 1985), and cysteine dioxygenase (McCoy et al., 2006). Although the chemical oxidation of Thp produces both sulfoxide diastereomers (Kanai et al., 2002), turnover by P4H produces only the 4*R* sulfoxide, which has the same relative stereochemistry as the natural product, Hyp.

α -Ketoglutarate is decarboxylated when Thp, flp, or Pro is the substrate in a reaction with P4H (Table 2.2). The relative rates of the decarboxylation reactions (Thp < flp < Pro) is similar to that for Hyp production (Table 2.1). These data are consistent with Thp, flp, and Pro being turned over by the same mechanism.

Role of Proline Conformation in Substrate Binding by P4H. Proline is the only proteinogenic amino acid whose *trans* peptide bond isomer is favored only slightly over the *cis* isomer (Fischer, 2000). In other peptide bonds, the *trans* conformation is favored greatly. Proline is also the only proteinogenic amino acid to contain a saturated ring. The five-membered pyrrolidine ring of proline primarily adopts two conformations, C^γ -*endo* or C^γ -*exo* (Figure 2.2). The pucker of its pyrrolidine ring and the isomerization of its peptide bond are related attributes of proline (Bretscher et al., 2001; DeRider et al., 2002). Electronegative atoms such as fluorine (*e.g.*, $\chi_H = 2.1$; $\chi_F = 4.0$ (Pauling, 1960)) alter the ring pucker via the *gauche* effect. A C^γ -*exo* pucker allows for an $n \rightarrow \pi^*$ interaction between the oxygen of the Xaa–Pro peptide bond and the carbon of the Pro–Gly peptide bond. In turn, this $n \rightarrow \pi^*$ interaction stabilizes a *trans* peptide bond. The absence of a significant $n \rightarrow \pi^*$ interaction in the C^γ -*endo* pucker favors a *cis* peptide bond (Bretscher et al., 2001; DeRider et al., 2002).

The proline analogues used in this study differ in their predominant ring puckers (Figure 2.2). Hyp and Flp adopt a C^{γ} -*exo* pucker and have a high *trans/cis* ratio, whereas flp and Pro adopt a C^{γ} -*endo* pucker and have a low *trans/cis* ratio (DeRider et al., 2002; Improta et al., 2001). The ring puckers of Kep and Thp have not been examined directly. Nonetheless, the *trans/cis* ratio of Kep is higher than that of Pro (Thomas et al., 2005), and the *trans/cis* ratio of Thp is lower than that of Pro (Kern et al., 1997). We find that Pro, flp and Thp are substrates for P4H, and that Hyp, Flp, and Kep are neither substrates nor inhibitors of the enzyme. We therefore propose that proline analogues that favor the C^{γ} -*endo* pucker and have a low *trans/cis* ratio bind to the active site of P4H, and those that favor the C^{γ} -*exo* pucker and have a high *trans/cis* ratio do not.

To test our proposal, we determined the ability of P4H to recognize Mtp (Figure 2.2). In essence, this analogue has *two* C-4 carbons and thus always displays both ring puckers (Krow and Cannon, 2004). Mtp has a *trans/cis* ratio that is between that of Pro and flp (Jenkins et al., 2004). We find that P4H does not hydroxylate either C-4 carbon of Mtp, and that Mtp does not inhibit P4H activity. These results refine our proposal by suggesting a proline residue with C^{γ} -*exo* pucker does not bind to the active site of the enzyme.

Our data suggest a means by which P4H diminishes product inhibition, which is often detrimental to enzyme function (Cook and Cleland, 2007; Walter and Frieden, 1963). The stereoelectronic consequences of catalysis by P4H convert the Pro substrate into a Hyp product, which has a greater preference for a C^{γ} -*exo* ring pucker and *trans* peptide bond. These changes discourage P4H from binding previously hydroxylated collagen strands. A similar mechanism to avert product inhibition is employed by oligosaccharyl transferase, which catalyzes the formation of a *cis* amide bond during the *N*-glycosylation of asparagine residues. Isomerization to the *trans*

conformations takes place after product release and prevents the product from binding again to the enzyme (Peluso et al., 2002).

A summarial model for substrate recognition by P4H is shown in Figure 2.7. This model is based on the data herein, as well as known structures of Pro and Hyp. Previously, we used X-ray diffraction analysis to determine the three-dimensional structures of crystalline AcProOMe and AcHypOMe (Panasik et al., 1994). In these structures, AcProOMe has a *cis* peptide bond and *C^γ-endo* ring pucker, and AcHypOMe has a *trans* peptide bond and *C^γ-exo* ring pucker. In the model, the iron(IV)-oxo species is proximal to the pro*R* hydrogen on C-4 of Pro, which is the single atom on the pyrrolidine ring that is farthest from the main chain of the peptide substrate. Thus, the hydrogen atom to be abstracted reaches most deeply into the enzymic active site. Hydroxylation then promotes a conformational change in the product that the P4H active site is unable to accommodate.

2.6 Conclusions

We have discovered two novel substrates for P4H: flp and Thp. These analogues demonstrate for the first time that perturbations can be made at the 4-position of a proline substrate with retention of high P4H activity. Other proline analogues investigated, Flp, Kep and Mtp, are not recognized by P4H. A comparison of the conformational preferences of these five proline analogues, along with Pro and Hyp, shows that proline analogues that prefer conformations similar to Pro, the substrate, are accepted by P4H, whereas analogues that are similar to Hyp, the product, are not. Therefore, the P4H-catalyzed formation of Hyp promotes conformational changes that limit product inhibition. Moreover, each analogue that binds to the active site of P4H is also a substrate, indicating that the enzyme discriminates at the level of substrate binding

rather than substrate turnover. The highly reactive iron(IV)-oxo species that forms in the active site of P4H is perhaps the most powerful oxidizing agent in biology (Krebs et al., 2007), and must therefore be sequestered deeply within its enzymic active site. Hence, P4H has an imperative for forming a snug complex with its substrate, and appears to do so.

Table 2.1Comparison of PEG-Gly-Tyr-Yaa-GlyOEt as substrates or inhibitors for P4H^a

Yaa	Substrate or Inhibitor	k_{cat} (min^{-1})	K_{M} (mM)	$k_{\text{cat}}/K_{\text{M}}$ ($10^3 \text{ M}^{-1} \text{ s}^{-1}$)
Pro	Substrate	360 ± 7	0.58 ± 0.03	10 ± 0.6
flp	Substrate	250 ± 21	1.6 ± 0.3	2.7 ± 0.5
Thp	Substrate	470 ± 59	6.4 ± 1.4	1.2 ± 0.3
Flp	Neither	—	—	—
Hyp	Neither ^b	—	—	—
Kep	Neither	—	—	—
Mtp	Neither	—	—	—

^a Reaction components and conditions were as described in the text. Values represent the mean (\pm SE) of three replicates. ^b Data from ref. (Atreya and Ananthanarayanan, 1991).

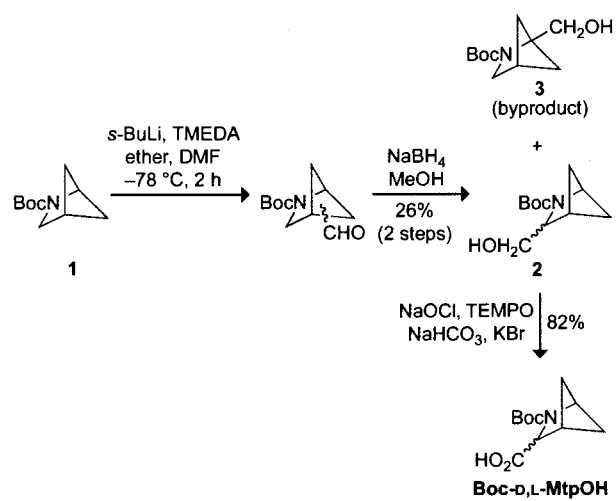
Table 2.2

Comparison of [^{14}C]CO₂ release from
[1- ^{14}C]α-ketoglutarate in P4H assays with
PEG-Gly-Tyr-Yaa-GlyOEt as the substrate^a

Yaa	v (μM/min)	% of Pro
Pro	1.9 ± 0.7	100
Flp	0.5 ± 0.1	26
Thp	0.1 ± 0.1	6

^a Reactions were run in duplicate. The rate of decarboxylation in the absence of the peptide substrate was subtracted from each measurement. Components and conditions were as described in the text. Substrate concentrations were 58, 160, and 640 μM, for Pro, flp, and Thp, respectively, which are 10% of each K_M value determined in the HPLC-based assay.

Scheme 1.



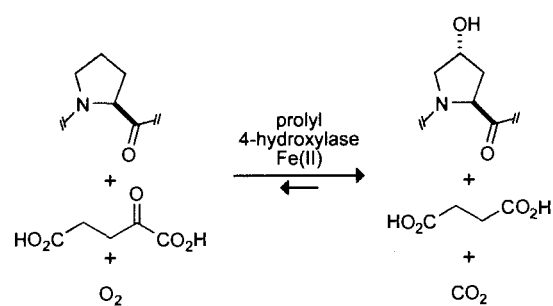


Figure 2.1 Reaction catalyzed by prolyl 4-hydroxylase.

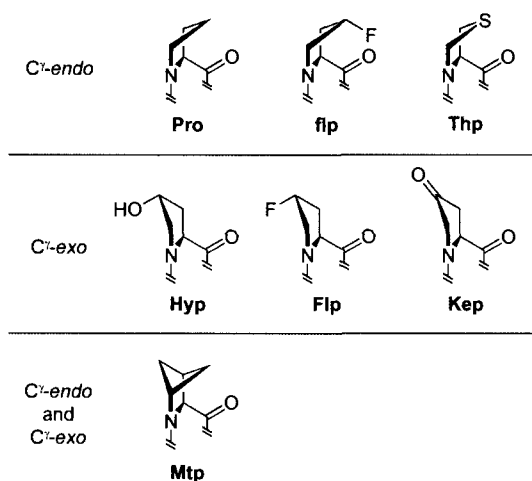


Figure 2.2 Structure of proline and its analogues studied herein, in their predominant ring pucker. Each amino acid was incorporated into a tetrapeptide and examined as a substrate or inhibitor of P4H.

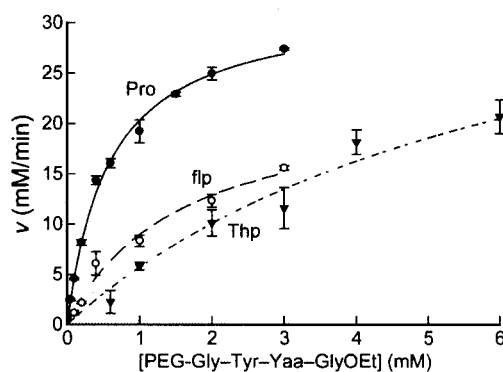


Figure 2.3 Comparison of PEG-Gly-Tyr-Yaa-GlyOEt peptides, where Yaa = Pro (●), flp (○), or Thp (▼), as substrates of P4H. Plot shows the rate of product formation at varying substrate concentrations. Reactions were performed as described in the Experimental Methods section. Reactions were initiated by the addition of tetrapeptide substrate (0.05–3 mM), and run at 30 °C for 5 min. Product formation was analyzed by HPLC. Individual points are the average (\pm SE) of three independent reactions. Data were fitted to the Michaelis–Menten equation to obtain kinetic parameters.

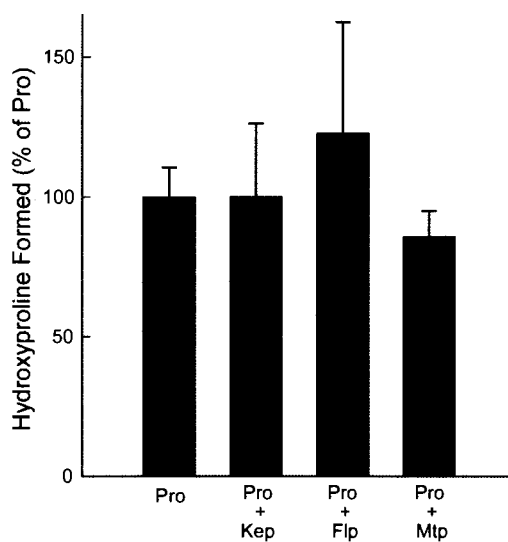


Figure 2.4 Peptides containing Flp, Kep, or Mtp do not inhibit Hyp formation from Pro by P4H. Graph shows the hydroxyproline formed from proline in the presence of no proline analogue, Kep, Flp, or Mtp. The reaction with no proline analogue was designated as 100%. Reactions contained 0.6 mM PEG-Gly-Tyr-Yaa-GlyOEt, where Yaa = Kep, Flp or Mtp, and 0.06 mM PEG-Gly-Tyr-Pro-GlyOEt. Individual points are the average (\pm SE) of three independent reactions. HPLC conditions are described in the Experimental Procedures.

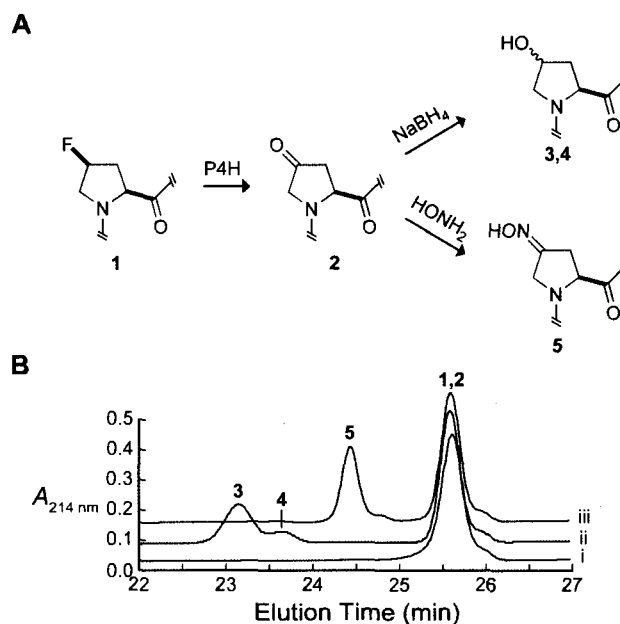


Figure 2.5 Characterization of the products for the turnover of PEG-Gly-Tyr-flp-GlyOEt by P4H. (A) Kep (2) is formed from flp (1) by P4H. Kep is reduced with sodium borohydride (10 equiv) for 30 min to form Hyp and hyp (3, 4), or Kep was converted to an oxime (5) by reaction with hydroxylamine (10 equiv) in 250 mM sodium phosphate buffer, pH 5.0, for 1 h at 100 °C. (B) HPLC analysis of PEG-Gly-Tyr-flp-GlyOEt reactions. (i) Turnover by P4H; (ii) turnover by P4H and treatment with sodium borohydride; (iii) turnover by P4H and treatment with hydroxylamine. Peak 1,2 is from the coeluting peptides containing flp or Kep; peaks 3 and 4 are from the peptides containing Hyp or hyp; peak 5 is from the peptide of containing the oxime of Kep.

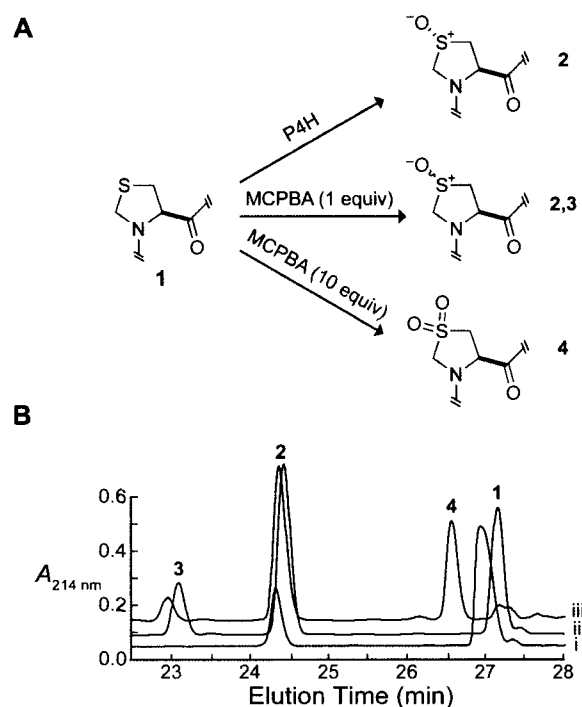


Figure 2.6 Characterization of the products for the turnover of PEG-Gly-Tyr-Thp-GlyOEt by P4H. **(A)** Thp(O) (**2**) is formed from Thp (**1**) upon catalysis by P4H. Thp(O) and thp(O) (**2,3**) are formed from Thp upon reaction with MCPBA (1 equiv) in chloroform for 2.5 h. Thp(O,O) (**4**) is formed from Thp upon reaction with MCPBA (10 equiv) in chloroform for 5 h. **(B)** HPLC analysis of reactions of PEG-Gly-Tyr-Thp-GlyOEt. (i) Turnover by P4H; (ii) treatment with MCPBA (1 equiv); (iii) treatment with MCPBA (10 equiv). Peak **1** is from the peptide containing Thp peptide; peaks **2** and **3** are for the peptides containing Thp(O) or thp(O); peak **4** is for the peptide containing Thp(O,O).

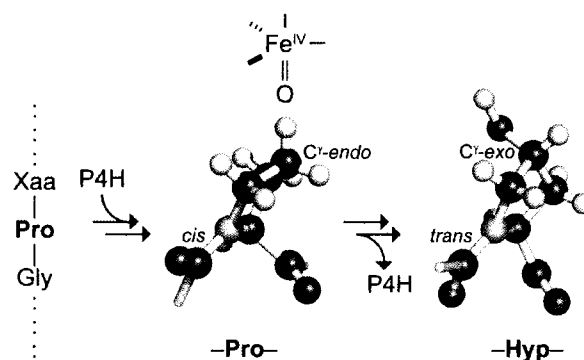


Figure 2.7 Putative model for substrate recognition by P4H. The Pro residue of the substrate has a *cis* peptide bond and *C^γ-endo* ring pucker; the iron(IV)-oxo species is in a position to remove the *proR* hydrogen on C-4 (Fujita et al., 1964). Hydroxylation changes the preferred conformation, and the Hyp residue of the product has a *trans* peptide bond and *C^γ-exo* ring pucker. The structures of Pro and Hyp are actual fragments of the crystal structures of AcProOMe and AcHypOMe (Panasik et al., 1994), and are aligned such that the nitrogen and its three pendant carbons are in spatial alignment.

Chapter Three

Direct and Continuous Assay for Prolyl 4-Hydroxylase

This chapter was published as

K. L. Gorres and R. T. Raines (2009) *Anal. Biochem.* 386:181-185.

3.1 Abstract

Prolyl 4-hydroxylase (P4H) is a non-heme iron dioxygenase that catalyzes the post-translational hydroxylation of (2*S*)-proline (Pro) residues in protocollagen strands. The resulting (2*S*,4*R*)-4-hydroxyproline (Hyp) residues are essential for the folding, secretion, and stability of the collagen triple helix. P4H uses α -ketoglutarate and O₂ as co-substrates, and forms succinate and CO₂ as well as Hyp. Described herein is the first assay for P4H that continuously and directly detects turnover of the proline-containing substrate. This assay is based on (2*S*,4*S*)-4-fluoroproline (flp), a proline analogue that is transformed into (2*S*)-4-ketoproline (Kep) and inorganic fluoride by P4H. The fluoride ion, and thus turnover by P4H, is detected by a fluoride ion-selective electrode. Using this assay, steady-state kinetic parameters for the human P4H-catalyzed turnover of a flp-containing peptide were determined and found to be comparable to those obtained with a discontinuous HPLC-based assay. In addition, this assay can be used to characterize P4H variants, as demonstrated by a comparison of catalysis by D414A P4H and the wild-type enzyme. Finally, the use of the assay to identify small-molecule inhibitors of P4H was verified by an analysis of catalysis in the presence of 2,4-pyridine dicarboxylate, an analogue of α -ketoglutarate. Thus, the assay described herein could facilitate biochemical analyses of this essential human enzyme.

3.2 Introduction

Collagens are the major structural proteins of the extracellular matrix. Collagens undergo a number of post-translational modifications during biosynthesis. One major modification is the hydroxylation of certain prolyl residues. The repeating amino-acid sequence of collagen, where every third residue is a glycine (Gly): Xaa–Yaa–Gly, is rich in (2*S*)-proline (Pro). Indeed, Pro is the most common amino acid found in the Xaa position (Prockop and Kivirikko, 1995). In the Yaa position, the most common amino acid is (2*S*,4*R*)-4-hydroxyproline (Hyp). Hyp is formed by a post-translational modification of proline catalyzed by prolyl 4-hydroxylase (P4H; EC 1.14.11.2).

Hyp is necessary for the stable formation of the triple-helical structure of collagen. Collagen with decreased levels of Hyp is defective in folding and secretion under physiological conditions (Berg and Prockop, 1973c; Bulleid et al., 1996; Chopra and Ananthanarayanan, 1982). P4H activity is required for viability of the nematode *Caenorhabditis elegans* (Friedman et al., 2000; Winter and Page, 2000) and the mouse *Mus musculus* (Holster et al., 2007). *In vitro*, P4H has been studied by availing enzyme via heterologous expression in insect cells, yeast and (only recently (Kersteen et al., 2004; Neubauer et al., 2005)) bacteria.

P4H is a member of the α -ketoglutarate-dependent, non-heme iron(II) dioxygenase family of enzymes (Fox, 1998; Guzman, 1998; Kivirikko and Pihlajaniemi, 1998). These enzymes require iron(II), α -ketoglutarate, and O₂ for catalysis (Figure 3.1). To accomplish difficult oxidizing reactions, these enzymes employ a highly reactive iron(IV)-oxo species (Costas et al., 2004; Hoffart et al., 2006). In P4H, this species abstracts the 4-*proR* hydrogen atom from a Pro substrate (Fujita et al., 1964), and then transfers a hydroxyl radical to form the Hyp product. The key iron(IV)-oxo species is formed by the oxidative decarboxylation of α -ketoglutarate, which

also results in the formation of succinate and CO_2 (Rhoads and Udenfriend, 1968). During the reaction, one atom of molecular oxygen is incorporated into Hyp and the other into succinate (Cardinale et al., 1971). The turnover of α -ketoglutarate can occur without the formation of Hyp. This uncoupling of α -ketoglutarate decarboxylation from substrate hydroxylation leads to inactivated P4H. Ascorbate can reactivate the enzyme (de Jong et al., 1982; Myllyla et al., 1978; Myllyla et al., 1984).

Many non-heme iron dioxygenases catalyze subtle changes to large substrates, making their enzymatic activity difficult to assay. P4H is not an exception. Historically, its activity has been monitored by assays that employ radioactivity. Hydroxylation of the collagen substrate can be monitored directly by the detection of [^{14}C]Hyp formed in collagen containing [^{14}C]Pro after acid hydrolysis of the product (Peterkofsky and Udenfriend, 1965). A more rapid assay involves radiolabeling collagen with [3,4- ^3H]proline and detecting [^3H]H $_2\text{O}$ after hydroxylation (Hutton et al., 1966a). These assays are, however, not only discontinuous, but also require the time-consuming production of radiolabeled collagen.

P4H activity has been measured indirectly by monitoring the turnover of α -ketoglutarate. Assays have been developed to detect residual α -ketoglutarate substrate or the incipient succinate or CO_2 product. Perhaps the most often used P4H assay quantifies the [^{14}C]CO $_2$ product of [1- ^{14}C] α -ketoglutarate decarboxylation (Kivirikko and Myllyla, 1982; Rhoads and Udenfriend, 1968). Trapping of the [^{14}C]CO $_2$ can, however, be inefficient. The turnover of [5- ^{14}C] α -ketoglutarate produces [1- ^{14}C]succinate, which must be separated from unreacted α -ketoglutarate prior to analysis. Methods of separation by column chromatography (Cunliffe et

al., 1986) and chemical precipitation of α -ketoglutarate (Kaule and Gunzler, 1990) have been reported previously.

Assays for P4H activity have also been developed that do not require radioactivity. The consumption of the co-substrate oxygen has been monitored using an O₂-sensing electrode (Nietfeld and Kemp, 1980). An assay developed more recently quantifies unreacted α -ketoglutarate by a post-reaction derivatization that forms a fluorescent product (McNeill et al., 2005). Another assay couples the formation of succinate to that of NAD⁺ formation via succinyl-coenzyme A synthetase, pyruvate kinase and lactate dehydrogenase (Luo et al., 2006). Although these assays are of general use for all α -ketoglutarate-dependent dioxygenases, they have the marked disadvantage of being indirect—they do not report on the hydroxylation of a substrate. Direct, but discontinuous, assays that monitor proline hydroxylation without using radioactivity have been developed with peptide substrates. In these assays, the product Hyp-containing peptide is separated from the substrate by thin-layer chromatography (Tandon et al., 1998) or by HPLC (Gorres et al., 2008; Kersteen et al., 2004).

Herein, we present the first assay for P4H that is both direct *and* continuous. The assay utilizes an alternative substrate, (2*S*,4*S*)-4-fluoroproline (flp), that upon turnover by P4H forms (2*S*)-4-ketoproline (Kep) and releases inorganic fluoride ion (Gorres et al., 2008). The rate of substrate turnover is monitored continuously by using a fluoride ion-selective electrode, which has been employed previously in assays for alkaline phosphatase and chondroitin AC lyase (Rye and Withers, 2002; Venetz et al., 1990). Our assay for P4H yields kinetic parameters comparable to those from a discontinuous HPLC-based assay using the same peptide substrate. The assay is also able to reveal the effect of altering an active-site residue and that of a small-molecule

inhibitor. Accordingly, the assay is likely to have a substantial impact on biochemical analyses of this essential enzyme.

3.3 Materials and methods

Materials. Boc-FlpOH and Boc-flpOH were from OmegaChem (Lévis, Québec). All other reagents were of reagent grade or better and were used without further purification.

Production and purification of P4H. P4H was produced and purified by using procedures reported previously (Kersteen et al., 2004). cDNA encoding the D414A variant of P4H was created by oligonucleotide-mediated site-directed mutagenesis using the pBK1.PDI1.P4H7 plasmid described previously. P4H D414A was produced and purified by the method used for wild-type P4H.

Synthesis of PEGylated peptides. PEG-Gly-Tyr-Yaa-GlyOEt peptides, with Yaa = Flp or Pro, were synthesized as described previously (Gorres et al., 2008). The peptide with Yaa = flp was also synthesized as described, except for the use of Wang resin (Novabiochem, Gibbstown, NJ).

Fluoride ion-based assay of enzymatic activity. Assays were performed at room temperature, which was $(23 \pm 2) ^\circ\text{C}$, in glass vials with stirring. Assay solutions were 0.30 ml of 50 mM Tris-HCl buffer, pH 7.8, containing bovine serum albumin (1.0 mg/mL), catalase (0.10 mg/mL), dithiothreitol (0.10 mM), ascorbate (2.0 mM), FeSO_4 (0.050 mM), P4H (90 nM), α -ketoglutarate (0.50 mM), and sodium fluoride (0.040 mM). The reaction mixture was allowed to equilibrate, and the tetrapeptide substrate was added from a stock solution in ethanol to initiate the reaction. The change in the concentration of fluoride ion was monitored with an Orion fluoride ion-selective electrode (Thermo Scientific, Waltham, MA) interfaced with a computer via an

electrode amplifier (Vernier, Beaverton, OR). The data were fitted by linear-regression analysis to obtain initial rates. Fluoride ion concentrations were calculated by comparison to a standard curve with sodium fluoride in 50 mM Tris–HCl buffer, pH 7.8, which was found to generate the same signal as sodium fluoride in the assay solution described above.

HPLC-based assay of enzymatic activity. An HPLC-based assay described previously (Kersteen et al., 2004) was used to confirm product formation by P4H. Assays were performed for 5 min at room temperature in 100 μ L total volume of 50 mM Tris–HCl buffer, pH 7.8, containing bovine serum albumin (1.0 mg/mL), catalase (0.10 mg/mL), dithiothreitol (0.10 mM), ascorbate (2.0 mM), FeSO₄ (0.050 mM), P4H (90 nM), and α -ketoglutarate (0.50 μ M).

3.4 Results

The P4H-catalyzed turnover of a flp-containing peptide produces a Kep-containing peptide and a fluoride ion (Figure 3.1). Hydroxylation of flp by P4H produces an α -fluorohydrin, which is known to fragment in an exothermic reaction (Seppelt, 1977). The fluoride ion released, and thus P4H activity, is monitored by utilization of a fluoride ion-selective electrode. Sodium fluoride is added to the reaction mixture to put the initial fluoride concentration within the linear range of the fluoride ion-selective electrode. As measured by the HPLC-based assay, the enzymatic activity of P4H was found to be unaffected by the presence of salts, including 10 mM sodium fluoride (data not shown).

To assess the substrate specificity of the assay, we compared peptide substrates containing Pro, flp, or (2*S*,4*R*)-4-fluoroproline (Flp). Pro, the natural substrate is turned over, but does not produce fluoride and therefore is not detected by the fluoride ion-selective electrode (data not shown). Flp is a fluoride-containing proline analogue that is not a substrate of P4H (Gorres et al.,

2008). Assays in which 0.50 mM PEG-Gly-Tyr-Flp-GlyOEt is added produce no change in signal, as in assays that lack peptide (Figure 3.2). In contrast, addition of 0.50 mM of the analogous flp-containing peptide to the assay causes a large increase in signal.

To confirm that the defluorination of flp is due to P4H, the reaction rates were determined at varying P4H concentrations. Duplicate reactions were performed under standard reaction conditions containing 1.0 mM PEG-Gly-Tyr-flp-GlyOEt. With no P4H added to the reaction, there was no increase in fluoride ion concentration. In reactions including 10, 20, 45, or 90 nM P4H, the initial velocity of the reactions correlated linearly with the P4H concentration (Figure 3.3).

The fluoride ion-selective electrode assay was compared to the previously developed HPLC-based assay by determining the steady-state kinetic parameters for the P4H-catalyzed turnover of flp. The rate of fluoride ion formation was determined at varying concentrations of the flp-containing peptide (Figure 3.4, Table 3.1). Assays were performed in triplicate under standard conditions, and data were fitted to the Michaelis–Menten equation to obtain kinetic parameters. The k_{cat}/K_M value was determined to be $(9.8 \pm 2.8) \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ (Table 3.1, Figure 3.4). This value is similar to that from the previously developed HPLC-based assay.

A variant of P4H and an inhibitor of P4H were used to probe the utility of this assay. Under standard assay conditions, the D414A variant of P4H showed no detectable catalytic activity at a concentration of 900 nM, which is 10-fold greater than normal (Figure 3.5A). The effect of the competitive inhibitor 2,4-pyridine dicarboxylate on the rate of reaction was investigated as well. In assays with 0.20 mM flp-containing peptide under standard conditions, 10, 33, and 100 μM concentrations of the inhibitor were shown to inhibit P4H increasingly (Figure 3.5B).

Considering that the reported K_M value of α -ketoglutarate is 22 μM (Myllyla et al., 1977) and its

concentration in the assays herein is 0.50 mM, the inhibition determined by the decrease in the rate of fluoride ion production by P4H in the presence of the inhibitor is consistent with the reported K_i value of 2 μ M (Majamaa et al., 1986; Majamaa et al., 1984).

3.5 Discussion

The assay for P4H activity described herein monitors directly and continuously the turnover of a peptide substrate (Figures 3.2–4; Table 3.1). The steady-state kinetic parameters of a peptide substrate containing flp are comparable to those for the analogous peptide containing Pro, the natural substrate, making flp a suitable substrate (Gorres et al., 2008). The peptide substrate is readily accessible, as flp is available from commercial vendors or accessible by a facile synthetic route (Chorghade et al., 2008). The assay has the advantages of being direct and continuous, in addition to using inexpensive instrumentation, avoiding the use of radioactivity, and not requiring additional enzymes.

Assays of enzymatic activity have identified a number of residues critical for P4H activity. Like most other non-heme iron(II) dioxygenases, the active site of P4H has a two histidine/one carboxylate motif that binds iron. Asp414 has been identified as the source of the carboxylate (Lamberg et al., 1995; Myllyharju and Kivirikko, 1997). Recently, a subclass of non-heme iron(II) dioxygenases that perform halogenation reactions, instead of hydroxylations, has been identified (Krebs et al., 2007). These halogenases lack an active-site carboxylate, containing an alanine residue instead of the canonical aspartate or glutamate. The D414A P4H variant studied herein emulates the active site of a halogenase, though no halogenated product has been detected in a solution of high halide ion concentration (unpublished data). The D414A variant also lacks

hydroxylase activity (Figure 3.5A), confirming the requirement of the aspartate for hydroxylase activity and validating the competence of the assay for reporting on P4H variants.

P4H plays a major role in the biosynthesis of collagen. Excessive collagen formation causes a number of fibrotic diseases, and P4H has been put forth as a target for beneficial intervention with chemotherapeutic agents (Franklin, 1997; Myllyharju, 2008; Shimizu, 2001). Analogues of α -ketoglutarate are known to inhibit P4H competitively with respect to α -ketoglutarate (Friedman et al., 2000; Majamaa et al., 1986; Majamaa et al., 1984). The assay is able to reveal inhibition by one such analogue, 2,4-pyridine dicarboxylate (Figure 3.5B). We anticipate that the assay could be adapted to a high-throughput format with the use of appropriate fluorescent or colorimetric fluoride-sensing reagents. Such reagents are under development (Gunnlaugsson et al., 2006).

In addition to stabilizing collagen, the hydroxylation of proline residues also plays a critical role in the sensing of molecular oxygen (Kaelin and Ratcliffe, 2008; Myllyharju, 2008). For example, the formation of Hyp in the transcription factor hypoxia inducible factor (HIF) is catalyzed by the proline hydroxylase domain proteins (PHDs), which are distinct from the collagen P4H. Incorporating flp into a HIF-derived peptide substrate for PHD enzymes could enable the assay described herein to be used in the study of those enzymes as well.

3.6 Conclusions

In summary, we have developed a new assay for the activity of P4H that utilizes flp as the substrate. Upon turnover, the release of fluoride ion is monitored by a fluoride ion-selective electrode. This assay for P4H is the first that is both direct and continuous. We have demonstrated the utility of the assay in characterizing P4H variants and identifying inhibitors.

We anticipate the modification of this assay into a high-throughput format suitable for the discovery of novel inhibitors of this important enzyme.

Table 3.1

Comparison of the turnover of PEG-Gly-Tyr-flp-GlyOEt by P4H (90 nM) determined by using the fluoride detection assay or the HPLC-based assay^a

Assay	V_{\max} ($\mu\text{M}\cdot\text{min}^{-1}$)	k_{cat} (min^{-1})	K_{M} (mM)	$k_{\text{cat}}/K_{\text{M}}$ ($10^3 \text{ M}^{-1}\text{s}^{-1}$)
fluoride	24 ± 3	270 ± 33	0.5 ± 0.1	9.8 ± 2.8
HPLC	32 ± 5	360 ± 53	1.2 ± 0.3	5.1 ± 1.4

^a Reaction components and conditions were as described in the Materials and methods section. Values represent the mean (\pm SE) of three replicates.

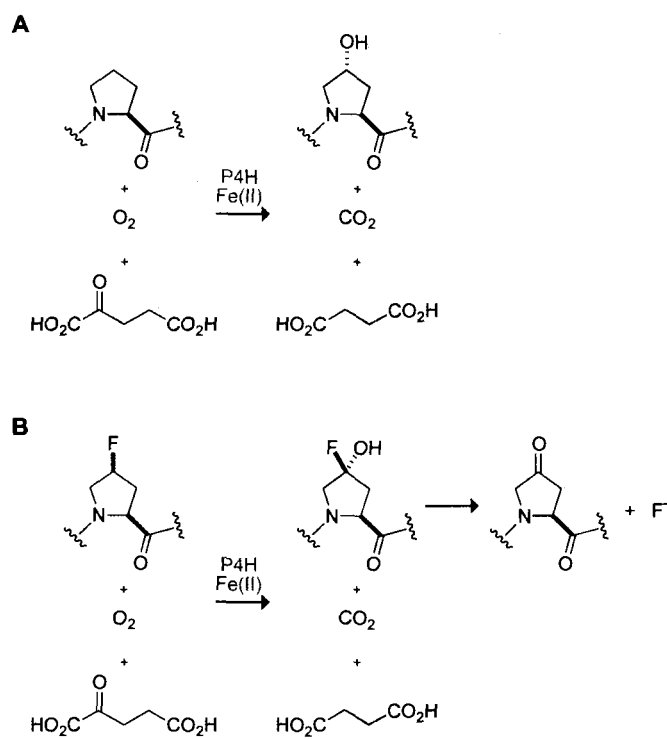


Figure 3.1 Reaction catalyzed by prolyl 4-hydroxylase when (A) a Pro-containing peptide or (B) a flp-containing peptide is the substrate.

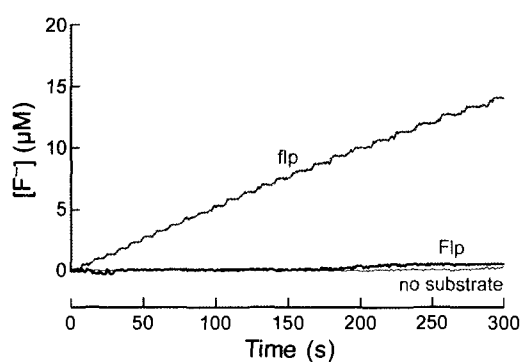


Figure 3.2 Fluoride ion-detection assay for P4H (90 nM). Assays were performed in the presence or absence of PEG-Gly-Tyr-Yaa-GlyOEt, where Yaa = flp (1.0 mM) or Flp (0.50 mM). The dashed line shows the fit of the Yaa = flp data by linear regression analysis. Assay conditions are as described in the Materials and Methods section.

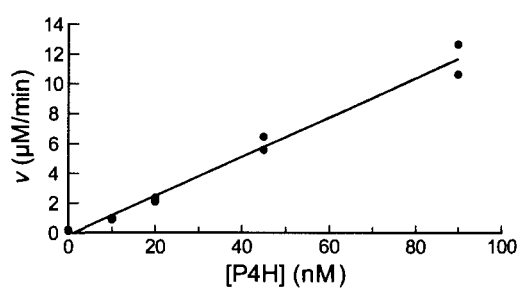


Figure 3.3 Dependence of the rate of fluoride-ion release from PEG-Gly-Tyr-flp-GlyOEt (1.0 mM) on the concentration of P4H. Assay conditions are as described in the Materials and methods section. Reactions were performed in duplicate. Data were fitted by linear regression analysis. $v = \partial[\text{F}^-]/\partial t$.

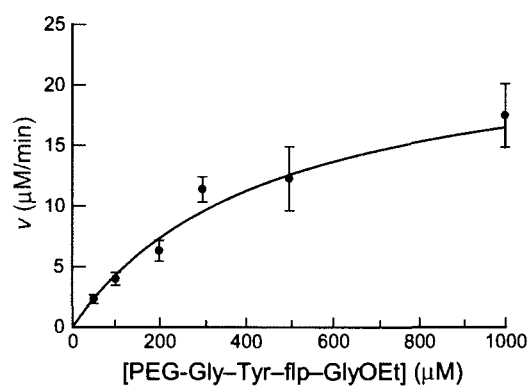


Figure 3.4 Catalysis of fluoride ion-release from PEG-Gly-Tyr-flp-GlyOEt by P4H (90 nM). Assay conditions are as described in the Materials and methods section. Individual points are the average (\pm SE) of three reactions. Data were fitted to the Michaelis–Menten equation: $v = \partial[F^-]/\partial t = k_{\text{cat}}[\text{P4H}][\text{peptide}]/(K_M + [\text{peptide}])$.

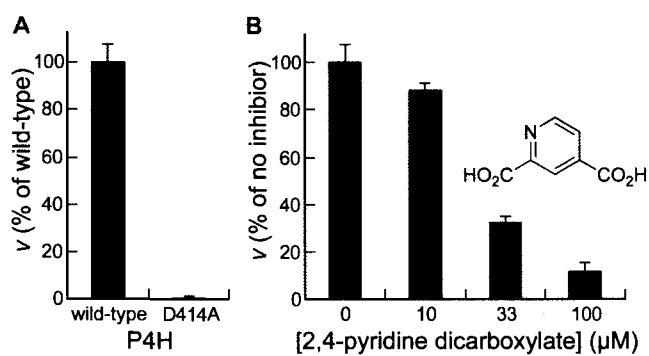


Figure 3.5 Catalysis of fluoride ion-release from PEG-Gly-Tyr-flp-GlyOEt (0.20 mM) by a P4H variant or in the presence of a small-molecule inhibitor. (A) Catalysis by wild-type P4H (90 nM) and its D414A variant (0.90 μM). (B) Inhibition of catalysis by 2,4-pyridine dicarboxylate. Assay conditions are as described in the Materials and methods section. $v = \partial[\text{F}^-]/\partial t$.

Chapter Four

Stringency of the 2-His-1-Asp Active-Site Motif in Prolyl 4-Hydroxylase

4.1 Abstract

The non-heme iron(II) dioxygenase family of enzymes contain a common 2-His-1-carboxylate iron-binding motif. These enzymes are known to catalyze a wide variety of oxidative reactions, such as the hydroxylation of aliphatic C–H bonds. Prolyl 4-hydroxylase (P4H) is an α -ketoglutarate-dependent iron(II) dioxygenase that catalyzes the post-translational hydroxylation of proline residues in collagen leading to stabilization of its triple-helical structure. P4H residues His412, Asp414, and His483 have been identified as the 2-His-1-carboxylate active-site motif. Recently, enzymes related to P4H that carry out oxidative halogenation have been discovered. The active site of these halogenases contains the two histidine residues, but the carboxylate ligand is replaced with a halide that coordinates the iron.

Here, we investigate the iron-binding active site residues of P4H. P4H Asp414 is replaced with alanine to mimic the active site of a halogenase. The D414A variant does not, however, convert P4H to a halogenase. Furthermore, the hydroxylase activity of P4H D414A cannot be rescued with small molecules. In addition, rearrangement of the relative positions of the 2-His-1-Asp residues in the active site of P4H causes the loss of hydroxylase activity. Overall, these results demonstrate the stringency of the iron-binding residues in the P4H active site.

4.2 Introduction

Metals are used as cofactors in biology, and iron is often found in enzymes that utilize molecular oxygen to perform oxidative reactions. In addition to the heme-containing iron enzymes, there is a class of non-heme, mononuclear iron(II) enzymes that catalyze a variety of reactions including hydroxylation of aliphatic C–H bonds, dihydroxylation of arene double bonds, epoxidation of C–C double bonds, heterocyclic ring formation, and oxidative aromatic ring opening. The iron in these non-heme dioxygenases is commonly coordinated by a 2-His-1-carboxylate motif (Koehtop et al., 2005). Two histidine (His) and one aspartate (Asp) or glutamate (Glu) residues bind the iron at the vertices of one triangular face of the octahedral metal center, forming a 2-His-1-carboxylate facial ligand triad (Figure 4.1A). This arrangement leaves three coordination sites on the iron open for oxygen, substrates, and cosubstrates to bind, which contributes to the versatility of these enzymes.

A subset of non-heme iron(II) enzymes utilizes α -ketoglutarate as a cosubstrate. These dioxygenases also contain the 2-His-1-carboxylate iron-binding motif. The primary substrate binds near the active site, while α -ketoglutarate binds directly to the iron in a bidentate manner. The reaction mechanism involves oxidative decarboxylation of α -ketoglutarate to produce succinate, CO_2 , and a high energy Fe(IV)-oxo intermediate (Costas et al., 2004). In a hydroxylation reaction, one atom of oxygen is incorporated into the product, and the other into succinate, illustrated in the reaction catalyzed by prolyl 4-hydroxylase (P4H) (Figure 4.2).

Prolyl 4-hydroxylase (P4H) is an α -ketoglutarate-dependent, iron(II) dioxygenase, and catalyzes the posttranslational hydroxylation of proline residues (Pro) during collagen biosynthesis (Myllyharju, 2003). This hydroxylation reaction that forms (2*S*,4*R*)-4-

hydroxyproline (Hyp) is vital for the proper folding of the collagen triple helix. The active site of P4H is within the α subunit of the $\alpha_2\beta_2$ enzyme tetramer. The β subunits (protein disulfide isomerase) keep the α subunits soluble and retained within the endoplasmic reticulum (Vuori et al., 1992b; Vuori et al., 1992c). The 2-His-1-carboxylate iron-binding residues of P4H α , identified by site-directed mutagenesis, are His412, Asp414, and His483 (Lamberg et al., 1996; Myllyharju and Kivirikko, 1997).

A recently discovered subclass of α -ketoglutarate-dependent, iron(II) dioxygenases catalyzes halogenation reactions. Halogenases play an important role in halogenated natural product biosynthesis. The halogenase SyrB2, found in *Pseudomonas syringae*, catalyzes the conversion of L-Thr to 4-Cl-L-Thr in syringomycin E biosynthesis. The three-dimensional structure of SyrB2 revealed a chloride ion present in place of a carboxylate ligand in the facial triad (Blasiak et al., 2006). The Asp or Glu present in other mononuclear iron enzymes is replaced by alanine (Ala) in SyrB2, which creates space in the active site for the binding of a chloride ligand to iron (Figure 4.1B). The structure of halogenase CytC3 also contained a 2-His motif rather than the 2-His-1-carboxylate facial triad (Wong et al., 2009).

Despite the difference in iron-binding residues, halogenation is thought to follow a mechanism similar to P4H and other α -ketoglutarate-dependent iron(II) dioxygenases (Figure 4.3). Decarboxylation of α -ketoglutarate generates a reactive Fe(IV)-oxo intermediate, which abstracts a hydrogen atom of the substrate C–H bond (Galic et al., 2007; Hoffart et al., 2006). This reaction yields a substrate radical and a Fe(III)-OH complex. Hydroxylated products result from the recombination of the substrate radical with the iron-coordinated hydroxyl radical. In contrast, the radical substrate intermediate of halogenases combines with the chlorine atom

coordinated to the iron, instead of the hydroxyl, to produce a chlorinated product. This discovery begs the question of how product formation, either hydroxylation or halogenation, is controlled.

Herein, we test the stringency of the iron-binding ligands in P4H. We test whether modification of the P4H active to mimic that of a halogenase can convert P4H hydroxylase activity into halogenase activity. Additionally, we study the active site of P4H by changing the relative positions of the residues which constitute the 2-His-1-carboxylate facial triad.

4.3 Materials and methods

Materials. *Escherichia coli* strain Origami B (DE3) were obtained from Novagen (Madison, WI). Enzymes used for DNA manipulation were from Promega (Madison, WI) and DNA oligonucleotides for mutagenesis and sequencing were obtained from Integrated DNA Technologies (Coralville, IA). DNA sequences were elucidated by capillary arrays on an Applied Bioscience automated sequencing instrument at the University of Wisconsin-Madison Biotechnology Center. Poly(L-proline) was obtained from Sigma Chemical (St. Louis, MO). Luria-Bertani (LB) medium contained tryptone (10 g), yeast extract (5 g) and NaCl (10 g). Terrific broth (TB) medium contained tryptone (12 g), yeast extract (24 g), K_2HPO_4 (72 mM), KH_2PO_4 (17 mM) and glycerol (4 mL). All media were prepared in deionized, distilled water, and autoclaved.

Instrumentation. UV absorbance measurements were made with a Cary 50 spectrophotometer from Varian (Palo Alto, CA). Fast protein liquid chromatography (FPLC) was performed with an AKTA system from Amersham-Pharmacia (Piscataway, NJ) and analyzed with the UNICORN Control System. High-performance liquid chromatography (HPLC) was carried out using a system from Waters (Milford, MA). This system was controlled with the

manufacturer's Millennium32 (Version 3.20) software. MALDI-MS was performed with a Perkin-Elmer (Wellesley, MA) Voyager MALDI-TOF mass spectrometer at the University of Wisconsin—Madison Biophysics Instrumentation Facility.

Production and purification of P4H active site variants. The pBK1.PDI1.P4H7 plasmid that encodes P4H α (I) and β subunits served as a template for mutagenesis (Kersteen et al., 2004). The QuikChange site-directed mutagenesis kit (Stratagene) was used to make point mutations in the gene encoding the P4H α subunit. The P4H variants were produced and purified by using procedures reported previously. Briefly, P4H expression in *E. coli* Origami B (DE3) cells was induced at 18 °C for 18 h. Cells were harvested, lysed by sonication, and fractionated by ammonium sulfate precipitation. P4H variants were bound to immobilized to a poly(L-proline)-affinity column, and then eluted with free poly(L-proline). Bound fractions were further purified by anion-exchange chromatography, followed by gel filtration chromatography.

Production and purification of P4H with a His₆ tag. DNA encoding six consecutive His residues at the C-terminus of P4H α (I) was inserted into the pBK1.PDI1.P4H7 plasmid by site-directed mutagenesis. The P4H α (I) with a C-terminal hexaHis tag along with the β subunit (P4H-His₆) was produced as described above. P4H-His₆ was purified by Ni²⁺-affinity chromatography.

HPLC-based enzyme activity assay. P4H activity was monitored using an HPLC-based assay described previously (Kersteen et al., 2004). Assays were conducted for 20 min at 30 °C in 100 μ L of 50 mM Tris-HCl buffer, pH 7.8, containing dithiothreitol (100 μ M), catalase (100 μ g/mL), ascorbate (2 mM), FeSO₄ (50–300 μ M), α -ketoglutarate (0.5–25 mM), bovine serum albumin (1 mg/mL) and P4H (0.09–1.5 μ M). The tetrapeptide substrate (dansyl-Gly-Phe–

Pro-GlyOEt) was added to initiate the reaction. A reversed-phased analytical C18 column was used to separate the substrate and product peptides. Substrate and product peptides from the enzyme reactions were also analyzed by MALDI-MS.

4.4 Results

The difference between hydroxylases and halogenases is more than Asp → Ala. P4H, an α -ketoglutarate-dependent iron (II) dioxygenase, contains a 2-His-1-carboxylate motif that binds iron in the active site. In related enzymes that catalyze halogenation reactions, instead of hydroxylations, the carboxylate-containing residue is replaced with an alanine that provides space for a halide ion to bind iron. If the only difference between hydroxylation and halogenation is an aspartate versus an alanine, then exchanging these residues should interconvert the activities. The halogenation reaction requires the presence of halide ions. Accordingly, we first investigated the activity of P4H in the presence of salts. Up to 100 mM KF, NaCl, NaBr or NaI had no effect on P4H activity (Figure 4.4). Higher salt concentrations (500 mM) decreased P4H activity by ~80%.

The P4H D414A variant, which has an active site analogous to that of SyrB2, was produced, and, as expected, lost hydroxylase activity. Still, no halogenated product from reaction mixtures containing D414A P4H was detected by HPLC or MALDI-MS even in the presence of 100 mM sodium fluoride, chloride, or bromide (Table 4.1). Subsequently, a D414G P4H variant was made to provide more space for a halide ion in the active site. This variant also showed neither halogenase nor hydroxylase activity.

Small molecules cannot rescue the function of Asp in the P4H active site. Functional groups missing from amino-acid side chains can be complemented by small molecules mimicking that

molecule and rescue the protein function (Qiao et al., 2006). In the D414A and D414G variants of P4H, the carboxylate functional group is deleted. Azide, formate, or acetate sodium salts (100 mM) were added to D414A P4H reaction mixtures, but did not result in the formation of hydroxylated product. Nor was the activity of D414G rescued by nitrate or nitrite.

A 3-His active site is not functional in P4H. A 3-His active site is found in cysteine dioxygenase, a non-heme, iron(II) dioxygenase related to P4H that does not utilize α -ketoglutarate. Rather than eliminating the carboxylate-contributing amino acid, as in the D414A and D414G variants, Asp was changed to His. The P4H D414H variant, however, was unable to form a hydroxylated product detectable by the HPLC-based enzyme assay or MALDI-MS.

The spatial orientation of the 2-His-1-Asp triad is critical for activity. The locations of the 2-His-1-Asp triad are found on one face of the iron. The relative position of these residues varies with respect to the locations in which α -ketoglutarate and oxygen bind the iron (Clifton et al., 2006). The locations of the two histidine and aspartate residues were shuffled in the H412D/D414H and D414H/H483D variants. No hydroxylated product was detected in reactions with either of these P4H variants by the HPLC-based enzyme assay or MALDI-MS, which leads to the conclusion that the relative positions of the 2-His-1-carboxylate facial triad in the active site of P4H are important in conferring P4H its hydroxylase activity.

P4H with a histidine tag. The purification of P4H is laborious and involves ammonium sulfate precipitation, poly(L-proline)-affinity chromatography, anion-exchange and size exclusion chromatography. To simplify the purification process, six histidine residues were added to the C-terminus of P4H α . The affinity-tag strategy was previously utilized for the P4H

α (III) isoform (Van Den Diepstraten et al., 2003). The P4H α (I)-His₆ and the β subunit were produced and purified successfully by Ni²⁺-affinity chromatography. P4H-His₆ showed hydroxylase activity in the HPLC-based assay. Though the activity of P4H-His₆ was only ~20% that of non-His-tagged P4H, the P4H-His₆ contained a significant amount of contaminant proteins. Further purification could be achieved by ion-exchange or gel filtration chromatography. This experiment does demonstrate that the His₆ tag did not abolish hydroxylase activity, and this purification strategy can be useful in future studies of P4H.

4.5 Discussion

The 2-His-1-carboxylate motif is common to many α -ketoglutarate-dependent iron(II) dioxygenases. The related halogenases differ in that the iron is coordinated by only 2-His. In halogenases, alanine is located in place of the Asp/Glu, and a chloride binds the iron in place of the carboxylate ligand. In an attempt to convert P4H to perform halogenation reactions, catalysis by the D414A and D414G variants of P4H was analyzed in the presence of halides. While these variants are known to lack hydroxylase activity, no gain of halogenase activity was detected. Similar experiments have been conducted in other dioxygenases to explore the conversion of hydroxylases to halogenases, and vice versa. Replacing alanine with aspartate or glutamate in the halogenase SyrB2 abolished halogenase activity with no observable hydroxylase activity (Blasiak et al., 2006). Substitution of aspartate for alanine in hydroxylase taurine dioxygenase (TauD) abolished hydroxylase activity with no gain in halogenase activity (Grzyska et al., 2007). The D101A variant bound iron, though not as efficiently as did the wild-type enzyme, but it was not known if sufficient chloride was binding in the active site. The association of D414A and D414G P4H with iron was not determined because even wild-type P4H does not bind iron with

high affinity (K_M value of only 5 μM (Tuderman et al., 1977)). The P4H variants made here were able to be produced and purified like the wild-type enzyme, including binding to a poly(L-proline)-affinity column, suggesting that global protein conformation was unaffected by the amino-acid substitutions.

All data thus far shows that the difference between Fe(II) and α -ketoglutarate-dependent hydroxylases and halogenases is greater than a one-residue change, that is, the presence or absence carboxylate side chain. Analysis of structural data of the halogenases CytC3 and SyrB2 suggests that residues that do not make a direct contact with the active-site iron are essential for binding α -ketoglutarate and the chloride (Blasiak et al., 2006; Wong et al., 2009). Residues that surround the active site form a hydrogen-bonding network, which might contribute to chloride binding in CytC3 and SyrB2. Additionally, the presence of a suitable hydrophobic pocket could be important for chloride binding in the active site. Altogether, the selective activity for halogenation, as opposed to hydroxylation, has been proposed to depend on a hydrogen-bonding network and a hydrophobic environment around the halide-binding site, in addition to allowing space for the halide by altering the protein ligands to iron. Interconversion of hydroxylase and halogenase activities will need to take these factors into account.

As the addition of halides did not result in halogenase activity in D414A P4H, the hydroxylase activity could not be rescued by a number of small molecules. Chemical rescue of the residues within the 2-His-1-carboxylate motif of α -ketoglutarate-dependent iron(II) dioxygenases has been demonstrated with other enzymes. The activity of the H174A variant of phytanoyl-CoA 2-hydroxylase was rescued with the addition of imidazole (Searls et al., 2005). Interestingly, the same mutation at the other iron-binding His residue could not be rescued.

Deleterious substitutions of the carboxylate ligand have also been rescued by small molecules, as demonstrated by TauD D101A rescue by formate (Grzyska et al., 2007). These data provide clues about the accessibility of the enzymic active sites, as well as the tolerance to deviation in ligand position.

Although the 2-His-1-carboxylate motif is common to many iron(II) α -ketoglutarate-dependent dioxygenases, there are variations of the metal coordination motif (Straganz and Nidetzky, 2006). Whereas the 2-His-1-carboxylate dioxygenases are quite selective for iron, enzymes containing three histidines and one glutamate (3-His-1-Glu) coordinate additional metals. Moreover, some dioxygenases do not utilize a carboxylate ligand at all. The halogenases, discussed previously, contribute just two histidine residues for iron binding. Enzymes with four histidine (4-His) and three histidine (3-His) ligands have also been identified. A comparison of the 3-His and 2-His-1-carboxylate metal sites revealed an overall structural similarity with slight differences in metal-ligand distances (Leitgeb and Nidetzky, 2008). Conversion of a 2-His-1-carboxylate to a 3-His metal center was accomplished in tyrosine hydroxylase (Fitzpatrick et al., 2003). In P4H, the 3-His motif (that is, D414H P4H) did not retain detectable enzymatic activity. Thus far, glutamate is the only amino acid that can replace Asp414 in P4H and maintain activity (Myllyharju and Kivirikko, 1997).

Instead of removing the carboxylate ligand, a second Asp/Glu could be added to the facial triad or the carboxylate could be relocated. His675 in aspartyl (asparaginy) β -hydroxylase was replaced with Asp or Glu, and the enzyme retained 20 and 12% of its activity (McGinnis et al., 1996). Gln or Glu could be substituted for His255 in TauD with 81% and 33% retention of its activity. In P4H, however, no activity was found in H412E or H483E variant (Myllyharju and Kivirikko, 1997). Rearrangement of the His and Asp ligands in the P4H active site (P4H variants

H412D/D414H and D414H/H483D) did not result in enzymatic activity, which demonstrates that the native orientation of the 2-His-1-carboxylate facial triad is critical for P4H catalytic ability.

In conclusion, the 2-His-1-Asp facial triad in the active site of P4H is resistant to variation, including alternative iron-binding residues and the spatial positioning of the residues. Mutational analysis of the iron-binding residues within non-heme, iron(II) dioxygenases shows some flexibility at these locations, though no pattern has emerged as to which amino acids can functionally substitute at a particular position. P4H has proven to be especially recalcitrant to change. Although focus has been placed on the common 2-His-1-carboxylate iron-binding motif, differences among the non-heme iron(II) dioxygenases point toward the importance of residues that do not bind the iron directly. This secondary coordination sphere presumably alters the steric and electrostatic environments of the facial triad and influences the reaction mechanism. Three-dimensional structural information for P4H, which has thus far eluded crystallization, will identify these additional residues, and could provide more information about the mechanism of P4H and enable the design of P4H variants with transformed activity.

Table 4.1

Characterization of P4H variants for hydroxylase and halogenase activity.

P4H variant	Activity	
	Hydroxylase	Halogenase
Wild-type	+	–
D414A	–	–
D414G	–	–
D414H	–	nd
D414H/H412D	–	nd
D414H/H438D	–	nd

nd Not determined.

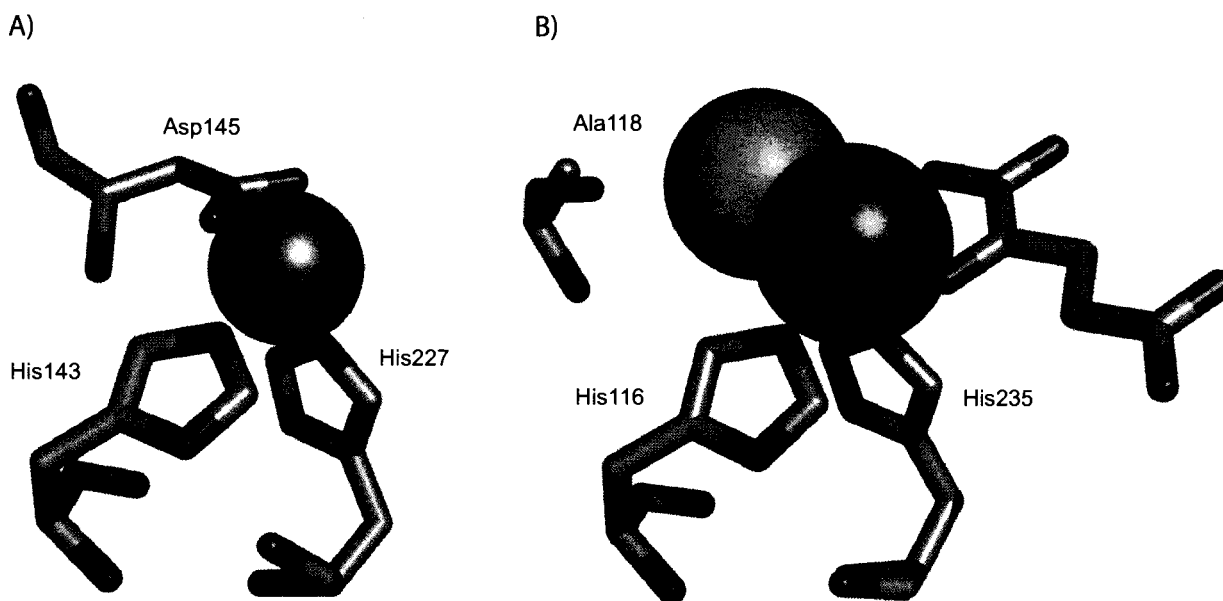


Figure 4.1 Active sites of α -ketoglutarate-dependent iron(II) dioxygenases. **A)** The *Chlamydomonas reinhardtii* prolyl 4-hydroxylase contains a 2-His-1-carboxylate facial triad comprised of His143, His227, and Asp145. A zinc atom (gray) is found in place of iron in the crystal structure (Koski et al., 2009). **B)** The SyrB2 halogenase has two histidine residues (His116, His235) that coordinate to the iron (orange). Alanine (Ala118) is found in place of the carboxylate-containing residue found in hydroxylases, which allows space for the chloride ion (green) (Blasiak et al., 2006). α -Ketoglutarate is also shown bound to iron.

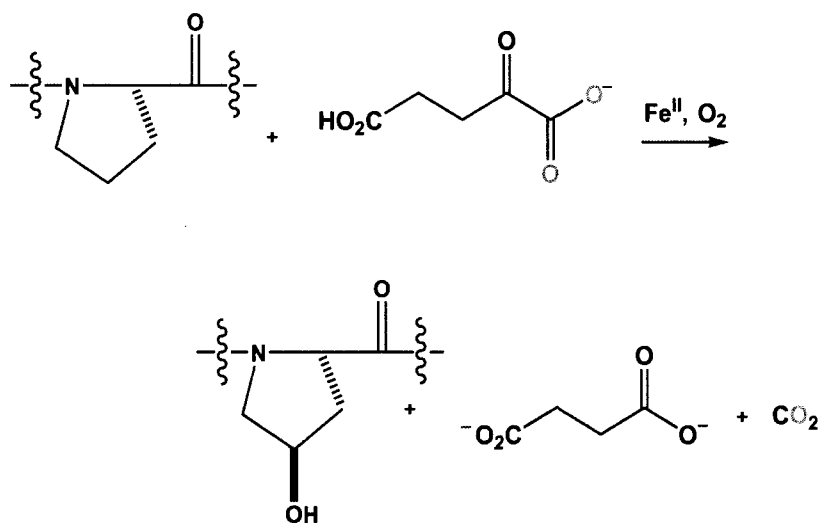
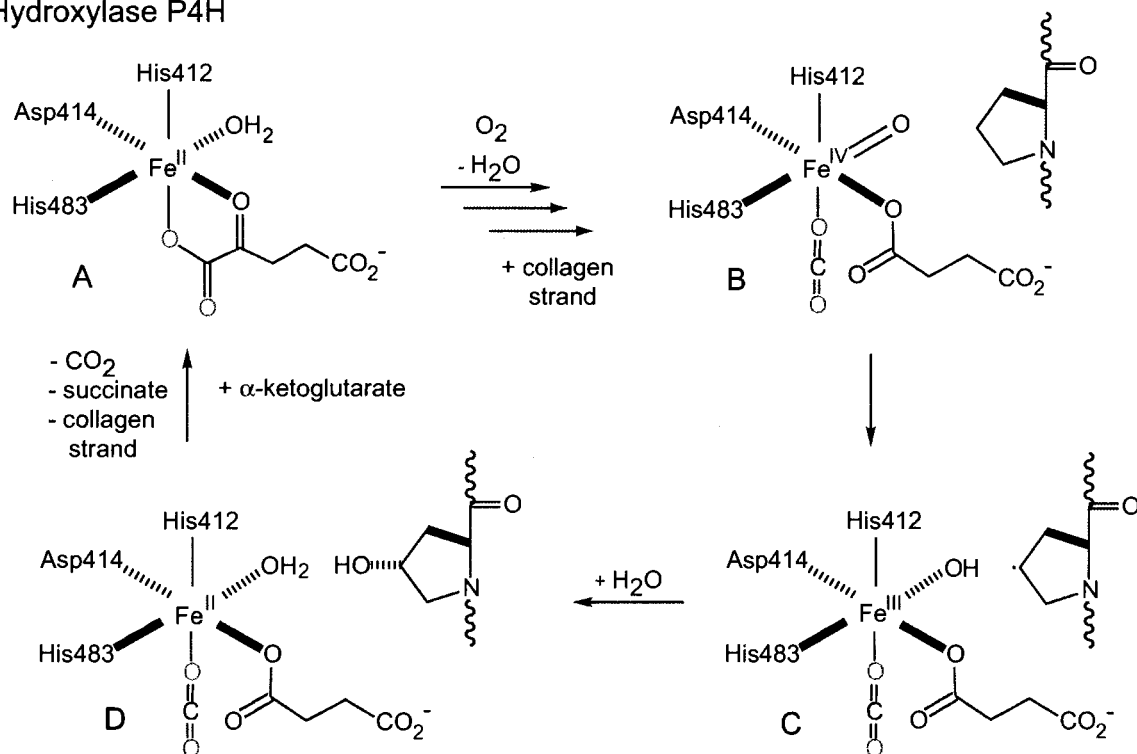


Figure 4.2 Reaction catalyzed by P4H.

A) Hydroxylase P4H



B) Halogenase SyrB2

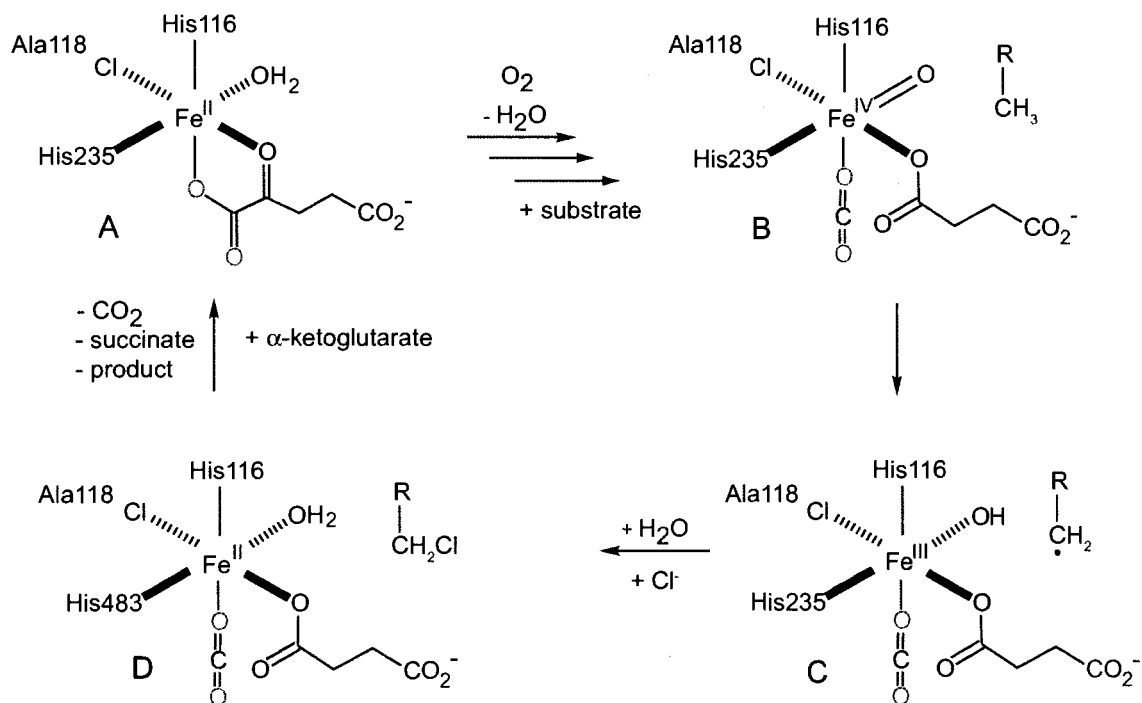


Figure 4.3 Proposed reaction mechanisms for the hydroxylase P4H (A) and halogenase SyrB2 (B). In both mechanisms, the intermediates are labeled **A–D**. **A** Iron(II) in the active site is bound by a 2-His-1-Asp motif in P4H, but a 2-His-1-chloride in SyrB2. **B** The reactive iron(IV)-oxo species is formed upon decarboxylation of α -ketoglutarate. **C** The iron(IV)-oxo species abstracts a hydrogen atom from the substrate producing a radical intermediate. **D** In the hydroxylase reaction, the substrate radical recombines with the hydroxyl moiety. In the halogenase reaction, the substrate radical reacts with the chloride ion.

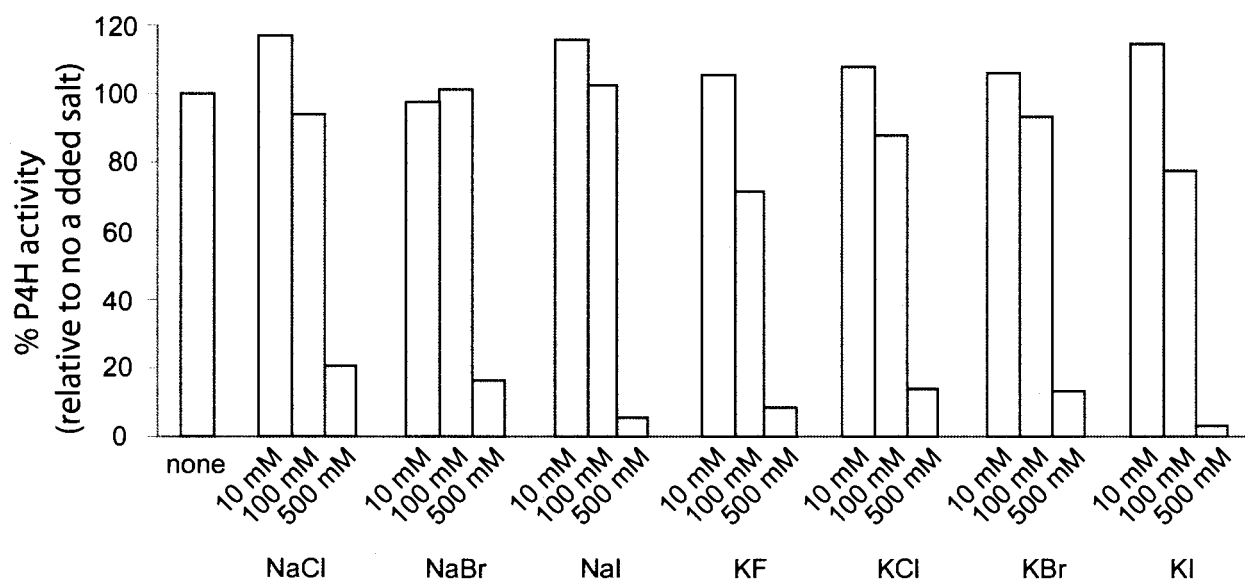


Figure 4.4 Tolerance of P4H activity to salts. P4H activity is shown in the absence and presence of increasing concentrations of sodium chloride (NaCl), sodium bromide (NaBr), sodium iodide (NaI), potassium fluoride (KF), potassium chloride (KCl), potassium bromide (KBr), and potassium iodide (KI). Reaction mixture contents are described in the Materials and methods section.

Chapter Five

Future Directions

5.1 Three-dimensional structure of P4H

Three-dimensional structural data for a protein provide a plethora of information that can explain previous observations, as well as generate new hypotheses. Although the structure of collagen prolyl 4-hydroxylase (P4H) remains to be determined, the three-dimensional structure of a number of enzymes in the α -ketoglutarate-dependent, iron(II) dioxygenase family have been solved. The structure of deacetoxycephalosporin C synthase (DAOCS) was the first to be solved (Valegard et al., 1998), followed by those of clavamate synthase (Zhang et al., 2000), taurine/ α -ketoglutarate dioxygenase (TauD) (Elkins et al., 2002), alkylsulfatase (AtsK) (Muller et al., 2004), and factor-inhibiting hypoxia-inducible factor 1 (FIH-1) (Elkins et al., 2003).

Structures of dioxygenases that catalyze the hydroxylation of proline residues, though not in collagen, are also known. The bacterial proline 3-hydroxylase adds a hydroxyl group to proline prior to its incorporation into antimicrobial, nonribosomal peptides (Clifton et al., 2001); prolyl hydroxylase domain protein (PHD2) is involved in oxygen sensing (McDonough et al., 2006); and the P4H from the alga *Chlamydomonas reinhardtii* (Cr-P4H-1) is related to plant P4Hs implicated in the synthesis of cell-wall glycoproteins (Koski et al., 2007). Recently, the structures of two prolyl hydroxylases complexed with peptide substrates have been solved, namely, Cr-P4H-1 with a (Ser-Pro)₅ peptide (Koski et al., 2009) and PHD2 with a peptide derived from hypoxia inducible factor α (Chowdhury et al., 2009). These enzymes share a common jelly roll fold, and the 2-His-1-carboxylate motif in the active sites are all positioned in a similar manner. Outside the jelly roll motif, the structures among this family of enzymes vary greatly due to the wide variety of substrates that each enzyme binds. These structures show that

Cr-P4H-1 and PHD2 bind the prolyl substrate in the C^γ-endo ring pucker conformation. This is also true for the collagen P4H, as presented in CHAPTER TWO.

The collagen P4H is unique in that it is a heterooligomeric protein. The active site is located in the C-terminal portion of the α subunit of the $\alpha_2\beta_2$ tetramer. No structural information is available for the active-site region, though it is predicted to be located within a jelly roll motif structure. The structure of the peptide-substrate-binding domain (Phe144–Ser244) of human P4H α (I) is comprised of tetratricopeptide repeat motifs. This region includes a concave surface groove lined by tyrosine residues that is a putative binding site for proline-rich peptides (Pekkala et al., 2004). None of the tyrosines was necessary for assembly with the β subunit. The β subunit is identical to protein disulfide isomerase (PDI). PDI is a member of the thioredoxin (Trx) superfamily that catalyzes the formation, reduction, and isomerization of disulfide bonds. Nevertheless, the active-site cysteine residues in PDI are not required for P4H tetramer assembly or activity (Vuori et al., 1992c). PDI is composed of four Trx-like domains, abb'a', however the b'a' domains are the minimum requirement for the formation of an active $\alpha_2\beta_2$ P4H (Pirneskoski et al., 2001). Although a structure of PDI from yeast is available (Tian et al., 2006), there is no P4H in yeast and the structure offers little as to how human PDI is arranged in the P4H tetramer. The structure of human PDI is yet to be determined.

The study of P4H has been aided by the production of active P4H tetramers in *E. coli* (Kersteen et al., 2004; Neubauer et al., 2005), and we have attempted to optimize the purification protocol by adding an affinity tag. High-throughput crystallization screens, however, have not yielded adequate crystals for structure determination. More extensive methods will need to be employed, such as reductive methylation of lysine residues, limited proteolysis, or crystallization

under anaerobic conditions. Additionally, P4H α isoforms α (II) and α (III), which also form $\alpha_2\beta_2$ tetramers, splice variants of P4H, or P4H variants that stabilize a particular protein conformation and the tetramer could yield suitable crystals.

5.2 P4H recognition of prolyl ring pucker

The recent structures of peptide substrate bound Cr-P4H-1 (Koski et al., 2009) and PHD2 (Chowdhury et al., 2009) revealed the substrate prolyl residue in the C $^{\gamma}$ -endo ring pucker conformation. These structures could be examined to determine residues that are important for ring pucker recognition, and mutagenesis of those residues could alter the conformational preference of the enzyme. Sequence analysis of the collagen P4H could identify the corresponding residues important for substrate conformation in that enzyme. These structural data could also be analyzed to identify the second sphere residues that contribute to substrate binding, including the active site iron. A comparison of these structures with available structures of α -ketoglutarate-dependent, iron(II) halogenases might reveal motifs necessary for chloride ion binding in the halogenases.

5.3 P4H Xaa position substrate preference

The typical minimum sequence requirement for hydroxylation by P4H is Xaa–Pro–Gly (Kivirikko et al., 1992). Free proline, Pro–Gly–Xaa, and Gly–Xaa–Pro tripeptides are not substrates. Though Xaa–Pro–Ala sequences are hydroxylated occasionally, and Gly can be replaced by β -alanine, the Pro–Gly sequence is the principal recognition motif. Additionally, the Xaa position plays a role in substrate recognition, as the rates of prolyl hydroxylation are

influenced by Xaa. Identifying the amino acids in the Xaa position preferred by P4H from peptides reported in the literature, however, is difficult because those peptides differed outside of the Xaa–Pro–Gly region in length and composition, and reaction conditions and enzyme preparations also varied.

Polytripeptides (Xaa–Pro–Gly)_n, where Xaa is Pro, are reported to produce the highest rate of hydroxylation by P4H. Only the Pro preceding Gly is hydroxylated, not the Pro in the Xaa position. Alanine, leucine, arginine, valine, and glutamate in the Xaa position give lower rates of hydroxylation, in that order. A similar trend was observed for the efficacy of peptides containing the inhibitory proline analogue, 5-oxaproline. The extent of inhibition was influenced by the amino acid preceding 5-oxaproline, where phenylalanine was the most effective, followed by alanine, valine, and glutamate (Gunzler et al., 1988). When Xaa is Gly in polytripeptides, no hydroxylation occurs, but some collagens do contain Gly–Hyp–Gly sequences. Charged amino acids also decrease the rate of reaction by P4H, as Arg in the Xaa position of pentapeptides yields one third and Glu one twentieth the rate of Pro.

Overall, the analysis of P4H preferences for residues in the Xaa position is incomplete. About half of the proteinogenic amino acids have not been investigated in Xaa–Pro–Gly peptides as P4H substrates. Of those investigated, some amino acids affect the peptide structure and hydroxylation by P4H when in the Xaa position (Rapaka et al., 1978). Aromatic amino acids influence the cis/trans ratio of the Xaa–Pro bond (Thomas et al., 2006). The conformation of the peptide bond might influence prolyl hydroxylation. Trp has not been investigated in the Xaa position of P4H substrates, and direct comparisons have not been made among Tyr, Phe, and Pro. Cys, along with Trp, does not occur in the triple-helical region of Type 1 collagen, and also has not been tested in a P4H substrate peptide. A comprehensive screen all 20 amino acids in the

Xaa position of peptide substrates for P4H would provide information about the structure of the P4H active site, as well as guide the design of inhibitors of P4H.

5.4 Inhibitors of P4H

Excessive production of collagen causes a number of fibrotic diseases that affect organs, such as the lungs, liver, and kidneys. Due to its essential role in collagen biosynthesis, P4H is a major target for medical intervention (Franklin, 1997; Myllyharju, 2008). P4H can be inhibited by limiting the co-substrates of the reaction, iron, O₂, and α -ketoglutarate. Inhibitors of P4H that chelate iron or sequester oxygen are useful in the laboratory, though certainly are not viable in living organisms. Bivalent cations inhibit P4H competitively with respect to iron(II), the most effective being zinc(II). Analogues of α -ketoglutarate competitively inhibit P4H. Two of the most potent inhibitors are pyridine 2,5-dicarboxylate with a K_i value of 0.8 μ M, as well as oxalylglycine and its cell-permeable form, dimethyloxalylglycine, which inhibit P4H at low micromolar concentrations (Baader et al., 1994).

The most well-studied peptide inhibitors are polymers of proline. The K_i value of polyproline, like the K_m of collagen substrates, decreases with increasing peptide chain length. Analogues of proline incorporated into peptides can also inhibit P4H, such as 3,4-dehydroproline, in which a double bond is incorporated between C ^{β} and C ^{γ} in the prolyl ring (Nolan et al., 1978), and 3-methyleneproline (Tandon et al., 1998). Also, the proline analogue containing an oxygen atom in place of the fifth carbon in the ring, 5-oxaproline, is a suicide inhibitor (Gunzler et al., 1988). A mixture of isomers of 3-fluoroproline was found to be turned over by P4H (Tandon et al., 1998), but which isomer and its kinetic constants remains to be

determined. Other proline analogues that could be investigated as substrates or inhibitors of P4H include 4-methyleneproline, (2*S*,4*S*)-4-hydroxyproline, and 4,4-difluoroproline.

Creating inhibitors utilizing proline analogues likely requires incorporating these residues into peptides or proteins. Even though some of the analogues are taken up by cells, proline residues in all proteins would be replaced with the analogue. Many of these proline analogues and others can be incorporated cotranslationally into proteins in a bacterial expression system, yet there is limited efficiency (Kim et al., 2004). Importantly, the peptide or protein inhibitor must be of significant length since the strength of peptide binding increases drastically with increasing peptide chain length. Given these parameters, and the challenges associated with polypeptide drug administration, small molecule inhibitors of P4H are desirable. Inhibitors could be sought out by high-throughput screening, perhaps by adapting the assay described in CHAPTER THREE, in combination with inhibitors developed utilizing the growing knowledge about the P4H active site and substrate conformational preferences.

Appendix One

Structure and Function of *Bacillus subtilis* YphP, a Novel Prokaryotic Protein Disulfide Isomerase with a CXC Catalytic Motif

Portions of this work have been published as

U. Derewenda, T. Boczek, K. L. Gorres, M. Yu, L. Hung, D. Cooper, A. Joachimiak, R.
T. Raines Z. S. Derewenda (2009) *Biochemistry*, doi: 10.1021/bi900437z.

This work was done in collaboration with Prof. Zygmunt S. Derewenda and coworkers, who were responsible for the production, crystallization, and solving the structure of YphP.

A1.1 Abstract

The DUF1094 family contains over 100 bacterial proteins of unknown function, all containing a conserved CXC motif. We solved the crystal structure of the *Bacillus subtilis* representative, the product of the *yphP* gene. This protein shows remarkable structural similarity to thioredoxins, with a canonical $\alpha\beta\alpha\beta\alpha\beta\alpha$ topology, despite low amino-acid sequence identity to thioredoxin. The CXC motif is found in the loop immediately downstream of the first β -strand, in a location equivalent to the CXXC motif of thioredoxins, with the first Cys occupying a position equivalent to the first Cys in canonical thioredoxin. The experimentally determined reduction potential of YphP is $E^{\circ'} = -130$ mV, significantly higher than that of thioredoxin and consistent with disulfide isomerase activity. Functional assays confirmed that the protein displays isomerase activity. We propose a mechanism by which the members of this family catalyze isomerization using the novel CXC catalytic site.

A1.2 Introduction

The *Bacillus subtilis* *yphP* gene codes for a member of a superfamily of over 100 prokaryotic, highly conserved proteins (DUF1094). These proteins are found predominantly in *Firmicutes* such as *Staphylococcus* and *Bacillus*, including *S. aureus* and *B. anthracis*, important human pathogens. No function has been assigned to this family and to date there is no structure described for any of its representatives. Tertiary structure prediction using 3DJury (Ginalski et al., 2003; von Grotthuss et al., 2003) suggests that YphP has a core domain with high similarity to the canonical thioredoxin (Trx) fold, as exemplified by the *Trypanosoma brucei* Trx (1r26) or *Malassezia sympodialis* Trx (2j23), even though amino-acid sequence identity level is very low (~ 15%). Trx is the prototypical member of the thiol–disulfide oxidoreductase family of enzymes that catalyze the formation, reduction, and/or isomerization of disulfide bonds. A similar fold in YphP raises the possibility that members of the DUF1094 proteins constitute a novel family of oxidoreductases. The DUF1094 proteins, however, lack the canonical CXXC sequence (where X denotes any amino acid) of the active site, and instead contain a CGC motif, so far only found in a family of enzymes with putative oxidoreductase activity (Ferguson et al., 2006).

The ubiquitous Trx superfamily includes proteins with a unique fold, in which the central, mixed β -sheet is sequestered between several short α -helices (Pan and Bardwell, 2006). These oxidoreductases contain a fingerprint CXXC sequence, in which the two cysteines alternate between the reduced and oxidized states, allowing them to participate in disulfide-exchange reactions. The location of this active-site sequence is highly conserved, at the end of the canonical β 2 strand and into the following α -helix, so that the N-terminal Cys is in an extended conformation, while the remaining three amino acids assume α -helical secondary structure. Interestingly, the sequence of the XX dipeptide varies among the superfamily members, and is

important for modulating the redox properties of the protein (Quan et al., 2007). For example, in DsbA, which catalyzes the formation of disulfide bonds in proteins secreted into the periplasmic space of *Escherichia coli*, a CPHC motif results in the highest known reduction potential ($E^{\circ'} = -121$ mV) matching the oxidizing environment of the periplasm (Grauschopf et al., 1995). In contrast, cytosolic thioredoxins contain a canonical CGPC motif with an $E^{\circ'}$ value that is 160 mV lower, i.e., -270 mV, which indicates a preference for an oxidized state and matches the reducing environment of the cytoplasm (Aslund et al., 1997). Glutaredoxins contain a CPYC motif and their reduction potentials vary depending on structural context, so that the *E. coli* glutaredoxins 1 and 3 have reduction potentials of -233 mV and -198 mV, respectively (Aslund et al., 1997). Finally, the protein disulfide isomerase (PDI) of the yeast endoplasmic reticulum has a CGHC sequence and a reduction potential of -180 mV (Laboissière et al., 1995).

In spite of this functional diversity, the general architecture of the active-site motif is highly conserved, so that no member of the thioredoxin superfamily has been found with fewer or more than two intervening residues between the cysteines. Consequently, the functional importance of the number of the residues between the cysteines for the reduction potential has been studied in synthetic peptides (Zhang and Snyder, 1989). These experiments revealed that for peptides $C(X)_nC$, where $n \leq 5$, those with a single intervening residue have the least stable disulfide. It has also been shown that a CGC-NH₂ tripeptide has a reduction potential of -167 mV, which is close to that of PDI, and that it exhibits disulfide isomerization activity (Woycechowsky and Raines, 2003). A CXC motif engineered into the active site of an *E. coli* thioredoxin by deleting the Pro residue from the CGPC active site showed likewise a destabilized disulfide with a reduction potential of ≥ -200 mV. In its reduced form, this variant displayed disulfide isomerase activity

that was 25-fold greater than that of the synthetic CGC peptide (Woycechowsky and Raines, 2003).

Surprisingly, amino-acid sequence alignments showed that YphP and all other members of the DUF1094 family contain a CGC motif. To gain insight into the structure and function of the DUF1094 family, we determined the crystal structure of the *B. subtilis* YphP at 2.1 Å resolution. The refined model confirmed that the protein is structurally closely related to the thioredoxin superfamily. The cysteine residues of the CGC motif in YphP, Cys53 and Cys55, are in a location analogous to the active site of thioredoxin. Cys53 occupies a site corresponding to the canonical N-terminal cysteine in the CXXC motif, while Cys55 is located in place of the Pro in Trx. To better understand the function of YphP, we determined its reduction potential and analyzed its isomerase activity by using a continuous assay that measures the rate of isomerization of non-native to native disulfide bonds in tachyplesin I (Kersteen et al., 2005). We then used two single site variants, C53A and C55A, to investigate the role of each of the two cysteine residues in the active site. Finally, we compared the activity of YphP to that of ΔP34 thioredoxin in the fluorescence assay. The results confirm that YphP represents a new family of oxidoreductases, the first with a natural CXC motif within its active site. Both the reduction potential and functional assays are consistent with protein disulfide isomerase activity.

A1.3 Materials and methods

Protein production and crystallization. The *B. subtilis* YphP (also denoted APC1446) was originally selected as one of the targets for high-throughput structure determination by the Midwest Center for Structural Genomics (<http://www.mcsg.anl.gov>). The open reading frame was amplified from genomic DNA with a recombinant KOD HiFi DNA polymerase (Novagen)

from *Thermococcus kodakaraensis* using conditions and reagents provided by the vendor (Novagen). The gene was cloned into a pMCSG7 vector (Stols et al., 2002) using a modified ligation independent cloning protocol (Dieckman et al., 2002). This construct produced a fusion protein with an N-terminal His₆ tag and a recognition site for the tobacco etch virus (TEV) protease. Unfortunately, further progress was impeded by the recalcitrance of the protein to crystallization. To circumvent this problem we used the entropy reduction approach (Derewenda, 2004; Derewenda and Vekilov, 2006; Goldschmidt et al., 2007). We identified three clusters with high conformational entropy residues: cluster 1, E39,K40,E42; cluster 2, K113,E114; and cluster 3, Q100,E101. Variants were generated in which residues in these clusters were replaced with either Tyr or Ala.

Wild-type protein was expressed with high yield in *E. coli* BL21(DE3)RIPL. The protein was purified using standard methods and a combination of nickel affinity chromatography (Ni-NTA agarose column, Qiagen), followed by rTEV proteolysis to cleave the His tag, and gel filtration in the presence of reducing agents (1 mM DTT, 5 mM β -mercaptoethanol) to circumvent aggregation caused by intermolecular disulfide formation. The protein was concentrated to 8-15 mg/ml. Variant proteins were purified in a similar manner.

A proprietary set of conditions (Cooper et al., 2007) was used for crystallization screening, against a reservoir of precipitant and an alternative reservoir containing salt (Newman, 2005). The cluster 1 Ala variant, and cluster 3 Tyr variant yielded poorly reproducible crystals under a variety of conditions. In contrast, the cluster 3 Ala variant (i.e., Q100A,E101A) gave reproducible crystals in two different conditions: 3.2 M (NH₄)₂SO₄, 0.1 M citric acid buffer at pH 5.0 and 2.0 M (NH₄)₂SO₄. Following optimization, the best crystals were obtained from

2.6 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M HEPES buffer at pH 7.6 with added 30% trimethylamine N-oxide; 5 mM β -mercaptoethanol in the reservoir was necessary for crystals to reach maximal dimensions.

To generate the selenomethionine (SeMet) labeled protein samples, the protein was expressed in *E. coli* B834 cells. Seed cultures of 20 mL were grown in Luria Broth for 4 h at 37 °C. The seed cultures were spun down at 3,000 r.p.m. and used to inoculate auto-induction medium containing SeMet. The cultures were grown for 4.5 h at 37 °C, followed by ~16 h at 20 °C and another 24 h at 10 °C. Harvested cultures yielded ~35 g of wet mass per 2 L. Protein was purified using the procedure described above and yielded more than 100 mg of pure SeMet-labeled protein. Protein was concentrated to ~25 mg/ml in a buffer containing 25 mM NaCl, 25 mM Tris buffer at pH 8.0, and 1 mM DTT. The structure was determined from SeMet-labeled crystal grown from 2 M $(\text{NH}_4)_2\text{SO}_4$, 100 mM Tris buffer at pH 6.5, and 5% PEP (pentaerythritol propoxylate) 629.

All single- and multiple-site mutations were introduced using the QuikChange™ site-directed mutagenesis kit (Stratagene) and confirmed by direct DNA sequencing. The respective protein variants were purified as described above for the wild-type.

Δ P34 *E. coli* thioredoxin was produced and purified as we described previously (Woycechowsky and Raines, 2003).

Crystallography. The crystals of the Q100A, E101A variant are tetragonal, $P4_12_12$, with unit cell $a=b=68.16$ Å, $c=302.02$ Å, and contain four molecules in the asymmetric unit. The crystals contain ~55% solvent. For data collection, a crystal was frozen in mother liquor containing additional 20% glycerol. Data were collected at beamline 5.0.2 of the Advanced Light Source

(ALS). The multiwavelength anomalous diffraction (MAD) data, used to solve the structure, were collected at three wavelengths: 0.9795 Å (peak), 0.9796 Å (inflection point) and 0.9393 Å (high energy remote). The data extended to 2.3–2.5 Å with the merging *R* factors ranging between 11.9 and 13.6% (see Table A1.1). Positions of selenium atoms were identified with SHELXD (Schneider and Sheldrick, 2002). Crystallographic phasing was done with SOLVE (Terwilliger, 2003). Phases were improved by iterative model building and refinement in RESOLVE (Terwilliger, 2003) followed by ARP-WARP (Morris et al., 2003) and manual rebuilding with O (Jones et al., 1991) and COOT (Emsley and Cowtan, 2004). Refinement was carried out with REFMAC5 (Murshudov et al., 1997) and with PHENIX (Zwart et al., 2008) using the TLS (translation/libration/screw) approximation of thermal motion (Winn et al., 2001). Solvent accessibility of the Cys residues was calculated using the PISA server (Krissinel and Henrick, 2007).

Isomerase Activity. Assays of disulfide isomerase activity were performed as described previously (Kersteen et al., 2005). This assay is continuous and uses a homogeneous peptide substrate. The substrate is a modified version of tachyplesin I, in which a dansyl fluorophore is added and the disulfide bonds are rearranged into a non-native conformation (Fig A1.1). Disulfide bond isomerization was assayed by monitoring the increase in fluorescence emission at 465 nm upon excitation at 280 nm over time. The final concentration of enzyme was 1 μM. Kinetic parameters were determined by nonlinear least-squares regression analysis using eq 1:

$$F = F_0 + (F_{\max} - F_0)(1 - e^{-kt}) \quad (1)$$

Disulfide-bond Reduction Potential. The reduction potential of the disulfide bond in YphP was determined from the thiol–disulfide exchange equilibrium between YphP and glutathione.

YphP (10 μ M) was incubated in 50 mM sodium phosphate, pH 7.6, containing EDTA (1 mM) and various ratios of reduced/oxidized glutathione, $E^{\circ'} = -0.252$ V (Lees and Whitesides, 1993). The reaction mixture was blanketed with Ar(g), and the reaction was allowed to proceed for 60 min at 30 $^{\circ}$ C, after which it was quenched and YphP was precipitated by the addition of trichloroacetic acid to 10%. The precipitated protein was washed with acetone, and resuspended in 100 mM sodium phosphate, pH 7.6, containing 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB; 0.10 mg/mL). The concentration of thiols was quantified by the ensuing absorbance at 412 nm. The amount of reduced YphP was compared to that of YphP that had been incubated with DTT (1 mM) to obtain the fraction of reduced YphP (f). The standard reduction potential of YphP was calculated by nonlinear least-squares analysis of the data with the Nernst equation (Eq. 2),

$$f = f_{\min} + \frac{f_{\max} - f_{\min}}{1 + e^{-(E - E^{\circ})(nF/RT)}} \quad (2)$$

where n is the number of electrons, F is the Faraday constant (96,485 J V $^{-1}$ mol $^{-1}$), R is the gas constant (8.314 J K $^{-1}$ mol $^{-1}$), T is the temperature (here, 303 K), E is the reduction potential of the solution, and E° is the reduction potential of YphP.

A1.4 Results and discussion

Crystal structure. The crystal structure of YphP was solved using the multiwavelength anomalous dispersion (MAD) method and the atomic model was refined to an R -factor of 19.4% (Table A1.1). The final model contains 584 amino acid residues (from four independent protein molecules), 374 water molecules, and twelve SO $_4^{2-}$ ions. 92.1% of the residues are in the most favored regions of the Ramachandran plot with the remaining residues in additionally allowed

regions. Three out of the four molecules show interpretable electron density for all 144 residues with additional cloning artifact of three residues (SNA) on the N-terminus. The fourth molecule is slightly less well defined, and only residues 1 to 143 were placed in the electron density. A number of electron density peaks in the solvent region were attributed to sulfate ions. Details of the data collection, refinement and model quality statistics are listed in Table A1.1.

The four molecules in the asymmetric unit are arranged into a dimer of dimers in which the major contacts are mediated by an N-terminal helix, unique to YphP and not found in a canonical thioredoxin fold (Fig A1.2A). A comparison of the tertiary structure of YphP to other known structures using DALI (Holm et al., 2006) shows that, with the exception of the N-terminal helix, the protein is very similar to a typical thioredoxin, with the highest similarity (Z score of 11.6) to the human thioredoxin (Andersen et al., 1997). The canonical $\alpha\beta\alpha\beta\beta\alpha$ topology is well conserved, although there are differences in the mutual disposition and lengths of the individual helices and loops (Fig A1.2B). The tetrameric ensemble appears to constitute a crystallographic artifact, because gel filtration data (not shown) show that the protein is monomeric in solution.

The four crystallographically independent molecules display a degree of conformational variability with respect to the N-terminal α -helix, suggesting that the latter is flexible with respect to the core of the molecule. This may account, at least in part, for the recalcitrance of the wild-type protein to crystallization. Several intermolecular contacts define the molecular packing. Importantly, three out of four YphP molecules in the asymmetric unit form crystal contacts directly *via* the low-entropy patches generated by the Q100A, E101A substitutions. This pattern suggests that an oligomeric ensemble may form first transiently in solution, and is then incorporated into the crystal *via* surface patches introduced by mutagenesis. Similar mechanism is observed for other proteins crystallized by this approach (Derewenda, Z.S., unpublished).

The CXC motif (Fig A1.3) is located within a loop following the first β -strand, as predicted by 3D-Jury (Ginalski et al., 2003) based on the amino-acid sequence. Cys53 is in an extended conformation ($\phi = \sim -70$, $\psi = \sim 165$; the dihedral angles are similar in all four molecules), and its position within the active-site loop is analogous to the first Cys of the canonical CXXC motif. The residue Gly54 ($\phi = \sim -45$, $\psi = \sim -40$) is the first amino acid of the $\alpha 1$ helix, while Cys55 occupies the position of the second X (Pro in thioredoxin). Of the two cysteines, Cys53 is more solvent exposed ($\sim 42 \text{ \AA}^2$) than Cys55 ($\sim 17 \text{ \AA}^2$), although there is some variability among the four molecules and there is evidence of static disorder of the side chain of Cys55. The distances between the two S^γ atoms in the four molecules range from 3.6 \AA to 5.9 \AA , ruling out intramolecular disulfide bonds in the crystal structure; thus, the crystal structure represents the reduced form of the enzyme, as expected from the presence of reducing agents. The sulfhydryl of Cys53 is involved in two hydrogen bonds: with the hydroxyl O^γ of Ser51 and with $N^{\eta 1}$ of Arg121 (Fig. A1.3), although specific distances vary depending on the side chain conformation. Both these residues are completely conserved within the family, suggesting a functional role. Interactions of arginines with thiolates are known in other redox-regulated proteins. For example, in the hydroperoxide resistance protein Ohr, the active-site cysteine Cys60, which is directly involved in peroxide reduction, accepts a hydrogen bond from $N^{\eta 1}$ of Arg18 at 3.4 \AA (Lesniak et al., 2002). In the *B. subtilis* chaperone Hsp33, the Zn^{2+} -binding Cys235 is within 3.4 \AA of $N^{\eta 1}$ of Arg86 (Janda et al., 2004). By contrast, Cys55 has no similar interactions. These different environments suggest that the pK_a value of Cys53 is lower than that of Cys55 and that Cys53 could be the key nucleophile in the active site.

In spite of the obvious similarity of the CGC motifs in YphP and the $\Delta P34$ variant of Trx (Woycechowsky and Raines, 2003), the active sites of these two proteins need not be structurally

equivalent. In Trx, the deletion of Pro34 might cause a compensating rotation of the α -helix, although there is no X-ray structure available to prove this notion. Also, given the amphipathic character of this helix, it is difficult to see how a deletion of Pro34 could be accomplished without disturbing the protein's stability. In fact, the thermal stability of the Δ P34 variant has been shown to be lower than that of the wild-type protein by $\sim 10^\circ\text{C}$ (Woycechowsky and Raines, 2003). In YphP, Cys55 simply occupies the place of Pro34 in Trx, while the position immediately downstream is occupied by an Ala, thus avoiding the strain that might characterize the Δ P34 Trx variant.

DUF1094 family. Figure A1.4 shows an alignment of representative amino-acid sequences of proteins in the DUF1094 family, as listed in the Pfam database (Finn et al., 2008). This family comprises 102 sequences in 62 species. The majority (80 sequences) are found among the *Firmicutes*, mostly in the *Bacillus* and *Staphylococcus* genera. Interestingly, in these organisms there are nearly always two homologous genes of this family in any species. The remaining sequences are in *Acidobacteria* and *Bacteroidetes*, and there is only a single gene in each species. The family shows a relatively high level of amino-acid conservation with several completely conserved motifs in addition to the CGCA active site. Topologically, all these conserved motifs are clustered primarily in loops around the active site, suggesting a possible role in substrate recognition. The atypical, N-terminal α -helix is the least conserved fragment in the DUF1094 family. We were unable to identify any homologues of these proteins in eukaryotic genomes.

Catalytic activity. The overall structural similarity of YphP to thioredoxins, and the presence of the CGC motif in the putative active site suggest that the protein, as well as its homologues in

the DUF1094 family, constitute a novel family of oxidoreductases. To confirm this hypothesis we carried out functional assays.

First, we determined the reduction potential, $E^{\circ'}$, of YphP to be (-130 ± 5) mV (Fig. A1.5). As pointed out above, for the known oxidoreductases with CXXC motifs in their active sites, the reduction potentials range from -270 mV in the *E. coli* *thioredoxin* (Aslund et al., 1997), a cytosolic reductant, to -122 mV in DsbA, which catalyzes oxidation of proteins secreted into the periplasm (Grauschopf et al., 1995). The stability of the active-site disulfide bond is tuned, in part, by the amino-acid residues between the active-site cysteines, *i.e.* the XX residues in the CXXC motif. The reduction potential of YphP was determined to be $E^{\circ'} = (-130 \pm 5)$ mV, which is notably high, and consistent with YphP being an isomerase/oxidase rather than a reductase. The only other protein containing a CXC motif within an intact protein, for which $E^{\circ'}$ has been determined, is the $\Delta P34$ variant of Trx; in that case $E^{\circ'} \geq -200$ mV, at least 70 mV higher than in wild-type Trx (Woycechowsky and Raines, 2003). It seems that decreasing the number of intervening residues between the cysteines from two (CXXC) to one (CXC), provides at least one mechanism to increase the reduction potential of a thiol–disulfide oxidoreductase. It is noteworthy, however, that for a synthetic CysGlyCysNH₂ peptide the reduction potential $E^{\circ'} = -167$ mV (Woycechowsky and Raines, 2003). Thus, the structural scaffold of YphP is also responsible for a further increase of $E^{\circ'}$ by about 40 mV.

We also showed that YphP was able to catalyze the isomerization of non-native to native disulfide bonds using tachyplesin I peptide as a substrate, yielding a k_{cat}/K_M value of $(4.7 \pm 0.03) \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$. This activity is $\sim 3\%$ that of protein disulfide isomerase (PDI) (Kersteen et al., 2005) and 42% that of $\Delta P34$ thioredoxin (Table A1.2). To investigate the roles of the cysteine thiols of YphP, we replaced each of the active-site cysteine residues with alanine, thus generating

single-site C53A and C55A variants. Unexpectedly, both of these variants catalyzed the isomerization of disulfide bonds in the scrambled tachyplesin substrate. The C53A and C55A variants had k_{cat}/K_M values of $(2.6 \pm 0.2) \times 10^3$ and $(3.3 \pm 0.1) \times 10^3 \text{ M}^{-1}\text{s}^{-1}$, respectively, only slightly lower than that of the wild-type protein. Yet, when *both* active-site cysteine residues were replaced with alanines, resulting in the C53A/C55A variant, isomerase activity was abolished completely (Table A1.2).

The catalytic mechanism. The canonical mechanism of disulfide-bond isomerization by PDI involves an attack on the substrate disulfide bond by the thiolate of the N-terminal cysteine of the CXXC motif (Kersteen and Raines, 2003; Wilkinson and Gilbert, 2004; Woycechowsky and Raines, 2000). A mixed disulfide is formed between PDI and the substrate, enabling the thiol and disulfides of the substrate to react until native disulfide bonds are formed in the substrate, so that PDI is no longer tethered to it by a mixed disulfide. Replacing the N-terminal Cys in PDI with a Ser eliminates enzymatic activity (Vuori et al., 1992a). In contrast, replacing the C-terminal Cys eliminates the ability of the enzyme to catalyze substrate reduction and oxidation, but not isomerization (Laboissière et al., 1995; Walker and Gilbert, 1997). During the rearrangement of disulfides in the substrate protein, a stable mixed disulfide between PDI and the substrate could prevent the release of PDI. The C-terminal cysteine of PDI is thought to provide an “escape” mechanism in which a disulfide is formed within the active site of PDI to allow for substrate release (Kersteen et al., 2005; Walker and Gilbert, 1997; Woycechowsky and Raines, 2003; Zhang and Snyder, 1989). A similar role has been ascribed to the C-terminal cysteine residue of thioredoxin (Chivers et al., 1996).

The crystal structure of YphP suggests that the protein functions as a disulfide isomerase using a mechanism similar to that of protein disulfide isomerase (Fig A1.6). In the reduced form,

Cys53 is most likely the attacking nucleophile. The structure predicts that the pK_a of this cysteine is lowered by hydrogen bonding with Arg121 and Ser31. These serine and alanine residues are conserved among the DUF1094 proteins. When a stable intermolecular disulfide forms, an escape route is provided by Cys55 which, we suggest, also becomes activated by Arg121 swinging into a new position, so that an intramolecular disulfide can form. This putative mechanism involves a crucial role assigned to Arg121, which is absolutely conserved in the DUF1094 family (Fig. A1.4). We suggest that the escape route involves transient formation of a Cys53-Cys55 disulfide, although we have not confirmed this experimentally.

Oxidized CXC sequences, which are unusual 11-member ring structures, are extremely rare in nature. The only examples in the PDB of high resolution structures containing such motifs are the thermophilic Ak.1 protease from *B. subtilis*, 1DBI (Smith et al., 1999), the L40C mutant of the NADH peroxidase, 1NHR (Miller et al., 1995), and Erv2p, 1JR8 (Gross et al., 2002) (Fig A1.7). A similar disulfide is also reported in the *Mengo encephalomyelitis* virus coat protein (Krishnaswamy and Rossmann, 1990), although low resolution precludes detailed stereochemical analysis. Finally, there is evidence that a CXC disulfide plays a functional role in the chaperone Hsp33 (Jakob et al., 1999), but no structure of the oxidized protein is available. In all three cases where high-resolution data are available, the torsion angles of the disulfide-containing ring suggest significant strain. It is interesting to note that the backbone conformations in the 11-member rings are different: in the Ak.1 protease and in Erv2P the residue in the X position is in the left-handed helical conformation, while in the NADH peroxidase mutant a serine occupying the equivalent position is in the right-handed helical conformation, very similar to Gly in YphP (Fig. A1.7). Based on these structural comparisons, it

appears that a disulfide bond in YphP is stereochemically possible, with a slight conformational adjustment to Cys55.

The observed activity of the two variants with only one of either of the active-site cysteines is easily rationalized in view of the fact that the escape mechanism is not necessary for the enzyme to function in assays with peptides, as described elsewhere (Kersteen et al., 2005). In the wild-type protein, Cys53 is likely to be the key nucleophile that attacks the substrate. It is more solvent exposed than Cys55, and the proximity of Arg121 makes its thiolate form more stable. However, in the absence of Cys53, e.g, in the C53A variant, Arg121 could easily swing towards Cys55, which is still sufficiently exposed to be reactive. This arrangement is different from that in *S. cerevisiae* PDI, where the C-terminal Cys is buried and unable to catalyze isomerization in the absence of the N-terminal Cys. Thus, the C53A and C55A variants can both catalyze isomerization.

A1.5 Conclusions

The *B. subtilis* YphP is a member of a novel oxidoreductase superfamily of enzymes (DUF1094), with a natural CXC motif within the active site and a tertiary fold closely reminiscent of the thioredoxins. The experimentally determined reduction potential, $E^{\circ'}$, is (-130 ± 5) mV. This is the first value reported for a CXC motif within a wild-type enzyme. Functional assays show that YphP isomerizes disulfide bonds in a peptide substrate. The enzymes in this family are highly conserved and occur primarily in the *Staphylococcus* and *Bacillus* species. All these organisms contain canonical genes that encode thioredoxins, which are cellular reducing agents that have only been crystallized in an oxidized state. It is possible,

that unlike thioredoxins, the DUF1094 enzymes are cellular oxidizing agents or *in vivo* catalysts of oxidative protein folding. Further biological studies will be needed to answer this question.

Table A1.1

Crystallographic data			
<i>Data Collection Statistics</i>			
	Remote	Edge	Peak
Wavelength (Å)	0.9393	0.9796	0.9795
Resolution (Å)	75.0–2.30 (2.39–2.30) *	75.0–2.50 (2.59–2.50)	75.0–2.30 (2.39–2.30)
Total Reflections	148,455	167,160	203,269
Unique Reflections	30,085	25,434	33,433
Redundancy	4.93	6.57	6.08
Completeness (%)	90.2 (59.1)	94.3 (71.5)	97.2 (82.9)
R _{merge} (%) **	13.6 (42.3)	12.4 (48.2)	11.9 (49.1)
Average I/σ (I)	17.4 (2.1)	19.3 (1.9)	19.7 (2.2)
<i>Refinement Statistics</i>			
Resolution (Å)	75.0–2.30 (2.37–2.30) *		
Reflections (working)	31,840 (2461)		
Reflections (test)	1,593 (126)		
R _{work} (%) §	19.0 (31.7)		
R _{free} (%) §	24.6 (37.3)		
Number of protein atoms	4484		
Number of water oxygens	249		
Number of water sulfate ions	12		
Average B Factors (Å ²) †			
Main Chain	50.7		
Side Chain	56.3		
Waters	50.6		
R.m.s. deviation from ideal geometry			
Bonds (Å)	0.006		
Angles (°)	1.02		
Ramachandran plot – favored (%)	97.9		
Ramachandran plot – outliers (%)	0.0		

* The numbers in parentheses describe the relevant value for the highest resolution shell.

** $R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum I_i$ where I_i is the intensity of the i -th observation and $\langle I \rangle$ is the mean intensity of the reflections. The values are for unmerged Friedel pairs.

§ $R = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$, crystallographic R factor, and $R_{\text{free}} = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$ where all reflections belong to a test set of randomly selected data; $R_{\text{overall}} =$ crystallographic R factor calculated for all data after a final cycle of refinement

† B-factors were refined using TLS approximation (see Methods)

Table A1.2**Values of $k_{\text{cat}}/K_{\text{M}}$ for catalysis of disulfide bond isomerization.**

Enzyme	Variant	$k_{\text{cat}}/K_{\text{M}}$ ($10^3 \text{ M}^{-1} \text{ s}^{-1}$)
YphP	wild-type	4.7 ± 0.1
YphP	C53A	2.6 ± 0.2
YphP	C55A	3.3 ± 0.1
YphP	C53A/C55A	0.3 ± 0.2
Trx	ΔP34	11.2 ± 0.8
PDI	wild-type	$170 \pm 50^{\text{a}}$
PDI	CGHA/CGHA	$86 \pm 35^{\text{a}}$

^a Data from ref. (Kersteen et al., 2005)

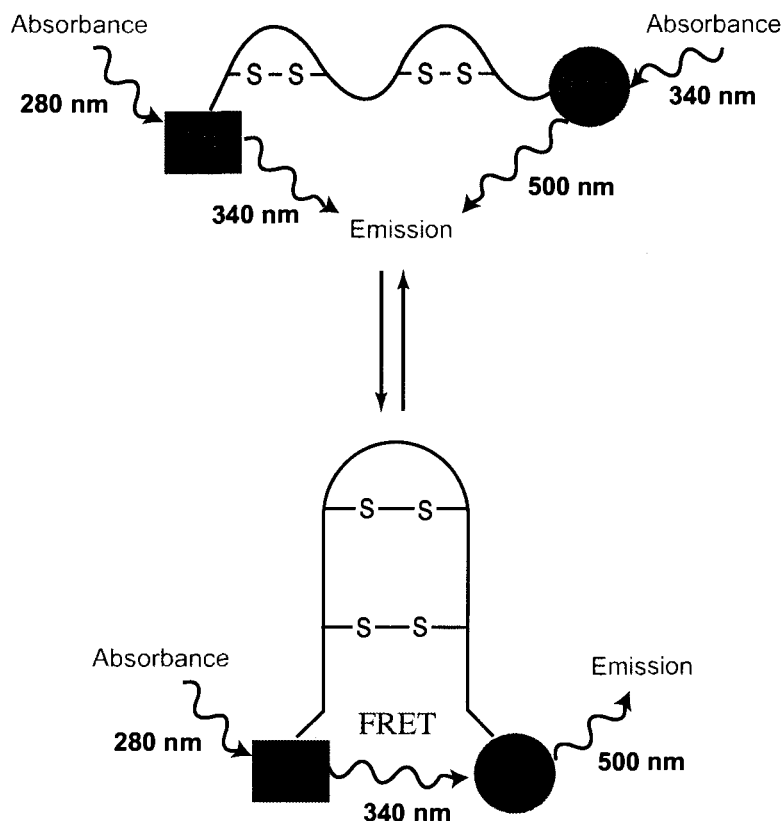


Figure A1.1 FRET-based assay for disulfide bond isomerization. Tachyplesin I (TI) folds into a β -sheet stabilized by six hydrogen bonds and two disulfide bonds (Cys3–Cys16 and Cys7–Cys 12). For use in the disulfide bond isomerization assay, Arg17 was changed to a lysine residue, and the dansyl acceptor fluorophore was attached to the ϵ -amino group of Lys17. Trp2 is the donor fluorophore. The non-native disulfide bonds in scrambled TI (sTI), where Cys3 and Cys7 form one disulfide and Cys12 and Cys16 form a second disulfide, are isomerized to the native structure.

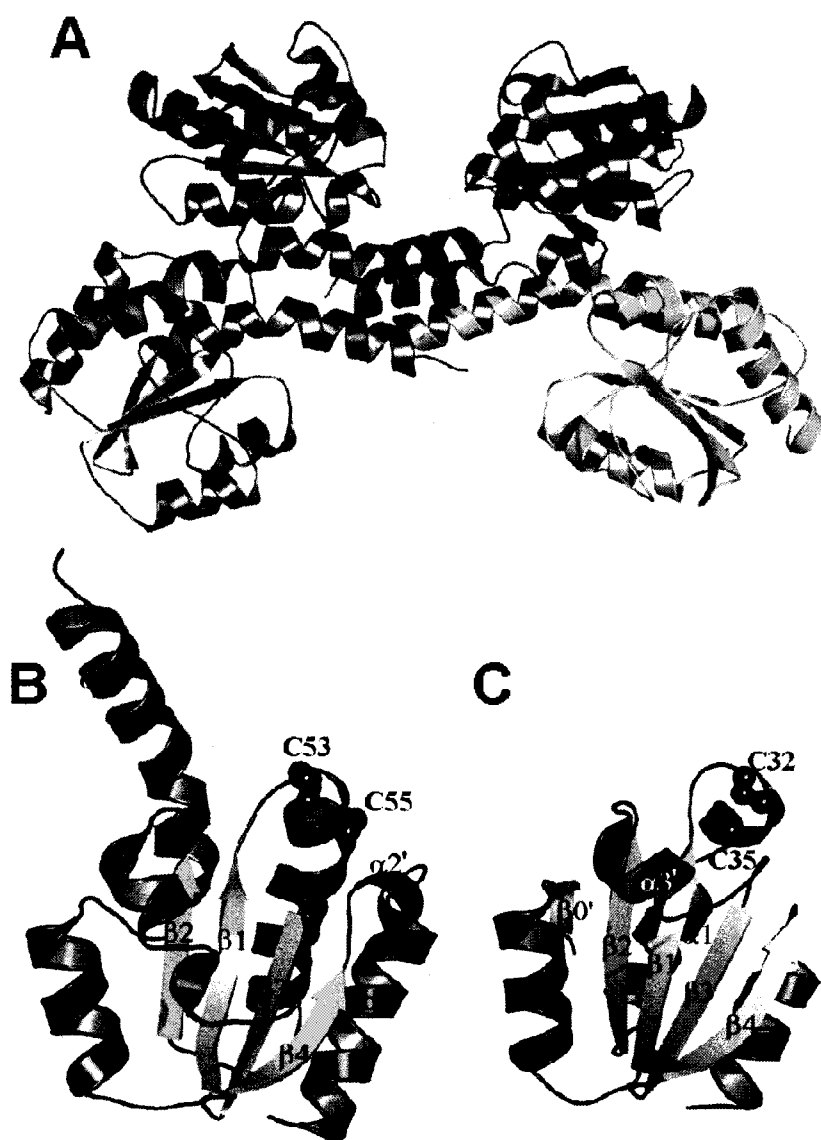


Figure A1.2 Ribbon diagrams of YphP. (A) The packing of four crystallographically independent molecules of YphP is colored by each molecule. (B) A single molecule of YphP is shown with the secondary structure elements identified (α -helices red, and β -strands yellow) and the two Cys side chains of the CXC motif shown as spheres and labeled. The additional N-terminal α -helix unique to the DUF1094 family is shown in cyan. (C) An analogous view of the human thioredoxin (PDB code 1AUC) is shown for comparison.

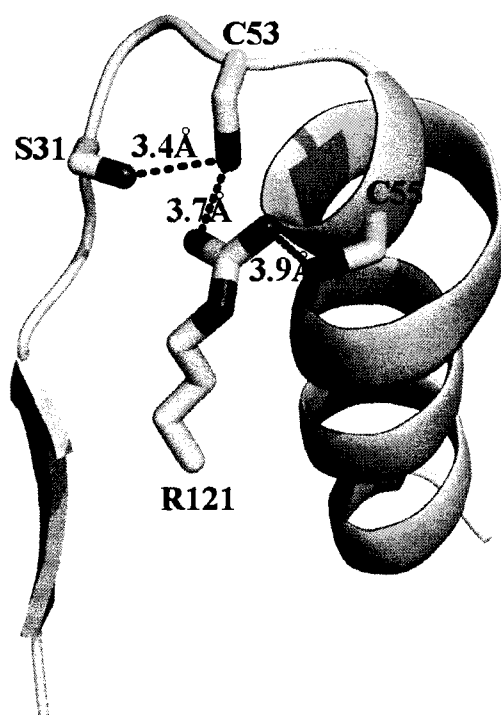


Figure A1.3 Structure of the catalytic loop of YphP. The coordinates used in this diagram are those of chain D, but are representative of all four chains in the structure.

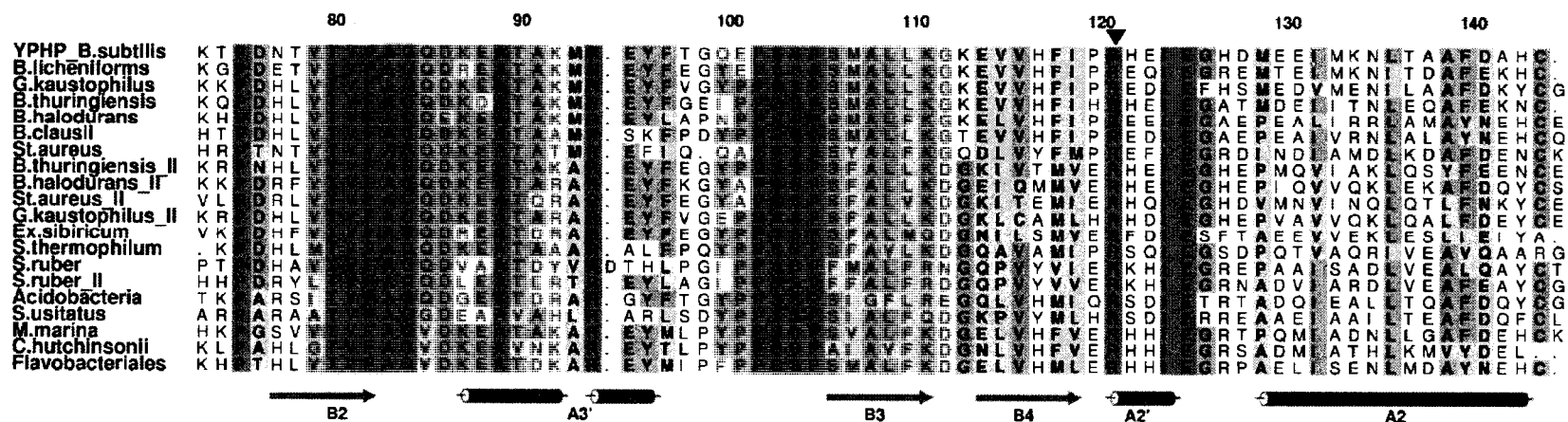
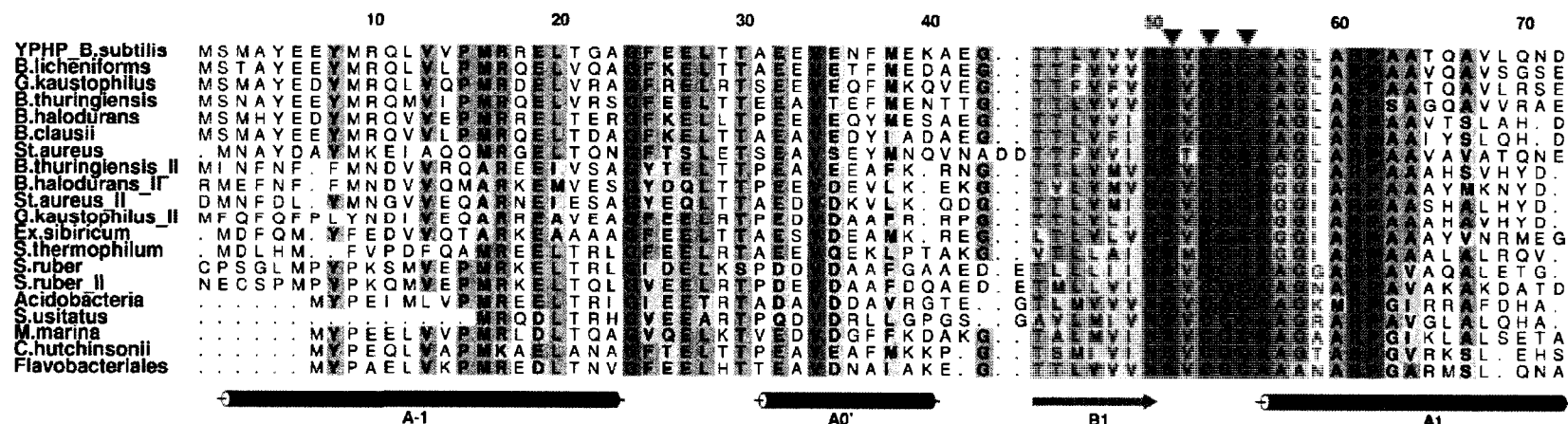


Figure A1.4 Sequence alignment of representative members of the DUF1094 family. Black triangles show the four residues that we identify as the catalytic amino acids (further details in the text). Semi- and fully conserved residues are indicated by darkening blue color. The secondary structure elements are shown below the alignment as cylinders (α -helices) and arrows (β -sheets).

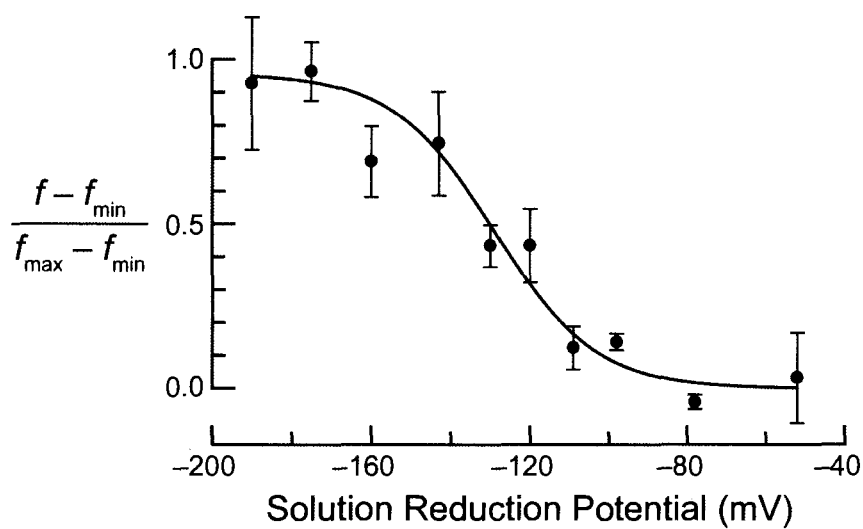


Figure A1.5 Reduction potential of YphP. The fraction of reduced YphP (f) is plotted as a function of the solution reduction potential, which was established at 30 °C by using reduced and oxidized glutathione. Values are the mean (\pm SE) from three experiments. The data were fitted to eq 2 to give $E^{\circ'} = (-130 \pm 5)$ mV.

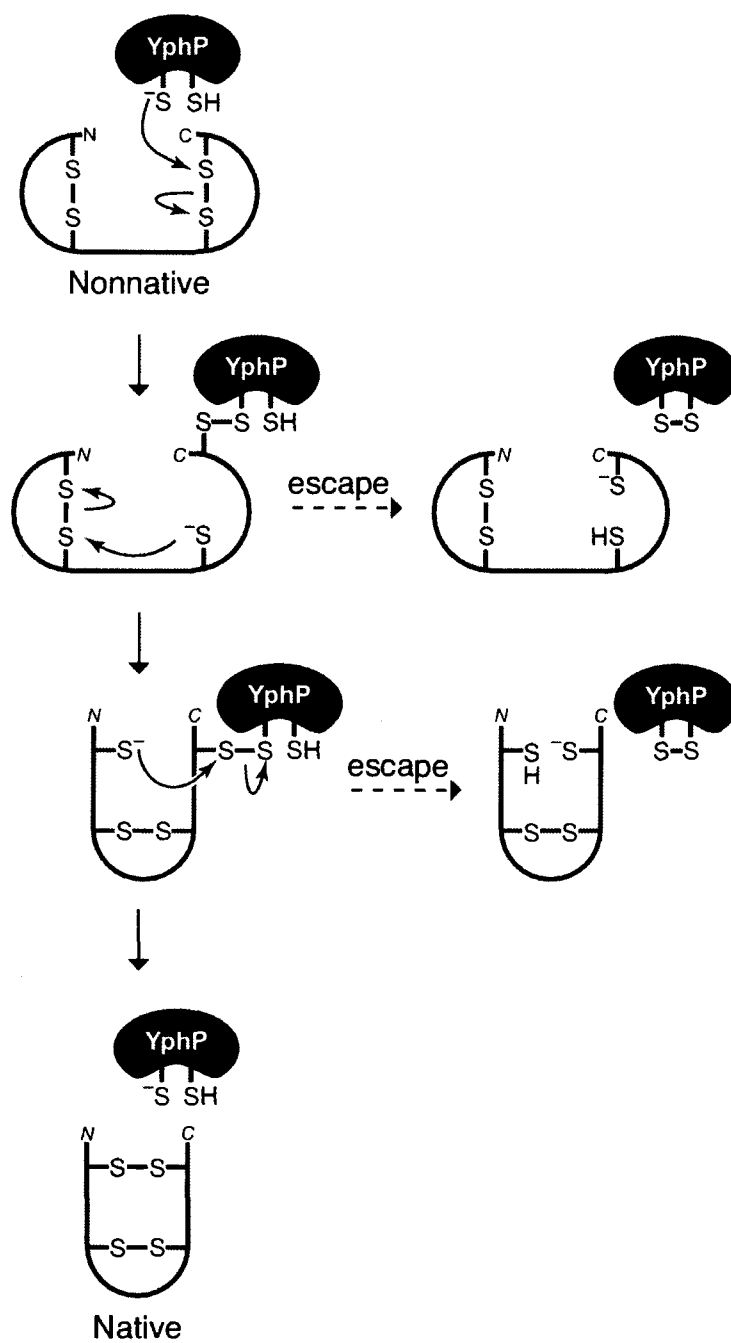


Figure A1.6 A diagram of the proposed reaction mechanism catalyzed by YphP and its homologues.

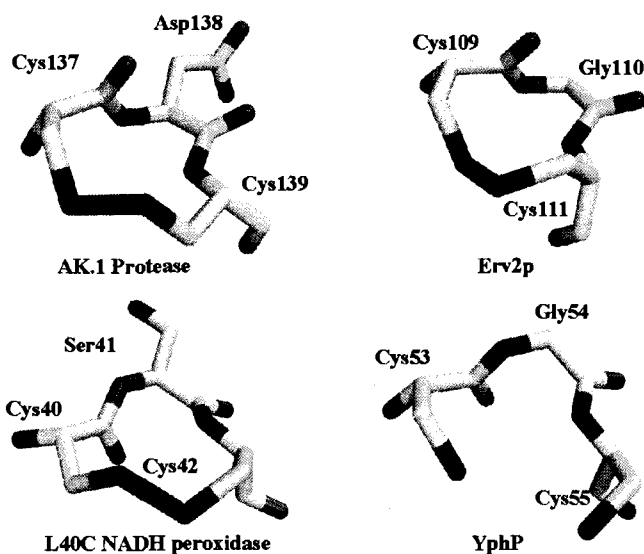


Figure A1.7 Structures of known oxidized CXC motifs and the reduced CXC motif of YphP; for details see text. The colors reflect atom type: yellow: carbon, blue: nitrogen, red: oxygen, green: sulfur.

Appendix Two

Towards Triple Helices in *Escherichia coli*

A2.1 Abstract

Collagen, the major structural protein of multicellular organisms, consists of three polypeptide strands that coil into a characteristic triple-helical structure. Each strand is comprised of a repeating, three amino-acid sequence, in which a glycine is located at every third residue, (Xaa–Yaa–Gly)_n. Most often, the Xaa and Yaa positions in the chain are (2*S*)-proline (Pro) and (2*S*,4*R*)-4-hydroxyproline (Hyp). The stability of the triple helix has been studied in vitro using collagen model peptides, (Pro–Pro–Gly)_n. Herein, we describe the design of a genetic screen for investigating collagen triple-helix formation in *E. coli*. The genetic screen makes use of the bacteriophage lambda repressor protein (cI), which controls the expression of a *lacZ* reporter gene. Fusion proteins of the N-terminal domain of the lambda repressor (NcI) and collagen-like peptides reconstitute a functional repressor when triple helices are formed. Like collagen, the foldon domain from the T4 fibrillin protein is a trimeric protein. Foldon trimer formation and variants that disrupt trimerization are also studied by using the genetic screen. This screen will have utility in the study of triple helix formation in vivo, including investigating amino acids sequences that alter triple-helix formation and the effects of proteins that interact with collagen.

A2.2 Introduction

Collagen is a major structural protein, comprising one quarter of all protein in the human body. The characteristic structure of collagen contains three polypeptide strands, each a left-handed polyproline type II helix, that coil together into a right-handed triple helix (Myllyharju and Kivirikko, 2004). Each strand is approximately 1000 residues in length and is composed of a repeating three amino acid sequence in which every third residue is glycine, (Xaa–Yaa–Gly). Most often, the Xaa and Yaa positions in the chain are (2*S*)-proline (Pro) and (2*S*,4*R*)-4-hydroxyproline (Hyp), respectively (Prockop and Kivirikko, 1995). Hyp is produced by a post-translational modification of proline catalyzed by the enzyme prolyl 4-hydroxylase (P4H). This hydroxylation is crucial for the folding and stability of the collagen triple helix under physiological conditions (Berg and Prockop, 1973c; Bulleid et al., 1996). Over 25 types of collagen have been identified, and each type has unique characteristics, including polymeric structure and physiological function and location. The most abundant types of collagen form fibrils in connective tissues. Type I collagen, the major protein in bone and skin, is a heterotrimer comprised of two $\alpha 1(I)$ and one $\alpha 2(I)$ polypeptide chains.

In addition to its structural role, collagen is also important for wound-healing and scar formation. In some cases, excessive amounts of scar tissue from injury or fibrotic disease can accumulate and cause damage to the afflicted tissue (O'Leary et al., 2002). Methods to control the production of collagen could be useful in alleviating these conditions. Conversely, production of stable, recombinant collagens is of interest for medical applications such as sutures, wound dressings and artificial heart valves (Olsen et al., 2003; Ramshaw et al., 1996).

Herein, we describe the design of a genetic screen for collagen triple-helix formation in *Escherichia coli*. The genetic screen is based on the bacteriophage lambda repressor protein. The

lambda repressor (cI) controls lambda phage life cycles by altering gene expression. The repressor is a dumbbell-shaped protein (26 kDa) with distinct N- and C-terminal domains. The N-terminal domain (NcI; residues 1-131) is the DNA-binding domain, and the C-terminal domain (CcI) functions as the dimerization domain. In its homodimeric form, cI binds to the operator regions of lambda DNA to repress transcription ((Ptashne, 2004), Figure A2.1a). Because the domains are discrete, replacement of the C-terminal domain with other naturally occurring dimerization domains produces a functional repressor (Hu et al., 1990). Protein–protein interactions, including leucine zipper dimerization, HIV protease dimerization, and ribonuclease A–ribonuclease inhibitor binding, have been studied using NcI-fusion proteins (Hu et al., 1990; Park and Raines, 2000; Park and Raines, 2002).

We make use of the lambda repressor screen to study the formation of the collagen triple helix. Collagen-mimic peptides, (Pro–Pro–Gly)_n, are fused to the C-terminus of NcI in place of the natural dimerization domain. Triple-helix formation of (Pro–Pro–Gly)_n forms functional repressors that bind the λO_{R1} of the lambda promoter (λP_R) and repress the expression of a *lacZ* reporter gene (Figure A2.1b).

Like collagen triple helices, the foldon domain of bacteriophage protein T4 fibrin also forms stable trimers. In fact, foldon has been used to template collagen (Frank et al., 2003). We investigate the oligomerization of foldon in *E. coli* using the lambda repressor screen. Fusion proteins of NcI and foldon (NcI-foldon) are found to repress expression of the *lacZ* reporter gene. Additionally, variants of foldon are screened for disruption of the trimer. The three-dimensional structure of foldon has been solved (Figure A2.2) (Guthe et al., 2004). Based on this structure, residues Glu5 and Arg15 are proposed to form an intersubunit salt bridge, and Trp20 is

thought to form the hydrophobic core of the trimer. Variants of foldon are made at these key residues.

A2.3 Materials and methods

Materials. Enzymes used for DNA manipulation were from Promega (Madison, WI). DNA sequences and oligonucleotides for mutagenesis and sequencing were obtained at the University of Wisconsin–Madison Biotechnology Center. Luria–Bertani (LB) medium contained tryptone (10 g), yeast extract (5 g), and NaCl (10 g).

Instrumentation. UV absorbance measurements were made with a Cary 50 spectrophotometer from Varian (Palo Alto, CA). MALDI–MS was performed with a Perkin-Elmer (Wellesley, MA) Voyager MALDI–TOF mass spectrometer at the University of Wisconsin–Madison Biophysics Instrumentation Facility.

Construction of pNcl-(PPG)_n. The pSH58 plasmid, derived from a pACYC184 vector, encodes the Ncl-RI fusion protein under the control of a *lacUV5* promoter and followed by a T₇ϕ transcription terminator (Park and Raines, 2002). This plasmid also contains the reporter cassette, the *lacZ* gene under the control of the lambda promoter (λP_R) and followed by an *rrnBT*₂ transcription terminator. To create a gene encoding Ncl-(PPG)_n fusion proteins, restriction sites *Xho*I and *Bgl*II were added by site-directed mutagenesis to the pSH58 plasmid at beginning and end of the DNA sequence encoding RI. A double-stranded fragment encoding (Pro–Pro–Gly)₇ was created by annealing two synthesized, single-stranded oligonucleotides, each with a 5' phosphate. Equimolar amounts of the oligonucleotides were mixed in STE buffer (10 mM Tris, pH 8.0, NaCl (50 mM), and EDTA (1 mM)), heated to 95 °C, and cooled gradually to permit annealing. The annealed oligonucleotides were designed to contain 5' and 3' overhangs

that are complementary to the *Xho*I and *Bgl*II restriction sites. This fragment was ligated into *Xho*I/*Bgl*II-digested pSH58. DNA encoding the FLAG epitope tag (DYKDDDDK) and a (GS)₃ linker was inserted by sequential site-directed mutagenesis using the QuikChange kit (Stratagene) to create pNcI-(PPG)₇ (Figure A2.3). Site-directed mutagenesis was also used to make insertions or deletions in the DNA encoding the collagen-like peptides to construct plasmids, pNcI-(PPG)_n, that direct the expression of NcI-FLAG-(GS)₃-(Pro-Pro-Gly)_n, where $n = 0, 3, 5, 7, 9, \text{ or } 11$.

Construction of pNcI-(XYG)₂₀. The pNcI-(XYG)₂₀ plasmid encodes NcI fused to a 27 amino-acid sequence from human collagen, (XYG)₉, where X and Y amino acids vary, plus (PPG)₁₁. The fragment from the gene encoding collagen was amplified from total RNA from human skin (Stratagene) using the OneStep reverse-transcription PCR kit (Qiagen, Valencia, CA). Due to the repetitive sequence encoding collagen, the primers used amplified a number of cDNAs, including the (XYG)₉-encoding sequence used here, even though that was not the targeted sequence. An *Apa*I restriction site (including a codon for glycine) was inserted into pNcI-(PPG)₁₁ between the coding regions for the (GS)₃ linker and the collagen-like peptide. The PCR fragment was digested with the *Apa*I restriction enzyme, and ligated into the *Apa*I-digested pNcI-(PPG)₁₁ to form a plasmid encoding NcI-(XYG)₉-(PPG)₁₁ (pNcI-(XYG)₂₀).

Production and purification of NcI-fusion proteins. The plasmids pNcI-(PPG)_n, where $n = 3, 5, 7 \text{ or } 9$, were transformed into *E. coli* Origami B (DE3) cells (Novagen, Madison, WI) by electroporation. Transformants were grown in LB medium containing chloramphenicol (40 µg/mL), and cDNA expression was induced by the addition of isopropyl-1-thio-β-D-galactopyranoside (IPTG, 30 µM). The bacterial cultures were pelleted, resuspended in TBS buffer (50 mM Tris, pH 7.4, NaCl (150 mM)), and lysed by sonication. The proteins were

purified by immunoprecipitation using anti-FLAG M2 agarose affinity gel (Sigma, St Louis, MO), following the manufacturer's protocol. Purified proteins were eluted from the resin with 0.1 M glycine, pH 3.5. Fusion proteins were analyzed by SDS-PAGE and by MALDI-MS.

Construction of pNcI-foldon. A double-stranded fragment encoding the 27 amino-acid foldon was created by annealing two synthesized, single-stranded oligonucleotides, as described above. This fragment was inserted into *HindIII/BglII*-digested pSH58. DNA encoding the FLAG tag was then inserted by site-directed mutagenesis to produce pNcI-foldon.

Foldon variants. Designed mutations were created by site-directed mutagenesis using the QuikChange kit (Stratagene). A library of mutations in the foldon gene at Glu5, Arg15, or Trp20 was made using primers containing a degenerate codon and the QuikChange Multi kit (Stratagene). The plasmid library was transformed into competent *E. coli* ElectroTen-Blue cells (Stratagene), and plated on LB agar containing chloramphenicol (40 µg/mL) and X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; 40 µg/mL). Colonies underwent blue/white screening to ensure the retention of the *lacZ* reporter construct. Colonies harboring the library of plasmids were grown in LB medium containing chloramphenicol (40 µg/mL) in deep-well 96-well plates. DNA sequencing reactions were performed directly on crude lysates (Smith and Raines, 2006).

Oligomerization assay. The ability of the NcI-fusion proteins to form trimers, and thus repress the expression of the *lacZ* reporter gene, was measured by quantification of cellular β-galactosidase (β-gal) activity (Miller, 1992). Plasmids encoding the NcI-fusion proteins and control plasmids, NcI (NcI-(PPG)₀) and cI (pSH28 (Park and Raines, 2000)), were transformed

into electrocompetent *E. coli* AG1688 cells (AG1688 = MC1061 F'128 *lacI^f lacZ::Tn5*). Three colonies of each transformation were picked and grown in LB media containing chloramphenicol (40 µg/mL) overnight. Fresh bacterial cultures were started with a 1:100 dilution of the overnight culture, and grown to OD₆₀₀ 0.4–0.9. Cultures were cooled to 4 °C to limit growth, and cell density was measured by the absorbance at 600 nm. Cells (25 µl) were added to 270 µL Z buffer (100 mM sodium phosphate, pH 7.0, KCl (10 mM), MgSO₄ (1 mM), and β-mercaptoethanol (50 mM)), 5 µl SDS (0.1 %), and 10 µl chloroform. The cells were mixed by vortexing to release the β-gal from the cells. The *o*-nitrophenyl β-D-galactopyranoside (ONPG) substrate (50 µl; 4 mg/mL) was added, and the reaction was allowed to proceed at 25 °C. When yellow color was visible, the reaction was quenched by the addition of 125 µl Na₂CO₃ (1 M). The amount of cleaved ONPG of was measured by absorbance at 420 nm, and β-gal activity was calculated by using eq 1:

$$\beta\text{-gal activity (Miller units)} = 1000 \times A_{420} / (A_{600} \times t \times v) \quad (1)$$

where t = reaction time (min), and v = volume of bacterial culture added (mL). The absorbance at 550 nm typically measured for cell debris was found to be negligible and was omitted.

A2.4 Results

Design of fusion proteins. The collagen-like (Pro–Pro–Gly)_n sequence or the foldon domain is connected to the C-terminus of NcI through an epitope tag and a flexible linker (Figure A2.2). The FLAG epitope tag was chosen because it also contains the enterokinase cleavage site, which could be used in to separate the (Pro–Pro–Gly)_n peptide from the NcI domain. The (Gly–Ser)₃

linker provides flexibility. The gene encoding NcI-(PPG)_n is under the control of the *lacUV5* promoter, which permits the expression of appropriate levels of the fusion protein in *E. coli* AG1688 cells for maximal assay sensitivity. Overexpression of either the fusion proteins or NcI alone represses *lacZ* expression (data not shown).

Fusion protein production and purification. Initially, we were concerned that expressing the repetitive collagen-like sequence with high proline content would be difficult in *E. coli*. The NcI-(PPG)_n fusion proteins, where $n = 3, 5, 7$, or 9 , were overexpressed in *E. coli* Origami B (DE3) cells and purified using an anti-FLAG affinity resin. Bacterial cultures expressing the cI protein were used as a negative control, as cI does not contain a FLAG tag, and was not purified by this protocol. Protein purification was analyzed by SDS-PAGE (Figure A2.4), and MALDI-MS (Table A2.1).

Screen for (Pro-Pro-Gly)_n trimerization. In vivo trimerization of the collagen-like sequence within the NcI-(PPG)₁₁ fusion proteins was examined using the *lacZ* reporter oligomerization assay. The fusion proteins were compared to NcI (NcI-(PPG)₀), which does not oligomerize, and to the wild-type lambda repressor, cI. Expression of NcI-(PPG)₁₁ in *E. coli* grown at 37 °C did not repress *lacZ* expression, suggesting that (PPG)₁₁ does not form trimers (Figure A2.5). Decreasing the temperature at which the bacteria were grown, however, showed greater repression of *lacZ* expression. The β-gal activity in bacterial cultures grown at 22, 17, and 12 °C decreased with lower temperatures, suggesting increased triple-helix formation at these temperatures.

Glycine in every third position is a characteristic of the collagen amino-acid sequence, and is critical for triple-helix formation. To test whether triple-helix formation of the (Pro-Pro-Gly)₁₁ sequence in NcI-(PPG)₁₁ was responsible for oligomerization and repression of *lacZ*, we

replaced the glycine in the sixth PPG triplet with an alanine residue (NcI-(PPG)₅(PPA)(PPG)₅). Replacing a glycine residue within the collagen-like sequence with an alanine residue should abolish triple-helix formation. Surprisingly, expression of *lacZ* in the presence of NcI-(PPG)₅(PPA)(PPG)₅ at 12 °C was repressed to a level similar to that in the presence of NcI-(PPG)₁₁ (data not shown). There is evidence that the amino acids surrounding the glycine-to-alanine substitution influence the detrimental effects of the substitution (Hyde et al., 2006). The content of proline residues within the sequence, which is quite high in NcI-(PPG)₅(PPA)(PPG)₅, and the propensity of the sequence to form a triple-helical structure can result in the formation of a triple helix distorted at the site of glycine substitution rather than destroying the triple-helical structure.

Screen for foldon trimerization. The ability of a NcI-foldon fusion protein to trimerize in vivo and to form a functional repressor was investigated using the oligomerization assay. NcI-foldon repressed *lacZ* expression similarly to wild-type lambda repressor, cI (Figure A2.6). Replacing Ser45 with leucine in the N-terminal domain of the lambda repressor disrupts DNA binding, even if the repressor is in the dimeric state (Nelson and Sauer, 1986). Accordingly, the S45L substitution in cI or NcI-foldon eliminated *lacZ* repression. In addition, the FLAG-affinity tag had no effect on repression of *lacZ* by NcI-foldon.

Based on the three-dimensional structure of foldon, residues Glu5 (amino acid numbering from the start of the foldon) and Arg15 are predicted to form a salt bridge, and Trp20 is thought to form a hydrophobic core. Little mutational analysis has been conducted on foldon, though the E5R foldon variant is known to prevent trimerization (Habazettl et al., 2009). To investigate the importance of these residues in foldon, we screened variants of foldon for trimerization using the oligomerization assay (Figure A2.7). The NcI-foldon variants were compared to NcI, which

does not repress of *lacZ* expression (high β -gal activity), and NcI-foldon wild-type, which repressed *lacZ* and resulted in β -gal activity similar to that in the presence of cI (Figure A2.6). The NcI-foldon variants were expressed in *E. coli* AG1688 cells grown at 37 °C, and the cultures were assayed for β -gal activity.

As confirmation of earlier work, the E5R foldon variant, as well as E5K, did not show evidence of trimerization, as measured by *lacZ* repression. The E5D foldon variant was the only substitution that retained wild-type repressor activity (Figure A2.6A). Only this most conservative change was permitted, confirming the significance of the negative charge at this position as part of the salt bridge. The positive charge at Arg15 was not permissible to change. None of the foldon variants tested formed functional repressors, including R15K, in which the positive charge is maintained (Figure A2.6B). The importance of the hydrophobic core was demonstrated, as none of the Trp20 foldon variants tested formed functional repressors, not even upon replacing tryptophan with tyrosine, another aromatic amino acid (Figure A2.6C). Substitutions at both residues involved in the salt bridge did not result in the formation of functional repressors. Although the E5K/R15K foldon variant contains two positive charges and was not expected to form trimers, the E5K/R15D foldon variant also did not form a functional repressor. (Figure A2.6D). The E5K/R15D variant still contains residues with positive and negative charges, but the location is exchanged and the identity of the residues is different. It is possible that the salt bridge could be recreated with other combinations of charged amino acid residues.

A2.5 Discussion

The native lambda repressor, cI, forms dimers in order to form a functional repressor. Genetic screens based on the lambda repressor replace the dimerization domain with a protein of interest to look for dimerization, or with proteins of interest to look for a protein–protein interaction, typically involving two proteins. Here, we are investigating proteins that form homotrimers. Although dimerization of either the collagen-like peptides or foldon would result in repression of *lacZ*, there is no evidence in the literature that collagen or foldon can form a dimer. Therefore, we call the assay described here an oligomerization assay, where NcI-fusion proteins that form trimers are functional repressors of *lacZ*, and result in decreased β -gal activity within cells.

Screening for oligomerization and the formation of a functional repressor can be done in a more high-throughput fashion by colorimetric plate screening or phage immunity assays. β -gal activity can be qualitatively detected by growing colonies on plates containing X-gal, a β -gal substrate that forms a blue precipitate upon cleavage. Functional repressors created by oligomerization can also perform the function of wild-type bacteriophage repressors in binding the λP_R promoter and repressing lytic gene expression. The phage λ KH54 (Hu et al., 1990) has a deletion in cI, so lytic genes are expressed, and plaques are formed in bacteria infected with λ KH54. Bacteria will grow, however, in the presence of an NcI-fusion protein that forms a functional repressor.

To investigate collagen triple-helix formation in *E. coli*, we used peptides that mimic collagen, (Pro-Pro-Gly)_n. The length, n, of the peptide was chosen based on the reported triple helix melting temperatures (T_m) of (Pro-Pro-Gly) polymers in vitro. The T_m value of a (Pro–

Pro-Gly)₇ triple helix is 6–7 °C (Hodges and Raines, 2005). A triple helix of this size is unlikely to form in bacteria grown near physiological temperatures. The (Pro-Pro-Gly)₁₀ triple helix, with a T_m value of 41 °C (Holmgren et al., 1998), should be more stable at growth conditions. The data presented here show that NcI-(PPG)₁₁ did not form a stable triple helix at 37 °C, which illustrates the differences in protein folding in vitro versus in vivo. Longer collagen-like sequences are needed in order for stable triple helices to form in bacteria at physiological temperature.

Alternatives to increasing peptide length for increased triple-helix stability include adding a trimer template, hydroxylating prolyl residues, or a collagen-binding protein to stabilize the triple helix. In this work, we studied the trimerization of foldon. Foldon has been used in vitro as a template for collagen (Frank et al., 2003). We look to identify foldon variants that form heterotrimers, and then utilize foldon as a scaffold for creating heterotrimeric collagens.

The hydroxylation of prolines preceding glycine residues in collagen stabilizes the triple-helical structure. Prolyl hydroxylation is catalyzed by the enzyme P4H, which is not found in bacteria. Recombinant P4H has been produced in *E. coli* (Kersteen et al., 2004; Neubauer et al., 2005), though P4H activity within bacteria has not been demonstrated. Coexpression of P4H along with the NcI-(PPG)_n proteins might increase triple-helix formation and repression of *lacZ*. The levels of P4H required for detectable activity are not known, and overexpression of P4H will require selective induction of P4H without altering levels of NcI-fusion proteins, so as not to decrease the sensitivity of the assay. If P4H is found to be active, hydroxylated collagen could be produced in bacteria.

The collagen triple helix might also be stabilized by collagen-binding proteins that hold the three strands together. For example, the binding of the HSP47 chaperone to collagen was

characterized using a yeast-two hybrid screen (Koide et al., 2000). Collagen-binding proteins are also found in bacteria and used to gain entry into a host. The adhesin protein, found on the surface of *Staphylococcus aureus*, is necessary and sufficient for mediating bacterial adherence to host collagen triple helices (Patti et al., 1992; Speziale et al., 1986). The collagen-binding domain (CBD) of adhesin forms a complex with multiple sites in bovine type II collagen with apparent K_d values of 3 μ M and 30 μ M (Patti et al., 1993). The genetic oligomerization screen could be useful in finding ways to disrupt the CBD–collagen interaction, which would interfere with a major virulence factor of *S. aureus* and prevent bacterial infection.

In conclusion, we have designed a genetic screen for collagen triple-helix formation in bacteria and have studied residues important for foldon trimerization. Amino acid substitutions within collagen-like peptides have been studied extensively in vitro. This genetic screen will permit a comparison of the in vitro data with in vivo triple helix formation. The screen might also be used to study proteins that interact with collagen, such as P4H and bacterial collagen-binding proteins. Randomly mutagenized P4H or adhesin could be analyzed to identify amino acid residues crucial for function. In parallel, the collagen-like sequence could be modified by making libraries of varying amino acids in the Xaa and Yaa positions in the (Xaa–Yaa–Gly)_n sequence to investigate the role of non-proline residues in the recognition of collagen by P4H or adhesin. Finally, small molecules could be screened for inhibition of P4H activity or the adhesin–collagen interaction, which would have many medical applications.

Table A2.1

Values of m/z determined by MALDI-MS of purified
NcI-(PPG)_n proteins.

NcI-(Pro-Pro-Gly) _n n	m/z	
	Calculated	Observed
3	16854	16857
5	17357	17364
7	17859	17852
9	18362	18377

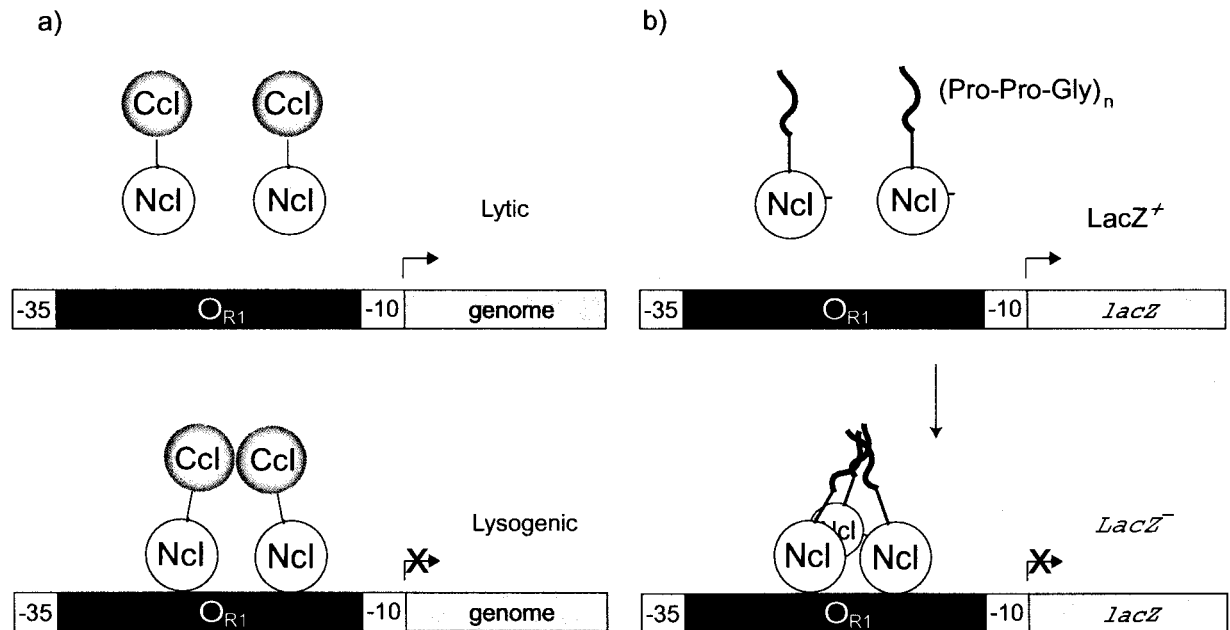


Figure A2.1 Schematic of the lambda repressor (cI) and Ncl-fusion protein control of gene expression. **a)** The wild-type (N- and C-terminal domains) lambda repressor (cI) dimer forms a functional repressor of gene expression. **b)** The N-terminus of the lambda repressor is fused to a repeating Pro-Pro-Gly sequence (PPG)_n. Upon triple helix formation a functional repressor is created, which controls expression of a *lacZ* reporter gene.

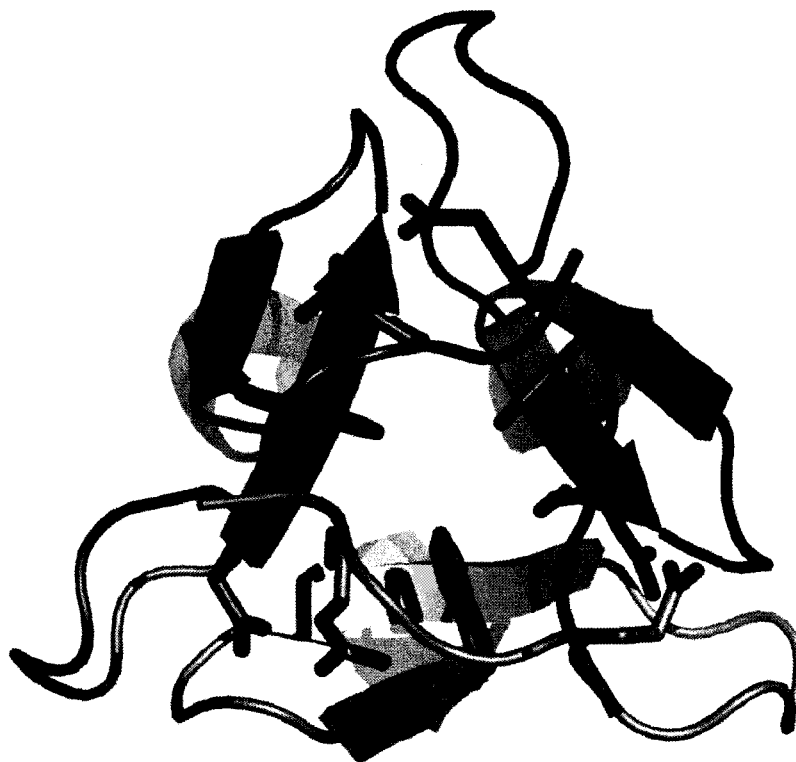


Figure A2.2 Structure of foldon. Each subunit of foldon is illustrated in a different color, green, orange, and purple (pdb 1RFO). Based on the three-dimensional structure, residues Glu5 and Arg15 form an intersubunit salt bridge. These residues are shown as sticks with oxygen atoms colored red and nitrogen atoms blue. Trp20 (brown) is thought to form the hydrophobic core of the trimer.

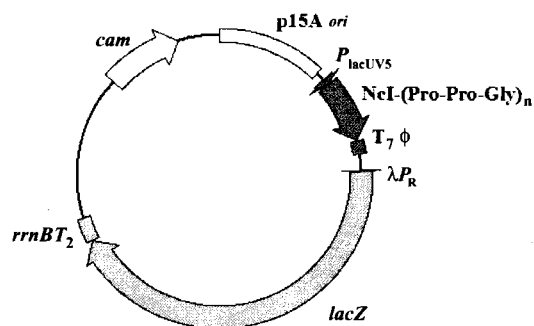


Figure A2.3 Map of plasmid pNcI-(PPG)_n. Plasmid pNcI-(PPG)_n contains the gene for a NcI-(Pro-Pro-Gly)_n fusion protein under the control of the *lacUV5* promoter (P_{lacUV5}) and followed by a $T_7\phi$ transcription terminator. The plasmid, with a p15A origin of replication (*p15A ori*), also includes the reporter cassette (λP_R -*lacZ*-*rrnBT_2*) and a chloramphenicol resistance gene (*cam*).

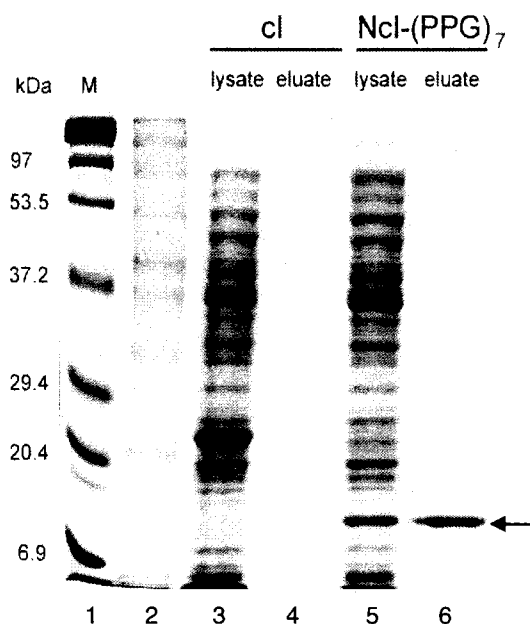


Figure A2.4 Purification of NcI-(PPG)₇. The purified protein is analyzed by SDS-PAGE (12 %). *Lanes 1 and 2* Molecular weight markers, *Lane 3* Cell lysate expressing cI, *Lane 4* Background elution from anti-FLAG resin, *Lane 5* Cell lysate expressing NcI-(PPG)₇, *Lane 6* NcI-(PPG)₇ eluted from anti-FLAG resin. Arrow indicates NcI-(PPG)₇ near the calculated weight of 17859 Da.

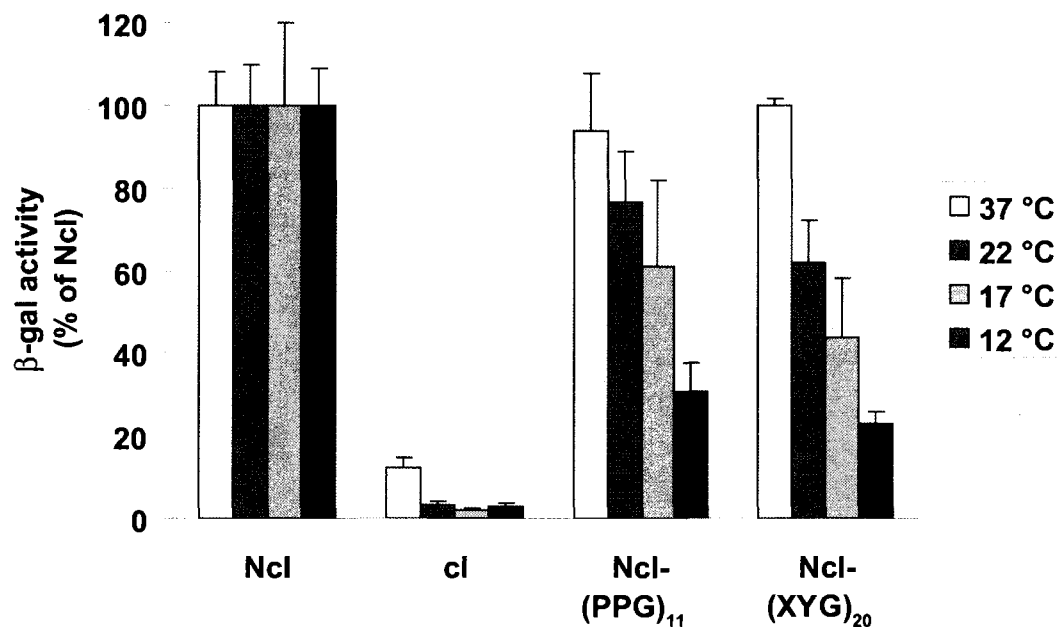


Figure A2.5 Dependence of triple-helix formation on temperature. The amount of β -gal repressed, compared to Ncl, by Ncl-(PPG)₁₁ and Ncl-(XYG)₂₀ is dependent on the temperature. Cultures were grown at 37, 22, 17, and 12 °C, and then assayed for β -gal activity. At lower temperatures, β -gal activity decreases, meaning repression of *lacZ* expression and increased triple helix formation.

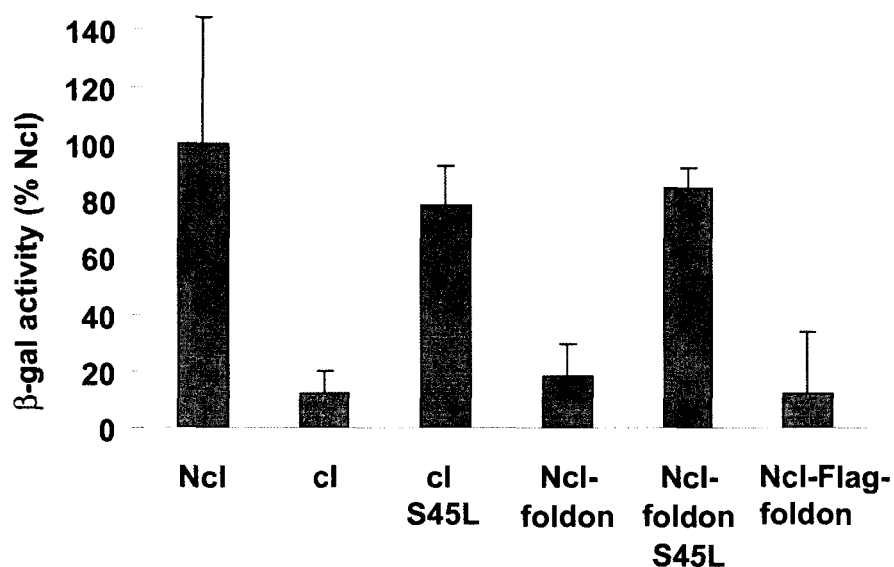


Figure A2.6 Ncl-foldon represses *lacZ* expression. The β -gal activity was measured in *E. coli* AG1688 cells grown at 37 °C that harbor plasmids containing the *lacZ* reporter cassette and a gene encoding Ncl, cI, or Ncl-foldon. The level of β -gal activity in the presence of Ncl-foldon was comparable to that in the presence of the wild-type lambda repressor, cI. The S46L variant of cI or Ncl-foldon, a mutation known to abolish DNA binding, eliminated repression. The FLAG-affinity tag between the Ncl and foldon domains did not affect repression.

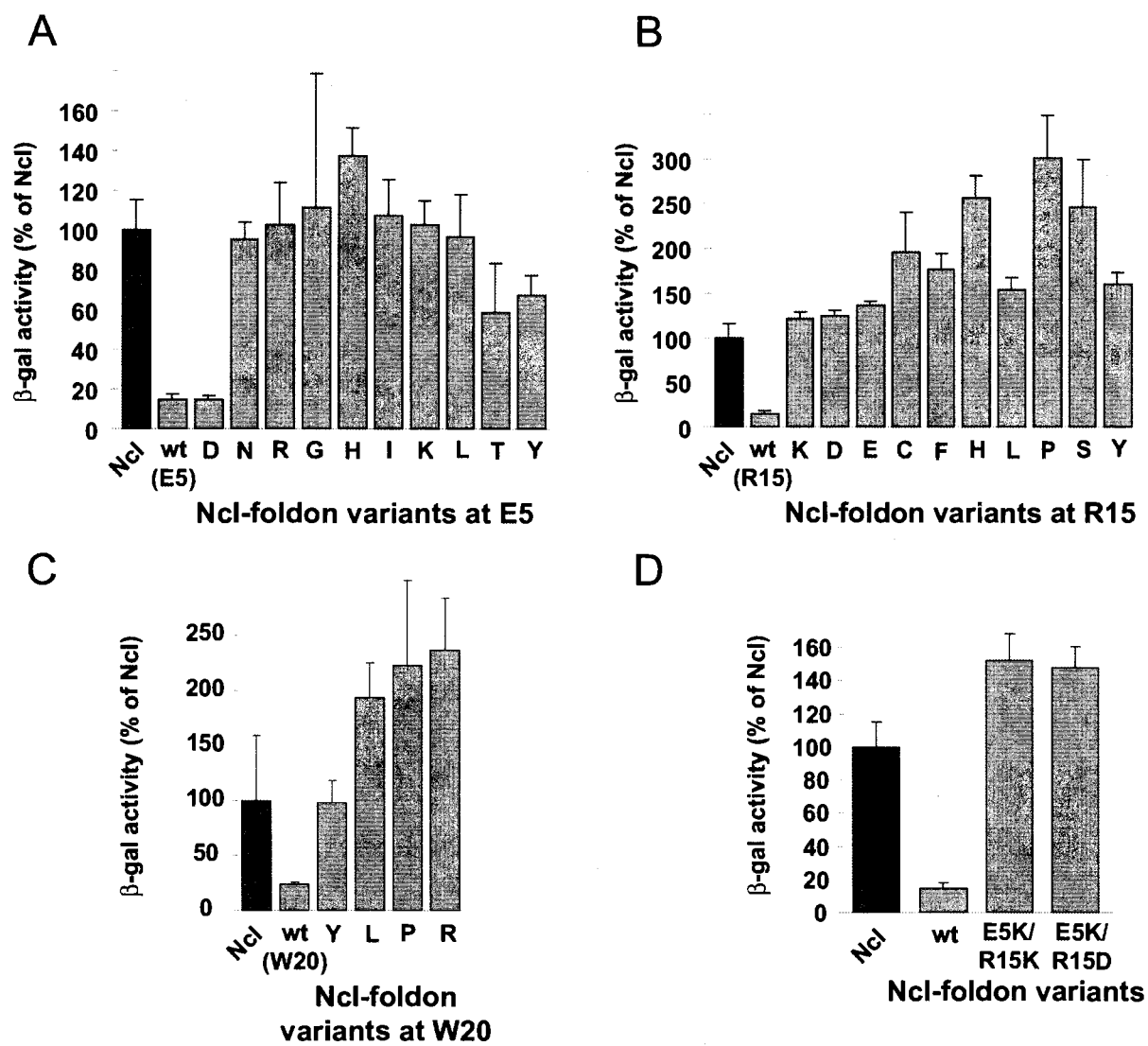


Figure A2.7 Oligomerization of NcI-foldon variants. Variants of NcI-foldon at Glu5, Arg15, or Trp20 (numbering from the start of the foldon).were expressed in *E. coli* AG1688 cells grown at 37 °C. The cultures were then assayed for β -gal activity. The NcI-foldon variants were compared to NcI, which does not repress of *lacZ* expression (100% β -gal activity), and wild-type NcI-foldon, which repressed *lacZ* and resulted in β -gal activity similar to that in the presence of cI (Figure A2.6).

A) The E5D foldon variant was the only variant that retained wild-type repressor activity. B) None of the Arg15 foldon variants tested formed functional repressors. C) None of the Trp20 foldon variants tested formed functional repressors. D) Neither the E5K/R15K nor the E5K/R15D foldon variant formed a functional repressor.

Appendix Three

Delta Program: Integrating Research, Teaching, and Learning

A3.1 Description of the Delta Program

The Delta Program (from www.delta.wisc.edu) is a research, teaching and learning community for faculty, staff, post-docs, and graduate students that will help current and future faculty succeed in the changing landscape of science, engineering, and math higher education. Through three core ideas: teaching-as-research, learning-through-diversity, and learning community, the Delta Program in Research, Teaching and Learning supports current and future science, technology, engineering and math (STEM) faculty in their ongoing improvement of student learning. The Delta Program is founded on three interrelated core ideas, as follows.

Teaching-as-Research: By applying research methods—idea, experiment, observation, analysis, improvement—to the challenge of teaching, Delta brings the skills of research faculty to the ongoing investigation of student learning, promotes innovation in teaching and measurement of student learning, and advances the role of instructors in the ongoing improvement of teaching practices.

Learning Community: Through collaborative activities and programs Delta will create a community of graduate students, post-docs, and faculty that will support and validate growth in teaching and learning and create a foundation for institutional change.

Learning-through-Diversity: Recognizing the common challenges in teaching and learning and the strength in bringing together diverse views, Delta is interdisciplinary, serving all science, engineering, and mathematics departments; cross-generational, bringing together graduate students, post-docs, and both new and experienced faculty; comprehensive, providing knowledge, practice, and community; responsive, reflecting the broad range of responsibilities that face today's faculty; and inclusive, welcoming for a multifaceted and diverse group of people.

A3.2 “All (content) is not lost”

Do students learn technical content in a biomedical engineering course taught through social, political and ethical issues?

Delta Internship Summative Report by Kelly Gorres and Prof. Kristyn Masters

A3.2.1 Abstract

Coursework in science and engineering often lacks study of the social, political, and ethical aspects of the topic. Most educators agree that these are important topics for students, both future scientists and future citizens, to think about. One main concern about adding these topics to any course is the fear that course content will have to be sacrificed. The main questions addressed by the research presented herein are “Does presenting a course in biomedical engineering in a social/political/ethical context increase student learning of the broader issues?” and “Do students also learn course content?” This work compares two courses in Biomedical Engineering (BME). One course is traditionally taught, while the other is taught from a science, technology and society (STS) approach. Comparisons were made of students before and after the STS course, and between students in each of the two courses. This work provides evidence that students in a context-based course do learn the technical content.

A3.2.2 Introduction

There has been a great movement in science education to improve “scientific literacy”. AAAS Scientific literacy can be defined as the ability to understand science-related news presented by the media or the ability to form an opinion on a new scientific breakthrough. However defined, this call for science education for the general public most often puts more emphasis on how science relates to society rather than scientific facts. This emphasis is also being brought into the classroom. Science curricula are including interactions among science, technology and society (STS). Course materials are presented in the context of real world issues (Middlecamp and Baldwin, 2006). While many science and engineering courses often contain “just the facts”, more teachers are introducing scientific material through some common topic. By putting technical concepts into a relevant context, students can make personal connections. Contextualized learning often takes place in a topic-based context using real world examples and case studies (Bennett et al., 2007).

The effects of context-based courses have been reviewed by Bennett, et al. Students in courses taught by the context-based approach have more positive attitudes toward science classes than conventionally taught courses. The learning of scientific concepts is overall similar when comparing traditional courses with context-based courses, although there is some conflicting data on this topic possibly due to different assessment techniques. Students in context-based courses tended to perform better on evaluations that were also context-based.

The project described herein investigates a biomedical engineering course that incorporates topics not addressed in typical science courses: ethics, science communication, and politics. Scientific research is dependent on funding that is in part determined by politics and public perception of science. Scholars have reasoned that including political arguments in science

curricula by discussing real world issues may be the best tool for revolutionizing science education and improving science literacy (Hodson, 1999; Hodson, 2003). Hodson asserts,

“By grounding content in socially and personally relevant contexts, an issues-based approach can provide the motivation that is absent from current abstract, de-contextualized approaches and can form a base for students to construct understanding that is personally relevant, meaningful, and important.” (Hodson, 2003)

A3.2.3 Project design

The strategy for the project was to compare two courses with some overlapping technical content taught by the same instructor. One class, Biomedical Engineering 510 “Introduction to Tissue Engineering”, was taught in a traditional manner with lectures, literature analysis and laboratory activities. The emphasis in this course was technical content. The other class, Biomedical Engineering 601 “Current Issues in Biomedical Engineering”, was approached from a social/political/ethical perspective. Topics were introduced through these issues and students learned how scientific research is influenced by politics, ethics, and public perception. To demonstrate to the students the diverse community of professionals that are involved in scientific issues, these topics were often introduced by guest speakers that have included Representatives from the WI State Legislature, a U.S. Congresswoman, prominent bioethicists and professional science journalists. The students read articles addressing topics in biomedical engineering found in mass media and technical journals and then participated in discussions and debates, wherein they learned the technical material in order to understand the issue and support their statements. The students were also asked to form opinions and deduce public perceptions in a lesson based on political cartoons. The ultimate goal was to provide students with a solid background in biomedical engineering issues and communication skills that will enable them to be active in the

global community.

A pre-course survey gathered student background information and demographics. Students were also asked questions on three topics taught in both courses: tissue engineering, gene therapy, and stem cells. After each unit in the BME 601 course, students were given post-unit quizzes which corresponded to questions in the pre-course survey. Each student response was categorized as either a technical (content) or non-technical (political/ethical/social issue), and scored by the number of responses given in each category.

- What are some of the main challenges facing implementation of tissue engineering strategies?
- What are the biggest obstacles and limitations currently facing gene therapy?
- Please list as many types of gene delivery vectors as you can.
- What are the different types of stem cells that you know of, and how are they different from each other? Please explain briefly.

Assessment of student attitudes toward science and technology was accomplished by using the VOSTS (Views on Science-Technology-Society) survey (Aikenhead and Ryan, 1992). The survey was created based on written responses by Canadian high school students, and converted into a multiple-choice format. The number of questions was greatly pared down to decrease survey fatigue and reduce redundancy.

Demographic information was also collected to allow for comparisons among different groups of students, such as comparing gender, race, and level of experience (undergraduate versus graduate-level education).

A3.2.4 Results and Discussion

Demographics. Of the 29 students in the BME 510 course, 28 responded to the survey, 16

were undergraduates and 12 were graduate students. Twenty students responded as biomedical engineering majors, while seven students were of related majors (mechanical engineering, pharmacology and toxicology, materials science, chemical and biological engineering, and biological system engineering).

The BME 601 course was comprised of 9 undergraduate students and 2 graduate students, 36% male and 64% female, and all biomedical engineering majors. Four of the students marked being partially a minority race.

Content. Student learning of biomedical engineering content (technical information) and issues related to science–technology–society (non-technical) was measured by the number and types of responses given by the students to questions given as part of a quiz.

When asked “What are some of the main challenges facing implementation of tissue engineering strategies?” in a pre-test of students in BME 601, 71% gave a response categorized as technical, and 29% gave non-technical (ethical or legal issues or research funding) responses (Table A3.1). No post-course data was collected. To the same question in a pre-test, all 29 students in BME 510 gave a technical response. In addition to a technical response, 11/29 (38%) also gave a non-technical responses. No post-course data was collected. Before either course, a larger percentage of students could provide an answer that contained technical information, but only about 1/3 of students in either course thought of non-technical issues.

The breadth of student learning in BME 601 was investigated by looking at both technical and non-technical responses to the question, “What are the biggest obstacles and limitations currently facing gene therapy?”. As shown in Figure A3.1, in the pre-test 8/14 students gave technical responses and 5/14 gave non-technical responses including cost, ethical issues, government regulation, and public perception. In the post-test 10/14 students gave technical

responses and 11/14 gave non-technical responses, most notably including an increase in the number of students who gave an answer categorized as public perception from four to nine.

These results demonstrate an increase in student awareness of societal issues.

To further address the amount of content students learn in a context-based course, students in BME 601 were given two technical questions both pre- and post-course. These results are found in Table A3.2. In the first question, “Please list as many types of gene delivery vectors as you can.” Students gave an average of 2.0 ± 2.0 answers/student with five students giving no response in the pre-test, whereas in the post-test all students gave answers, and there was an average of 3.8 ± 1.8 answers/student. Similar results were observed when students were asked questions about stem cells. In pre-test responses to “What are the different types of stem cells that you know of, and how are they different from each other? Please explain briefly.”, students gave an average of 2.9 ± 1.8 answers/student. In post-test responses to “Please describe three different types/classes of stem cells, focusing on how they are different from each other (i.e. their maturity/potency, their source).”, students gave an average of 5.0 ± 1.6 answers/student. These results provide evidence for student learning of technical content in a course taught by a science–technology–society approach.

Attitudes. A portion of the VOSTS survey was given to the BME 601 students at the beginning of the course. The students’ responses to summarized survey statements are given below. Overall, most respondents choose options that included some sort of interaction between scientists and the non-scientific community.

- When asked their position on the statement **A country’s politics affect that country’s scientists**, 10/11 replied that scientists ARE affected by their country’s politics, but the reasons given varied among because of funding (1), policy and money (1), policy (5), and scientists are part of society (5).

- Responses to the statement **We always have to make trade-offs (compromises) between the positive and negative effects of science and technology**, included seven saying there are always tradeoffs, and two saying not always.
- When asked if **More money should be spent on science and technology in the US even though this money will not be available for other things, such as social programs, education, business incentives and lower taxes**, 3 said more money, 5 said spending should be balanced, and 0 said less.
- To the statement “**Community or government agencies should tell scientists what to investigate; otherwise scientists will investigate what is of interest only to them**”, 3 students replied government should decide for important public problems otherwise scientists, 3 choose scientists with government advice, and 4 thought solely scientists should decide.
- All 11 students agreed with the statement: **Science and technology offer a great deal of help in resolving such social problems as pollution and overpopulation**, but the answers were divided into Science and technology can certainly help to resolve these problems. (2); Science and technology can help resolve some social problems but not others. (2); Science and technology solve many social problems, but science and technology also cause many of these problems. (3); and It’s not a question of science and technology helping, but rather it’s a question of people using science and technology wisely. (4).

A3.2.5 Lessons learned

In this project, the assessments using open-ended questions that asked for the limitations or challenges of a technology worked well because they allowed the students to respond without restrictions on the scope of answer and without alluding to any particular problem. The responses were easily classified as technical or non-technical, and comparisons could be made. Enrollment was less than anticipated, which made gender, race, and educational level comparison difficult. Also, some students were in both classes, so the data for the second course may be skewed because of these overlapping students.

Based on the results of the VOSTS survey and the pre-course survey, many of the students enrolled in BME 601 were interested and aware of many political, ethical and societal issues,

which is presumably why they enrolled in the course. To investigate the utility of context-based courses, it may be better to compare courses populated by less advanced students.

A3.2.6 Conclusions

Biomedical Engineering 601 “Current Issues in Biomedical Engineering” is a course that was designed to provide a setting in which students can discuss social, political, and ethical issues in biomedical engineering, and through those discussions learn the scientific information related to those topics. The main questions focused upon by the research presented herein are “Does presenting a course in biomedical engineering in a social/political/ethical context increase student learning of the broader issues?” and “Do students also learn course content?”. The students enrolled in the course were generally aware of the relationship between scientists and society prior to enrolling in the course, as evidenced in the Views on Science-Technology-Society survey. Prior to the course, student knowledge of technical content on tissue engineering, stem cells, and gene therapy varied, but a comparison of pre- and post-tests showed an overall increase in the number of technical responses given by students, suggesting an increase of technical knowledge. This study provides support for context-based or Science-Technology-Society approached courses, and that technical information typically emphasized in science courses is still learned by students.

Table A3.1

Number of students that gave technical and non-technical responses/total number of students enrolled to the pre-course question, "What are some of the main challenges facing implementation of tissue engineering strategies?"

<u>Course</u>	<u>Technical answers</u>	<u>Non-technical answers</u>
Social/Ethical Issues (BME 601)	10/14 (71%)	4/14 (29%)
Technical course (BME 510)	29/29 (100%)	11/29 (38%)

Table A3.2

The technical knowledge, measured by the number of responses given to questions asked about gene delivery or stem cells, increased over the BME 601 course.

	Average number of responses per student	
	<u>Gene delivery</u>	<u>Stem cells</u>
Pre-test	2.0 ± 2.0	2.9 ± 1.8
Post-test	3.8 ± 1.8	5.0 ± 1.6

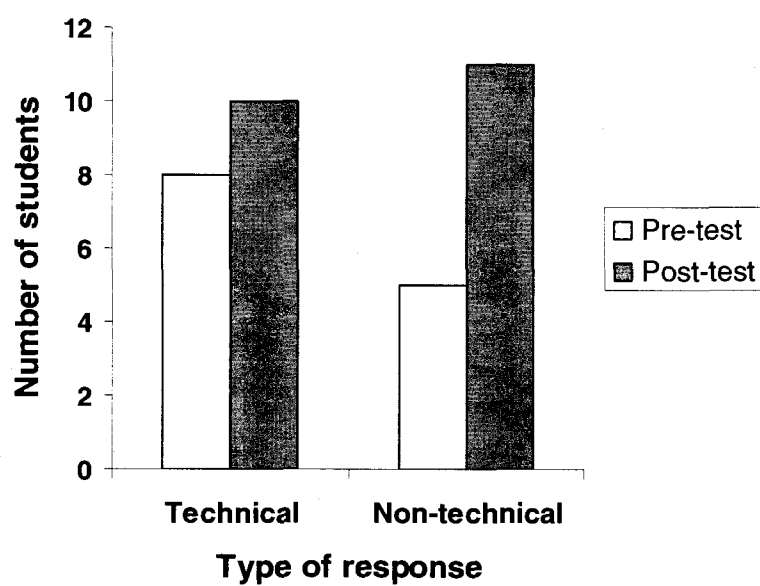


Figure A3.1 Non-technical responses from BME 601 students increased from pre-course to post-course when asked, “What are the biggest obstacles and limitations currently facing gene therapy?”.

References

- Adefarati, A. A., Giacobbe, R. A., Hensens, O. D., and Tkacz, J. S. (1991). Biosynthesis of L-671,329, and echinocandin-type antibiotic produced by *Zalerion arboricola*: Origins of some of the unusual amino acids and the dimethylmyristic acid side chain. *J. Am. Chem. Soc.* 113:3542-3545.
- Aikenhead, G. S., and Ryan, A. G. (1992). The development of a new instrument: Views on science-technology-society (VOSTS). *Sci. Educ.* 76:477-491.
- Aldunate, R., Casar, J. C., Brandan, E., and Inestrosa, N. C. (2004). Structural and functional organization of synaptic acetylcholinesterase. *Brain Res. Rev.* 47:96-104.
- Andersen, J. F., Sanders, D. A., Gasdaska, J. R., Weichsel, A., Powis, G., and Montfort, W. R. (1997). Human thioredoxin homodimers: Regulation by pH, role of aspartate 60, and crystal structure of the aspartate 60 → asparagine mutant. *Biochemistry* 36:13979-13988.
- Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T., and Kivirikko, K. I. (1997). Cloning of the human prolyl 4-hydroxylase α subunit isoform α (II) and characterization of the type II enzyme tetramer. *J. Biol. Chem.* 272:17342-17348.
- Aslund, F., Berndt, K. D., and Holmgren, A. (1997). Redox potentials of glutaredoxins and other thiol-disulfide oxidoreductases of the thioredoxin superfamily determined by direct protein-protein redox equilibria. *J. Biol. Chem.* 272:30780-30786.
- Atreya, P. L., and Ananthanarayanan, V. S. (1991). Interaction of prolyl 4-hydroxylase with synthetic peptide substrates: A conformational model for collagen proline hydroxylation. *J. Biol. Chem.* 266:2852-2858.
- Baader, E., Tschank, G., Baringhaus, K. H., Burghard, H., and Gunzler, V. (1994). Inhibition of prolyl 4-hydroxylase by oxalyl amino acid derivatives in vitro, in isolated microsomes and in embryonic chicken tissues. *Biochem. J.* 300 (Pt 2):525-530.
- Bächinger, H. P. (1987). The influence of peptidyl-prolyl *cis-trans* isomerase on the in vitro folding of type III collagen. *J. Biol. Chem.* 262:17144-17148.
- Baldwin, J. E., Field, R. A., Lawrence, C. C., Lee, V., Robinson, J. K., and Schofield, C. J. (1994). Substrate specificity of proline 4-hydroxylase: Chemical and enzymatic synthesis of 2*S*,3*R*,4*S*-epoxyproline. *Tetrahedron Lett.* 35:4649-4652.
- Baldwin, J. E., Field, R. A., Lawrence, C. C., Merritt, K. D., and Schofield, C. J. (1993). Proline 4-hydroxylase: Stereochemical course of the reaction *Tetrahedron Lett.* 34:7489-7492.

- Bella, J., Eaton, M., Brodsky, B., and Berman, H. M. (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* 266:75-81.
- Bennett, J., Lubben, F., and Hogarth, S. A. (2007). A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Sci. Educ.* 91:347-370.
- Berg, R. A., and Prockop, D. J. (1973a). Affinity column purification of procollagen proline hydroxylase from chick embryos and further characterization of the enzyme. *J. Biol. Chem.* 218:1175-1182.
- Berg, R. A., and Prockop, D. J. (1973b). Purification of [^{14}C]procollagen and its hydroxylation by prolyl-hydroxylase. *Biochemistry* 12:3395-3401.
- Berg, R. A., and Prockop, D. J. (1973c). The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple helix of collagen. *Biochem. Biophys. Res. Comm.* 52:115-120.
- Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2001). Crystal structure of a collagen-like polypeptide with repeating sequence Pro-Hyp-Gly at 1.4 Å resolution: Implications for collagen hydration. *Biopolymers* 56:8-13.
- Berisio, R., Vitagliano, L., Mazzarella, L., and Zagari, A. (2002). Crystal structure of the collagen triple helix model [(Pro-Pro-Gly) $_{10}$] $_3$. *Protein Sci.* 11:262-270.
- Bhatnagar, R. S., Rapaka, R. S., and Urry, D. W. (1978). Interaction of polypeptide models of elastin with prolyl hydroxylase. *FEBS Lett.* 95:61-64.
- Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T., and Drennan, C. L. (2006). Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* 440:368-371.
- Bornstein, P. (1967a). Comparative sequence studies of rat skin and tendon collagen. I. Evidence for incomplete hydroxylation of individual prolyl residues in the normal proteins. *Biochemistry* 6:3082-3093.
- Bornstein, P. (1967b). The incomplete hydroxylation of individual prolyl residues in collagen. *J Biol Chem* 242:2572-2574.
- Brahmachari, S. K., and Ananthanarayanan, V. S. (1979). β -turns in nascent procollagen are sites of posttranslational enzymatic hydroxylation of proline. *Proc. Natl. Acad. Sci. USA* 76:5119-5123.
- Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L., and Raines, R. T. (2001). Conformational stability of collagen relies on a stereoelectronic effect. *J. Am. Chem. Soc.* 123:777-778.

- Bruick, R. K., and McKnight, S. L. (2001). A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* 294:1337-1340.
- Buczek, O., Bulaj, G., and Olivera, B. M. (2005). Conotoxins and the posttranslational modification of secreted gene products. *Cell. Mol. Life Sci.* 62:3067-3079.
- Bulleid, N. J., Wilson, R., and Lees, J. F. (1996). Type-III procollagen assembly in semi-intact cells: Chain association, nucleation and triple-helix folding do not require formation of inter-chain disulphide bonds but triple-helix nucleation does require hydroxylation. *Biochem. J.* 317:195-202.
- Cabral, W. A., Chang, W., Barnes, A. M., Weis, M., Scott, M. A., Leikin, S., Makareeva, E., Kuznetsova, N. V., Rosenbaum, K. N., Tifft, C. J., Bulas, D. I., Kozma, C., Smith, P. A., Eyre, D. R., and Marini, J. C. (2007). Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta. *Nat. Genet.* 39:359-365.
- Cardinale, G. J., Rhoads, R. E., and Udenfriend, S. (1971). Simultaneous incorporation of ¹⁸O into succinate and hydroxyproline catalyzed by collagen proline hydroxylase. *Biochem. Biophys. Res. Commun.* 43:537-543.
- Cardinale, G. J., and Udenfriend, S. (1974). Prolyl hydroxylase. *Adv. Enzymol. Relat. Areas Mol. Biol.* 41:245-300.
- Chen, I., and Ting, A. Y. (2005). Site-specific labeling of proteins with small molecules in live cells. *Curr. Opin. Biotechnol.* 16:35-40.
- Chivers, P. T., Laboissiere, M. C., and Raines, R. T. (1996). The CXXC motif: Imperatives for the formation of native disulfide bonds in the cell. *EMBO J.* 15:2659-2667.
- Chopra, R. K., and Ananthanarayanan, V. S. (1982). Conformational implications of enzymatic proline hydroxylation in collagen. *Proc. Natl. Acad. Sci. USA* 79:7180-7184.
- Chorghade, M. S., Mohapatra, D. K., Sahoo, G., Gurjar, M. K., Mandlecha, M. V., Bhoite, N., Moghe, S., and Raines, R. T. (2008). Practical syntheses of 4-fluoroprolines. *J. Fluor. Chem.* 129:781-784.
- Chowdhury, R., Hardy, A., and Schofield, C. J. (2008). The human oxygen sensing machinery and its manipulation. *Chem. Soc. Rev.* 37:1308-1319.
- Chowdhury, R., McDonough, M. A., Mecinovic, J., Loenarz, C., Flashman, E., Hewitson, K. S., Domene, C., and Schofield, C. J. (2009). Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure* 17:981-989.
- Clifton, I. J., Hsueh, L.-C., Baldwin, J. E., Harlos, K., and Schofield, C. J. (2001). Structure of proline 3-hydroxylase. *Eur. J. Biochem.* 268:6625-6636.

Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., and Schofield, C. J. (2006). Structural studies on 2-oxoglutarate oxygenases and related double-stranded β -helix fold proteins. *J. Inorg. Biochem.* 100:644-669.

Cook, P., and Cleland, W. W., eds. (2007). (New York, NY).

Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., and Derewenda, Z. (2007). Protein crystallization by surface entropy reduction: Optimization of the SER strategy. *Acta Crystallogr. D: Biol. Crystallogr.* 63:636-645.

Costas, M., Mehn, M. P., Jensen, M. P., and Que, L., Jr. (2004). Dioxygen activation at mononuclear nonheme iron active sites: Enzymes, models, and intermediates. *Chem. Rev.* 104:939-986.

Counts, D. F., Cardinale, G. J., and Udenfriend, S. (1978). Prolyl Hydroxylase Half Reaction - Peptidyl Prolyl-Independent Decarboxylation of Alpha-Ketoglutarate. *Proceedings of the National Academy of Sciences of the United States of America* 75:2145-2149.

Cummins, E. P., Berra, E., Comerford, K. M., Ginouves, A., Fitzgerald, K. T., Seeballuck, F., Godson, C., Nielsen, J. E., Moynagh, P., Pouyssegur, J., and Taylor, C. T. (2006). Prolyl hydroxylase-1 negatively regulates I κ B kinase-beta, giving insight into hypoxia-induced NF κ B activity. *Proc. Natl. Acad. Sci. U.S.A.* 103:18154-18159.

Cunliffe, C. J., Franklin, T. J., and Gaskell, R. M. (1986). Assay of prolyl 4-hydroxylase by the chromatographic determination of [14 C]succinic acid on ion-exchange minicolumns. *Biochem. J.* 240:617-619.

Dann, C. E., 3rd, Bruick, R. K., and Deisenhofer, J. (2002). Structure of factor-inhibiting hypoxia-inducible factor 1: An asparaginyl hydroxylase involved in the hypoxic response pathway. *Proc Natl Acad Sci U S A* 99:15351-15356.

de Jong, L., Albracht, S. P., and Kemp, A. (1982). Prolyl 4-hydroxylase activity in relation to the oxidation state of enzyme-bound iron. The role of ascorbate in peptidyl proline hydroxylation. *Biochim. Biophys. Acta.* 704:326-332.

de Jong, L., and Kemp, A. (1984). Stoichiometry and Kinetics of the Prolyl 4-Hydroxylase Partial Reaction. *Biochim. Biophys. Acta* 787:105-111.

de Jong, L., van der Kraan, I., and de Waal, A. (1991). The kinetics of the hydroxylation of procollagen by prolyl 4-hydroxylase. Proposal for a processive mechanism of binding of the dimeric hydroxylating enzyme in relation to the high kcat/Km ratio and a conformational requirement for hydroxylation of -X-Pro-Gly- sequences. *Biochim Biophys Acta* 1079:103-111.

de Waal, A., and de Jong, L. (1988). Processive action of the two peptide binding sites of prolyl 4-hydroxylase in the hydroxylation of procollagen. *Biochemistry* 27:150-155.

- Deprez, P., Inestrosa, N. C., and Krejci, E. (2003). Two different heparin-binding domains in the triple-helical domain of ColQ, the collagen tail subunit of synaptic acetylcholinesterase. *J. Biol. Chem.* 278:23233-23242.
- Derewenda, Z. S. (2004). Rational protein crystallization by mutational surface engineering. *Structure* 12:529-535.
- Derewenda, Z. S., and Vekilov, P. G. (2006). Entropy and surface engineering in protein crystallization. *Acta Crystallogr. D: Biol. Crystallogr.* 62:116-124.
- DeRider, M. L., Wilkens, S. J., Waddell, M. J., Bretscher, L. E., Weinhold, F., Raines, R. T., and Markley, J. L. (2002). Collagen stability: Insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations. *J. Am. Chem. Soc.* 124:2497-2505.
- Dieckman, L., Gu, M., Stols, L., Donnelly, M. I., and Collart, F. R. (2002). High throughput methods for gene cloning and expression. *Protein Expr. Purif.* 25:1-7.
- Diegelmann, R. F., Ondregjickova, O., and Katz, E. (1969). Oxygen-18 and fluoroproline studies on the synthesis of hydroxyproline and oxoproline in actinomycin. *Arch. Biochem. Biophys.* 131:276-287.
- Dunn, D. M., and Franzblau, C. (1982). Effects of ascorbate on insoluble elastin accumulation and cross-link formation in rabbit pulmonary artery smooth muscle cultures. *Biochemistry* 21:4195-4202.
- Ehrismann, D., Flashman, E., Genn, D. N., Mathioudakis, N., Hewitson, K. S., Ratcliffe, P. J., and Schofield, C. J. (2007). Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *Biochem. J.* 401:227-234.
- Elkins, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. J. (2003). Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* 278:1802-1806.
- Elkins, J. M., Ryle, M. J., Clifton, I. J., Dunning Hotopp, J. C., Lloyd, J. S., Burzlaff, N. I., Baldwin, J. E., Hausinger, R. P., and Roach, P. L. (2002). X-ray crystal structure of *Escherichia coli* taurine/ α -ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry* 41:5185-5192.
- Emsley, P., and Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D: Biol. Crystallogr.* 60:2126-2132.

Engel, J., Chen, H.-T., Prockop, D. J., and Klump, H. (1977). The triple helix-coil conversion of collagen-like polytripeptides in aqueous and nonaqueous solvents. Comparison of the thermodynamic parameters and the binding of water to (L-Pro-L-Pro-Gly)_n and (L-Pro-L-Hyp-Gly)_n. *Biopolymers* 16:601-622.

Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K. I. (1999). Evidence for 4-hydroxyproline in viral proteins. Characterization of a viral prolyl 4-hydroxylase and its peptide substrates. *J Biol Chem* 274:22131-22134.

Ferguson, A. D., Labunskyy, V. M., Fomenko, D. E., Arac, D., Chelliah, Y., Amezcua, C. A., Rizo, J., Gladyshev, V. N., and Deisenhofer, J. (2006). NMR structures of the selenoproteins Sep15 and SelM reveal redox activity of a new thioredoxin-like family. *J. Biol. Chem.* 281:3536-3543.

Finn, R. D., Tate, J., Mistry, J., Coghill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2008). The Pfam protein families database. *Nucleic Acids Res.* 36:D281-288.

Fischer, G. (2000). Chemical aspects of peptide bond isomerisation. *Chem. Soc. Rev.* 29:119-127.

Fitzpatrick, P. F., Ralph, E. C., Ellis, H. R., Willmon, O. J., and Daubner, S. C. (2003). Characterization of metal ligand mutants of tyrosine hydroxylase: Insights into the plasticity of a 2-histidine-1-carboxylate triad. *Biochemistry* 42:2081-2088.

Fox, B. G. (1998). Catalysis by non-heme iron. In *Comprehensive Biological Catalysis*, M. Sinnott, ed. (New York, Academic Press), pp. 261-348.

Frank, S., Boudko, S., Mizuno, K., Schulthess, T., Engel, J., and Bachinger, H. P. (2003). Collagen triple helix formation can be nucleated at either end. *J. Biol. Chem.* 278:7747-7750.

Franklin, T. J. (1997). Therapeutic approaches to organ fibrosis. *Int. J. Biochem. Cell Biol.* 29:79-89.

Franzke, C. W., Bruckner, P., and Bruckner-Tuderman, L. (2005). Collagenous transmembrane proteins: Recent insights into biology and pathology. *J. Biol. Chem.* 280:4005-4008.

Friedman, L., Higgin, J. J., Moulder, G., Barstead, R., Raines, R. T., and Kimble, J. (2000). Prolyl 4-hydroxylase is required for viability and morphogenesis in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 97:4736-4741.

Fujita, Y., Gottlieb, A., Peterkofsky, B., Udenfriend, S., and Witkop, B. (1964). The preparation of *cis*- and *trans*-4-³H-L-prolines and their use in studying the mechanism of enzymatic hydroxylation in chick embryos. *J. Am. Chem. Soc.* 86:4709-4716.

- Galonic, D. P., Barr, E. W., Walsh, C. T., Bollinger, J., J.M., and Krebs, C. (2007). Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nat. Chem. Biol.* 3:113-116.
- Giacovazzo, C., Monaco, H. L., Artioli, G., Viterbo, D., Ferraris, G., Gilli, G., Zanotti, G., and Catti, M. (2002). Fundamentals of Crystallography, 2nd edn. (Oxford, UK, Oxford University Press).
- Gill, A. C., Ritchie, M. A., Hunt, L. G., Steane, S. E., Davies, K. G., Bocking, S. P., Rhie, A. G., Bennett, A. D., and Hope, J. (2000). Post-translational hydroxylation at the N-terminus of the prion protein reveals presence of PPII structure in vivo. *EMBO J.* 19:5324-5331.
- Gill, S. C., and von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182:319-326.
- Ginalski, K., Elofsson, A., Fischer, D., and Rychlewski, L. (2003). 3D-Jury: A simple approach to improve protein structure predictions. *Bioinformatics* 19:1015-1018.
- Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., and Eisenberg, D. (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci.* 16:1569-1576.
- Gorres, K. L., Edupuganti, R., Krow, G. R., and Raines, R. T. (2008). Conformational preferences of substrates for human prolyl 4-hydroxylase. *Biochemistry* 47:9447-9455.
- Gottlieb, A. A., Fujita, Y., Udenfriend, S., and Witkop, B. (1965). Incorporation of cis- and trans-4-fluoro-L-prolines into proteins and hydroxylation of trans isomer during collagen biosynthesis. *Biochemistry* 4:2507-2513.
- Grauschopf, U., Winther, J. R., Korber, P., Zander, T., Dallinger, P., and Bardwell, J. C. A. (1995). Why Is DsbA Such an Oxidizing Disulfide Catalyst? *Cell* 83:947-955.
- Gross, E., Sevier, C. S., Vala, A., Kaiser, C. A., and Fass, D. (2002). A new FAD-binding fold and intersubunit disulfide shuttle in the thiol oxidase Erv2p. *Nat. Struct. Biol.* 9:61-67.
- Groves, J. T., and McClusky, G. A. (1976). Aliphatic hydroxylation via oxygen rebound. Oxygen transfer catalyzed by iron. *J. Am. Chem. Soc.* 98:859-861.
- Gryder, R. M., Lamon, M., and Adams, E. (1975). Sequence position of 3-hydroxyproline in basement membrane collagen. *J. Biol. Chem.* 250:2470-2474.
- Grzyska, P. K., Muller, T. A., Campbell, M. G., and Hausinger, R. P. (2007). Metal ligand substitution and evidence for quinone formation in taurine/alpha-ketoglutarate dioxygenase. *J. Inorg. Biochem.* 101:797-808.

- Gunnlaugsson, T., Glynn, M., Tocci, G. M., Kruger, P. E., and Pfeffer, F. M. (2006). Anion recognition and sensing in organic and aqueous media using luminescent and colorimetric sensors. *Coord. Chem. Rev.* 250:3094-3117.
- Gunzler, V., Brocks, D., Henke, S., Myllylä, R., Geiger, R., and Kivirikko, K. I. (1988). Syncatalytic inactivation of prolyl 4-hydroxylase by synthetic peptides containing the unphysiologic amino acid 5-oxaproline. *J. Biol. Chem.* 263:19498-19504.
- Guthe, S., Kapinos, L., Moglich, A., Meier, S., Grzesiek, S., and Kiefhaber, T. (2004). Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *J. Mol. Biol.* 337:905-915.
- Guzman, N. A., ed. (1998). Prolyl Hydroxylase, Protein Disulfide Isomerase, and Other Structurally Related Proteins (New York, Marcel Dekker).
- Habazettl, J., Reiner, A., and Kiefhaber, T. (2009). NMR structure of a monomeric intermediate on the evolutionarily optimized assembly pathway of a small trimerization domain. *J. Mol. Biol.* 389:103-114.
- Hanson, E. S., Foot, L. M., and Leibold, E. A. (1999). Hypoxia post-translationally activates iron-regulatory protein 2. *J. Biol. Chem.* 274:5047-5052.
- Hanson, E. S., Rawlins, M. L., and Leibold, E. A. (2003). Oxygen and iron regulation of iron regulatory protein 2. *J. Biol. Chem.* 278:40337-40342.
- Hara, R., and Kino, K. (2009). Characterization of novel 2-oxoglutarate dependent dioxygenases converting L-proline to *cis*-4-hydroxy-L-proline. *Biochem. Biophys. Res. Commun.* 379:882-886.
- Helaakoski, T., Annunen, P., Vuori, K., Macneil, I. A., Pihlajaniemi, T., and Kivirikko, K. I. (1995). Cloning, baculovirus expression, and characterization of a second mouse prolyl 4-hydroxylase α -subunit isoform: Formation of an $\alpha_2\beta_2$ tetramer with the protein disulfide-isomerase/ β subunit. *Proc. Natl. Acad. Sci. USA* 92:4427-4431.
- Helaakoski, T., Vuori, K., Myllylä, R., Kivirikko, K. I., and Pihlajaniemi, T. (1989). Molecular cloning of the α -subunit of human prolyl 4-hydroxylase: The complete cDNA-derived amino acid sequence and evidence for alternative splicing of RNA transcripts. *Proc. Natl. Acad. Sci. USA* 86:4392-4396.
- Hieta, R., and Myllyharju, J. (2002). Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from *Arabidopsis thaliana*. Effective hydroxylation of proline-rich, collagen-like, and hypoxia-inducible transcription factor α -like peptides. *J. Biol. Chem.* 277:23965-23971.
- Hirsila, M., Koivunen, P., Gunzler, V., Kivirikko, K. I., and Myllyharju, J. (2003). Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J. Biol. Chem.* 278:30772-30780.

- Hodges, J. A., and Raines, R. T. (2005). Stereoelectronic and steric effects in the collagen triple helix: Toward a code for strand association. *J. Am. Chem. Soc.* 127:15923-15932.
- Hodson, D. (1999). Going beyond cultural pluralism: Science education for sociopolitical action. *Sci. Educ.* 83:775-796.
- Hodson, D. (2003). Time for action: Science education for an alternative future. *Int. J. Sci. Educ.* 25:645-670.
- Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger, J. M., Jr., and Krebs, C. (2006). Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci. USA* 103:14738-14743.
- Holm, L., Kaariainen, S., Wilton, C., and Plewczynski, D. (2006). Using Dali for structural comparison of proteins. *Curr. Protoc. Bioinformatics* Chapter 5:Unit 5 5.
- Holmgren, S. K., Bretscher, L. E., Taylor, K. M., and Raines, R. T. (1999). A hyperstable collagen mimic. *Chem. Biol.* 6:63-70.
- Holmgren, S. K., Taylor, K. M., Bretscher, L. E., and Raines, R. T. (1998). Code for collagen's stability deciphered. *Nature* 392:666-667.
- Holster, T., Pakkanen, O., Soininen, R., Sormunen, R., Nokelainen, M., Kivirikko, K. I., and Myllyharju, J. (2007). Loss of assembly of the main basement membrane collagen, Type IV, but not fibril-forming collagens and embryonic death in collagen prolyl 4-hydroxylase I null mice. *J. Biol. Chem.* 282:2512-2519.
- Hon, W. C., Wilson, M. I., Harlos, K., Claridge, T. D., Schofield, C. J., Pugh, C. W., Maxwell, P. H., Ratcliffe, P. J., Stuart, D. I., and Jones, E. Y. (2002). Structural basis for the recognition of hydroxyproline in HIF-1 alpha by pVHL. *Nature* 417:975-978.
- Hu, J. C., O'Shea, E. K., Kim, P. S., and Sauer, R. T. (1990). Sequence requirements for coiled-coils: Analysis with lambda repressor-GCN4 leucine zipper fusions. *Science* 250:1400-1403.
- Huang, F., and Du, W. (2009). Solution structure of Hyp10Pro variant of conomorphin, a cysteine-free and D-amino-acid containing conopeptide. *Toxicon*.
- Hutton, J. J., Jr., Trappel, A. L., and Udenfriend, S. (1966a). A rapid assay for collagen proline hydroxylase. *Anal. Biochem.* 16:384-394.
- Hutton, J. J., Jr., and Udenfriend, S. (1966). Soluble collagen proline hydroxylase and its substrates in several animal tissues. *Proc Natl Acad Sci U S A* 56:198-202.

- Hutton, J. J., Jr., Witkop, B., Kurtz, J., Berger, A., and Udenfriend, S. (1968). Synthetic polypeptides as substrates and inhibitors of collagen proline hydroxylase. *Arch. Biochem. Biophys.* 125:779-785.
- Hutton, J. J., Kaplan, A., and Udenfriend, S. (1967). Conversion of the amino acid sequence Gly-Pro-Pro in protein to Gly-Pro-Hyp by collagen proline hydroxylase. *Arch. Biochemistry Biophys.* 121:384-391.
- Hutton, J. J., Tappel, A. L., and Udenfriend, S. (1966b). Requirements for α -ketoglutarate, ferrous iron and ascorbate by collagen proline hydroxylase. *Biochemistry Biophys. Res. Commun.* 24:179-184.
- Hyde, T. J., Bryan, M. A., Brodsky, B., and Baum, J. (2006). Sequence dependence of renucleation after a Gly mutation in model collagen peptides. *J. Biol. Chem.* 281:36937-36943.
- Improta, R., Benzi, C., and Barone, V. (2001). Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* 123:12568-12577.
- Inouye, K., Sakakibara, S., and Prockop, D. J. (1976). Effects of the stereo-configuration of the hydroxyl group in 4-hydroxyproline on the triple-helical structures formed by homogenous peptides resembling collagen. *Biochim. Biophys. Acta* 420:133-141.
- Ishikawa, Y., Wirz, J., Vranka, J. A., Nagata, K., and Bachinger, H. P. (2009). Biochemical characterization of the prolyl 3-hydroxylase 1/CRTAP/cyclophilin B complex. *J. Biol. Chem.*
- Jaakkola, P., Mole, D. R., Tian, Y. M., Wilson, M. I., Gielbert, J., Gaskell, S. J., Kriegsheim, A., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001). Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O_2 -regulated prolyl hydroxylation. *Science* 292:468-472.
- Jakob, U., Muse, W., Eser, M., and Bardwell, J. C. (1999). Chaperone activity with a redox switch. *Cell* 96:341-352.
- Janda, I., Devedjiev, Y., Derewenda, U., Dauter, Z., Bielnicki, J., Cooper, D. R., Graf, P. C., Joachimiak, A., Jakob, U., and Derewenda, Z. S. (2004). The crystal structure of the reduced, Zn^{2+} -bound form of the *B. subtilis* Hsp33 chaperone and its implications for the activation mechanism. *Structure* 12:1901-1907.
- Jenkins, C. L., Bretscher, L. E., Guzei, I. A., and Raines, R. T. (2003). Effect of 3-hydroxyproline residues on collagen stability. *J. Am. Chem. Soc.* 125:6422-6427.
- Jenkins, C. L., Lin, G., Duo, J., Rapolu, D., Guzei, I. A., Raines, R. T., and Krow, G. R. (2004). Substituted 2-azabicyclo[2.1.1]hexanes as constrained proline analogues: Implications for collagen stability. *J. Org. Chem.* 69:8565-8573.

- Jenkins, C. L., and Raines, R. T. (2002). Insights on the conformational stability of collagen. *Nat. Prod. Rep.* 19:49-59.
- Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47 (Pt 2):110-119.
- Kaelin, W. G. (2005). Proline hydroxylation and gene expression. *Annu. Rev. Biochem.* 74:115-128.
- Kaelin, W. G., Jr., and Ratcliffe, P. J. (2008). Oxygen sensing by metazoans: The central role of the HIF hydroxylase pathway. *Mol. Cell* 30:393-402.
- Kanai, K., Podanyi, B., Bokotey, S., Hajdu, F., and Hermecz, I. (2002). Stereoselective sulfoxide formation from a thioproline derivative. *Tetrahedron Asymmetry* 13:491-495.
- Kaska, D. D., Gunzler, V., Kivirikko, K. I., and Myllyla, R. (1987). Characterization of a low-relative-molecular-mass prolyl 4-hydroxylase from the green alga *Chlamydomonas reinhardtii*. *Biochem. J.* 241:483-490.
- Katz, E., Kamal, F., and Mason, K. (1979). Biosynthesis of trans-4-hydroxy-L-proline by *Streptomyces griseoviridis*. *J. Biol. Chem.* 254:6684-6690.
- Katz, E., Prockop, D. J., and Udenfriend, S. (1962). Precursors of the hydroxyproline and ketoproline in actinomycin. *J. Biol. Chem.* 237:1585-1588.
- Kaule, G., and Gunzler, V. (1990). Assay for 2-oxoglutarate decarboxylating enzymes based on the determination of [1-¹⁴C]succinate: Application to prolyl 4-hydroxylase. *Anal. Biochem.* 184:291-297.
- Ke, Q., and Costa, M. (2006). Hypoxia-inducible factor-1 (HIF-1). *Mol. Pharmacol.* 70:1469-1480.
- Kern, D., Schutkowski, M., and Drakenberg, T. (1997). Rotational barriers of cis/trans isomerization of proline analogues and their catalysis by cyclophilin. *J. Am. Chem. Soc.* 119:8403-8408.
- Kersteen, E. A., Barrows, S. R., and Raines, R. T. (2005). Catalysis of protein disulfide bond isomerization in a homogeneous substrate. *Biochemistry* 44:12168-12178.
- Kersteen, E. A., Higgin, J. J., and Raines, R. T. (2004). Production of human prolyl 4-hydroxylase in *Escherichia coli*. *Protein Exp. Purif.* 38:279-291.

- Kersteen, E. A., and Raines, R. T. (2003). Catalysis of protein folding by protein disulfide isomerase and small-molecule mimics. *Antioxid. Redox Signal.* 5:413-424.
- Kerwar, S. S., Marcel, R. J., and Salvador, R. A. (1976). Studies on the effect of L-3,4-dehydroproline on collagen synthesis by chick embryo polysomes. *Arch. Biochem. Biophys.* 172:685-688.
- Keskiaho, K., Hieta, R., Sormunen, R., and Myllyharju, J. (2007). Chlamydomonas reinhardtii has multiple prolyl 4-hydroxylases, one of which is essential for proper cell wall assembly. *Plant Cell* 19:256-269.
- Kieliszewski, M. J. (2001). The latest hype on Hyp-O-glycosylation codes. *Phytochemistry* 57:391-323.
- Kieliszewski, M. J., and Lampion, D. T. A. (1994). Extensin: Repetitive motifs, functional sites, post-translational codes, and phylogeny. *Plant J.* 5:157-172.
- Kieliszewski, M. J., and Shpak, E. (2001). Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins. *Cell. Mol. Life Sci.* 58:1386-1398.
- Kim, W., and Conticello, V. P. (2007). Protein engineering methods for investigation of structure-function relationships in protein-based elastomeric materials. *Polym. Rev.* 47:93-119.
- Kim, W., George, A., Evans, M., and Conticello, V. P. (2004). Cotranslational incorporation of a structurally diverse series of proline analogues in an *Escherichia coli* expression system. *ChemBioChem* 5:928-936.
- Kim, W., McMillan, R. A., Snyder, J. P., and Conticello, V. P. (2005). A stereoelectronic effect on turn formation due to proline substitution in elastin-mimetic polypeptides. *J. Am. Chem. Soc.* 127:18121-18132.
- Kishore, U., Greenhough, T. J., Waters, P., Shrive, A. K., Ghai, R., Kamran, M. F., Bernal, A. L., Reid, K. B., Madan, T., and Chakraborty, T. (2006). Surfactant proteins SP-A and SP-D: Structure, function and receptors. *Mol. Immunol.* 43:1293-1315.
- Kivirikko, K. I., Kishida, Y., Sakakibara, S., and Prockop, D. J. (1972). Hydroxylation of (X-Pro-Gly)_n by procollagen proline hydroxylase. Effect of chain length, helical conformation and amino acid sequence in the substrate. *Biochim. Biophys. Acta* 271:347-356.
- Kivirikko, K. I., and Myllyla, R. (1982). Posttranslational enzymes in the biosynthesis of collagen: Intracellular enzymes. *Methods Enzymol.* 82 Pt A:245-304.

- Kivirikko, K. I., Myllyla, R., and Pihlajaniemi, T. (1989). Protein hydroxylation: Prolyl 4-hydroxylase, an enzyme with four cosubstrates and a multifunctional subunit. *FASEB J.* 3:1609-1617.
- Kivirikko, K. I., Myllyla, R., and Pihlajaniemi, T. (1992). Hydroxylation of proline and lysine residues in collagens and other animal and plant proteins. In *Post-translational modifications of proteins*, J. J. Harding, and C. M. J. C., eds. (Boca Raton, FL, CRC Press).
- Kivirikko, K. I., and Pihlajaniemi, T. (1998). Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv. Enzymol. Relat. Areas Mol. Biol.* 72:325-398.
- Kivirikko, K. I., Suga, K., Kishida, Y., Sakakibara, S., and Prockop, D. J. (1971). Asymmetry in the hydroxylation of (Pro-Pro-Gly) 5 by procollagen proline hydroxylase. *Biochem Biophys Res Commun* 45:1591-1596.
- Koditz, J., Nesper, J., Wottawa, M., Stiehl, D. P., Camenisch, G., Franke, C., Myllyharju, J., Wenger, R. H., and Katschinski, D. M. (2007). Oxygen-dependent ATF-4 stability is mediated by the PHD3 oxygen sensor. *Blood* 110:3610-3617.
- Koehn, F. E., Longley, R. E., and Reed, J. K. (1992). Microcolins A and B, new immunosuppressive peptides from the blue-green alga *Lyngbya majuscula*. *J. Nat. Prod.* 55:613-619.
- Koehntop, K. D., Emerson, J. P., and Que, L., Jr. (2005). The 2-His-1-carboxylate facial triad: A versatile platform for dioxygen activation by mononuclear non-heme iron(II) enzymes. *J. Biol. Inorg. Chem.* 10:87-93.
- Koide, T., Aso, A., Yorihuzi, T., and Nagata, K. (2000). Conformational requirements of collagenous peptides for recognition by the chaperone protein HSP47. *J. Biol. Chem.* 275:27957-27963.
- Koivu, J., and Myllyla, R. (1986). Protein Disulfide-Isomerase retains procollagen prolyl 4-hydroxylase structure in its native conformation. *Biochemistry* 25:5982-5986.
- Koivu, J., Myllyla, R., Helaakoski, T., Pihlajaniemi, T., Tasanen, K., and Kivirikko, K. I. (1987). A single polypeptide acts both as the beta subunit of prolyl 4-hydroxylase and as a protein disulfide-isomerase. *J. Biol. Chem.* 262:6447-6449.
- Koivunen, P., Hirsila, M., Gunzler, V., Kivirikko, K. I., and Myllyharju, J. (2004). Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *J. Biol. Chem.* 279:9899-9904.
- Koivunen, P., Tiainen, P., Hyvarinen, J., Williams, K. E., Sormunen, R., Klaus, S. J., Kivirikko, K. I., and Myllyharju, J. (2007). An endoplasmic reticulum transmembrane prolyl 4-hydroxylase

is induced by hypoxia and acts on hypoxia-inducible factor alpha. *J. Biol. Chem.* 282:30544-30552.

Koski, M. K., Hieta, R., Bollner, C., Kivirikko, K. I., Myllyharju, J., and Wierenga, R. K. (2007). The active site of an algal prolyl 4-hydroxylase has a large structural plasticity. *J. Biol. Chem.* 282:37112-37123.

Koski, M. K., Hieta, R., Hirsila, M., Ronka, A., Myllyharju, J., and Wierenga, R. K. (2009). The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J. Biol. Chem.* doi:10.1074/jbc.M109.014050.

Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B., and Berman, H. M. (1999). Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol.* 6:454-457.

Kramer, R. Z., Vitagliano, L., Bella, J., Berisio, R., Mazzarella, L., Brodsky, B., Zagari, A., and Berman, H. M. (1998). X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *J. Mol. Biol.* 280:623-638.

Krebs, C., Galonic Fujimori, D., Walsh, C. T., and Bollinger, J. M., Jr. (2007). Non-heme Fe(IV)-oxo intermediates. *Acc. Chem. Res.* 40:484-492.

Krishnaswamy, S., and Rossmann, M. G. (1990). Structural refinement and analysis of Mengo virus. *J. Mol. Biol.* 211:803-844.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372:774-797.

Krow, G. R., and Cannon, K. C. (2004). Azabicyclo[2.1.1]hexanes. A review. *Heterocycles* 62:877-898.

Krow, G. R., Herzon, S. B., Lin, G., Qiu, F., and Sonnet, P. E. (2002). Temperature-dependent regiochemical diversity in lithiation-electrophilic substitution reactions on *N*-BOC-2-azabicyclo[2.1.1]hexane. 2,4- and 3,5-methanoproline. *Org. Lett.* 4:3151-3154.

Krow, G. R., Huang, Q., Lin, G., Centafont, R. A., Thomas, A. M., Gandla, D., Debrosse, C., and Carroll, P. J. (2006). 5-Carboxy-2-azabicyclo[2.1.1]hexanes as precursors of 5-halo, amino, phenyl, and 2-methoxycarbonyl ethyl methanopyrrolidines. *J. Org. Chem.* 71:2090-2098.

Krow, G. R., Lin, G., and Yu, F. (2005). The rearrangement route to 3-carboxy- and 3-hydroxymethyl-2-azabicyclo[2.1.1]hexanes: 3,5-methanoproline. *J. Org. Chem.* 70:590-595.

Kukkola, L., Hieta, R., Kivirikko, K. I., and Myllyharju, J. (2003). Identification and characterization of a third human, rat, and mouse collagen prolyl 4-hydroxylase isoenzyme. *J. Biol. Chem.* 278:47685-47693.

Kukkola, L., Koivunen, P., Pakkanen, O., Page, A. P., and Myllyharju, J. (2004). Collagen prolyl 4-hydroxylase tetramers and dimers show identical decreases in K_m values for peptide substrates with increasing chain length: mutation of one of the two catalytic sites in the tetramer inactivates the enzyme by more than half. *J Biol Chem* 279:18656-18661.

Kuznetsova, A. V., Meller, J., Schnell, P. O., Nash, J. A., Ignacak, M. L., Sanchez, Y., Conaway, J. W., Conaway, R. C., and Czyzyk-Krzeska, M. F. (2003). von Hippel-Lindau protein binds hyperphosphorylated large subunit of RNA polymerase II through a proline hydroxylation motif and targets it for ubiquitination. *Proc. Natl. Acad. Sci. U.S.A.* 100:2706-2711.

Laboissière, M. C. A., Sturley, S. L., and Raines, R. T. (1995). The essential role of protein-disulfide isomerase is to unscramble non-native disulfide bonds. *J. Biol. Chem.* 270:28006-28009.

Lamberg, A., Helaakoski, T., Myllyharju, J., Peltonen, S., Notbohm, H., Pihlajaniemi, T., and Kivirikko, K. I. (1996). Characterization of human type III collagen expressed in a baculovirus system. *J. Biol. Chem.* 271:11988-11995.

Lamberg, A., Pihlajaniemi, T., and Kivirikko, K. I. (1995). Site-directed mutagenesis of the α subunit of human prolyl 4-hydroxylase. Identification of three histidine residues critical for catalytic activity. *J. Biol. Chem.* 270:9926-9931.

Lancaster, D. E., McDonough, M. A., and Schofield, C. J. (2004). Factor inhibiting hypoxia-inducible factor (FIH) and other asparaginyl hydroxylases. *Biochem Soc Trans* 32:943-945.

Lassot, I., Segéral, E., Berlioz-Torrent, C., Durand, H., Groussin, L., Hai, T., Benarous, R., and Margottin-Goguet, F. (2001). ATF4 degradation relies on a phosphorylation-dependent interaction with the SCF(β TrCP) ubiquitin ligase. *Mol. Cell. Biol.* 21:2192-2202.

Lawrence, C. C., Sobey, W. J., Field, R. A., Baldwin, J. E., and Schofield, C. J. (1996). Purification and initial characterization of proline 4-hydroxylase from *Streptomyces griseoviridis* P8648: A 2-oxoacid, ferrous-dependent dioxygenase involved in etamycin biosynthesis. *Biochem. J.* 313:185-192.

Lee, C., Kim, S. J., Jeong, D. G., Lee, S. M., and Ryu, S. E. (2003). Structure of human FIH-1 reveals a unique active site pocket and interaction sites for HIF-1 and von Hippel-Lindau. *J. Biol. Chem.* 278:7558-7563.

Lees, W. J., and Whitesides, G. M. (1993). Equilibrium constants for thiol-disulfide interchange reactions: A coherent, corrected set. *J. Org. Chem.* 58:642-647.

Leitgeb, S., and Nidetzky, B. (2008). Structural and functional comparison of 2-His-1-carboxylate and 3-His metallocentres in non-haem iron(II)-dependent enzymes. *Biochem. Soc. Trans.* 36:1180-1186.

Lesniak, J., Barton, W. A., and Nikolov, D. B. (2002). Structural and functional characterization of the *Pseudomonas* hydroperoxide resistance protein Ohr. *EMBO J.* 21:6649-6659.

Li, D., Hirsila, M., Koivunen, P., Brenner, M. C., Xu, L., Yang, C., Kivirikko, K. I., and Myllyharju, J. (2004). Many amino acid substitutions in a hypoxia-inducible transcription factor (HIF)-1 α -like peptide cause only minor changes in its hydroxylation by the HIF prolyl 4-hydroxylases: Substitution of 3,4-dehydroproline or azetidine-2-carboxylic acid for the proline leads to a high rate of uncoupled 2-oxoglutarate decarboxylation. *J. Biol. Chem.* 279:55051-55059.

Linke, S., Stojkoski, C., Kewley, R. J., Booker, G. W., Whitelaw, M. L., and Peet, D. J. (2004). Substrate requirements of the oxygen-sensing asparaginyl hydroxylase factor-inhibiting hypoxia-inducible factor. *J. Biol. Chem.* 279:14391-14397.

Liu, J. D., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305:1437-1441.

Loenarz, C., Mecinovic, J., Chowdhury, R., McNeill, L. A., Flashman, E., and Schofield, C. J. (2009). Evidence for a stereoelectronic effect in human oxygen sensing. *Angew. Chem. Int. Ed.* 48:1784-1787.

Lopez-Vera, E., Walewska, A., Skalicky, J. J., Olivera, B. M., and Bulaj, G. (2008). Role of hydroxyprolines in the in vitro oxidative folding and biological activity of conotoxins. *Biochemistry* 47:1741-1751.

Luo, L. S., Pappalardi, M. B., Tummino, P. J., Copeland, R. A., Fraser, M. E., Grzyska, P. K., and Hausinger, R. P. (2006). An assay for Fe(II)/2-oxoglutarate-dependent dioxygenases by enzyme-coupled detection of succinate formation. *Anal. Biochem.* 353:69-74.

Ma, Y., Shida, H., and Kawasaki, T. (1997). Functional expression of human mannan-binding proteins (MBPs) in human hepatoma cell lines infected by recombinant vaccinia virus: Post-translational modification, molecular assembly, and differentiation of serum and liver MBP. *J. Biochem.* 122:810-818.

Majamaa, K., Gunzler, V., Hanauske-Abel, H. M., Myllyla, R., and Kivirikko, K. I. (1986). Partial identity of the 2-oxoglutarate and ascorbate binding sites of prolyl 4-hydroxylase. *J. Biol. Chem.* 261:7819-7823.

Majamaa, K., Hanauske-Abel, H. M., Gunzler, V., and Kivirikko, K. I. (1984). The 2-oxoglutarate binding site of prolyl 4-hydroxylase. Identification of distinct subsites and evidence for 2-oxoglutarate decarboxylation in a ligand reaction at the enzyme-bound ferrous ion. *Eur. J. Biochem.* 138:239-245.

- Marc, D., Mercey, R., and Lantier, F. (2007). Scavenger, transducer, RNA chaperone? What ligands of the prion protein teach us about its function. *Cell. Mol. Life Sci.* 64:815-829.
- McCaldon, P., and Argos, P. (1988). Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins: Struct. Funct. Genet.* 4:99-122.
- McCoy, J. G., Bailey, L. J., Bitto, E., Bingman, C. A., Aceti, D. J., Fox, B. G., and Phillips, G. N. (2006). Structure and mechanism of mouse cysteine dioxygenase. *Proc. Natl. Acad. Sci. U.S.A.* 103:3084-3089.
- McDonough, M. A., Li, V., Flashman, E., Chowdhury, R., Mohr, C., Lienard, B. M., Zondlo, J., Oldham, N. J., Clifton, I. J., Lewis, J., McNeill, L. A., Kurzeja, R. J., Hewitson, K. S., Yang, E., Jordan, S., Syed, R. S., and Schofield, C. J. (2006). Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc. Natl. Acad. Sci. U.S.A.* 103:9814-9819.
- McGinnis, K., Ku, G. M., VanDusen, W. J., Fu, J., Garsky, V., Stern, A. M., and Friedman, P. A. (1996). Site-directed mutagenesis of residues in a conserved region of bovine aspartyl (asparaginyl) β -hydroxylase: Evidence that histidine 675 has a role in binding Fe^{2+} . *Biochemistry* 35:3957-3962.
- McNeill, L. A., Bethge, L., Hewitson, K. S., and Schofield, C. J. (2005). A fluorescence-based assay for 2-oxoglutarate-dependent oxygenases. *Anal. Biochem.* 336:125-131.
- Middlecamp, C. H., and Baldwin, O. (2006). Chemistry, society, and civic engagement (Part 1): The SENCER project. *J. Chem. Ed.* 83:1308-1312.
- Mikhaylova, O., Ignacak, M. L., Barankiewicz, T. J., Harbaugh, S. V., Yi, Y., Maxwell, P. H., Schneider, M., Van Geyte, K., Carmeliet, P., Revelo, M. P., Wyder, M., Greis, K. D., Meller, J., and Czyzyk-Krzeska, M. F. (2008). The von Hippel-Lindau tumor suppressor protein and Egl-9-Type proline hydroxylases regulate the large subunit of RNA polymerase II in response to oxidative stress. *Mol. Cell. Biol.* 28:2701-2717.
- Miller, H., Mande, S. S., Parsonage, D., Sarfaty, S. H., Hol, W. G., and Claiborne, A. (1995). An L40C mutation converts the cysteine-sulfenic acid redox center in enterococcal NADH peroxidase to a disulfide. *Biochemistry* 34:5180-5190.
- Miller, J. H. (1992). A short course in bacterial genetics: A laboratory manual and handbook for *Escherichia coli* and related bacteria (Plainview, NY, Cold Spring Harbor Laboratory Press).
- Miller, M. A., Scott, E. E., and Limburg, J. (2008). Expression, purification, crystallization and preliminary X-ray studies of a prolyl-4-hydroxylase protein from *Bacillus anthracis*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 64:788-791.

- Min, J. H., Yang, H., Ivan, M., Gertler, F., Kaelin, W. G., Jr., and Pavletich, N. P. (2002). Structure of an HIF-1 α -pVHL complex: Hydroxyproline recognition in signaling. *Science* 296:1886-1889.
- Mocharla, R., Mocharla, H., and Leu, R. W. (1987). Effects of inhibitors of C1q biosynthesis on macrophage Fc receptor subclass-mediated antibody-dependent cellular cytotoxicity and phagocytosis. *Cell. Immunol.* 105:127-135.
- Morello, R., Bertin, T. K., Chen, Y., Hicks, J., Tonachini, L., Monticone, M., Castagnola, P., Rauch, F., Glorieux, F. H., Vranka, J., Bachinger, H. P., Pace, J. M., Schwarze, U., Byers, P. H., Weis, M., Fernandes, R. J., Eyre, D. R., Yao, Z., Boyce, B. F., and Lee, B. (2006). CRTAP is required for prolyl 3-hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell* 127:291-304.
- Mori, H., Shibasaki, T., Yano, K., and Ozaki, A. (1997). Purification and cloning of a proline 3-hydroxylase, a novel enzyme which hydroxylates free L-proline to *cis*-3-hydroxy-L-proline. *J. Bacteriol.* 179:5677-5683.
- Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* 374:229-244.
- Muller, I., Kahnert, A., Pape, T., Sheldrick, G. M., Meyer-Klaucke, W., Dierks, T., Kertesz, M., and Uson, I. (2004). Crystal structure of the alkylsulfatase AtsK: Insights into the catalytic mechanism of the Fe(II) α -ketoglutarate-dependent dioxygenase superfamily. *Biochemistry* 43:3075-3088.
- Muller, W., Hanauske-Abel, H., and Loos, M. (1978). Biosynthesis of the first component of complement by human and guinea pig peritoneal macrophages: evidence for an independent production of the C1 subunits. *J. Immunol.* 121:1578-1584.
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D: Biol. Crystallogr.* 53:240-255.
- Myllyharju, J. (2003). Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* 22:15-24.
- Myllyharju, J. (2008). Prolyl 4-hydroxylases, key enzymes in the synthesis of collagens and regulation of the response to hypoxia, and their roles as treatment targets. *Annals of Medicine* 40:402-417.
- Myllyharju, J., and Kivirikko, K. I. (1997). Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *EMBO J.* 16:1173-1180.

- Myllyharju, J., and Kivirikko, K. I. (2004). Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in Genetics* 20:33-43.
- Myllyharju, J., Kukkola, L., Winter, A. D., and Page, A. P. (2002). The exoskeleton collagens in *Caenorhabditis elegans* are modified by prolyl 4-hydroxylases with unique combinations of subunits. *J. Biol. Chem.* 277:29187-29196.
- Myllyla, R., Kuutti-Savolainen, E. R., and Kivirikko, K. I. (1978). The role of ascorbate in the prolyl hydroxylase reaction. *Biochem. Biophys. Res. Commun.* 83:441-448.
- Myllyla, R., Majamaa, K., Gunzler, V., Hanauske-Abel, H. M., and Kivirikko, K. I. (1984). Ascorbate is consumed stoichiometrically in the uncoupled reactions catalyzed by prolyl 4-hydroxylase and lysyl hydroxylase. *J. Biol. Chem.* 259:5403-5405.
- Myllyla, R., Tuderman, L., and Kivirikko, K. I. (1977). Mechanism of the prolyl hydroxylase reaction. 2. Kinetic analysis of the reaction sequence. *Eur. J. Biochem.* 80:349-357.
- Myllyla, R., Wang, C., Heikkinen, J., Juffer, A., Lampela, O., Risteli, M., Ruotsalainen, H., Salo, A., and Sipila, L. (2007). Expanding the lysyl hydroxylase toolbox: New insights into the localization and activities of lysyl hydroxylase 3 (LH3). *J. Cell. Physiol.* 212:323-329.
- Nagarajan, V., Kamitori, S., and Okuyama, K. (1998). Crystal structure analysis of collagen model peptide (Pro-Pro-Gly)₁₀. *J. Biochem.* 124:1117-1123.
- Nagarajan, V., Kamitori, S., and Okuyama, K. (1999). Structure analysis of a collagen-model peptide with a (Pro-Hyp-Gly) sequence repeat. *J. Biochem. (Tokyo)* 125:310-318.
- Nelson, H. C., and Sauer, R. T. (1986). Interaction of mutant λ repressors with operator and non-operator DNA. *J. Mol. Biol.* 192:27-38.
- Neubauer, A., Neubauer, P., and Myllyharju, J. (2005). High-level production of human collagen prolyl 4-hydroxylase in *Escherichia coli*. *Matrix Biology* 24:59-68.
- Newman, J. (2005). Expanding screening space through the use of alternative reservoirs in vapor-diffusion experiments. *Acta Crystallogr. D: Biol. Crystallogr.* 61:490-493.
- Nietfeld, J. J., and Kemp, A. (1980). Properties of prolyl 4-hydroxylase containing firmly-bound iron. *Biochim. Biophys. Acta* 613:349-358.
- Nietfeld, J. J., and Kemp, A. (1981). The function of ascorbate with respect to prolyl 4-hydroxylase activity. *Biochim. Biophys. Acta* 657:159-167.
- Nietfeld, J. J., van der Kraan, J., and Kemp, A. (1981). Dissociation and reassociation of prolyl 4-hydroxylase subunits after cross-linking of monomers. *Biochim. Biophys. Acta* 661:21-27.

- Nolan, J. C., Ridge, S., Oronsky, A. L., and Kerwar, S. S. (1978). Studies on the mechanism of reduction of prolyl hydroxylase activity by D,L-3,4-dehydroproline. *Arch. Biochem. Biophys.* 189:448-453.
- O'Leary, R., Wood, E. J., and Guillou, P. J. (2002). Pathological scarring: Strategic interventions. *Eur. J. Surg.* 168:523-534.
- Oehme, F., Ellinghaus, P., Kolkhof, P., Smith, T. J., Ramakrishnan, S., Hutter, J., Schramm, M., and Flamme, I. (2002). Overexpression of PH-4, a novel putative proline 4-hydroxylase, modulates activity of hypoxia-inducible transcription factors. *Biochem Biophys Res Commun* 296:343-349.
- Olsen, D., Yang, C., Bodo, M., Chang, R., Leigh, S., Baez, J., Carmichael, D., Perala, M., Hamalainen, E. R., Jarvinen, M., and Polarek, J. (2003). Recombinant collagen and gelatin for drug delivery. *Adv. Drug Deliv. Rev.* 55:1547-1567.
- Pan, J. L., and Bardwell, J. C. A. (2006). The origami of thioredoxin-like folds. *Protein Sci.* 15:2217-2227.
- Panasik, N., Jr., Eberhardt, E. S., Edison, A. S., Powell, D. R., and Raines, R. T. (1994). Inductive effects on the structure of proline residues. *Int. J. Pept. Protein Res.* 44:262-269.
- Park, S. H., and Raines, R. T. (2000). Genetic selection for dissociative inhibitors of designated protein-protein interactions. *Nat. Biotechnol.* 18:847-851.
- Park, S. H., and Raines, R. T. (2002). Genetic screen to dissect protein-protein interactions: Ribonuclease inhibitor-ribonuclease A as a model system. *Methods* 28:346-352.
- Pascal, R. A., Oliver, M. A., and Chen, Y. C. J. (1985). Alternate substrates and inhibitors of bacterial 4-hydroxyphenylpyruvate dioxygenase. *Biochemistry* 24:3158-3165.
- Patti, J. M., Boles, J. O., and Hook, M. (1993). Identification and biochemical characterization of the ligand binding domain of the collagen adhesin from *Staphylococcus aureus*. *Biochemistry* 32:11428-11435.
- Patti, J. M., Jonsson, H., Guss, B., Switalski, L. M., Wiberg, K., Lindberg, M., and Hook, M. (1992). Molecular characterization and expression of a gene encoding a *Staphylococcus aureus* collagen adhesin. *J. Biol. Chem.* 267:4766-4772.
- Pauling, L. (1960). *The Nature of the Chemical Bond*, 3rd ed. (Ithaca, NY, Cornell University Press).
- Pearce, G., Bhattacharya, R., Chen, Y. C., Barona, G., Yamaguchi, Y., and Ryan, C. A. (2009). Isolation and Characterization of Hydroxyproline-rich Glycopeptide Signals in Black Nightshade Leaves. *Plant Physiol.*

- Pekkala, M., Hieta, R., Bergmann, U., Kivirikko, K. I., Wierenga, R. K., and Myllyharju, J. (2004). The peptide-substrate-binding domain of collagen prolyl 4-hydroxylases is a tetratricopeptide repeat domain with functional aromatic residues. *J Biol Chem* 279:52255-52261.
- Peluso, S., de L. Ufret, M., O'Reilly, M. K., and Imperiali, B. (2002). Neoglycopeptides as inhibitors of oligosaccharyl transferase. *Chem. Biol.* 9:1323-1328.
- Peterkofsky, B., and Udenfriend, S. (1965). Enzymatic hydroxylation of proline in microsomal polypeptide leading to formation of collagen. *Proc. Natl. Acad. Sci. USA* 53:335-342.
- Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllyla, R., Huhtala, M. L., Koivu, J., and Kivirikko, K. I. (1987). Molecular cloning of the β -Subunit of human prolyl 4-hydroxylase. This subunit and protein disulfide isomerase are products of the same gene. *EMBO J.* 6:643-649.
- Pirneskoski, A., Ruddock, L. W., Klappa, P., Freedman, R. B., Kivirikko, K. I., and Koivunen, P. (2001). Domains b' and a' of protein disulfide isomerase fulfill the minimum requirement for function as a subunit of prolyl 4-hydroxylase. The N-terminal domains a and b enhances this function and can be substituted in part by those of ERp57. *J. Biol. Chem.* 276:11287-11293.
- Porter, R. R., and Reid, K. B. (1978). The biochemistry of complement. *Nature* 275:699-704.
- Prockop, D. J., and Juva, K. (1965). Synthesis of hydroxyproline in vitro by the hydroxylation of proline in a precursor of collagen. *Proc. Natl. Acad. Sci. USA* 53:661-668.
- Prockop, D. J., and Kivirikko, K. I. (1969). Effect of polymer size on the inhibition of protocollagen proline hydroxylase by polyproline II. *J. Biol. Chem.* 244:4838-4842.
- Prockop, D. J., and Kivirikko, K. I. (1995). Collagens: Molecular biology, diseases, and potentials for therapy. *Annu. Rev. Biochem.* 64:403-434.
- Ptashne, M. (2004). A genetic switch: Phage lambda revisited, 3rd edn (Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press).
- Qi, H. H., Ongusaha, P. P., Myllyharju, J., Cheng, D., Pakkanen, O., Shi, Y., Lee, S. W., Peng, J., and Shi, Y. (2008). Prolyl 4-hydroxylation regulates Argonaute 2 stability. *Nature* 455:421-424.
- Qiao, Y., Molina, H., Pandey, A., Zhang, J., and Cole, P. A. (2006). Chemical rescue of a mutant enzyme in living cells. *Science* 311:1293-1297.
- Quan, S., Schneider, I., Pan, J., Von Hacht, A., and Bardwell, J. C. A. (2007). The CXXC motif is more than a redox rheostat. *J. Biol. Chem.* 282:28823-28833.

- Raines, R. T. (2006). 2005 Emil Thomas Kaiser Award. *Protein Sci.* 15:1219-1225.
- Ramshaw, J. A., Werkmeister, J. A., and Glattauer, V. (1996). Collagen-based biomaterials. *Biotechnol. Genet. Eng. Rev.* 13:335-382.
- Ramshaw, J. A. M., Shah, N. K., and Brodsky, B. (1998). Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *J. Struct. Biol.* 122:86-91.
- Rao, N. V., and Adams, E. (1978). Partial reaction of prolyl hydroxylase. (Gly-Pro-Ala)_n stimulates α -ketoglutarate decarboxylation without prolyl hydroxylation. *J. Biol. Chem.* 253:6327-6330.
- Rapaka, R. S., Renugopalakrishnan, V., Urry, D. W., and Bhatnagar, R. S. (1978). Hydroxylation of proline in polytripeptide models of collagen: Stereochemistry of polytripeptide-prolyl hydroxylase interaction. *Biochemistry* 17:2892-2898.
- Renner, C., Alefelder, S., Bae, J. H., Budisa, N., Huber, R., and Moroder, L. (2001). Fluoroproline as tools for protein design and engineering. *Angew. Chem., Int. Ed.* 40:923-925.
- Rhoads, R. E., and Udenfriend, S. (1968). Decarboxylation of α -ketoglutarate coupled to collagen proline hydroxylase. *Proc. Natl. Acad. Sci. USA* 60:1473-1478.
- Rhodes, R. K., and Miller, E. J. (1978). Physicochemical characterization and molecular organization of the collagen A and B chains. *Biochemistry* 17:3442-3448.
- Risteli, M., Niemitalo, O., Lankinen, H., Juffer, A. H., and Myllyla, R. (2004). Characterization of collagenous peptides bound to lysyl hydroxylase isoforms. *J. Biol. Chem.* 279:37535-37543.
- Rosenbloom, J., and Cywinski, A. (1976a). Biosynthesis and secretion of tropoelastin by chick aorta cells. *Biochem. Biophys. Res. Commun.* 69:613-620.
- Rosenbloom, J., and Cywinski, A. (1976b). Inhibition of proline hydroxylation does not inhibit secretion of tropoelastin by chick aorta cells. *FEBS Lett.* 65:246-250.
- Ryan, C. A., and Pearce, G. (2003). Systemins: A functionally defined family of peptide signals that regulate defensive genes in Solanaceae species. *Proc. Natl. Acad. Sci.* 100:14577-12580.
- Rye, C. S., and Withers, S. G. (2002). Development of an assay and determination of kinetic parameters for chondroitin AC lyase using defined synthetic substrates. *Anal. Biochem.* 308:77-82.
- Sakakibara, S., Inouye, K., Shudo, K., Kishida, Y., Kobayashi, Y., and Prockop, D. J. (1973). Synthesis of (Pro-Hyp-Gly)_n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline. *Biochim. Biophys. Acta.* 303:198-202.

- Salvador, R. A., Tsai, I., Marcel, R. J., Felix, A. M., and Kerwar, S. S. (1976). The in vivo inhibition of collagen synthesis and the reduction of prolyl hydroxylase activity by 3,4-dehydroproline. *Arch. Biochem. Biophys.* 174:381-392.
- Schneider, T. R., and Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr. D: Biol. Crystallogr.* 58:1772-1779.
- Schofield, C. J., and Zhang, Z. (1999). Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr. Opin. Struct. Biol.* 9:722-731.
- Searls, T., Butler, D., Chien, W., Mukherji, M., Lloyd, M. D., and Schofield, C. J. (2005). Studies on the specificity of unprocessed and mature forms of phytanoyl-CoA 2-hydroxylase and mutation of the iron binding ligands. *J. Lipid Res.* 46:1660-1667.
- Seppelt, K. (1977). Trifluoromethanol, CF₃OH. *Angew. Chem. Int. Ed.* 16:322-323.
- Sheehan, J. C., Mania, D., Nakamura, S., Stock, J. A., and Maeda, K. (1968). The structure of telomycin. *J. Am. Chem. Soc.* 90:462-470.
- Shimizu, I. (2001). Antifibrogenic therapies in chronic HCV infection. *Curr. Drug Targets Infect. Disord.* 1:227-240.
- Shimizu, M., Igasaki, T., Yamada, M., Yuasa, K., Hasegawa, J., Kato, T., Tsukagoshi, H., Nakamura, K., Fukuda, H., and Matsuoka, K. (2005). Experimental determination of proline hydroxylation and hydroxyproline arabinogalactosylation motifs in secretory proteins. *Plant J.* 42:877-889.
- Shoulders, M. D., Hodges, J. A., and Raines, R. T. (2006). Reciprocity of steric and stereoelectronic effects in the collagen triple helix. *J. Am. Chem. Soc.* 128:8112-8113.
- Shoulders, M. D., and Raines, R. T. (2009). Collagen Structure and Stability. *Annu. Rev. Biochem.*
- Smith, B. D., and Raines, R. T. (2006). Genetic selection for critical residues in ribonucleases. *J. Mol. Biol.* 362:459-478.
- Smith, C. A., Toogood, H. S., Baker, H. M., Daniel, R. M., and Baker, E. N. (1999). Calcium-mediated thermostability in the subtilisin superfamily: The crystal structure of *Bacillus* Ak.1 protease at 1.8 Å resolution. *J. Mol. Biol.* 294:1027-1040.
- Speziale, P., Raucci, G., Visai, L., Switalski, L. M., Timpl, R., and Hook, M. (1986). Binding of collagen to *Staphylococcus aureus* Cowan 1. *J. Bacteriol.* 167:77-81.
- Steiner, T., Hess, P., Bae, J. H., Wiltschi, B., Moroder, L., and Budisa, N. (2008). Synthetic biology of proteins: Tuning GFPs folding and stability with fluoroproline. *PLoS One* 3:e1680.

- Steinmann, B., Bruckner, P., and Superti-Furga, A. (1991). Cyclosporin A slows collagen triple-helix formation in vivo: Indirect evidence for a physiological role of peptidyl-prolyl cis-trans-isomerase. *J. Biol. Chem.* 266:1299-1303.
- Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F. R., and Donnelly, M. I. (2002). A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr. Purif.* 25:8-15.
- Straganz, G. D., and Nidetzky, B. (2006). Variations of the 2-His-1-carboxylate theme in mononuclear non-heme Fe^{II} oxygenases. *Chembiochem* 7:1536-1548.
- Takamatsu, N., Ohba, K., Kondo, J., Kondo, N., and Shiba, T. (1993). Hibernation-associated gene regulation of plasma proteins with a collagen-like domain in mammalian hibernators. *Mol. Cell. Biol.* 13:1516-1521.
- Takeuchi, T., and Prockop, D. J. (1969). Biosynthesis of abnormal collagens with amino acid analogues. I. Incorporation of L-azetidine-2-carboxylic acid and *cis*-4-fluoro-L-proline into protocollagen and collagen. *Biochim. Biophys. Acta* 175:142-155.
- Tanaka, M., Sato, K., and Uchida, T. (1981). Plant prolyl hydroxylase recognizes poly(L-proline) II helix. *J. Biol. Chem.* 256:11397-11400.
- Tandon, M., Wu, M., and Begley, T. P. (1998). Substrate specificity of human prolyl-4-hydroxylase. *Bioorg. Med. Chem. Lett.* 8:1139-1144.
- Terwilliger, T. C. (2003). SOLVE and RESOLVE: Automated structure solution and density modification. *Methods Enzymol.* 374:22-37.
- Thomas, K. M., Naduthambi, D., Tririya, G., and Zondlo, N. J. (2005). Proline editing: A divergent strategy for the synthesis of conformationally diverse peptides. *Org. Lett.* 7:2397-2400.
- Thomas, K. M., Naduthambi, D., and Zondlo, N. J. (2006). Electronic control of amide cis-trans isomerism via the aromatic-prolyl interaction. *J. Am. Chem. Soc.* 128:2216-2217.
- Thornburg, L. D., Lai, M. T., Wishnok, J. S., and Stubbe, J. (1993). A non-heme iron protein with heme tendencies: An investigation of the substrate specificity of thymine hydroxylase. *Biochemistry* 32:14023-14033.
- Tiainen, P., Myllyharju, J., and Koivunen, P. (2005). Characterization of a second *Arabidopsis thaliana* prolyl 4-hydroxylase with distinct substrate specificity. *J. Biol. Chem.* 280:1142-1148.
- Tiainen, P., Pasanen, A., Sormunen, R., and Myllyharju, J. (2008). Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, an enzyme modifying the basement membrane collagen IV. *J. Biol. Chem.* 283:19432-19439.

- Tian, G., Xiang, S., Noiva, R., Lennarz, W. J., and Schindelin, H. (2006). The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell* 124:1085-1088.
- Tuderman, L., Myllyla, R., and Kivirikko, K. I. (1977). Mechanism of the prolyl hydroxylase reaction. 1. Role of co-substrates. *Eur. J. Biochem.* 80:341-348.
- Valegard, K., van Scheltinga, A. C., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H. J., Baldwin, J. E., Schofield, C. J., Hajdu, J., and Andersson, I. (1998). Structure of a cephalosporin synthase. *Nature* 394:805-809.
- van de Wetering, J. K., van Golde, L. M., and Batenburg, J. J. (2004). Collectins: Players of the innate immune system. *Eur. J. Biochem.* 271:1229-1249.
- Van Den Diepstraten, C., Papay, K., Bolender, Z., Brown, A., and Pickering, J. G. (2003). Cloning of a novel prolyl 4-hydroxylase subunit expressed in the fibrous cap of human atherosclerotic plaque. *Circulation* 108:508-511.
- van der Wel, H., Ercan, A., and West, C. M. (2005). The Skp1 prolyl hydroxylase from Dictyostelium is related to the hypoxia-inducible factor- α class of animal prolyl 4-hydroxylases. *J Biol Chem* 280:14645-14655.
- Venetz, W. P., Mangan, C., and Siddiqi, I. W. (1990). Kinetic determination of alkaline-phosphatase activity based on hydrolytic cleavage of the P-F bond in monofluorophosphate and fluoride ion-selective electrode. *Anal. Biochem.* 191:127-132.
- von Grotthuss, M., Pas, J., Wyrwicz, L., Ginalska, K., and Rychlewski, L. (2003). Application of 3D-Jury, GRDB, and Verify3D in fold recognition. *Proteins* 53:418-423.
- Voss, T., Eistetter, H., Schafer, K. P., and Engel, J. (1988). Macromolecular organization of natural and recombinant lung surfactant protein SP 28-36. Structural homology with the complement factor C1q. *J. Mol. Biol.* 201:219-227.
- Vranka, J. A., Sakai, L. Y., and Bachinger, H. P. (2004). Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. *J. Biol. Chem.* 279:23615-23621.
- Vuori, K., Myllyla, R., Pihlajaniemi, T., and Kivirikko, K. I. (1992a). Expression and site-directed mutagenesis of human protein disulfide isomerase in *Escherichia coli*. This multifunctional polypeptide has two independently acting catalytic sites for the isomerase activity. *J. Biol. Chem.* 267:7211-7214.

- Vuori, K., Pihlajaniemi, T., Marttila, M., and Kivirikko, K. I. (1992b). Characterization of the human prolyl 4-hydroxylase tetramer and its multifunctional protein disulfide-isomerase subunit synthesized in a baculovirus expression system. *Proc. Natl. Acad. Sci. USA* 89:7467-7470.
- Vuori, K., Pihlajaniemi, T., Myllyla, R., and Kivirikko, K. I. (1992c). Site-directed mutagenesis of human protein disulphide isomerase: Effect on the assembly, activity and endoplasmic reticulum retention of human prolyl 4-hydroxylase in *Spodoptera frugiperda* insect cells. *EMBO J.* 11:4213-4217.
- Walker, K. W., and Gilbert, H. F. (1997). Scanning and escape during protein-disulfide isomerase-assisted protein folding. *J. Biol. Chem.* 272:8845-8848.
- Walsh, C. T., Garneau-Tsodikova, S., and Gatto, G. J., Jr. (2005). Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* 44:7342-7372.
- Walter, C., and Frieden, E. (1963). The prevalence and significance of the product inhibition of enzymes. *Adv. Enzymol. Relat. Areas Mol. Biol.* 25:167-274.
- Wang, J., Chen, G., Lee, J., and Pantopoulos, K. (2008). Iron-dependent degradation of IRP2 requires its C-terminal region and IRP structural integrity. *BMC Mol. Biol.* 9:15.
- Wang, J., Chen, G., Muckenthaler, M., Galy, B., Hentze, M. W., and Pantopoulos, K. (2004). Iron-mediated degradation of IRP2, an unexpected pathway involving a 2-oxoglutarate-dependent oxygenase activity. *Mol. Cell. Biol.* 24:954-965.
- West, C. M., Van Der Wel, H., Sassi, S., and Gaucher, E. A. (2004). Cytoplasmic glycosylation of protein-hydroxyproline and its relationship to other glycosylation pathways. *Biochim Biophys Acta* 1673:29-44.
- West, C. M., van der Wel, H., and Wang, Z. A. (2007). Prolyl 4-hydroxylase-1 mediates O₂ signaling during development of *Dictyostelium*. *Development* 134:3349-3358.
- West, C. M., Wel, H. v. d., and Blader, I. J. (2006). Detection of cytoplasmic glycosylation associated with hydroxyproline. *Methods Enzymol.* 417:389-404.
- Wilkinson, B., and Gilbert, H. F. (2004). Protein disulfide isomerase. *Biochim. Biophys. Acta* 1699:35-44.
- Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D: Biol. Crystallogr.* 57:122-133.

- Winter, A. D., and Page, A. P. (2000). Prolyl 4-hydroxylase is an essential procollagen-modifying enzyme required for exoskeleton formation and the maintenance of body shape in the nematode *Caenorhabditis elegans*. *Mol. Cell Biol.* 20:4084-4093.
- Wong, C., Fujimori, D. G., Walsh, C. T., and Drennan, C. L. (2009). Structural analysis of an open active site conformation of nonheme iron halogenase CytC3. *J. Am. Chem. Soc.* 131:4872-4879.
- Woycechowsky, K. J., and Raines, R. T. (2000). Native disulfide bond formation in proteins. *Curr. Opin. Chem. Biol.* 4:533-539.
- Woycechowsky, K. J., and Raines, R. T. (2003). The CXC motif: A functional mimic of protein disulfide isomerase. *Biochemistry* 42:5387-5394.
- Wu, H., Graaf, B. d., Mariani, C., and Cheung, A. Y. (2001). Hydroxyproline-rich glycoproteins in plant reproductive tissues: Structure, functions and regulation. *Cell. Mol. Life Sci.* 58:1418-1429.
- Yuasa, K., Toyooka, K., Fukuda, H., and Matsuoka, K. (2005). Membrane-anchored prolyl hydroxylase with an export signal from the endoplasmic reticulum. *Plant J.* 41:81-94.
- Zhang, R. M., and Snyder, G. H. (1989). Dependence of formation of small disulfide loops in two-cysteine peptides on the number and types of intervening amino acids. *J. Biol. Chem.* 264:18472-18479.
- Zhang, X.-M. (1998). Radical substituent effects of an α -fluoro and α -trifluoromethyl groups. *J. Org. Chem.* 63:3590-3594.
- Zhang, Z., Ren, J., Stammers, D. K., Baldwin, J. E., Harlos, K., and Schofield, C. J. (2000). Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat. Struct. Biol.* 7:127-133.
- Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., McKee, E., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., Storoni, L. C., Terwilliger, T. C., and Adams, P. D. (2008). Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* 426:419-435.